JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

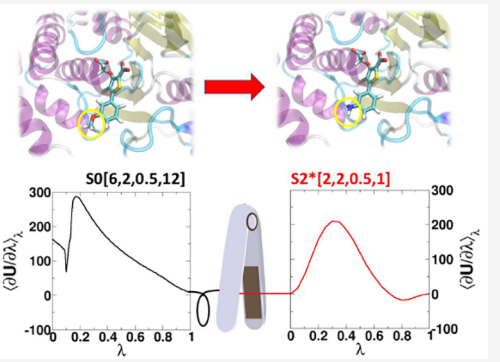
AMBER Free Energy Tools: A New Framework for the Design of Optimized Alchemical Transformation Pathways

Hsu-Chun Tsai, Tai-Sung Lee, Abir Ganguly, Timothy J. Giese, Maximilian CCJC Ebert, Paul Labute, Kenneth M. Merz, Jr., and Darrin M. York*

Cite This: J. Chem. Theory Comput. 2023, 19, 640−658 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We develop a framework for the design of optimized alchemical transformation pathways in free energy simulations using nonlinear mixing and a new functional form for so-called "softcore" potentials. We describe the implementation and testing of this framework in the GPU-accelerated AMBER software suite. The new optimized alchemical transformation pathways integrate a number of important features, including (1) the use of smoothstep functions to stabilize behavior near the transformation end points, (2) consistent power scaling of Coulomb and Lennard-Jones (LJ) interactions with unitless control parameters to maintain balance of electrostatic attractions and exchange repulsions, (3) pairwise form based on the LJ contact radius for the effective interaction distance with separation-shifted scaling, and (4) rigorous smoothing of the potential at the nonbonded cutoff boundary. The new softcore potential form is combined with smoothly transforming nonlinear λ weights for mixing specific potential energy terms, along with flexible λ-scheduling features, to enable robust and stable alchemical transformation pathways. The resulting pathways are demonstrated and tested, and shown to be superior to the traditional methods in terms of numerical stability and minimal variance of the free energy estimates for all cases considered. The framework presented here can be used to design new alchemical enhanced sampling methods, and leveraged in robust free energy workflows for large ligand data sets.



1. INTRODUCTION

Alchemical free energy simulations are an integral part of computer aided drug design, particularly at the stage of lead refinement where they are used to rank the binding affinity of compounds to their targets, and in some cases make predictions about selectivity and off-target effects.[1−8] As a result, there is great activity in the field to develop a wide range of methods for which to improve the accuracy, robustness and throughput of these simulations.[2,9−13] Alchemical free energy methods leverage the property that the free energy is a state function to enable nonphysical thermodynamic pathways to be constructed that are amenable to computation. So while, in principle, the free energy is independent of transformation pathway, in practice the choice of pathway is crucial. For example, whereas the direct calculation of the binding free energy of a ligand along a physical pathway can be considerably challenging and computationally intensive to achieve with high precision (although there has been much progress in this area[14−16]), it is often easier to compute the "alchemical" transformation between two similar ligands, both in the bound and unbound states, to arrive at the desired free energy difference (Figure 1). In this way, the relative binding free energy[17] (RBFE) can be computed with sampling requirements that are often less intensive and more precise than the calculation of an absolute binding free energy (ABFE) whether along a physical pathway that would provide additional information about binding kinetics, or alchemical pathway where the ligand is "annihilated" in the bound state and in solution. Analogously, the relative and absolute solvation free energy (RSFE and ASFE, respectively) can be computed considering transformations in the gas phase and in solution. These cycles are illustrated in Figure 1.

Alchemical free energy simulations typically require atoms to be created and/or annihilated during the transformation process—or more precisely, transformed into so-called "dummy atoms".[18] Dummy atoms are placeholders that are designed to interact with the real atoms of the physical system only through select bonded interactions such that they do not alter the relative free energy (i.e., they do not introduce a net potential of mean force on any of the real atoms). Transformations of real atoms into dummy atoms can be especially challenging if there is poor phase space overlap of neighboring states along the transformation coordinate.[18−22]

Received: July 12, 2022
Published: January 9, 2023

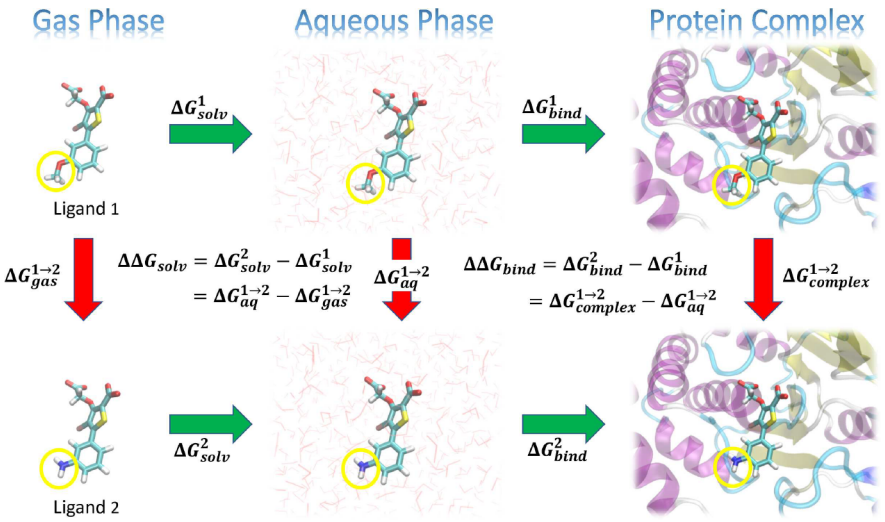


Figure 1. Illustration of a thermodynamic cycle for a relative solvation free energy, RSFE ($\Delta\Delta G_{solv}$) and relative binding free energy, RBFE ($\Delta\Delta G_{bind}$) for two ligands ("Ligand 1" and "Ligand 2"). The green arrows represent the absolute solvation free energy, ASFE (ΔG_{solv}) and absolute binding free energy, ABFE (ΔG_{bind}) of each ligand (indicated by superscripts) that involve changing their environment from gas to aqueous phase, or from unbound in the aqueous phase to bound in a complex with the protein target, respectively. These quantities are experimentally measurable, but are challenging to directly compute as the change in the environment can be considerably complicated. The red arrows represent alchemical transformations where Ligand 1 is mutated into a similar Ligand 2 in a fixed environment. These transformations are frequently more amenable to practical computations. The yellow circles in the figures indicate the region of each ligand that undergoes the most significant changes in the alchemical transformation, and would likely be modeled using a so-called "softcore potential" during the transformation.

This issue is not restricted to only transformations of real atoms into dummy atoms (which is an extreme case), it can also occur in transformations between two real atoms that have significantly different force field parameters.

A number of strategies have been explored to develop robust, stable transformation pathways in alchemical free energy simulations, including use of so-called "softcore potentials" with separation-shifted scaling,[23−25] parameter interpolation,[26] short-range switching,[27] or capping the short-range interactions,[28,29] and nonlinear mixing of the end point potentials.[30−34] Recent studies have shown that adverse effects of these problematic transitions can lead to large variance and in some cases order/disorder transitions along the alchemical path that can hinder sampling and convergence of free energy estimates.[29] Very recently, a λ-enveloping distribution sampling method,[35−37] which is related to approaches to optimize minimum variance pathways in alchemical transformations,[38−40] has been explored as an alternative coupling scheme to more conventional λ-intermediate states.

There are many strategies for performing free energy simulations,[2,4] and in turn choosing appropriate alchemical transformation pathways. In the present work, we will develop robust alchemical pathways (including new softcore potentials) that are particularly useful for "concerted" alchemical transformations (sometimes referred to as "one-step" or "unified" procedures) where all nonbonded terms (e.g., electrostatic and Lennard-Jones terms) occur synchronously. This differs from "stepwise" transformations (sometimes referred to as "multi-step" or "split" procedures) where transformations of electrostatic and Lennard-Jones terms occur asynchronously; for example, in a 3-step "decharge/LJ/recharge" transformation. Concerted transformations are particularly useful in relative binding free energy calculations as they avoid weakly bound states that may require additional restraints. Nonetheless, the methods presented here are not restricted to purely concerted transformations, and as will be shown below, have been integrated into a powerful λ-scheduling framework in AMBER that enables the design of customized alchemical transformations for different energy terms that are seamlessly integrated with existing enhanced sampling tools such as replica exchange.

In the present work, we develop new highly robust alchemical transformation pathways that are designed to overcome commonly encountered classes of problems[23,27,30,31,33] designated here as end point catastrophes, particle collapses, and large gradient-jumps. This work builds on our previously developed smoothstep softcore potentials,[41] but introduces a number of important improvements, including (1) consistent power scaling of Coulomb (Coul) and Lennard-Jones (LJ) interactions with unitless control parameters to maintain balance of electrostatic attractions and exchange repulsions, (2) introduction of a pairwise form based on the LJ contact radius for the effective interaction distance with separation-shifted scaling, and (3) rigorous smoothing of the softcore potential at the nonbonded cutoff boundary. We test the new alchemical transformation pathway on a number of systems in both the context of thermodynamic integration (TI) and free energy perturbation (FEP) and compare results to other commonly used pathways and softcore potentials.

The remainder of the manuscript is organized as follows. Section 2 outlines the key theoretical background. Section 3 provides details of the computational methods. Section 4 presents results and discussion for a series of test cases and comparison with other methods. Section 5 concludes with a summary of main points and direction for next developments. The methods presented here are available for beta testing in the Drug Discovery Boost upgrade to the AMBER software suite and will be integrated into the next official AMBER release.

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

2. THEORY

This section summarizes the theoretical background for the alchemical transformation pathway framework and provides implementation-level details. In particular, we develop alchemical transformation pathways that combine nonlinear mixing of different potential energy terms and a new smoothstep softcore potential to enable robust, stable free energies to be computed. Section 2.1 summarizes the necessary theoretical background and establishes a notational convention that enables precise definition of specific energy terms, transforming regions and interacting sets of atoms that form the λ-dependent potential energy and transformation pathway. Section 2.2 describes flexible, stable forms and λ-scheduling of the weight functions used to mix different potential energy terms. Section 2.3 describes the new functional form for the softcore potentials. Together, the weights along with the softcore potentials form the foundation of the new alchemical transformation pathways that are developed and tested in the Results and Discussion.

2.1. Background and Notation. This section presents the necessary theoretical background as well as develops a notation that enables precise identification of the energy terms and interacting sets of atoms that form the basis of different alchemical transformation pathways. The discussion of the alchemical transformation pathway here will be in the context of thermodynamic integration (TI)[42,43] using a hybrid single-dual-topology[34] implementation in AMBER22. Nonetheless, the improvement in phase space overlap between states along the pathway are transferable to other FEP methods with traditional BAR,[44,45] MBAR,[46,47] and formally equivalent unbinned weighted histogram analysis methods (UWHAM),[48] as well as their recent extensions that enable large-scale network-wide analysis using a constrained variational approach (BARnet and MBARnet).[49]

In practice, TI and FEP methods generally require taking incremental steps along an alchemical transformation pathway parametrized by a coordinate λ that varies between 0 and 1. The end states at λ values of 0 and 1 are generally physical "real states", i.e., chemically distinct molecules with distinct compositions. Alternatively, the continuum of states for $0 < \lambda < 1$ are nonphysical "alchemical states". The free energy change, $\Delta A_{0\rightarrow 1}$, between states "0" and "1" can be achieved through integration of the thermodynamic derivative as

$$\Delta A_{0\rightarrow 1} = \int_0^1 \mathrm{d}\lambda \cdot \left(\frac{\mathrm{d}A}{\mathrm{d}\lambda}\right) = \int_0^1 \mathrm{d}\lambda \cdot \left\langle \frac{\partial U(\mathbf{r}^N;\lambda)}{\partial \lambda} \right\rangle_\lambda \approx \sum_{k=1}^{M} w_k \cdot \left\langle \frac{\partial U(\mathbf{r}^N;\lambda)}{\partial \lambda} \right\rangle_{\lambda_k} \quad (1)$$

where $\mathrm{r}^N = \mathrm{r}_1, \mathrm{r}_2, ..., \mathrm{r}_N$ represents the Cartesian positions of each particle, the second integral involves the derivative of the potential energy U with respect to the parameter λ that smoothly connects the end states $\lambda = 0$ and $\lambda = 1$, and the sum indicates numerical integration over M quadrature points (λ_k, for $k = 1, ..., M$) with associated weights w_k. While the free energy is a state function and formally is invariant to the pathway connecting states, in practical simulations the thermodynamic averages in eq 1 are extremely sensitive to the pathway, and similar issues arise for FEP methods. The λ-dependent total potential energy $U(\mathrm{r}^N;\lambda)$ can be written as

$$U(\mathbf{r}^N;\lambda) = \sum_t W_{0,t}(\lambda)\cdot U_{0,t}(\mathbf{r}^N;\lambda) + W_{1,t}(\lambda)\cdot U_{1,t}(\mathbf{r}^N;\lambda) \quad (2)$$

where the individual state energies, $U_0(\mathrm{r}^N;\lambda)$ and $U_1(\mathrm{r}^N;\lambda)$ are given in terms of their energy term components (indexed by t) as

$$U_0(\mathbf{r}^N;\lambda) = \sum_t U_{0,t}(\mathbf{r}^N;\lambda) \quad (3)$$

and similarly for $U_1(\mathrm{r}^N;\lambda)$. In eq 3, the subscript t sums over different potential energy terms summarized in Table 1. These individual state energy terms in general can have an explicit nonlinear λ-dependence that arises from the form of the softcore potential that is used, as will be described in section 2.3. In addition, in the expression for the λ-dependent total potential energy $U(\mathrm{r}^N;\lambda)$, each of the energy terms $U_{0/1,t}(\mathrm{r}^N;\lambda)$ has a λ-dependent weight ("mixing term") associated with it, $W_{0/1,t}(\lambda)$, the form of which will be described in more detail in section 2.2 below.

Table 1. Definition of Potential Energy Terms and Their Abbreviations Used as Subscripts[a]

energy term	index	description	collective term
U_{bond}	7	Bond stretch	Bonded (b) $U_b = U_{bond} + U_{ang} + U_{tor}$
U_{ang}	6	Angle bend	
U_{tor}	5	Torsion rotate (proper/improper)	
U_{LJ}	4	Lennard-Jones	Nonbonded (nb) $U_{nb} = U_{dir} + U_{1-4Ele} + U_{LJ} + U_{1-4LJ}$
U_{1-4LJ}	3	1−4 Lennard-Jones	
U_{dir}	2	PME direct/real space	
U_{1-4Ele}	1	1−4 Electrostatic	
U_{rec}	0	PME reciprocal space	

[a]The absence of an energy term subscript indicates all energy terms (i.e., summation over all energy terms, index $t = 0, ..., 7$). The bonded and nonbonded terms are "short" and "intermediate" ranged, respectively, and kept track of with array lists, the latter for which is set by a distance cutoff and updated dynamically when needed. Note under these definitions, the 1−4 Ele and 1−4 LJ terms are considered part of the nonbonded terms, and the total electrostatic energy (U_{Ele}) is not purely a "nonbonded" term as it contains also the nonlocal reciprocal space term, i.e., $U_{Ele} = U_{dir} + U_{1-4Ele} + U_{rec}$.

Each of the energy terms in Table 1, with the exception of U_{rec}, involves a straightforward summation over the relevant sets of atoms to compute 2-body, 3-body, or 4-body interactions. In the case of free energy simulations, we need to further distinguish between contributions to the energy that are made from different nonoverlapping sets of atoms. Specifically, we need to subdivide the system into two main subdivisions: one region is alchemically transforming, whereas the remainder of the system is immutable (I), i.e., not transforming. Within the hybrid single-dual-topology, the immutable region is represented by a single "topology" and set of coordinates. The transforming region of the system is represented by a formal dual topology and separate sets of coordinates for each state, and is further subdivided into constrained coordinate/common core (CC) and the separable-coordinate/softcore (SC) regions (in previous work[13] we used the abbreviations TC and TS, but feel that SC and CC are more straightforward). The SC, CC, and I regions are illustrated in Figure 2. The CC region has corresponding

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

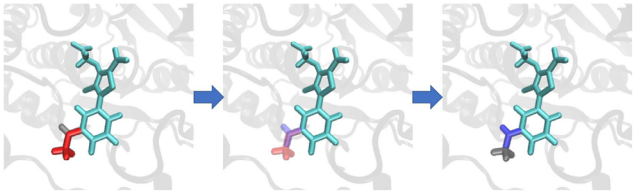


Figure 2. Illustration of a transformation between two ligands (L66 and L67) bound to a protein target (PTP1B) showing the softcore (SC, blue and red), common core (CC, cyan) and immutable (I, gray) regions. There are two SC regions (red for L67 on the left, and blue for L66 on the right) that represent functional groups that are annihilated/created during the transformation and involve separate topologies and coordinates. In the end states where the atoms of the SC region are transformed into noninteracting "dummy" atoms, their colors are grayed out. The middle panel depicts an intermediate λ value for which the SC regions of L66 and L67 are partially transformed, as depicted by their partial translucency. The CC region is transforming as well (as the force field parameters are changing from those of L67 to L66), but the atoms share a common set of coordinates within a single topology in order to reduce the amount of phase space sampling. The immutable region is the protein and surrounding solvent that are not transforming. The coordinates of the SC, CC, and I regions are all treated dynamically at every stage of the alchemical transformation which is conducted in a series of simulations ("λ windows") each at different values of the alchemical transformation coordinate λ that varies from 0 to 1 and controls the potential energy.

atoms in each topology constrained to have the same positions in order to facilitate phase space overlap between states during the alchemical transformation. The SC region, on the other hand, has separable independent coordinates for each topology that can adopt different conformations and do not directly interact with one another.

For example, if two drug molecules involved in an alchemical transformation (e.g., as shown in Figure 2) share a common chemical core of atoms such as an aromatic ring and differ only by certain attached substituents for which atoms between the topologies cannot easily be mapped, then the CC region would contain the atoms of the common aromatic ring and the SC region would contain the atoms of the different substituents. In the alchemical transformation of λ: 0 → 1, the SC atoms of state "0" are "turned off" by mutating the real atoms of state "0" into so-called "dummy atoms",[18] while at the same time the SC atoms of state "1" are being "turned on" in a synchronous counter-transformation. The dummy atoms do not interact with their environment, with the exception of certain bonded interactions that must obey the constraint conditions that they introduce no net potential of mean force onto any nondummy atom. Often this separable dual-coordinate approach requires the introduction of "softcore potentials",[24,41] i.e., explicit nonlinear λ-dependent terms to "soften" the interaction of these atoms with their surroundings. These are most often employed for nonbonded interactions such as LJ and Ele (or in the case of PME electrostatics, often just the U_{dir} term), but other forms have also been developed for bonds and other energy terms.[23,24,33]

Thus, the system can be divided into nonoverlapping regions described in Table 2: I (immutable), CC (transforming constrained/common core), and SC (transforming separable/softcore). The I region has the same atomic coordinates, parameters, and internal potential energy for both states 0 and 1. The CC region can have different parameters between states 0 and 1, but the coordinates of mapped atoms are constrained to be the same. The SC region also can have different parameters between states 0 and 1, but unlike the CC region each state has its own separable set of atomic coordinates.

To facilitate development of the framework for enabling improved alchemical transformation pathways, we now introduce a system whereby we use subscripts to indicate the state ("0" or "1") and (optionally) the specific term in the potential energy as described in Table 1, and superscripts to indicate the specific atoms involved in the energy term interaction. The state of the system and specific energy terms follow the general form $U^{X}_{\{0/1\},t}$ or $U^{X/Y}_{\{0/1\},t}$ where the state is indicated as either 0 or 1, the energy term is designated an appropriate abbreviation index (t) as indicated in Table 1, and the superscript "X" indicates an internal potential energy for region "X" and "X/Y" indicates the interaction energy between regions "X" and "Y" as indicated in Table 2.

Using this notation, the individual state "0" energy, $U_0(\mathrm{r}^N;\lambda)$ can be expanded as

Table 2. Energy Decomposition Terms Based on Non-Overlapping Sets of Atoms Comprising the Separable-Coordinate/Softcore (SC), Constrained Coordinate/Common Core (CC) and Immutable (I) Regions[a]

energy	region/interactions	description
U^{SC}	Internal energy of the SC region	Each of the contributing bonded or nonbonded internal energy terms arises from a set of atoms that are contained within indicated region; i.e., all atoms of the term belong to the indicated region.
U^{CC}	Internal energy of the CC region	
U^{I}	Internal energy of the I region	
$U^{(CC+I)}$	Internal energy of the CC+I region	
$U^{SC/(CC+I)}$	Interaction energy between SC and (CC+I) regions: $U^{SC/(CC+I)} = U^{SC/CC}+U^{SC/I}$	Each of the contributing bonded or nonbonded interaction energy terms arises from a set of atoms that span the SC and combined (CC+I) regions; i.e., some belong to the SC region, while others in the same term belong to the (CC+I) region.

[a]The three regions are as follows: (1) (SC) Separable coordinate/softcore region that is transforming with λ and is treated with separable coordinates within the dual topology framework; (2) (CC) Constrained coordinate/common core region that is transforming with λ but that shares the same coordinates within the dual topology framework (making it effectively a single topology); and (3) (I) Immutable (not transforming with λ). Regions can be combined as a union of atom sets with the "+" operator; e.g., (SC+CC) combines the SC and CC regions, and (SC+CC+I) would imply all regions (and hence atoms) of the system. The energy decomposition involves using the superscripts U^{X} to indicate either an internal energy within the region "X", or an interaction energy $U^{X/Y}$ between regions "X" and "Y". Thus, in its most expanded form, the total potential energy can be written as $U = U^{SC} + U^{CC} + U^{I} + U^{SC/CC} + U^{SC/I} + U^{CC/I}$. Note the absence of subscripts indicates a summation over all energy terms in Table 1; however, the superscript notation can also be applied to energy terms individually. It is assumed, for the convenience and simplification of our notation, that the SC, CC, and I regions are defined in such a way that no individual 3-body or 4-body term spans all three regions (e.g., there is no angle bending term that has one atom in each of the SC, CC, and I regions).

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

$$U_0(\mathbf{r}^N;\lambda) = \sum_t U_{0,t}^{SC}(\mathbf{r}^N;\lambda) + U_{0,t}^{SC/(CC+I)}(\mathbf{r}^N;\lambda) + U_{0,t}^{(CC+I)}(\mathbf{r}^N;\lambda) \quad (4)$$

and similarly for $U_1(\mathrm{r}^N;\lambda)$. Note that $U_0(\mathrm{r}^N;\lambda)$ is the potential energy of the state "0" topology written as (possibly) having an explicit λ dependence (e.g., through a softcore potential or parameter interpolation[26] form). The general expanded form of the λ-dependent total potential energy $U(\mathrm{r}^N;\lambda)$ can be written as

$$\begin{aligned} U(\mathbf{r}^N;\lambda) = {} & \sum_t W_{0,t}^{SC}(\lambda)\cdot U_{0,t}^{SC}(\mathbf{r}^N;\lambda) + W_{1,t}^{SC}(\lambda)\cdot U_{1,t}^{SC}(\mathbf{r}^N;\lambda) \\ & + \sum_t W_{0,t}^{SC/(CC+I)}(\lambda)\cdot U_{0,t}^{SC/(CC+I)}(\mathbf{r}^N;\lambda) \\ & + W_{1,t}^{SC/(CC+I)}(\lambda)\cdot U_{1,t}^{SC/(CC+I)}(\mathbf{r}^N;\lambda) \\ & + \sum_t W_{0,t}^{(CC+I)}(\lambda)\cdot U_{0,t}^{(CC+I)}(\mathbf{r}^N;\lambda) \\ & + W_{1,t}^{(CC+I)}(\lambda)\cdot U_{1,t}^{(CC+I)}(\mathbf{r}^N;\lambda) \end{aligned} \quad (5)$$

The above equation is general in the sense that it assumes a possible explicit nonlinear λ-dependence for any of the energy terms. In the framework presented here that has been implemented into the AMBER Drug Discovery Boost package to AMBER22 (and will be officially available in the future AMBER release), only the terms that involve nonbonded interaction involving the SC region will potentially use softcore potentials, and hence have an explicit nonlinear λ dependence. Further, as mentioned previously, in the current presentation, the PME reciprocal space term is not decomposed into regional contributions. With these conditions, the specific λ-dependent total potential energy $U(\mathrm{r}^N;\lambda)$ can be written as

$$\begin{aligned} U(\mathbf{r}^N;\lambda) = {} & W_{0,rec}(\lambda)\cdot U_{0,rec}(\mathbf{r}^N) + W_{1,rec}(\lambda)\cdot U_{1,rec}(\mathbf{r}^N) \\ & + \sum_{t\neq rec} W_{0,t}^{(CC+I)}(\lambda)\cdot U_{0,t}^{(CC+I)}(\mathbf{r}^N) \\ & + W_{1,t}^{(CC+I)}(\lambda)\cdot U_{1,t}^{(CC+I)}(\mathbf{r}^N) \\ & + \sum_{t\neq rec} W_{0,t}^{SC/(CC+I)}(\lambda)\cdot U_{0,t}^{SC/(CC+I)}(\mathbf{r}^N;\lambda) \\ & + W_{1,t}^{SC/(CC+I)}(\lambda)\cdot U_{1,t}^{SC/(CC+I)}(\mathbf{r}^N;\lambda) \\ & + \sum_{t\neq rec} W_{0,t}^{SC}(\lambda)\cdot U_{0,t}^{SC}(\mathbf{r}^N;\lambda) + W_{1,t}^{SC}(\lambda)\cdot U_{1,t}^{SC}(\mathbf{r}^N;\lambda) \end{aligned} \quad (6)$$

2.2. λ-Dependent Weight Functions for Scaling Potential Energy Components. This section describes the λ-dependent weight functions and the control flags that allow their manipulation in free energy simulations in AMBER. The next subsection 2.2.1 introduces the family of smoothstep functions and describes their most relevant properties that will be exploited to develop the λ-dependent weight functions in the following subsection 2.2.2. In subsection 2.2.3, these weight functions are then generalized to operate on specific subranges of λ between 0 and 1 to form a flexible λ-scheduling framework.

2.2.1. Smoothstep Functions. Consider the family of smoothstep functions of orders P (P = 0, 1, 2, ⋯) defined as the polynomial functions (up to P = 4 shown):

for $0 \le x \le 1$:

$$S_0(x) = x,$$

$$S_1(x) = -2x^3 + 3x^2,$$

$$S_2(x) = 6x^5 - 15x^4 + 10x^3,$$

$$S_3(x) = -20x^7 + 70x^6 - 84x^5 + 35x^4,$$

$$S_4(x) = 70x^9 - 315x^8 + 540x^7 - 420x^6 + 126x^5,$$

and

$$S_P(x \le 0) = 0;\ S_P(x \ge 1) = 1,\ \forall\ P \in \mathbb{N} \quad (7)$$

The smoothstep functions are monotonically increasing functions that have desirable 0 and 1 end point values and vanishing end point derivative properties:

$$\left[\frac{\mathrm{d}^k S_P(x)}{\mathrm{d}x^k}\right]_{x=0} = \left[\frac{\mathrm{d}^k S_P(x)}{\mathrm{d}x^k}\right]_{x=1} = 0 \quad \forall\ k \in \mathbb{N},\ 0 < k \le \mathrm{P} \quad (8)$$

In addition, the smoothstep functions obey the symmetry condition

$$S_P(1-x) = 1 - S_P(x) \quad (9)$$

A smoothstep function with a higher order will have a smoother function curve and smaller derivatives near 0 and 1 but a larger derivative in between. The zero-order (P = 0) smoothstep function is in fact simply linear with constant slope, including at the end points, which can lead to end point catastrophe problems. As illustrated in previous work,[41] the second order smoothstep function (P = 2) overall offers a good balance between smooth vanishing derivatives at the end points, and modest derivatives for intermediate values of λ. AMBER22 offers the flexibility to choose different smoothstep functions through the λ-scheduling mechanism described below.

2.2.2. Form of the λ-Dependent Weight Functions. We now describe a general form for the weight functions $W(\lambda)$, where we only retain the 0 and 1 subscript to indicate the state. The weight functions are defined in terms of the smoothstep functions as

$$W_0(\lambda) = 1 - S_P(\lambda) = S_P(1-\lambda) \quad (10)$$

$$W_1(\lambda) = S_P(\lambda) \quad (11)$$

In the above equation, we drop the explicit superscripts and subscripts in eq 6 that can be controlled by different flags available to the user in AMBER22. Previous work has illustrated that use of smoothstep functions of order greater than 0 (i.e., a weight function that goes beyond the simple linear λ-dependence and has vanishing derivatives at the end points), affords improvement of the transformation pathway, particularly at the end points where large variation in $< \partial U/\partial\lambda >_\lambda$ can occur.[41] These weight functions both operate within the range $0 \le \lambda \le 1$ (they have constant end point values outside of this range), and satisfy the normalization condition:

$$W_0(\lambda) + W_1(\lambda) = 1 \quad (12)$$

and the symmetry condition:

$$W_0(1-\lambda) = W_1(\lambda) \quad (13)$$

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

2.2.3. λ-Scheduling of Weight Functions. In some cases, it is desirable to have the flexibility to apply more complicated λ schedules that operate over a subinterval of λ values between 0 and 1. The generalized λ scheduling weight for W_0 can be defined so that it is changing only within the interval $\lambda_{min} \le \lambda \le \lambda_{max}$ as

$$W_0(\lambda) = 1 - S_p(z(\lambda))$$

$$z(\lambda)\begin{cases} = 0, & \text{if } \lambda \le \lambda_{min} \\ = \dfrac{\lambda - \lambda_{min}}{\lambda_{max} - \lambda_{min}}, & \text{if } \lambda_{min} \le \lambda \le \lambda_{max} \\ = 1, & \text{if } \lambda_{max} \le \lambda \end{cases} \tag{14}$$

where $0 \le \lambda_{min} \le \lambda_{max} \le 1$. In the current framework, the complementary weight function $W_1(\lambda)$ can be selected to either satisfy the normalization condition (eq 12), or the symmetry condition (eq 13) above. Only if the interval $z(\lambda)$ is centered at $\lambda = 0.5$ are both the normalization and symmetry conditions simultaneously satisfied. The Drug Discovery Boost package in AMBER22 allows flexible λ scheduling of this form for different energy components. The detailed usage can be found in the updated AMBER22 manual.[50]

2.3. New Softcore Potential Form. With a flexible form of the weight functions in eq 6 described, we now turn to a presentation of a new functional form for the softcore potentials that provide the explicit λ-dependence in the potential energy terms in eq 4. In the current framework, softcore potentials are developed for both the LJ and nonbonded electrostatic interactions (i.e., the direct/real space component of the PME method). Hence, the main terms that are affected by the softcore potentials are those contained in the nonbonded interactions between the SC and (CC+I) regions, i.e., those terms present in $U_{nb}^{SC/(CC+I)}$. Formally, these terms can also be present in the internal energy of the SC region, if these terms are being "turned off" to form the "dummy state". In principle, the internal potential energy interactions in the dummy state are arbitrary so long as they are treated consistently in different legs of the thermodynamic cycle that are subtracted. However, in practice, choice of the interactions in the dummy state are important, and should be selected to minimize the volume of phase space needed to sample the dummy state while at the same time avoiding sampling traps (multiple distinct free energy basins separated by high barriers) that could lead to inconsistent results. In fact, the proper choice of potential energy interactions in the dummy state, together with well-established generalized ensemble methods such as Hamiltonian replica exchange,[51−53] can lead to powerful new alchemical enhanced sampling methods. These issues will be explored in the next paper in this series.[54] For the present paper, the dummy state was created by scaling (i.e., "turning off") electrostatic interactions, and in some cases also torsion angle and 1−4 LJ terms, but keeping other bonded and normal LJ terms in place. Hence, the new form of the softcore potential will affect mainly the $U_{nb}^{SC/(CC+I)}$ term, which, as results presented later in the paper will show, has a profound effect on the free energy estimates.

The LJ, Coulomb, and PME direct-space interactions for a set of interacting point particles i and j separated by a distance r_{ij} are given by

$$U_{LJ}(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{15}$$

$$U_{Coul}(r_{ij}) = \left(\frac{q_i q_j}{4\pi\varepsilon_0}\right)\frac{1}{r_{ij}} \tag{16}$$

and

$$U_{dir}(r_{ij}) = \left(\frac{q_i q_j}{4\pi\varepsilon_0}\right)\frac{\mathrm{erfc}(\kappa r_{ij})}{r_{ij}} \tag{17}$$

where σ_{ij} and ϵ_{ij} are the pairwise LJ contact distance and well depth, respectively, and q_i and q_j are the partial charges of particles i and j, respectively, erfc() is the complementary error function and κ is the Ewald coefficient.

To soften these pairwise particle−particle interactions, a parametric form for the separation-shifted scaling is used to modify the effective interaction distance. A commonly used form[23,24] is given by

$$r_{ij}^{LJ}(\lambda;\ \alpha) = [r_{ij}^n + \lambda\alpha\sigma_{ij}^n]^{1/n} \tag{18}$$

and

$$r_{ij}^{Coul}(\lambda;\ \beta) = [r_{ij}^m + \lambda\beta]^{1/m} \tag{19}$$

where n and m are positive integers and α and β are adjustable positive semidefinite parameters (note that α is unitless whereas β has units of distance raised to the power of m). The values of $n = 6$ and $m = 2$ are often used, and have been the default values in AMBER until the present work. We will demonstrate later in the manuscript how this softcore potential can lead to numerical instabilities, and discuss ongoing efforts and progress to improve the methods.

We introduce the following modified form of the separation-shifted scaling that leads to considerable improvement:

$$r_{ij}^{LJ*}(\lambda;\ \alpha^{LJ}) = [r_{ij}^n + \alpha^{LJ}\cdot f_{SW}(r_{ij})S_2(\lambda)\sigma_{ij}^n]^{1/n} \tag{20}$$

and

$$r_{ij}^{Coul*}(\lambda;\ \alpha^{Coul}) = [r_{ij}^m + \alpha^{Coul}\cdot f_{SW}(r_{ij})S_2(\lambda)\sigma_{ij}^m]^{1/m} \tag{21}$$

where α^{LJ} and α^{Coul} are the corresponding unitless parameters, S_2 is the second-order smoothstep function in eq 7, and $f_{SW}(r_{ij})$ is a switching function designed to smoothly return to the normal r_{ij}, and thus long-ranged behavior, by the end of the cutoff

$$f_{SW}(r_{ij}) \equiv 1 - S_2\left(\frac{r_{ij} - R_{cut,i}}{R_{cut,f} - R_{cut,i}}\right) \tag{22}$$

where $R_{cut,i}$ is the distance that the switching function begins switching and $R_{cut,f}$ is the final distance where the switching ends (returning the effective interaction distance to be r_{ij}). Henceforth, we will set $R_{cut,f} = R_{cut}$, and $R_{cut,i} = R_{cut} - 2$ Å, respectively.

The form of the new softcore potential is thus

$$U_{LJ}(r_{ij};\ \lambda) = U_{LJ}[r_{ij}^{LJ*}(\lambda;\ \alpha^{LJ})] \tag{23}$$

$$U_{Coul}(r_{ij};\ \lambda) = U_{Coul}[r_{ij}^{Coul*}(\lambda;\ \alpha^{Coul})] \tag{24}$$

and

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

$$U_{dir}(r_{ij};\ \lambda) = U_{dir}[r_{ij}^{Coul*}(\lambda;\ \alpha^{Coul})] \quad (25)$$

Most free energy simulations in the condensed phase are performed under periodic boundary conditions and use the PME method[55,56] to treat long-ranged electrostatic interactions, in which case the electrostatic softcore potentials described here apply to the U_{dir} term (eq 25). If, on the other hand, PME electrostatics is not used, then the electrostatic softcore potential apply to the U_{Coul} term (eq 24).

3. METHODS

Throughout the manuscript, we will explore alchemical free energy calculations using concerted transformation pathways. The new alchemical transformation pathway and softcore potential developed here will be compared to several previously developed methods, including the softcore potential method of Steinbrecher, Joung, and Case[24] with default parameters used in AMBER18, as well as modified parameters developed by Ebert and Labute,[57] and a modified smoothstep softcore potential[41] used in AMBER20. The present method builds on the latter, but deviates in functional form to include a universal pairwise interaction with consistent power scaling of Coulomb and Lennard-Jones interactions with unitless control parameters and rigorous smoothing of the potential at the nonbonded cutoff boundary. All simulations in the present work were performed with the pmemd.cuda module of AMBER Drug Discovery Boost package (AMBER DD Boost)[13] as a modified software patch to AMBER20 that now has been fully implemented and is available in AMBER22.[50]

3.1. Three Example Molecular Systems. We examine transformations involving three small molecules as test systems. The first test case (denoted "DPT/0") involves the vanishing transformation of 3,4-diphenyltoluene, a bulky hydrophobic molecule, into a dummy state. The second test case (denoted "Na^+/0") involves the vanishing transformation of Na^+, a small ion strongly interacting with its environment, into a dummy state. The third test case (denoted "L51c/h") involves the alchemical transformation between two factor Xa ligands, L51c → L51h in solution, which involves the migration of charged functional groups from one region of the ligand to another.[58] The hydration free energy simulations for DPT/0, Na^+/0, and L51c/h were modeled using the GAFF force field[11,59] with the AM1-BCC charges[60,61] and solvated with TIP3P[62] waters (DPT/0, Na^+/0, and L51c/h systems contained 2114, 6101, and 1511 water molecules, respectively). All the initial structures for gaseous simulations were prepared by stripping water from those equilibrated structures in the aqueous phase with a periodic box. Alchemical free energy calculations were performed for different softcore potential methods using single-step concerted transformations in a series of 21 alchemical states equally spaced along the λ dimension ranging from 0 to 1 ($\Delta\lambda$ = 0.05). The system in each transformation was first energy minimized with 1000 steps in which the steepest descent method was used, then the initial conformations for each λ window (total 21 windows) were sequentially generated with 5 ps pre-equilibration with NVT ensemble. Each window was run in the NVT ensemble at 298 K through a Langevin thermostat with a friction constant of 5.0 ps^{-1} for 5.2 ns with the first 200 ps discarded prior to analysis, in order to get 5 ns of production sampling. The long-range electrostatics were evaluated with the particle mesh Ewald (PME) method.[55,56] A cutoff of 10 Å was used for nonbonded interactions, including the direct space PME terms and particles interacting through softcore potentials. All simulations were performed using a 1 fs integration time step, and only the bonds and angle involving hydrogens of water molecules were constrained with the SHAKE algorithm.[63,64]

3.2. Relative Hydration and Binding Free Energy Simulations. We examine six possible transformations between four ligands (L66, L67, L74, and L75) that target binding to protein tyrosine phosphatase 1B (PTP1B)[65,66] with softcore regions for each transformation selected using a variant of the maximum common substructure search algorithm[67] as implemented in the Cheminformatics software RDKit.[68] Specifically, an "extended" MCS atom-mapping algorithm we developed is used that builds on the original MCS algorithm and excludes from the "maximum overlap" region atoms that differ either in chemical identity or hybridization, and further enforces the condition that for each ligand, the same softcore and common core regions are used consistently for every transformation in the network involving that ligand. We refer to this atom-mapping procedures as MCS-E_{nw}. Initial structures were taken from the published data[66] and simulations were prepared using the AMBER ff14SB,[69] GAFF2[70] force fields. Ligands and ligand−protein complexes were solvated using TIP4P-Ew[71] water and an initial buffer size of 20 and 16 Å, respectively. Any remaining net charge of the system was first neutralized and then solvated as 0.15 M ion concentration by the addition of Na^+ or Cl^- as appropriate. A minimization with Cartesian restraints relative to the starting structure was applied to all ligand and protein heavy atoms with a force constant of 5 kcal/mol/$Å^2$, and followed by another minimization without any restraint at λ = 0. After two steps of minimization (5000 steps of steepest descent for each minimization), two short 5 ps equilibrations were performed and followed by 500 ps equilibration with NPT ensemble. The system then was heated at a fixed volume with 300 k. After the heating stage, a 500 ps equilibration with NPT ensemble was performed, followed by 2 ns annealing. The annealing was heated to 600 K in the first 50 ps, then stayed with 600 K for another 100 ps, and eventually cooled down to 300 K in the last 50 ps. After the annealing stage, the restraint on the ligand and protein heavy atoms was reduced to zero in five steps over 1 ns. After the above procedure was performed at λ = 0, the sequential minimization, equilibration, and heating were performed, which means the initial structure for each λ window was taken from the equilibrated structure of the last λ window. Each λ window was minimized with 5000 steps for which the steepest descent method was used, then two short 5 ps equilibration were performed followed by a 500 ps equilibration with NPT ensemble and heated to 300 K. Alchemical free energy calculations were performed for different softcore potential methods using single-step concerted transformations in a series of 21 alchemical states equally spaced along the λ dimension ranging from 0 to 1 ($\Delta\lambda$ = 0.05). Each window was run independently (with different initial velocities) 4 times in the NPT ensemble at 300 K through a Langevin thermostat with a friction constant of 2.0 ps^{-1} for 5 ns with the first half of data discarded prior to analysis. The long-range electrostatics were evaluated with the particle mesh Ewald (PME) method.[55,56] A cutoff of 10 Å was used for nonbonded interactions, including the direct space PME terms and particles interacting through softcore

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

potentials. Only the bonds involving hydrogen were constrained with the SHAKE algorithm[63,64] except the atoms of ligands, and all simulations were performed using a 1 fs integration time step.

3.3. Absolute Hydration Free Energy Simulations. We examine absolute hydration free energy calculations for four molecules taken or modified from FreeSolv database.[72] Initial structures were taken from FreeSolv,[72] and simulations were prepared using the AMBER ff14SB,[69] GAFF force field,[11,59] with the AM1-BCC charges.[60,61] The systems were solvated with TIP3P[62] water and an initial buffer size of 20 Å. The equilibration procedure was in the same way as for relative binding free energy simulations. Alchemical free energy calculations were performed for different softcore potential methods using concerted scheme in a series of 21 alchemical states equally spaced along the λ dimension ranging from 0 to 1 ($\Delta\lambda$ = 0.05). Each window was run independently (with different initial velocities) 4 times in the NPT ensemble at 300 K through Langevin thermostat with a friction constant of 2.0 ps^{-1} for 2.7 ns with the first 0.2 ns of data discarded prior to analysis. The long-range electrostatics were evaluated with the particle mesh Ewald (PME) method.[55,56] A cutoff of 10 Å was used for nonbonded interactions, including the direct space PME terms and particles interacting through softcore potentials. Only the bonds involving hydrogen were constrained with the SHAKE algorithm[63,64] except the atoms of ligands, and all simulations were performed using a 1 fs integration time step.

3.4. Benchmark Reference Calculations Using Alchemical Enhanced Sampling (ACES). To generate the benchmark reference number for test systems, we introduced the alchemical enhanced sampling method (ACES) that will be described in detail elsewhere.[13,54] Initial structures and the equilibration procedure were performed in the same way as for relative binding free energy simulations, but were repeated in 16 independent runs and with nonuniform λ scheduling which led to 25 λ windows (0, 0.176834, 0.229764, 0.269379, 0.302697, 0.33229, 0.359436, 0.384886, 0.40913, 0.432518, 0.455318, 0.477748, 0.5, 0.522252, 0.544682, 0.567482, 0.59087, 0.615114, 0.640564, 0.66771, 0.697303, 0.730621, 0.770236, 0.823166, 1). These 25 λ points correspond to uniform spacing of the $S_2(\lambda)$ smoothstep function (eq 7) that is used for the nonlinear mixing weights in eq 6. As the $S_2(\lambda)$ function has vanishing gradients at the λ = 0 and 1 end-points, this schedule has the greatest density of points at λ = 0.5 rather than at the end-points. Optimal λ spacing for the new alchemical transformation pathway will be explored in more detail in future work. The benchmark reference calculations were sampled with 4 times as many independent trials as for the other production calculations (16 as opposed to 4 independent trials). There are two main requirements to perform efficient ACES calculations in AMBER22. The first requirement is that the targeted conformational barriers are reduced or eliminated in the "dummy state", which can be achieved by the use of the "gti_add_sc" control flag equal to 5. The second requirement is that the conformations in the "dummy state" need to be efficiently propagated to the real state end point, which can be achieved by using the Hamiltonian replica exchange[73−77] framework in AMBER. Alchemical free energy calculations were performed using concerted transformations in 25 alchemical states. Each window was run in NPT ensemble at 300 K through the Langevin thermostat with a friction constant of 2.0 ps^{-1} for 5 ns with the first half of data discarded prior to analysis. The long-range electrostatics were evaluated with the particle mesh Ewald (PME) method.[55,56] A cutoff of 10 Å was used for nonbonded interactions, including the direct space PME terms and particles interacting through softcore potentials. Only the bonds involving hydrogen were constrained with the SHAKE algorithm[63,64] except the atoms of ligands, and all simulations were performed using a 1 fs integration time step.

3.5. Simulations to Examine Stability and Energy Conservation. We examine the alchemical transformation from Mg^{2+} to Ca^{2+} in aqueous solution. The system was prepared using the AMBER ff14SB,[69] GAFF force field,[11,59] and solvated using TIP3P[62] waters with initial buffer size of 20 Å. Any remaining net charge of the system was first neutralized and then solvated as 0.15 M ion concentration by addition of Na^+ or Cl^- as appropriate. Alchemical free energy calculations were performed for different softcore potential methods using single-step concerted transformations in a series of 21 alchemical states equally spaced along the λ dimension ranging from 0 to 1 ($\Delta\lambda$ = 0.05). The system for each λ window was first energy minimized with 1000 steps which the steepest descent method was used, then followed by 5 ps heating stage. After the heating, the 2 ns pre-equilibration with NPT ensemble at 298 K through Langevin thermostat with a friction constant of 5.0 ps^{-1} was performed. Each window was run in the NVE ensemble at 298 K for 5.2 ns with the first 0.2 ns discarded prior to analysis to get 5 ns of production sampling. A cutoff of 10 Å was used for nonbonded interactions, including the direct space PME terms and particles interacting through softcore potentials. All simulations were performed using a 1 fs integration time step, and only the bonds and angle involving hydrogens of water molecules were constrained with the SHAKE algorithm.[63,64] The SHAKE tolerance was set to 10^{-8}.

4. RESULTS AND DISCUSSION

We present results for the development and validation of a new framework for improved alchemical transformation pathways in AMBER, provide discussion in the context of comparisons with other existing methods, and identify new directions for further advance. The next section (section 4.1) uses simple 2-particle models to illustrate examples of the end point catastrophe, particle collapse, and large gradient-jumps that the new softcore potential developed in section 4.2 is designed to address. Section 4.3 compares the methods developed in the current work to several established methods in the literature using three example molecular systems in solution. Section 4.4 provides validation tests against RSFE and RBFE calculations for a set of 4 ligands that target the protein tyrosine phosphatase 1B (PTP1B), and another validation tests against ASFE calculations for 4 small molecules that represent more stringent edge cases, with focus on comparison to highly converged benchmark reference calculations and theoretical cycle closure conditions. Section 4.5 ends the results and discussion with an analysis of energy fluctuations and conservation in the simulations using different alchemical transformation pathways. Finally, the manuscript concludes with a summary of key results, and outlook for the future.

Notation for specifying alternative alchemical transformation pathways. Through the remainder of the manuscript, we will make comparisons to several alternative alchemical transformation pathways (eq 2) that differ by their λ-dependent weight functions (section 2.2) and softcore potentials (section

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

Table 3. Abbreviations and Associated Forms of the Alchemical Transformation Pathways[a]

abbreviation	$W(\lambda)$	n	m	α/α^{LJ}	β/α^{Coul}	eqs	ref	comment
S0 Linear	S0							Linear mixing/no softcore
S0[6, 2, 0.5, 12]	S0	6	2	0.5	12	18/19	24	Default in AMBER18
S0[6, 2, 0.2, 17.3]	S0	6	2	0.2	17.3	18/19	57	Modified for improved stability
S2[6, 2, 0.2, 50]	S2	6	2	0.2	50	18/19	41	Default in AMBER20
S2*[2, 2, 0.5, 1]	S2	2	2	0.5	1.0	20/21		Present work, new form

[a]Indicated are the form of the dependent weight functions (in terms of smoothstep function SX, where X indicates the order of the smoothstep function); and n, m, α, and β parameters of eqs 18 and 19 or n, m, α^{LJ}, and α^{Coul} parameters of eqs 20 and 21. All α parameters are unitless, whereas the β parameter has units of Å^2. The S2*[2,2,0.5,1] softcore potential differs in functional form from the others, using eqs 20/21 for the separation-shifted scaling of the effective interaction distance with unitless α^{LJ} and α^{Coul} parameters rather than the more conventional form shown in eqs 18 and 19.

Table 4. Summary of Parameters Used in Transformations of Illustrative 2-Body Model Systems (Taken from Joung and Cheatham Ion Monovalent Parameters[78] and TIP3P Water[62])[a]

	λ = 0 state				→	λ = 1 state			
Transformation	$Q_i \cdot Q_j$	σ_{ij}	ε_{ij}	k		$Q_i \cdot Q_j$	σ_{ij}	ε_{ij}	k
$Na^+ \rightarrow 0$	−0.834	2.80135	0.16006	1.0	→				1.0
$Li^+ \rightarrow Cs^+$	−1.0	3.568	0.03482		→	−1.0	4.648	0.06782	
R → 0	−0.0834	5.0	1.0	1.0	→				1.0

[a]The 1-body parameters (Q in units of $|e|$, σ in units of Å, and ε in units of kcal/mol) for atoms are Cs^+ (Q = 1.0, σ = 1.888, ε = 0.3944318); Li^+ (Q = 1.0, σ = 0.808, ε = 0.103984); Cl^- (Q = −1.0, σ = 2.760, ε = 0.0116615); Na^+ (Q = 1.0, σ = 1.226, ε = 0.1684375); OW (Q = −0.834, σ = 1.57535, ε = 0.1521). The hydrophobic R group model was chosen as a LJ sphere of σ_{ij} = 5.0 Å and ε_{ij} = 1.0 kcal/mol, and Coulomb interaction between an OW atom and an R group charge Q = 0.1$|e|$. For the annihilation models, an additional quadratic "surface tension" term of the form $(1/2)k \cdot r^2$ was added (force constant k in units of kcal/mol·Å^2) to capture the effect that a real solution would resist formation of an empty cavity. Particles were constrained to remain with a 10 Å range.

2.3). The notation will take the general form: SX[n, m, α, β], where "X" indicates the order of the smoothstep function used in the λ-dependent weight function (i.e., mixing term). Specifically, S0 indicates linear (i.e., zeroth-order smoothstep function) and S2 indicates a second-order smoothstep weight functions/mixing terms, respectively. The integers n and m are positive integers and α and β are adjustable positive semidefinite parameters that control the behavior of the separation-shifted scaling of the effective interaction distance for the Lennard-Jones and Coulomb interactions in eqs 18 and 19. In the present work, we introduce a new functional form for the softcore potential that uses separation-shifted scaling of eqs 20 and 21 that contain unitless α^{LJ} and α^{Coul} parameters, and we will distinguish this softcore potential form by the label "S2*". The 5 alchemical transformation pathways compared in the current work are thus designated as S0 Linear, S0[6,2,0.5,12], S0[6,2,0.2,17.3], S2[6,2,0.2,50] and S2*-[2,2,0.5,1] (present work), and are summarized in Table 3.

4.1. End Point Catastrophes, Particle Collapses, and Large Gradient-Jumps. We examine three problems that commonly occur in alchemical simulations with concerted transformations that involve simultaneous changes in both nonbonded Lennard-Jones and Coulombic electrostatic interactions: end point catastrophes, particle collapses, and large gradient-jumps. These problems have been discussed elsewhere.[13,41] The *end point catastrophe* is the sharp divergence of the free energy that is prone to occur at the thermodynamic end points (λ becomes close to 0 or 1) and can largely be avoided by use of appropriate softcore potentials.[23,24] However, under certain circumstances, the use of softcore potentials can lead to large amplitude fluctuations or phase transition behavior along the λ dimension and result in new artificial minimum at intermediate λ states[29] due to an imbalance of Coulomb attraction and exchange repulsion.[24] This is referred to as the *particle collapse* problem, and can be treated by increasing the softcore exchange repulsions and/or decreasing the Coulomb attractions at short distances across the range of λ values to correct for unstable imbalance. Nonetheless, this treatment for particle collapses can lead to *large gradient-jumps*, i.e., large amplitude fluctuations of the thermodynamic derivatives that are sensitive to the parameters that control the balance of Coulomb attraction and exchange repulsion.[41] Note that while the behavior of *end point catastrophes* and *large gradient-jumps* might appear similar, their origins are different. The *end point catastrophes* are due to poor phase space overlap and often occur with linear alchemical transformations, and usually can be corrected with the use of softcore potentials. The *large gradient-jumps* can occur even with the use of softcore potentials, and often happen when large β values are required to adjust the softcore parameter ratio to solve the Coulomb-exchange imbalance problem. As will be seen later, the proposed nonlinear mixing and new functional form of the softcore potential alleviates all of these problems.

In order to provide simple illustrative examples, we examine transformations involving 2-body model systems for which accurate 1D profiles can be computed (integrated) numerically. Specifically, we consider cases of the annihilation of a Na^+ ion interacting with a TIP3P water oxygen, the transformation of a Li^+/Cl^- ion pair into a Cs^+/Cl^- ion pair, and the annihilation of a large spherical hydrophobic R group interacting with a TIP3P water oxygen. These are designated as "$Na^+ \rightarrow 0$", "$Li^+ \rightarrow Cs^+$", and "R → 0", respectively. Interactions were constrained to remain with a 10 Å interaction distance, and the interaction parameters for each model are summarized in Table 4.

For the annihilation models, an additional quadratic "surface tension" term $(1/2)k \cdot r^2$ is added (k = 1.0 kcal/mol/Å^2) to capture the effect that a real solution opposes formation of an empty cavity. Note that the purely heuristic value of k chosen

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

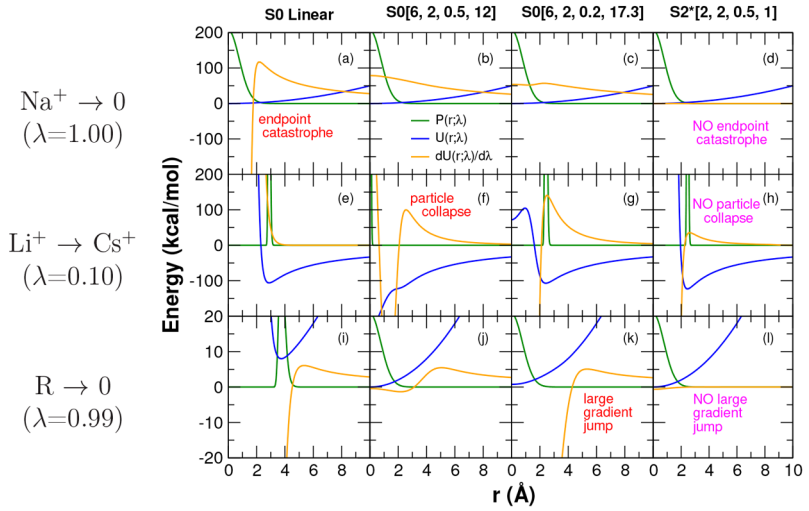


Figure 3. Illustrative examples of end point catastrophes, particle collapses, and large gradient-jumps in transformations of the 2-particle model systems summarized in Table 4. The 2-particle nature of the models facilitates comparison using 1D plots as a function of interparticle separation r. Shown are the potential energy $U(r; \lambda)$ (blue), derivative $dU(r; \lambda)/d\lambda$ (gold), and Boltzmann probability $P(r; \lambda)$ (green) scaled so as to be easier to view on the same plot. For the annihilation models ($Na^+ \rightarrow 0$ and $R \rightarrow 0$), an additional quadratic "surface tension" term was added (see Table 4). This surface tension term is meant to capture the effect that a real solution would resist formation of an empty cavity, and acts as a weak confining potential to keep particles in close proximity in the limit their interactions become very weak or vanish. Particles were further constrained to remain with a 10 Å range. These 2-particle models were designed to qualitatively mimic the behavior of the three example alchemical free energy simulations presented later in the discussion. *Top row*: Annihilation of Na^+ ion (interacting with a TIP3P water oxygen) at $\lambda = 1.00$ illustrating end point catastrophe. At $\lambda = 1.00$, the Na^+ Coulomb and LJ interactions vanish, and only the weak confining potential remains and is shown as the blue curve, as in this case it is the same as $U(r; 1)$. In a real condensed phase simulation, the annihilated Na^+ would be immersed in the water bath and unable to avoid passing on top of the many space-filling water molecules so as to produce an end point catastrophe. In this simple illustrative 2-particle model, the surface tension term produces the same qualitative effect. *Middle row*: $Li^+ \rightarrow Cs^+$ transformation with the ions interacting with a TIP3P water oxygen at $\lambda = 0.10$ illustrating particle collapse. *Bottom row*: Annihilation of large R group (interacting with a TIP3P water oxygen) at $\lambda = 0.99$ illustrating large gradient-jump. The R group is modeled as a 5.0-Å LJ sphere with partial charge of 0.1|*e*|, and also uses the same surface tension term as for the annihilation of Na^+ above. Note: the scale of the *y*-axis differs for this model by a factor of 10 relative to the ion models above.

for illustration here for 2-particle systems is much different than those used in conjunction with implicit solvent models of many-particle systems that are typically 2 orders of magnitude smaller.[79] As these are 2-body interaction models, this allows a simple 1D representation of the λ-dependent potential energy, $U(r; \lambda)$, and its derivative with respect to λ, $dU(r; \lambda)/d\lambda$, as well as the Boltzmann probability distribution, $P(r; \lambda) \sim \exp[-\beta U(r; \lambda)]$ at $T = 298.15$ K. When the probability distribution is examined, its sensitivity to λ and overlap with $dU(r; \lambda)/d\lambda$ provide insight into the stability of the transformation pathway and its integration. The illustrative examples are shown in Figure 3, and they examine specific λ states where problems are observed to occur and compare results for three transformation pathways (see Table 3): S0 Linear, or "linear mixing" (no softcore potential), traditional softcore potential S0[6,2,0.5,12],[24] and softcore potential with updated/modified parameters S0[6,2,0.2,17.3].[57] Also shown for completeness is the new softcore potential and nonlinear mixing scheme, designated S2*[2,2,0.5,1], that will be developed below in the sections that follow.

The end point catastrophe is illustrated in Figure 3(a) for the $Na^+ \rightarrow 0$ transformation at $\lambda = 1$ using the linear mixing scheme. In the linear mixing scheme, the $dU(r; \lambda)/d\lambda$ is independent of λ (although it does depend on r, it is the same function of r for all λ values), and blows up at the origin ($r = 0$); however, at the $\lambda = 1$ end point (with the Na^+ fully annihilated), the probability of OW is at a maximum at the origin, causing the $\langle \partial U/\partial \lambda \rangle_\lambda$ to diverge. This behavior is easily handled by the use of a softcore potential that results in a $dU(r; \lambda)/d\lambda$ profile that stably approaches a constant value at $r = 0$, as shown in Figure 3(b,c).

However, the use of softcore potentials can also become problematic under different circumstances, as illustrated by the $Li^+ \rightarrow Cs^+$ transformation at $\lambda = 0.1$ in Figure 3(e,f,g). Unlike the linear mixing scheme shown in Figure 3(e), the S0[6,2,0.5,12] softcore potential forms an intermediate λ potential that produces a deep artificial minima at $r = 0$ causing the Cl^- probability to collapse on the Li^+/Cs^+ transforming particle (Figure 3f). This produces a spike in $\langle \partial U/\partial \lambda \rangle_\lambda$. The origin of this particle collapse (which is exhibited over the range $0.1 \leq \lambda \leq 0.9$) is that the Coulomb attraction is able to overcome the exchange-repulsion at these intermediate λ points. Adjustment of the relative α and β parameters so as to make the Lennard-Jones exchange-repulsion "harder' (i.e., soften more slowly) by lowering the α value from 0.5 to 0.2, while at the same time making the Coulomb attraction term "softer" by raising the β value from 12 to 17.3 $Å^2$ leads to the S0[6,2,0.2,17.3] curve in Figure 3(g), which does not exhibit particle collapse. As will be shown later, this softcore potential does not guarantee elimination of particle collapse in all cases, but does reduce the occurrence in practice.

Unfortunately, the rebalancing of the exchange and Coulomb terms can lead to other adverse effects, as illustrated in Figure 3(i,j,k) for the annihilation of the bulky R group. This transformation is dominated by the exchange-repulsion term, which is sensitive to the α parameter, especially with softcore potentials that have a large value of $n = 6$ in eq 18. Whereas the S0[6,2,0.5,12] softcore potential is reasonably

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

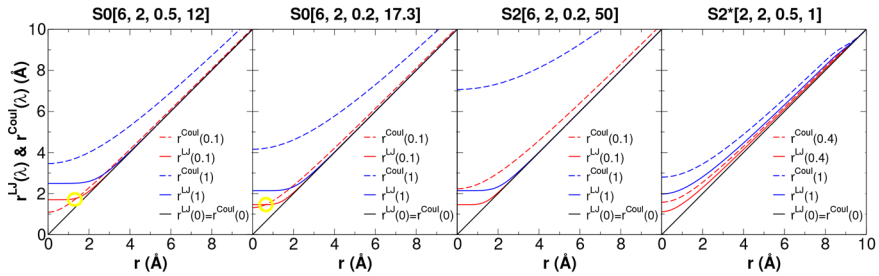


Figure 4. Separation-shifted scaling, $r^{Coul}(\lambda)$ and $r^{LJ}(\lambda)$, of Coulomb/LJ nonbonded interactions used in eqs 18 and 19 and eqs 20 and 21 shown for each of the softcore potentials summarized in Table 3. In addition to the end points (λ = 0, 1), an intermediate value of λ is selected to illustrate if/when the curves cross as indicated by a yellow circle. For the new S2*[2,2,0.5,1] softcore potential, the $r^{Coul}(\lambda)/r^{LJ}(\lambda)$ are nonintersecting over the entire range of λ [0,1].

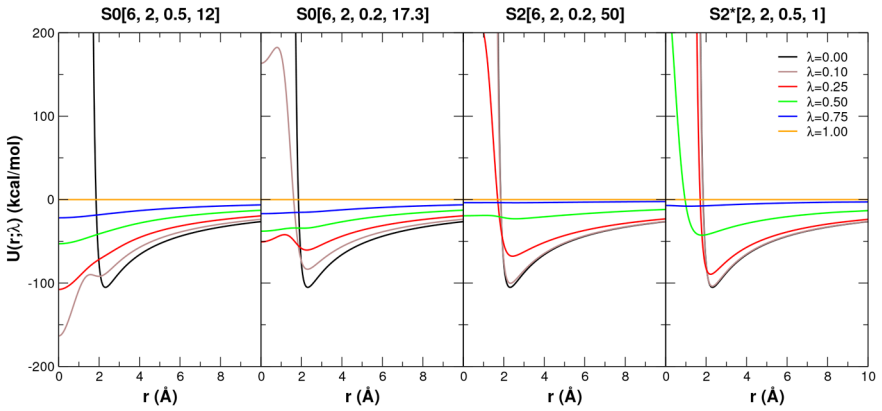


Figure 5. $U(r; \lambda)$ potential energy curve for the interaction of a Na^+ ion with a TIP3P water oxygen as a function of separation distance r for several values of λ in the $Na^+ \rightarrow 0$ transformation (Table 4). $U(r; \lambda)$ curves are shown for the alchemical transformation pathways/softcore potentials summarized in Table 3.

well-behaved near the end point at λ = 0.99 (Figure 3j), the decrease in the α parameter for the S0[6,2,0.2,17.3] softcore potential leads to large values of the $dU(r; \lambda)/d\lambda$ for small r and a resulting sharp increase in the magnitude of $\langle \partial U/\partial \lambda \rangle_\lambda$ as $\lambda \rightarrow 1$ (Figure 3k). This is known as a large gradient-jump, and although less severe, has similarities to the end point catastrophe that can occur with the S0 Linear transformation (Figure 3i) despite that the S0[6,2,0.2,17.3] transformation uses a softcore potential.

In this way, the form of the traditional S0 alchemical transformations cannot simultaneously eliminate the end point catastrophe, particle collapse, and large gradient-jump problems. Recent work to extend the alchemical transformation to use a smoothstep softcore potential,[41] S2-[6,2,0.2,50], further improves the ability to handle these problems, but does not eliminate them for some edge cases. As is illustrated in Figure 3(d,h,l) and will be shown more extensively below, the new form of the alchemical transformation pathway S2*[2,2,0.5,1] developed in the sections that follow has been designed to overcome these and other issues in a highly robust fashion, and has resolved the instability issues of all edge cases encountered thus far for the previous smoothstep softcore potential.[41] Nonetheless, there will likely be cases that are found where the new alchemical transformation is not well-suited, and recourse must be taken into more conservative approaches that involve λ-scheduling or stepwise decoupling of electrostatic and LJ terms, possibly with additional restraints.

4.2. New Smoothstep Softcore Potential with Balanced Coulomb and Exchange. We recently reported a new form of softcore potential that utilized a smoothstep function to enable stabilization of alchemical transformations,[41] while otherwise maintaining the functional form of the more traditional softcore potential of eqs 18 and 19. In the current work, we revise the functional form to introduce a new form of the softcore potential in eqs 20 and 21. Two features of the new functional form are (1) a pairwise σ_{ij} term is introduced to both Coulomb and LJ terms and is controlled by unitless α parameters, and (2) smooth switching of $r_{ij}^{LJ*}(\lambda; \alpha^{LJ})$ and $r_{ij}^{Coul*}(\lambda; \alpha^{Coul})$ at the cutoff boundary is introduced so as to eliminate discontinuities of the energy and forces and improve the stability of simulations.

One of the main considerations was to introduce a balanced scaling of the Coulomb and LJ exchange terms throughout the range of λ [0 to 1]. In this way, if the real particle−particle interaction potential at λ = 0, before scaling, is able to prevent the particles from collapsing onto one another, then by ensuring that the separation-shifted scaling of the exchange interaction is less than or equal to that of the Coulomb interaction, collapse at small r values can generally be prevented. This implies the condition:

$$r_{ij}^{Coul}(\lambda) > r_{ij}^{LJ}(\lambda) \quad \forall \ \lambda \in [0, 1] \tag{26}$$

One of the issues with the original softcore potential was that both the functional form in eqs 18 and 19 and the n and m scaling powers were different, making it the case that at some

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

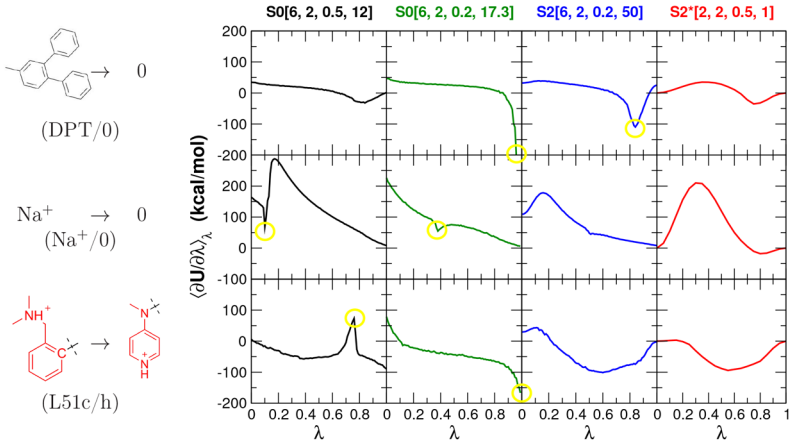


Figure 6. $\langle \partial U/\partial \lambda \rangle_\lambda$ profiles for alchemical free energy simulations of three molecular systems in aqueous solution using the concerted scheme and the alchemical transformation pathways summarized in Table 3: the annihilation of diphenyltoluene (DPT/0, upper rows), the annihilation of Na^+ ion (Na^+/0, middle rows), and the transformation between Factor Xa ligand[80] L51c to L51h (L51c/h, bottom rows). The L51c ligand has 65 atoms and L51h has 58 atoms, and the red-colored atoms shown are the defined softcore regions, whereas the atoms common to both ligands are not shown except the connecting carbon shown in black. Unstable points are highlighted by yellow circles and are discussed in the text.

intermediate λ values the $r_{ij}^{\mathrm{Coul}}(\lambda)$ and $r_{ij}^{\mathrm{LJ}}(\lambda)$ curves would cross (Figure 4), allowing the possibility that Coulomb attraction could overcome the exchange repulsion. In the new softcore potential developed in this work, the functional form of the separation-shifted scaling for LJ and Coulomb interactions of eqs 20 and 21 are the same when $n = m$. For this case, the only requirement is that

$$r_{ij}^{\mathrm{Coul}*}(\lambda;\, \alpha^{\mathrm{Coul}}) > r_{ij}^{\mathrm{LJ}*}(\lambda;\, \alpha^{\mathrm{LJ}})$$
$$\forall\ \lambda \in [0,\, 1] \text{ if } \alpha^{\mathrm{Coul}} > \alpha^{\mathrm{LJ}} \text{ for } n = m \qquad (27)$$

We experimented extensively with several values of $n = m$ and combinations of the unitless α^{LJ} and α^{Coul} parameters, and found that a very robust balance was achieved with $n = m = 2$ and $\alpha^{\mathrm{LJ}} = 0.5$ and $\alpha^{\mathrm{Coul}} = 1.0$ (see Figure S1 and Tables S1−S3 of the Supporting Information for comparisons). For these values of the parameters, $r^{\mathrm{Coul}*}(\lambda) > r^{\mathrm{LJ}*}(\lambda)$ (and are nonintersecting) over the entire range of λ [0,1] (Figure 4, rightmost panel). We designate this softcore potential S2*[2,2,0.5,1].

Figure 5 illustrates the potential energy curves for the interaction of a Na^+ ion with a TIP3P water oxygen as a function of separation distance r for several values of λ in the $Na^+ \rightarrow 0$ transformation (Table 4). With the S0[6,2,0.5,12] softcore potential, there are deep minima at the origin for even small λ values that exacerbate the particle collapse problem. With the modified S0[6,2,0.2,17.3] softcore potential, the potential is more repulsive for small values of λ, but still may have spurious minima that arise at the origin for $\lambda = 0.25$. The original smoothstep softcore potential S2[6,2,0.2,50] remains repulsive to slightly larger values of λ, but has behavior somewhat similar to the S0[6,2,0.2,17.3] potential, again producing spurious minima, which is due to the introduction of the softcore potential, at intermediate $\lambda = 0.5$ values and a deep minimum at $r = 0$. The new smoothstep softcore potential introduced in the current work, S2*[2,2,0.5,1], eliminates these spurious and deep minima at the origin, having a much more gradual reduction of the repulsive exchange interactions. In following sections we compare the behavior of the 4 softcore potentials in alchemical free energy simulations of increasing complexity using various assessment metrics discussed in the next section.

4.3. Comparison of $\langle \partial U/\partial \lambda \rangle_\lambda$ Profiles for Different Alchemical Transformation Pathways. Figure 6 $\langle \partial U/\partial \lambda \rangle_\lambda$ compares profiles for alchemical free energy simulations of three example molecular systems in aqueous solution using the concerted scheme and the alchemical transformation pathways summarized in Table 3: the annihilation of a bulky hydrophobic molecule (diphenyltoluene), the annihilation of a small ion (Na^+), and the transformation of a charged region in a large ligand (Factor Xa ligand[80] L51c to L51h). These are abbreviated DPT/0, Na^+/0, and L51c/h, respectively. These real systems loosely mirror, in a very qualitative sense, the simple 2-particle model systems used for illustration purposes in section 4.1.

Overall, the S0[6,2,0.5,12] pathway performs worst for the Na^+/0 transformation, since there is strong Coulomb attraction that overpowers the poorly balanced exchange interactions, similar to the particle collapse observed for the 2-particle $Li^+ \rightarrow Cs^+$ model system in Figure 3(f). The S0[6,2,0.2,17.3], on the other hand, better balances the Coulomb and exchange, but at the consequence of exhibiting large gradient-jumps for the DPT/0 transformation, similar to problems observed in the 2-particle R → 0 model system in Figure 3(k).

In addition, there are several instances where the $\langle \partial U/\partial \lambda \rangle_\lambda$ profiles exhibit instabilities at intermediate or end points, as indicated in Figure 6 by the yellow circles. These instabilities arise from intervals for which there is anomalously poor phase space overlap[19] and large standard error of the free energy estimate per λ-interval. An additional consideration for TI is the ability to accurately integrate the $\langle \partial U/\partial \lambda \rangle_\lambda$ profiles, and this can be estimated using quadrature stability analysis, and in particular the estimated quadrature RMSE index Q_{RMSE} that measures the sensitivity of the free energy to numerical integration using fewer discrete points (see Supporting Information for detailed discussion of quadrature stability). Specifically, Q_{RMSE} estimates the error of integrating the $\langle \partial U/\partial \lambda \rangle_\lambda$ profile using a cubic spline representation with roughly $1/e$ (~36.8%) fewer points. The larger is the value of Q_{RMSE}, the greater is the estimated thermodynamic integration error of the

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

Table 5. Examination of Standard Errors for TI Free Energy Estimates and Quadrature Stability Indexes for $\langle \partial U/\partial \lambda \rangle_\lambda$ Profiles for DPT/0, Na^+/0, and L51c/h Model Transformations[a]

	S0[6,2,0.5,12]	S0[6,2,0.2,17.3]	S2[6,2,0.2,50]	S2*[2,2,0.5,1]
Transformation	Std. Err/Q_{RMSE}	Std. Err/Q_{RMSE}	Std. Err/Q_{RMSE}	Std. Err/Q_{RMSE}
DPT/0	**0.06**/0.50	*0.14*/*2.44*	0.08/2.38	0.09/**0.44**
Na^+/0	0.05/7.33	*0.08*/*7.59*	**0.05**/2.29	0.07/**0.21**
L51c/h	*0.95*/*3.36*	0.18/2.01	**0.12**/2.17	0.18/**0.70**

[a]For each index, the smallest/largest value for a given transformation is highlighted in boldface/italics. All units are in kcal/mol.

$\langle \partial U/\partial \lambda \rangle_\lambda$ profile. Table 5 compares the TI standard errors and quadrature stability indexes for each transformation pathway (Table 3) for the DPT/0, Na^+/0, and L51c/h model transformations. In all cases except one, the standard error estimates were below 0.2 kcal/mol (the exception being the L51c/h transformation using S0[6,2,0.5,12] with std. err. of 0.95 kcal/mol). The quadrature stability indexes, Q_{RMSE}, on the other hand, with the new S2*[2,2,0.5,1] softcore potential/transformation pathway have universally the smallest values, ranging from 0.21 to 0.7 kcal/mol, whereas the other transformation pathways ranged from 0.5 to 7.6 kcal/mol. The smaller quadrature stability indexes for S2*[2,2,0.5,1] qualitatively reflect the generally improved smoothness of the $\langle \partial U/\partial \lambda \rangle_\lambda$ profiles in Figure 6. In the following section, the various transformation pathways will be examined in the context of relative solvation and binding free energies of ligand networks.

4.4. Validation against Solvation and Relative Binding Free Energies. In the previous sections, we analyzed in detail both 2-particle numerical models as well as a set of real model transformations in aqueous solution. Here we turn toward more complex systems and real-world applications. The goal here is not to evaluate the accuracy of the force field or to address challenging sampling issues (the examples were chosen so as to avoid such issues), but rather how the alchemical transformation pathway itself affects the accuracy of the free energy estimate. To achieve this goal, we will focus attention on comparisons to benchmark reference calculations to assess the reliability of different alchemical transformation pathways. We first establish a baseline by examining a well-studied set of ligand transformations in gaseous, aqueous, and complex environments for which the established methods typically do not fail, in order to establish that the new methods also perform well. Next we examine more challenging transformations that involve absolute hydration free energies of bulky hydrophobic, polar, and anionic/cationic systems where other methods are observed to break down.

Well-Studied Ligand Transformations in Gaseous, Aqueous and Protein Complex Environments. We first examine transformations between four ligands (L66, L67, L74, and L75) that target the protein tyrosine phosphatase 1B (PTP1B)[65,66] (Figure 7). A thermodynamic graph is constructed that involves all 6 possible alchemical transformations between ligands, producing 6 edges to the complete thermodynamic graph (Figure 7). We will consider transformations within this thermodynamic graph in three different environments: gas phase, aqueous solution, and complexed to PTP1B. This will enable cycles to be constructed for both RSFE and RBFE (Figure 1). In addition, we perform benchmark reference calculations using new alchemical advanced enhanced sampling techniques, as well as increased number of λ sampling windows and number of independent runs in order to increase the level of precision of the TI/BAR/MBAR free energy estimates (see Methods for details). Note that we report our benchmark reference data using the BAR analysis method as we find this to be generally the most robust protocol for the data set considered here, and avoids errors that arise from numerical quadrature with TI.

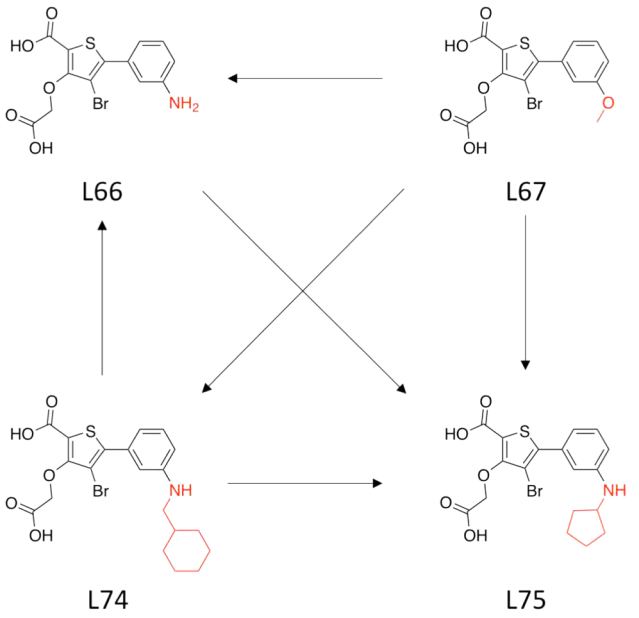


Figure 7. Illustration of a thermodynamic graph which involves 4 ligands (L66, L67, L74, and L75) that target the PTP1B.[65,66] The thermodynamic graph is constructed that involves all 6 possible alchemical transformations between ligands. The red-colored atoms shown are the defined softcore regions chosen in accord with the MCS-E_{nw} atom-mapping algorithm described in the Methods section.

Table 6 summarizes relative hydration free energy values from TI for 6 alchemical transformations between ligands. All of the error estimates are quite small with respect to the reference calculations (less than 0.2 kcal/mol), and all except for the S0[6,2,0.2,17.3] pathway have errors below 0.1 kcal/mol. The situation is overall similar for the relative binding free energy values shown in Table 7, with TI error estimates all below 0.2 kcal/mol. An independent assessment of reliability of the calculations can be ascertained through analysis of theoretical cycle closure conditions that the exact results must strictly obey. The cycle closure errors for the RSFE calculations using TI are around 0.1 kcal/mol or less for all methods with the exception of S0[6,2,0.2,17.3] which has TI cycle closure errors greater than 0.3 kcal/mol. The corresponding cycle closure errors for the RBFE calculations using TI are generally less than 0.3 kcal/mol, with the S2[6,2,0.2,50] being particularly small (less than 0.1 kcal/mol). In general, the cycle closure errors are even smaller with BAR and MBAR analysis (see Tables S4 and S5 of the Supporting Information for further details). Table 8 compares the TI standard errors

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

Table 6. Relative Hydration Free Energy Values (kcal/mol)[a] for Four Alchemical Transformation Pathways/Softcore Potentials

Transformation	S0[6,2,0.5,12]	S0[6,2,0.2,17.3]	S2[6,2,0.2,50]	S2*[2,2,0.5,1]	ref
L66 → L75	2.34(07)	2.20(08)	2.40(07)	2.39(06)	2.31(04)
L67 → L66	−8.17(07)	−8.08(06)	−8.07(07)	−7.98(08)	−8.00(03)
L67 → L74	−6.29(07)	−6.49(16)	−6.21(12)	−6.28(10)	−6.22(05)
L67 → L75	−5.74(08)	−5.55(12)	−5.80(07)	−5.59(10)	−5.70(04)
L74 → L66	−1.72(06)	−1.56(14)	−1.76(10)	−1.87(10)	−1.72(05)
L74 → L75	0.51(11)	0.35(19)	0.55(15)	0.59(13)	0.59(05)
MUE	0.07	0.17	0.06	0.06	
RMSE	0.08	0.17	0.06	0.08	

[a]The relative hydration free energy values are obtained by concerted scheme and analyzed using the TI method. Standard errors are shown in parentheses (multiplied by 10^2). Accurate reference values (ref) were obtained from ACES calculations, sampled with 16 independent trials (as opposed to 4 for other calculations) and 25 λ windows (as opposed to 21 for other calculations), and analyzed with BAR method (see Methods section 3.4 for details).

Table 7. Relative Binding Free Energy Values (kcal/mol)[a] for Four Alchemical Transformation Pathways/Softcore Potentials

Transformation	S0[6,2,0.5,12]	S0[6,2,0.2,17.3]	S2[6,2,0.2,50]	S2*[2,2,0.5,1]	ref
L66 → L75	−0.72(09)	−0.77(12)	−0.82(09)	−0.69(17)	−0.84(05)
L67 → L66	0.83(05)	0.75(06)	0.78(05)	0.72(05)	0.74(03)
L67 → L74	−0.24(26)	−0.17(18)	−0.04(12)	0.15(10)	−0.08(06)
L67 → L75	−0.32(28)	−0.17(14)	−0.12(09)	−0.05(13)	−0.06(05)
L74 → L66	0.77(08)	0.84(19)	0.92(14)	0.81(16)	0.88(06)
L74 → L75	0.19(15)	−0.27(51)	0.01(16)	0.21(14)	0.01(06)
MUE	0.15	0.10	0.03	0.11	
RMSE	0.15	0.12	0.04	0.13	

[a]Relative binding free energy values are obtained by concerted scheme and analyzed using the TI method. Standard errors are shown in parentheses (multiplied by 10^2). Accurate reference values (ref) were obtained from ACES calculations, sampled with 16 independent trials (as opposed to 4 for other calculations) and 25 λ windows (as opposed to 21 for other calculations), and analyzed with BAR method (see Methods section 3.4 for details).

Table 8. Examination of Standard Errors for TI Free Energy Estimates and Quadrature Stability Indexes for $\langle \partial U/\partial\lambda \rangle_\lambda$ Profiles for PTP1B Model Transformations in the Complex[a]

	S0[6,2,0.5,12]	S0[6,2,0.2,17.3]	S2[6,2,0.2,50]	S2*[2,2,0.5,1]
Transformation	Std. Err/Q_{RMSE}	Std. Err/Q_{RMSE}	Std. Err/Q_{RMSE}	Std. Err/Q_{RMSE}
L66 → L75	0.07/0.42	0.10/0.83	**0.06**/*0.94*	*0.16*/**0.13**
L67 → L66	0.04/0.34	*0.05*/*1.03*	0.04/0.68	**0.04**/**0.06**
L67 → L74	*0.25*/**0.21**	0.13/0.99	0.09/*1.16*	**0.07**/0.25
L67 → L75	*0.28*/0.25	0.10/*1.40*	**0.08**/0.87	0.12/**0.08**
L74 → L66	**0.07**/0.47	0.13/1.11	0.11/*1.21*	*0.14*/**0.38**
L74 → L75	0.11/0.25	*0.48*/*5.17*	0.11/1.62	**0.08**/**0.14**

[a]For each index, the smallest/largest value for a given transformation is highlighted in boldface/italics. All units are in kcal/mol.

and quadrature stability indexes for each transformation pathway for the 6 transformations. In most cases, the standard error estimates are below 0.2 kcal/mol, except the L67 → L74 and L67 → L75 using the S0[6,2,0.5,12] pathway. The S2*[2,2,0.5,1] pathway generally has the smallest quadrature stability indexes, in particular Q_{RMSE}, ranging from 0.06 to 0.38 kcal/mol, whereas the other alchemical transformation pathways ranged from 0.21 to 5.17 kcal/mol. Overall, the new S2*[2,2,0.5,1] performs very well and appears to be stable and robust for these transformations.

Absolute Hydration Free Energies of Bulky Hydrophobic, Polar, And Anionic/Cationic Systems. Next, we examine the absolute hydration free energies for phenanthrene, 7-cyclopentanylindole (7-CPI), phenoxide ion, and anilinium ion. These systems represent more stringent edge cases that combine large steric annihilation with neutral, negative, and positive charged systems. Similarly to the PTP1B ligands, we also perform benchmark reference calculations in order to make error estimates using new alchemical enhanced sampling techniques, as well as increased number of λ sampling windows and number of independent trials.

Figure 8 compares $\langle \partial U/\partial\lambda \rangle_\lambda$ profiles for alchemical free energy simulations of these four molecular systems in solution using the concerted scheme and the alchemical transformation pathways summarized in Table 3. Overall, the S0[6,2,0.5,12] pathway performs worst for the annihilation of phenoxide ion, similar to the particle collapse observed for the 2-particle $Li^+ \rightarrow Cs^+$ model system in Figure 3(f). Specifically, for the phenoxide ion, the profile is not smooth for λ values around 0.25. The origin of this issue is the particle collapse problem, where the softened exchange repulsions cannot balance the attractive softcore Coulomb attractions of the charged particles, causing them to collapse on top of one another. The S0[6,2,0.2,17.3] causes large gradient-jumps particularly for the larger phenanthrene and 7-cyclopentanylindole transformations, similar to problems observed in the 2-particle R →

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

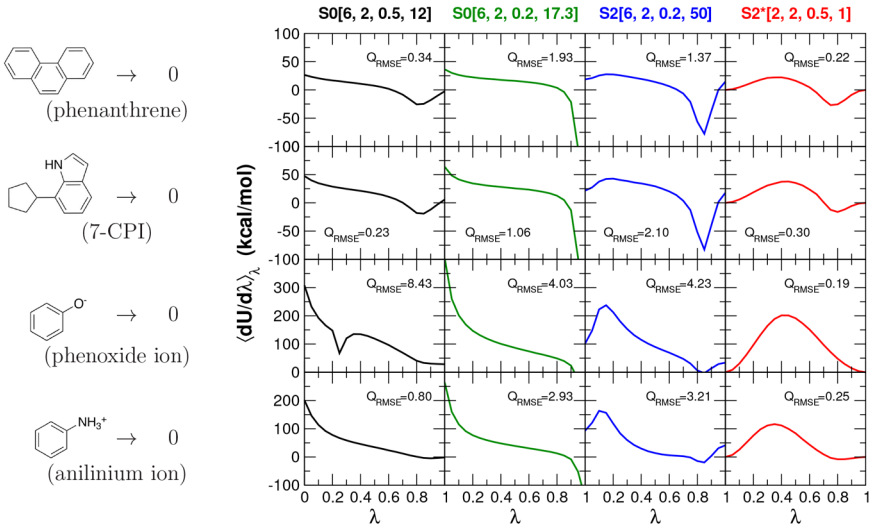


Figure 8. $\langle \partial U/\partial \lambda \rangle_\lambda$ profiles for alchemical free energy simulations of four molecular systems in aqueous solution using the concerted scheme and the alchemical transformation pathways summarized in Table 3: the annihilation of phenanthrene (upper rows), the annihilation of 7-cyclopentanylindole (7-CPI) (the second rows), the annihilation of phenoxide ion (the third rows), and the annihilation of anilinium ion (bottom rows).

Table 9. Absolute Hydration Free Energy Values (kcal/mol)[a] for Four Alchemical Transformation Pathways[b]

Transformation	S0[6,2,0.5,12]		S0[6,2,0.2,17.3]		S2[6,2,0.2,50]		S2*[2,2,0.5,1]		ref
phenanthrene →0	5.91(02)	**0.01**	6.36(07)	**0.44**	5.70(16)	**0.22**	5.85(05)	**0.07**	5.92(03)
7CPI[c] → 0	7.40(06)	**0.05**	8.20(06)	**0.75**	7.41(08)	**0.04**	7.50(05)	**0.05**	7.45(03)
phenoxide ion →0	79.50(09)	**8.01**	71.95(08)	**0.46**	71.56(05)	**0.07**	71.51(07)	**0.05**	71.49(02)
anilinium ion →0	50.89(06)	**0.04**	51.14(07)	**0.29**	50.84(06)	**0.01**	50.81(06)	**0.04**	50.85(03)
Avg. Δ[d]		2.03		0.49		0.09		0.05	

[a]The absolute hydration free energy values are obtained by concerted scheme and analyzed using the TI method. Standard errors are shown in parentheses (multiplied by 10^2). ref is the values obtained from ACES calculations, sampled 4X greater with more independent trials and λ windows, and analyzed with BAR method (see Methods section 3.4 for details). [b]The bold numbers are the differences between the free energy values with respect to the reference numbers. [c]7-cyclopentanylindole. [d]The average free energy differences with respect to the reference.

0 model system in Figure 3(k). The S2[6,2,0.2,50] exhibits instabilities between intermediate λ = 0.90 and 0.95 for the annihilation of phenanthrene and 7-cyclopentanylindole due to poor phase space overlap. The S2*[2,2,0.5,1] overall performs best and achieves stable $\langle \partial U/\partial \lambda \rangle_\lambda$ curves, as reflected also by improved quadrature stability (Q_{RMSE} is smallest, with the minor exception of the 7-CPI case where S0[6,2,0.5,12] is slightly lower).

Table 9 summarizes absolute hydration free energy values for four molecular systems calculated with different alchemical transformation pathways. The bold numbers are the error estimates between the free energy values with respect to the ACES reference results. The annihilation of phenoxide ion shows by far the largest error (8.01 kcal/mol) with S0[6,2,0.5,12]. The large difference matches the irregularity of the $\langle \partial U/\partial \lambda \rangle_\lambda$ profile, which has a sharp kink around λ = 0.25. The S0[6,2,0.5,12] results are overall the worst, with an average estimated error of 2.03 kcal/mol. The S0[6,2,0.2,17.3] has the next highest average estimated error of 0.49 kcal/mol, followed by S2[6,2,0.2,50] which is significantly reduced (0.09 kcal/mol). The S2*[2,2,0.5,1] has the smallest estimated error of 0.05 kcal/mol, with the highest value (0.07 kcal/mol) being for phenanthrene. This suggests that the new S2*[2,2,0.5,1] is robust in its ability to stably carry out these more challenging absolute hydration free energy simulations and provide reliable free energy estimates.

4.5. Energy Stability and Conservation. In order to examine the stability of the simulations (i.e., total energy fluctuations and conservation) using different functional forms and parameters for the various softcore potentials, we perform simulations for $Mg^{2+} \rightarrow Ca^{2+}$ (both ions treated using softcore potentials) with single precision model (SPFP) and the NVE ensemble at λ = 0.5 using four different softcore potentials. Note that for λ = 0.5, the value of the λ-dependent weight functions $W(\lambda)$ in eq 2 are all the same (0.5). The time evolution of the total energy and 1 ps energy drift is compared in Figure 9. In the 1 ps time domain, the distributions of energy drift values are normal (symmetric about the mean, unimodal and Gaussian-like) for all of the softcore potentials, and also similar to the end state simulations (data not shown). A statistically significant difference, however, is that the mean value is considerably larger for the S0[6,2,0.5,12] and S0[6,2,0.2,17.3] softcore potentials (0.19 and 0.12 kcal/mol, respectively) with respect to the smoothstep softcore potentials that have mean values of 0.01 kcal/mol or less. These differences are more evident by looking at the time evolution of the total energy, which illustrates there is a persistent, steady energy drift over 5 ns that gives rise to the mean values in the 1 ps histograms. This drift persists for simulations extended to 20 ns with very similar slope (data not shown). Formally, only the new S2*[2,2,0.5,1] softcore potential is rigorously smooth at the cutoff boundary and conserves energy. However, it appears that for the present

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

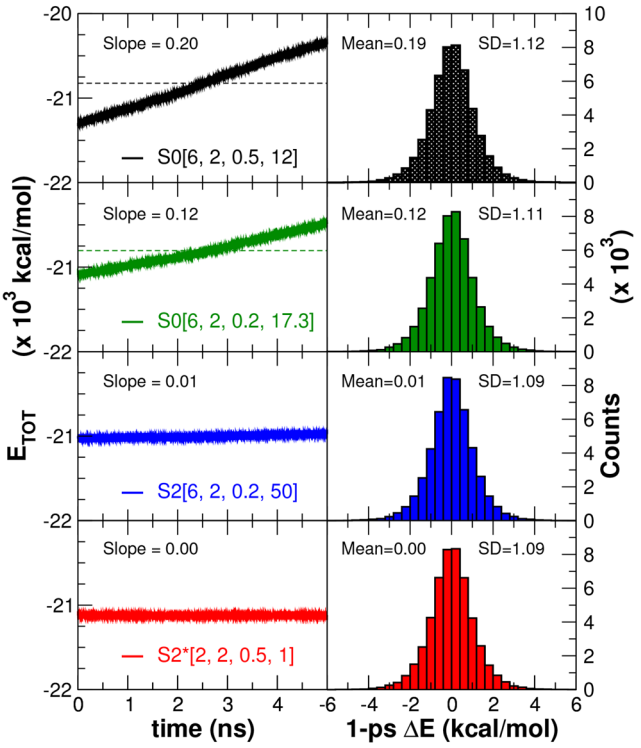


Figure 9. Potential energy drift in the relative hydration free energy calculations for $Mg^{2+} \rightarrow Ca^{2+}$ with single precision model (SPFP) and the NVE ensemble at $\lambda = 0.5$ for the alchemical transformation pathways/softcore potentials summarized in Table 3. The left column shows the time evolution of the total energy (dashed line indicates the average value in 10^3 kcal/mol, and slope is in kcal/mol/ps), whereas the right column shows the 1 ps energy drift histograms with mean values and standard deviations in kcal/mol.

example, the S2[6,2,0.2,50] simulation also conserves energy quite well. A more detailed preliminary analysis indicates that the origin of the trends in energy conservation derives mainly from short-ranged interactions, and relatively minimally from the lack of continuity (in cases other than S2*[2,2,0.5,1]) of the softcore potentials at the cutoff boundary. Specifically, instabilities can arise when "hard" collisions that occur for potentials where the short-ranged LJ are steeply changing due to the large exponent $n = 6$ in eq 18, and this behavior is exacerbated when these repulsions are countered by strong Coulomb interactions (i.e., smaller β values in eq 19). Thus, the magnitude of the energy drift is negligible for the new S2*[2,2,0.5,1] softcore potential, and that for the other softcore potentials follows the trend S0[6,2,0.5,12] > S0-[6,2,0.5,17.3] > S2[6,2,0.2,50] in accord with the increasing β values. It should be noted that none of the energy drifts are severe, and do not greatly impact results when used with an appropriate thermostat. Nonetheless, it is important to examine closely these issues, and engineer stable softcore potentials from the start in the design of next-generation alchemical transformation pathways. As a final comment, we note that we observed virtually no performance difference using the various softcore potentials discussed in the current work (see Table S6 of the Supporting Information).

5. CONCLUSION

The current work develops a robust framework for the design of alchemical transformation pathways using a new form for the λ-dependent mixing weight functions and smoothstep softcore potential. The latter is distinct from previous softcore potentials available in AMBER in that it uses smoothstep functions to stabilize behavior near the $\lambda = 0$ and 1 end points, consistent power scaling of Coulomb and LJ interactions with unitless control parameters to maintain balance of electrostatic attractions and exchange repulsions, pairwise form based on the LJ contact radius for the effective interaction distance with separation-shifted scaling, and rigorous smoothing of the potential at the nonbonded cutoff boundary. The new alchemical transformation pathway is illustrated to overcome commonly encountered end point catastrophe, particle collapse, and large gradient-jump problems in free energy simulations, and represents a considerable advance with respect to other alchemical transformations and softcore potentials. The current work provides the foundation for more robust free energy simulations, including the development of new alchemical enhanced sampling methods and new streamlined workflows for high-throughput simulation and analysis of ligand libraries and thermodynamic graph networks.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00725.

Comparison of the new softcore potential form with different values of n, m, $\alpha^{LJ}*$ and $\alpha^{Coul}*$; quadrature stability; validation against relative binding free energies; references (PDF)

AUTHOR INFORMATION

Corresponding Author

Darrin M. York − *Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0002-9193-7055; Email: Darrin.York@rutgers.edu

Authors

Hsu-Chun Tsai − *Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0001-7027-5649

Tai-Sung Lee − *Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0003-2110-2279

Abir Ganguly − *Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0002-0630-1109

Timothy J. Giese − *Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey 08854, United States;* orcid.org/0000-0002-0653-9168

Maximilian CCJC Ebert − *Congruence Therapeutics, Saint-Laurent, Quebec, Canada H4S 1Z9*

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658

Paul Labute − *Chemical Computing Group ULC, Montreal, Quebec, Canada H3A 2R7*
Kenneth M. Merz, Jr. − *Department of Chemistry and Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0001-9139-5893

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00725

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful for financial support provided by the National Institutes of Health (No. GM107485 to DMY). Computational resources were provided by the Office of Advanced Research Computing (OARC) at Rutgers, the MSU HPC, the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant ACI-1548562 (supercomputer Expanse at SDSC through allocation CHE190067), and by the Texas Advanced Computing Center (TACC) at the University of Texas at Austin (supercomputer Longhorn through allocation CHE20002). We gratefully acknowledge the support of the nVidia Corporation with the donation of several Pascal and Volta GPUs and the GPU-time of a GPU-cluster where the reported benchmark results were performed.

■ REFERENCES

(1) Cournia, Z.; Chipot, C.; Roux, B.; York, D. M.; Sherman, W. In *Free Energy Methods in Drug Discovery—Introduction*; ACS Symposium Series; Armacost, K. A., Thompson, D. C., Eds.; ACS, 2021; Vol. *1397*, pp 1−38.
(2) Mey, A. S. J. S.; Allen, B. K.; Macdonald, H. E. B.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best Practices for Alchemical Free Energy Calculations [Article v1.0]. *Living Journal of Computational Molecular Science* **2020**, *2*, 18378−18429.
(3) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911−2937.
(4) Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee, T. D.; Pearlman, D. A.; Radak, B. K.; Tao, Y.; Tsai, H.-C.; Xu, H.; Sherman, W.; York, D. M. Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5595−5623.
(5) Mobley, D. L.; Klimovich, P. V. Perspective: Alchemical free energy calculations for drug discovery. *J. Chem. Phys.* **2012**, *137*, 230901.
(6) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations. *J. Am. Chem. Soc.* **2017**, *139*, 946−957.
(7) Chodera, J.; Mobley, D.; Shirts, M.; Dixon, R.; Branson, K.; Pande, V. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150−160.
(8) Jorgensen, W. L. Efficient drug lead discovery and optimization. *Acc. Chem. Res.* **2009**, *42*, 724−733.
(9) Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50*, 1625−1632.
(10) Lee, T.-S.; Hu, Y.; Sherborne, B.; Guo, Z.; York, D. M. Toward Fast and Accurate Binding Affinity Prediction with pmemdGTI: An Efficient Implementation of GPU-Accelerated Thermodynamic Integration. *J. Chem. Theory Comput.* **2017**, *13*, 3077−3084.
(11) Lee, T.-S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018**, *58*, 2043−2050.
(12) Mermelstein, D. J.; Lin, C.; Nelson, G.; Kretsch, R.; McCammon, J. A.; Walker, R. C. Fast and flexible gpu accelerated binding free energy calculations within the amber molecular dynamics package. *J. Comput. Chem.* **2018**, *39*, 1354−1358.
(13) Lee, T.-S.; Tsai, H.-C.; Ganguly, A.; Giese, T. J.; York, D. M. In *Robust, Efficient and Automated Methods for Accurate Prediction of Protein-Ligand Binding Affinities in AMBER Drug Discovery Boost*; ACS Symposium Series; Armacost, K. A., Thompson, D. C., Eds.; ACS, 2021; Vol. *1397*, pp 161−204.
(14) Heinzelmann, G.; Henriksen, N. M.; Gilson, M. K. Attach-Pull-Release Calculations of Ligand Binding and Conformational Changes on the First BRD4 Bromodomain. *J. Chem. Theory Comput.* **2017**, *13*, 3260−3275.
(15) Henriksen, N. M.; Fenley, A. T.; Gilson, M. K. Computational Calorimetry: High-Precision Calculation of Host-Guest Binding Thermodynamics. *J. Chem. Theory Comput.* **2015**, *11*, 4377−4394.
(16) Yin, J.; Henriksen, N. M.; Slochower, D. R.; Gilson, M. K. The SAMPL5 host-guest challenge: computing binding free energies and enthalpies from explicit solvent simulations by the attach-pull-release (APR) metho. *J. Comput.-Aided Mol. Des.* **2017**, *31*, 133−145.
(17) Tembre, B. L.; McCammon, J. A. Ligand-receptor interactions. *Comput. Chem.* **1984**, *8*, 281−283.
(18) Fleck, M.; Wieder, M.; Boresch, S. Dummy Atoms in Alchemical Free Energy Calculations. *J. Chem. Theory Comput.* **2021**, *17*, 4403−4419.
(19) Giese, T. J.; York, D. M. Development of a Robust Indirect Approach for MM → QM Free Energy Calculations That Combines Force-Matched Reference Potential and Bennett's Acceptance Ratio Methods. *J. Chem. Theory Comput.* **2019**, *15*, 5543−5562.
(20) König, G.; Brooks, B. R.; Thiel, W.; York, D. M. On the convergence of multi-scale free energy simulations. *Mol. Simul.* **2018**, *44*, 1062−1081.
(21) Wu, D.; Kofke, D. A. Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation. *J. Chem. Phys.* **2005**, *123*, 54103.
(22) Wu, D.; Kofke, D. A. Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations. *J. Chem. Phys.* **2005**, *123*, 84109.
(23) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* **1994**, *222*, 529−539.
(24) Steinbrecher, T.; Joung, I.; Case, D. A. Soft-Core Potentials in Thermodynamic Integration: Comparing One- and Two-Step Transformations. *J. Comput. Chem.* **2011**, *32*, 3253−3263.
(25) Hornak, V.; Simmerling, C. Development of softcore potential functions for overcoming steric barriers in molecular dynamics simulations. *J. Mol. Graphics Model.* **2004**, *22*, 405−413.
(26) Giese, T. J.; York, D. M. A GPU-Accelerated Parameter Interpolation Thermodynamic Integration Free Energy Method. *J. Chem. Theory Comput.* **2018**, *14*, 1564−1582.
(27) Gapsys, V.; Seeliger, D.; de Groot, B. L. New Soft-Core Potential Function for Molecular Dynamics Based Alchemical Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2373−2382.
(28) Buelens, F. P.; Grubmüller, H. Linear-scaling soft-core scheme for alchemical free energy calculations. *J. Comput. Chem.* **2012**, *33*, 25−33.
(29) Pal, R. K.; Gallicchio, E. Perturbation potentials to overcome order/disorder transitions in alchemical binding free energy calculations. *J. Chem. Phys.* **2019**, *151*, 124116.
(30) Mezei, M. Polynomial path for the calculation of liquid state free energies from computer simulations tested on liquid water. *J. Comput. Chem.* **1992**, *13*, 651−656.

(31) Resat, H.; Mezei, M. Studies on free energy calculations. I. Thermodynamic integration using a polynomial path. *J. Chem. Phys.* **1993**, *99*, 6052−6061.
(32) Simonson, T. Free energy of particle insertion. *Mol. Phys.* **1993**, *80*, 441−447.
(33) Steinbrecher, T.; Mobley, D. L.; Case, D. A. Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Chem. Phys.* **2007**, *127*, 214108.
(34) Jiang, W.; Chipot, C.; Roux, B. Computing Relative Binding Affinity of Ligands to Receptor: An Effective Hybrid Single-Dual-Topology Free-Energy Perturbation Approach in NAMD. *J. Chem. Inf. Model.* **2019**, *59*, 3794−3802.
(35) König, G.; Glaser, N.; Schroeder, B.; Kubincová, A.; Hünenberger, P. H.; Riniker, S. An Alternative to Conventional λ-Intermediate States in Alchemical Free Energy Calculations: λ-Enveloping Distribution Sampling. *J. Chem. Inf. Model.* **2020**, *60*, 5407−5423.
(36) König, G.; Ries, B.; Hünenberger, P. H.; Riniker, S. Efficient Alchemical Intermediate States in Free Energy Calculations Using λ-Enveloping Distribution Sampling. *J. Chem. Theory Comput.* **2021**, *17*, 5805−5815.
(37) de Ruiter, A.; Petrov, D.; Oostenbrink, C. Optimization of Alchemical Pathways Using Extended Thermodynamic Integration. *J. Chem. Theory Comput.* **2021**, *17*, 56−65.
(38) Reinhardt, M.; Grubmuller, H. Determining Free-Energy Differences Through Variationally Derived Intermediates. *J. Chem. Theory Comput.* **2020**, *16*, 3504−3512.
(39) Blondel, A. Ensemble Variance in Free Energy Calculations by Thermodynamic Integration: Theory, Optimal "Alchemical" Path, and Practical Solutions. *J. Comput. Chem.* **2004**, *25*, 985−993.
(40) Pham, T. T.; Shirts, M. R. Optimal pairwise and non-pairwise alchemical pathways for free energy calculations of molecular transformation in solution phase. *J. Chem. Phys.* **2012**, *136*, 124120.
(41) Lee, T.-S.; Lin, Z.; Allen, B. K.; Lin, C.; Radak, B. K.; Tao, Y.; Tsai, H.-C.; Sherman, W.; York, D. M. Improved Alchemical Free Energy Calculations with Optimized Smoothstep Softcore Potentials. *J. Chem. Theory Comput.* **2020**, *16*, 5512−5525.
(42) Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* **1935**, *3*, 300−313.
(43) Straatsma, T. P.; Berendsen, H. J. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *J. Chem. Phys.* **1988**, *89*, 5876−5886.
(44) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(45) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(46) Shirts, M. R.; Pande, V. S. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J. Chem. Phys.* **2005**, *122*, 144107.
(47) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
(48) Tan, Z.; Gallicchio, E.; Lapelosa, M.; Levy, R. M. Theory of binless multi-state free energy estimation with applications to protein-ligand binding. *J. Chem. Phys.* **2012**, *136*, 144102.
(49) Giese, T. J.; York, D. M. Variational Method for Networkwide Analysis of Relative Ligand Binding Free Energies with Loop Closure and Experimental Constraints. *J. Chem. Theory Comput.* **2021**, *17*, 1326−1336.
(50) Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Berryman, J.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kaymak, M. C.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shajan, A.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A.*AMBER22*; University of California: San Francisco: San Francisco, CA, 2022.
(51) Meng, Y.; Sabri Dashti, D.; Roitberg, A. E. Computing alchemical free energy differences with Hamiltonian replica exchange molecular dynamics (H-REMD) simulations. *J. Chem. Theory Comput.* **2011**, *7*, 2721−2727.
(52) Roe, D. R.; Bergonzo, C.; Cheatham, T. E. Evaluation of enhanced sampling provided by accelerated molecular dynamics with Hamiltonian replica exchange methods. *J. Phys. Chem. B* **2014**, *118*, 3543−3552.
(53) Hahn, D. F.; König, G.; Hünenberger, P. H. Overcoming Orthogonal Barriers in Alchemical Free Energy Calculations: On the Relative Merits of λ-Variations, λ-Extrapolations, and Biasing. *J. Chem. Theory Comput.* **2020**, *16*, 1630−1645.
(54) Lee, T.-S.; Tsai, H.-C.; Ganguly, A.; York, D. M. ACES: Alchemically Enhanced Sampling. *J. Chem. Theory Comput.* **2022**, DOI: 10.1021/acs.jctc.2c00697.
(55) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(56) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(57) Labute, P.; Ebert, M. In *Optimizing Simulations Protocols for Relative Free Energy Calculations*; ACS Symposium Series; Armacost, K. A., Thompson, D. C., Eds.; ACS, 2021; Vol. *1397*, pp 227−245.
(58) Lee, Y.-K.; Parks, D. J.; Lu, T.; Thieu, T. V.; Markotan, T.; Pan, W.; McComsey, D. F.; Milkiewicz, K. L.; Crysler, C. S.; Ninan, N.; Abad, M. C.; Giardino, E. C.; Maryanoff, B. E.; Damiano, B. P.; Player, M. R. 7-fluoroindazoles as potent and selective inhibitors of factor Xa. *J. Med. Chem.* **2008**, *51*, 282−297.
(59) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(60) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. method. *J. Comput. Chem.* **2000**, *21*, 132−146.
(61) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.
(62) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(63) Miyamoto, S.; Kollman, P. A. SETTLE: An analytic version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952−962.
(64) 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.
(65) Wilson, D. P.; Wan, Z.-K.; Xu, W.-X.; Kirincich, S. J.; Follows, B. C.; Joseph-McCarthy, D.; Foreman, K.; Moretto, A.; Wu, J.; Zhu, M.; Binnun, E.; Zhang, Y.-L.; Tam, M.; Erbe, D. V.; Tobin, J.; Xu, X.; Leung, L.; Shilling, A.; Tam, S. Y.; Mansour, T. S.; Lee, J. Structure-Based Optimization of Protein Tyrosine Phosphatase 1B Inhibitors: From the Active Site to the Second Phophotyrosine Binding Site. *J. Med. Chem.* **2007**, *50*, 4681−4698.
(66) 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.
(67) Raymond, J. W.; Gardiner, E. J.; Willett, P. RASCAL: Calculation of Graph Similarity using Maximum Common Edge Subgraphs. *Comput. J.* **2002**, *45*, 631−644.
(68) http://www.rdkit.org, RDKit: Open-source cheminformatics (accessed Oct. 26th, 2022).
(69) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(70) He, X.; Man, V. H.; Yang, W.; Lee, T.-S.; Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* **2020**, *153*, 114502.
(71) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665−9678.
(72) Mobley, D. L.; Guthrie, J. P. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 711−720.
(73) Hritz, J.; Oostenbrink, C. Hamiltonian replica exchange molecular dynamics using soft-core interactions. *J. Chem. Phys.* **2008**, *128*, 144121.
(74) Jiang, W.; Roux, B. Free energy perturbation Hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. *J. Chem. Theory Comput.* **2010**, *6*, 2559−2565.
(75) Arrar, M.; de Oliveira, C. A. F.; Fajer, M.; Sinko, W.; McCammon, J. A. w-REXAMD: A Hamiltonian replica exchange approach to improve free energy calculations for systems with kinetically trapped conformations. *J. Chem. Theory Comput.* **2013**, *9*, 18−23.
(76) Itoh, S. G.; Okumura, H. Hamiltonian replica-permutation method and its applications to an alanine dipeptide and amyloid-β(29−42) peptides. *J. Comput. Chem.* **2013**, *34*, 2493−2497.
(77) Yang, M.; Huang, J.; MacKerell, A. D., Jr. Enhanced Conformational Sampling Using Replica Exchange with Concurrent Solute Scaling and Hamiltonian Biasing Realized in One Dimension. *J. Chem. Theory Comput.* **2015**, *11*, 2855−2867.
(78) Joung, I. S.; Cheatham, T. E., III Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.
(79) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discovery* **2015**, *10*, 449−461.
(80) Hu, Y.; Sherborne, B.; Lee, T.-S.; Case, D. A.; York, D. M.; Guo, Z. The importance of protonation and tautomerization in relative binding affinity prediction: a comparison of AMBER TI and Schrödinger FEP. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 533−539.

https://doi.org/10.1021/acs.jctc.2c00725
J. Chem. Theory Comput. 2023, 19, 640−658