

Molecular Dynamics with Very Large Time Steps for the Calculation of Solvation Free Energies

Charles R. A. Abreu* and Mark E. Tuckerman*

Cite This: *J. Chem. Theory Comput.* 2020, 16, 7314–7327

Read Online

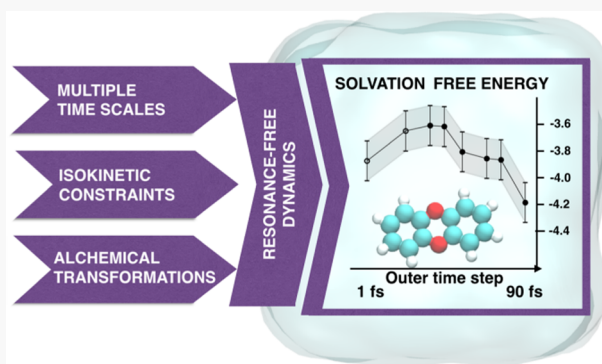
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: In multiple time scale molecular dynamics, the use of isokinetic constraints along with massive thermostating has enabled the adoption of very large integration steps, well beyond the limits imposed by resonance artifacts in standard algorithms. In this work, we present two new contributions to this topic. First, we investigate the velocity distribution and the temperature–kinetic energy relationship associated with the isokinetic Nosé–Hoover family of methods, showing how they depend on the number of thermostats attached to each atomic degree of freedom. Second, we investigate the performance of these methods in the calculation of solvation free energies, the determination of which is often key for understanding the partition of a chemical species among distinct environments. We show how one can extract this property from canonical (constant-NVT) simulations and compare the result to experimental data obtained at a specific pressure. Finally, we demonstrate that large time steps can, in fact, be used to improve the efficiency of these calculations and that attaching multiple thermostats per degree of freedom is beneficial for effectively exploring the configurational space of a molecular system.



1. INTRODUCTION

Physically meaningful states of molecular systems are often distinguished by integer numbers, such as the number of molecules of a given species. In computational chemistry, so-called alchemical methods enable us to link these states together through unphysical but continuous paths, making it possible to assess their relative stability. This information is essential for applications that involve the partition of species among different phases, from the design of industrial processes to the study of ecosystems and environmental science. For instance, the coupling/decoupling of a single molecule (solute) from its surrounding medium (solvent) can be evaluated by defining a parameter $0 \leq \lambda \leq 1$ that governs the magnitude of the solute–solvent interactions while keeping the intrasolvent and intra-solute interactions intact. This approach has been used for the calculation of hydration free energies of organic molecules,^{1,2} binding free energies of protein–ligand^{3–6} and other host–guest systems,^{7,8} activity coefficients,^{9,10} and various other thermodynamic properties. In the case of protein systems, alchemical methods have also been used to effect changes in the protonation state of residues¹¹ or in the chemical identity of ligands,¹² to mention just a few examples.

For a given model system, one can improve the calculation of alchemical free energies by devising techniques for efficiently sweeping across the λ domain and concomitantly rapidly exploring the configuration space. A large body of literature deals with the first issue of sampling the λ domain using adaptations of enhanced sampling algorithms. These include, for example,

concerted sampling of multiple λ states, as has been done via expanded ensemble¹³ or replica exchange¹⁴ methods, or by turning λ into a dynamical variable, such as is done in λ -metadynamics^{15,16} and λ -adiabatic free energy dynamics (AFED).^{17,18} In this article, we focus on the second approach by investigating the application of resonance-free multiple time step integration to the calculation of solvation free energies.

When simulating the dynamics of a flexible molecular model, the step size of a standard integration method, such as velocity Verlet is limited by the time scale of the fastest motions (for instance, the vibrations of chemical bonds that involve light atoms). This limit can be increased by classifying forces according to their time scales and applying an integration strategy in which the slowest forces are evaluated less frequently. As these are usually the most computationally expensive forces, notable gains in efficiency can be achieved. The reversible reference system propagator algorithm (RESPA)¹⁹ provides a rigorous framework for performing such a multiple time-scale integration in a time-reversible and symplectic (or, more generally, geometric) manner. In practice, however, resonance

Received: July 5, 2020

Published: November 16, 2020



artifacts^{20,21} can cause instability in the RESPA algorithm and restrict the attainable range of outer time steps to a more severe degree than the actual system dynamics would. Fortunately, an effective solution to this problem exists in the case of canonical or constant-NVT molecular dynamics (MD). The approach involves coupling a set of thermostats to each atomic degree of freedom via isokinetic constraints that involve the velocities of both the atoms and the thermostat variables.^{22–24} The stochastic isokinetic Nosé–Hoover RESPA method of ref 25, referred to as SIN(R) and which is an outgrowth of the original isokinetic Nosé–Hoover chain algorithm,²⁴ allows outer time steps approaching 100 fs to be used in simulations of molecular systems modeled by fully flexible and polarizable force fields,^{26–29} effectively eliminating resonance artifacts.

The aim of the present paper is 2-fold. We first seek a deeper understanding of the isokinetic thermostating method through a derivation of its stationary probability density for particle velocities. We then establish, for the first time, the resulting relationship between temperature and kinetic energy, which differs from that of a canonical ensemble and depends on the number of thermostats attached to each degree of freedom. We show, by practical examples, that increasing this number causes the sampled configurations to decorrelate more quickly, thus rendering the additional effort worthwhile in terms of statistical quality improvement. We then evaluate the performance of the SIN(R) method in the calculation of solvation free energies by alchemical transformations. For this evaluation, we carry out independent simulations with distinct values of the coupling parameter λ and use the multistate Bennett acceptance ratio (MBAR) method^{30,31} to extract the free energy profiles and to perform a careful uncertainty analysis. Since this type of calculation is usually done via isothermal–isobaric (constant-NPT) simulations to conform to experimental conditions, we determine the circumstances in which NVT ensemble simulations can be used instead. Our results demonstrate that the SIN(R) method is able to provide reliable solvation free energy estimates in a very efficient way with very large time steps.

2. METHODOLOGY

2.1. Stochastic Isokinetic Nosé–Hoover Method. In this section, we review the equations of motion of the Stochastic Isokinetic Nosé–Hoover method.²⁵ We derive their resulting stationary phase-space probability distribution, showing that the coordinates obey the desired Boltzmann–Gibbs distribution.^{24,25} In addition, we demonstrate that velocities follow a particular distribution that differs from the standard Maxwell–Boltzmann type and derive the method’s distinctive definition of temperature as a velocity-based ensemble average.

Consider a system of N particles in three dimensions. Each configurational degree of freedom i has coordinate q_i , velocity v_i , and associated mass m_i . The system is under the influence of a potential energy field $U(\mathbf{q})$, and therefore, the force acting on each degree of freedom is given by $F_i = -\frac{\partial U}{\partial q_i}$. Here, $\mathbf{q} \equiv (q_1, \dots, q_{3N})$ is the $3N$ -dimensional coordinate vector. In the isokinetic Nosé–Hoover method,²⁴ two sets of extended-system velocities $\mathbf{v}_1 = \{\{v_{1,ij}\}_{i=1}^{3N}\}_{j=1}^L$ and $\mathbf{v}_2 = \{\{v_{2,ij}\}_{i=1}^{3N}\}_{j=1}^L$ are defined and assigned inertial parameters Q_1 and Q_2 , respectively. These are used to couple a number L of stochastic Nosé–Hoover thermostats to each degree of freedom (massive thermostating) and, at the same time, to impose $3N$ isokinetic constraints involving each velocity v_i and the thermostat velocities $v_{1,ij}$. The equations of

motion are written in the form of an Itô stochastic differential equation (SDE) system as

$$dq_i = v_i dt \quad (1a)$$

$$dv_i = \frac{F_i}{m_i} dt - \alpha_i v_i dt \quad (1b)$$

$$dv_{1,ij} = -\alpha_i v_{1,ij} dt - v_{2,ij} v_{1,ij} dt \quad (1c)$$

$$dv_{2,ij} = \frac{Q_1 v_{1,ij}^2 - kT}{Q_2} dt - \gamma v_{2,ij} dt + \sqrt{\frac{2\gamma kT}{Q_2}} dw_{ij} \quad (1d)$$

where γ is a friction constant, dw_{ij} denotes an infinitesimal increment of a Wiener process,³² and α_i is a Lagrange multiplier, introduced as a means of imposing an isokinetic constraint on each degree of freedom i .^{24,25}

$$m_i v_i^2 + \frac{Q_1 L}{L+1} \sum_{j=1}^L v_{1,ij}^2 = LkT \quad (2)$$

By differentiating the constraint once with respect to time, we obtain $m_i v_i dv_i + \frac{Q_1 L}{L+1} \sum_{j=1}^L v_{1,ij} dv_{1,ij} = 0$. If we then substitute dv_i and $dv_{1,ij}$ from eq 1b and 1c, respectively, we obtain an expression for α_i as

$$\alpha_i = \frac{F_i v_i - \frac{Q_1 L}{L+1} \sum_{j=1}^L v_{2,ij} v_{1,ij}^2}{m_i v_i^2 + \frac{Q_1 L}{L+1} \sum_{j=1}^L v_{1,ij}^2} \quad (3)$$

The next step consists of determining the probability density $\rho(\mathbf{x})$ preserved by eqs 1a–1d, with $\mathbf{x}$ being the vector of all dynamical variables. It has been demonstrated²⁵ that $\rho(\mathbf{x})$ is equal to the equilibrium distribution of a deterministic version of eqs 1a–1d obtained by setting $\gamma = 0$, thus reducing the SDE system to a system of ordinary differential equations (ODEs)

$$\dot{q}_i = v_i \quad (4a)$$

$$\dot{v}_i = \frac{F_i}{m_i} - \alpha_i v_i \quad (4b)$$

$$\dot{v}_{1,ij} = -(\alpha_i + v_{2,ij}) v_{1,ij} \quad (4c)$$

$$\dot{v}_{2,ij} = \frac{Q_1 v_{1,ij}^2 - kT}{Q_2} \quad (4d)$$

Following the non-Hamiltonian statistical mechanical framework of refs 33 and 34, we must first compute the phase-space compressibility of the ODE system eqs 4a–4d. In general, if a system of ODEs takes the form $\dot{\mathbf{x}} = \boldsymbol{\eta}(\mathbf{x})$, then the compressibility is defined as $\kappa(\mathbf{x}) = \nabla_{\mathbf{x}} \cdot \boldsymbol{\eta}(\mathbf{x})$, where $\mathbf{x}$ is the vector of all dynamical variables. One of the necessary conditions for these equations to preserve a measure $e^{-w(\mathbf{x})} d\mathbf{x}$ in phase space is that $\kappa = \frac{dw(\mathbf{x})}{dt}$.^{33,34} In the present case, the phase-space compressibility of eq 4a–4d is

$$\kappa = - \sum_{i=1}^{3N} \left[\alpha_i + v_i \frac{\partial \alpha_i}{\partial v_i} + \sum_{j=1}^L \left(\alpha_i + v_{2,ij} + v_{1,ij} \frac{\partial \alpha_i}{\partial v_{1,ij}} \right) \right] \quad (5)$$

Performing the derivatives in eq 5 and simplifying, we obtain

$$\kappa = - \sum_{i=1}^{3N} \left(L \frac{F_i v_i - Q_i \sum_{j=1}^L v_{2,ij} v_{1,ij}^2}{m_i v_i^2 + \frac{Q_i L}{L+1} \sum_{j=1}^L v_{1,ij}^2} + \sum_{j=1}^L v_{2,ij} \right) \quad (6)$$

Using the constraint in eq 2, which is satisfied over any exact trajectory described by eqs 4a–4d, we obtain

$$\alpha_i = \frac{F_i v_i - \frac{Q_i L}{L+1} \sum_{j=1}^L v_{2,ij} v_{1,ij}^2}{LkT} \quad (7)$$

so that

$$\kappa = \frac{1}{kT} \sum_{i=1}^{3N} \left[-F_i v_i + \sum_{j=1}^L v_{2,ij} (Q_i v_{1,ij}^2 - kT) \right] \quad (8)$$

Since $\frac{dU}{dt} = - \sum_{i=1}^{3N} F_i v_i$ as a consequence of the chain rule, and $\frac{d}{dt} \left(\frac{1}{2} Q_i v_{2,ij}^2 \right) = v_{2,ij} (Q_i v_{1,ij}^2 - kT)$ due to eq 4d, we can conclude that the measure preserved in the extended phase space is $e^{-[U(\mathbf{q}) + \frac{1}{2} Q_2 \|\mathbf{v}_2\|^2]/kT} d\mathbf{x}$. Assuming ergodicity within the constrained manifold in which the dynamics occurs, the stationary phase-space distribution³⁴ is

$$\rho(\mathbf{x}) = \frac{1}{\Omega} e^{-[U(\mathbf{q}) + \frac{1}{2} Q_2 \|\mathbf{v}_2\|^2]/kT} \prod_{i=1}^{3N} \delta \left(m_i v_i^2 + \frac{Q_i L}{L+1} \sum_{j=1}^L v_{1,ij}^2 - LkT \right) \quad (9)$$

where δ is the Dirac delta function and Ω is the corresponding generalized partition function.³⁴ Thus, the marginal distribution of coordinates is clearly of the Boltzmann–Gibbs type, that is, $\rho_{\mathbf{q}} \propto \exp \left[-\frac{U(\mathbf{q})}{kT} \right]$. In turn, the thermostat-related velocities $\mathbf{v}_2$ obey a multivariate Gaussian distribution $\rho_{\mathbf{v}_2} \propto \exp \left(-\frac{Q_2 \|\mathbf{v}_2\|^2}{2kT} \right)$.

What is less evident, thus far, is how the particle velocities $\mathbf{v} \equiv (v_1, \dots, v_{3N})$ are distributed. To derive this distribution, we start by integrating out all components of the (extended) vector $\mathbf{v}_1$ and determining a corresponding marginal distribution. This can be achieved by resorting to a general identity for Dirac delta functions (see Hörmander,³⁵ Theorem 6.1.5), which is

$$\int \delta(f(\mathbf{z})) d\mathbf{z} = \int \frac{1}{\|\nabla_{\mathbf{z}} f\|} dS \quad (10)$$

where dS is the Euclidean measure on the hypersurface defined by the condition $f(\mathbf{z}) = 0$. The specific case of $f(\mathbf{z}) = \mathbf{z}^T \mathbf{z} - R^2$, with $\mathbf{z} \in \mathbb{R}^L$ and $R \geq 0$, defines the surface of a hypersphere with radius R in L dimensions. The gradient of f is $\nabla_{\mathbf{z}} f = 2\mathbf{z}$, thus making $\|\nabla_{\mathbf{z}} f\| = 2R$ for any $\mathbf{z}$ on this surface. Finally, the remaining integral $\int dS$ is taken over the surface area of the hypersphere³⁶ given by $\frac{2\pi^{L/2}}{\Gamma(L/2)} R^{L-1}$, where $\Gamma(\cdot)$ is the complete Gamma-function. Therefore,

$$\int \delta \left(\sum_{j=1}^L z_j^2 - R^2 \right) d\mathbf{z} = \frac{\pi^{L/2}}{\Gamma(\frac{L}{2})} R^{L-2} \quad (11)$$

which we can use to integrate $\rho(\mathbf{x})$ with respect to $v_{1,ij}$ for all i, j . For every degree of freedom i , we can set $R = (LkT - m_i v_i^2)^{1/2}$ and change variables from $v_{1,ij}$ to $z_j = \left(\frac{Q_i L}{L+1} \right)^{1/2} v_{1,ij}$ for all j .

Finally, we find that the desired marginal probability distribution is

$$\rho(\mathbf{q}, \mathbf{v}, \mathbf{v}_2) = \frac{1}{\Xi} e^{-[U(\mathbf{q}) + \frac{Q_2}{2} \|\mathbf{v}_2\|^2]/kT} \times \prod_{i=1}^{3N} (LkT - m_i v_i^2)^{(L-2)/2} \theta(LkT - m_i v_i^2) \quad (12)$$

where the new partition function Ξ incorporates all constant factors. Here, $\theta(x)$ is the Heaviside step function. Note that the isokinetic constraints make $\rho(\mathbf{q}, \mathbf{v}, \mathbf{v}_2) = 0$ if $v_i = \pm \sqrt{\frac{LkT}{m_i}}$ for any i when $L > 2$, and for this reason, they confine the dynamics inside the “hyperbox” defined by such limits. It is interesting to observe that the distribution becomes independent of $\mathbf{v}$ (hence uniform throughout hyperbox) when $L = 2$. By integrating $\mathbf{q}$ and $\mathbf{v}_2$, as well as all particle-related velocities except v_i , we deduce that $\rho(v_i) \propto \left(1 - \frac{m_i v_i^2}{LkT} \right)^{(L-2)/2} \theta \left(1 - \frac{m_i v_i^2}{LkT} \right)$. Figure 1 contains

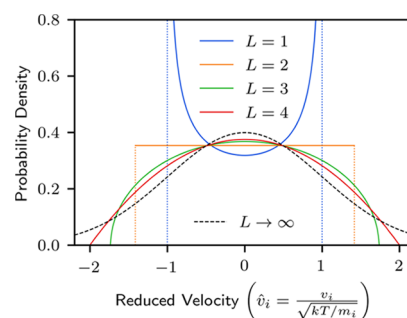


Figure 1. Probability density functions of a particle velocity component, theoretically determined for both the deterministic and stochastic versions of the Isokinetic Nosé–Hoover method, in which L is the number of thermostats attached to each degree of freedom. The black dashed line corresponds to the Maxwell–Boltzmann distribution, which is the limit that $L \rightarrow \infty$.

graphical representations of this marginal distribution, with proper normalization, for several values of L . This is in contrast with the standard Nosé–Hoover chain method,³⁷ in which the inclusion of more thermostats has no influence on the particle velocity distribution, at least after ergodicity issues due to hidden integrals of motion have been solved.³⁴ Interestingly, in the case of $L = 1$ (i.e., a single thermostat per degree of freedom), the velocities are concentrated near the boundaries of their accessible ranges rather than in the middle. Due to Bernoulli’s identity $e^x = \lim_{n \rightarrow \infty} (1 + x/n)^n$, it is clear that the Maxwell–Boltzmann (that is, Gaussian) distribution is approached as L increases.

To end this section, we consider the mean-square kinetic energy $\langle (1/2) m_i v_i^2 \rangle$ of each degree of freedom i and its relation to the system temperature. By employing the change of variables $v_i = z \sqrt{LkT/m_i}$ and realizing that the isokinetic constraint would make $z \in [-1, 1]$, we can write

$$\left\langle \frac{1}{2} m_i v_i^2 \right\rangle = \frac{LkT}{2} \frac{\int_{-1}^1 z^2 (1 - z^2)^{(L-2)/2} dz}{\int_{-1}^1 (1 - z^2)^{(L-2)/2} dz} = \frac{L}{L+1} \frac{kT}{2} \quad (13)$$

Double time scale SIN(R) equations:

$$dq_i = v_i dt$$

$$dv_i = \frac{F_i^{[2]}}{m_i} \left(1 - \frac{m_i v_i^2}{LkT}\right) dt + \frac{F_i^{[1]}}{m_i} \left(1 - \frac{m_i v_i^2}{LkT}\right) dt + \left[\frac{Q_1}{(L+1)kT} \sum_{j=1}^L v_{2,i,j} v_{1,i,j}^2 \right] v_i dt$$

$$dv_{1,i,j} = \left(-\frac{F_i^{[2]} v_i}{LkT} v_{1,i,j} \right) dt + \left(-\frac{F_i^{[1]} v_i}{LkT} v_{1,i,j} \right) dt + \left[\frac{Q_1}{(L+1)kT} \sum_{k=1}^L v_{2,i,k} v_{1,i,k}^2 - v_{2,i,j} \right] v_{1,i,j} dt$$

$$dv_{2,i,j} = \frac{Q_1 v_{1,i,j}^2 - kT}{Q_2} dt - \gamma v_{2,i,j} dt + \sqrt{\frac{2\gamma kT}{Q_2}} dW_{i,j}$$

Middle-RESPA integrator:

$$e^{\Delta t iL} = e^{\frac{\Delta t}{2} iL_B^{[2]}} \left[e^{\frac{\Delta t}{2n} iL_B^{[1]}} e^{\frac{\Delta t}{2n} iL_A} \underbrace{e^{\frac{\Delta t}{2n} iL_{\text{isoK}}} e^{\frac{\Delta t}{n} iL_{\text{DOU}}} e^{\frac{\Delta t}{2n} iL_{\text{isoK}}}}_{e^{\frac{\Delta t}{n} iL_{\text{bath}}}} e^{\frac{\Delta t}{2n} iL_A} e^{\frac{\Delta t}{2n} iL_B^{[1]}} \right]^n e^{\frac{\Delta t}{2} iL_B^{[2]}}$$

Figure 2. Stochastic isokinetic Nosé–Hoover (RESPA) equations of motion with a double time scale splitting of the forces and middle-scheme integration of the thermostat variables. Regarding the colored exponential propagators in the bottom, the effect of each one on the dynamical variables can be obtained by solving a modified system of equations in which the terms that share the same color remain, while all other colored terms are replaced by zero, as discussed in the main text.

Thus, kinetic energy equipartitions among all degrees of freedom in the isokinetic ensemble as it does in the standard canonical ensemble. However, particles are slightly slower on average than they would be in a canonical ensemble at the same temperature unless $L \rightarrow \infty$. As the example applications in Section 3.5 will show, small values of L ($\sim L = 1 - 4$) lead to an efficient sampling of the ensemble distribution while effectively avoiding resonances. Finally, because all individual kinetic energies are independent and identically distributed, we can sum up the expression above for all i and conclude that

$$\langle K \rangle = \frac{L}{L+1} \frac{3NkT}{2} \quad (14)$$

2.2. Multiple Time Scale Integration. By constructing a palindromic operator decomposition method, known as a Strang splitting³⁸ or Trotter–Suzuki splitting,^{39,40} it is possible to devise explicit, time-reversible integrators for the SIN(R) equations of motion. Leimkuhler et al.²⁵ presented two such integrators that achieve multiple time stepping by means of the reference system propagator algorithm (RESPA).¹⁹ These were based on the eXtended-system Outside RESPA (XO-RESPA) and eXtended-system Inside RESPA (XI-RESPA) schemes, both introduced previously by Martyna et al.⁴¹ They can be viewed as generalizations of the traditional, single time scale method in which a velocity Verlet step is surrounded by the integration of the thermostat variables and evaluation of their effect on the particle velocities.⁴² More recently, Zhang et al.^{29,43} showed that a different splitting, the “middle” scheme, generally yields a more accurate sampling of the coordinate distribution than does the traditional scheme. As its name suggests, in the newer scheme, the thermostat integration goes in the middle of each step, similar to the BAOAB integrator^{44,45} used for Langevin dynamics. The impressive success of BAOAB in reproducing $\rho_q \propto \exp\left[-\frac{U(q)}{kT}\right]$ is due to a property known as “superconvergence”.⁴⁴ In certain circumstances, e.g., at the high-friction limit, the leading term of the discretization error on ρ_q cancels out.⁴⁵ Particularly for systems with quadratic Hamiltonians, such as coupled harmonic oscillators, there is no such error at all.^{45,46} Zhang et al.^{29,43} also observed that, like BAOAB, the middle scheme degrades the sampling of velocities, whose distribution departs from theoretical prediction for the specified

temperature. This is not often a drawback, especially with methods like SIN(R) that are devised for configurational sampling rather than actual dynamical simulation.

In the present work, we employ a variant of the recently introduced middle-SIN(R) integrator, which has also shown superior performance in comparison to the original formulations.²⁹ For this, we start by splitting the force on each degree of freedom i into a sum of M terms as $F_i = \sum_{k=1}^M F_i^{[k]}$. By convention, the characteristic time scale of each component increases with index k . If we write the equations of motion as $\dot{\mathbf{x}} = iL\mathbf{x}$, where iL represents a non-Hamiltonian extension of the Liouville operator,³³ we can introduce a partitioning of the operator as

$$iL = iL_A + \sum_{k=1}^M iL_B^{[k]} + iL_{\text{bath}} \quad (15)$$

where iL_A is the universal kinetic component that generates changes in the particle coordinates, $iL_B^{[k]}$ depends on the forces with index k , and iL_{bath} generates transformations in the thermostat variables as well as their effects on the particle velocities. In the usual notation with exponential operators, a step of the middle RESPA scheme can be written in a recursive manner as

$$e^{\Delta t iL} = \mathcal{G}_M(\Delta t) \quad (16)$$

where $\mathcal{G}_M$ is the endpoint of a sequence of nested operators whose members are defined recursively as

$$\mathcal{G}_k(\delta t) = e^{(\delta t/2) iL_B^{[k]}} \left[\mathcal{G}_{k-1} \left(\frac{\delta t}{n_k} \right) \right]^{n_k} e^{(\delta t/2) iL_B^{[k]}}, \text{ if } k > 1 \quad (17)$$

meaning that every step of size δt taken at a given time scale k will involve n_k substeps of size $\delta t/n_k$. The recursive process comes to an end when operator $\mathcal{G}_2$ invokes its nested operator $\mathcal{G}_1$, which is defined as

$$\mathcal{G}_1(\delta t) = e^{(\delta t/2) iL_B^{[1]}} e^{(\delta t/2) iL_A} e^{\delta t iL_{\text{bath}}} e^{(\delta t/2) iL_A} e^{(\delta t/2) iL_B^{[1]}} \quad (18)$$

In the case of SIN(R), the innermost propagator $e^{\delta t iL_{\text{bath}}}$ requires further splitting. The scheme we employ in the present work is

$$e^{\delta t L_{\text{bath}}} = e^{(\delta t/2) i L_{\text{IsoK}}} e^{\delta t L_{\text{DOU}}} e^{(\delta t/2) i L_{\text{IsoK}}} \quad (19)$$

In this equation, IsoK refers to isokinetic transformations exclusively induced by the thermostats (that is, with no physical forces involved) and DOU denotes a drifted Ornstein–Uhlenbeck process performed for each $v_{2,i,j}$. The latter is the source of randomness of an otherwise deterministic dynamics.

To provide a panoramic view of the whole splitting scheme, we present, in Figure 2, a colored diagram containing the SIN(R) equations of motion and operator splitting for the case of a middle RESPA integrator with two time scales. Note that α_j , as in the simplified form of eq 7, has already been substituted back into the equations of motion.

In order to proceed, the action of $e^{\delta t L_{\text{b}}^{[k]}}$ is obtained by analytically solving the ODE system composed of

$$\dot{v}_i = \frac{F_i^{[k]}}{m_i} \left(1 - \frac{m_i v_i^2}{LkT} \right) \quad (20a)$$

$$\dot{v}_{1,i,j} = -\frac{F_i^{[k]}}{LkT} v_{1,i,j} \quad (20b)$$

along with the trivial equations $\dot{q}_i = 0$ and $\dot{v}_{2,i,j} = 0$. For this system, in particular, we work out a solution here whose form differs from the scheme presented originally^{22,23} and employed afterward.^{24–29} Aside from being mathematically equivalent, the new solution is computationally more robust than the original one, as it dispenses with the conditional treatment of a numerical indeterminacy (see ref 25). Clearly, the propagator $e^{\delta t L_{\text{b}}^{[k]}}$ leaves the coordinates $\mathbf{q}$ and, consequently, all forces $F_i^{[k]}$ unchanged. The same occurs for the thermostat variables $\mathbf{v}_2$. To facilitate the solution, let us define dimensionless variables for each degree of freedom i as $y_i = \sqrt{\frac{m_i}{LkT}} v_i$ and $y_{j+1} = \sqrt{\frac{Q_1}{(L+1)kT}} v_{1,i,j}$, so that the isokinetic constraint, eq 2, becomes

$$\sum_{j=1}^{L+1} y_j^2 = \|\mathbf{y}\|^2 = 1 \quad (21)$$

meaning that the dimensionless vector $\mathbf{y} \in \mathbb{R}^{L+1}$ is constrained to have unit norm and be restricted to the surface of a unit hypersphere. After performing the variable transformation, eqs 20a and 20b becomes

$$\dot{y}_1 = b(1 - y_1^2) \quad (22a)$$

$$\dot{y}_j = -b y_1 y_j \quad \text{for } j \in [2, L+1] \quad (22b)$$

where $b = \frac{F_i^{[k]}}{\sqrt{Lm_i kT}}$. It is assumed that the initial conditions, $y_j(0) = y_j^0$ for all j , satisfy eq 21. Note that eq 22a can be solved for $y_1(t)$ independent of the other equations. A straightforward solution is given by $\text{arctanh } y_1(t) - \text{arctanh } y_1^0 = bt$. Using hyperbolic trigonometric identities and the fact that $1 - (y_1^0)^2 = \sum_{j=2}^{L+1} (y_j^0)^2$, we can rewrite it as

$$y_1(t) = \frac{y_1^0 \cosh(bt) + \sinh(bt)}{\sqrt{[y_1^0 \cosh(bt) + \sinh(bt)]^2 + \sum_{j=2}^{L+1} (y_j^0)^2}} \quad (23)$$

Then, once $y_1(t)$ is known, the solution of eq 22b can be determined as $y_j(t) = y_j^0 e^{-b \int_0^t y_1(s) ds}$. Note that the same scaling factor applies for all j 's > 1 . It can be obtained by solving eq 21

after substituting $y_j(t)$ for all j 's. From this procedure, it turns out that for $j \in [2, L+1]$,

$$y_j(t) = \frac{y_j^0}{\sqrt{[y_1^0 \cosh(bt) + \sinh(bt)]^2 + \sum_{j=2}^{L+1} (y_j^0)^2}} \quad (24)$$

The common denominator in the equations above is the norm of a vector $\mathbf{y}^*$ defined such that $y_1^* = y_1^0 \cosh(bt) + \sinh(bt)$ and $y_j^* = y_j^0$ for $j > 1$. Thus, the solution of eqs 22a and 22b can be written as

$$\mathbf{y} = \frac{\mathbf{y}^*}{\|\mathbf{y}^*\|} \quad (25)$$

This gives the solution an “update and renormalize” form, which is computationally attractive, as it guarantees that the isokinetic constraints will remain free of accumulated rounding errors. By recasting the solution in terms of dimensional variables, with initial condition $\mathbf{v}(0) = \mathbf{v}^0$ and $\mathbf{v}_1(0) = \mathbf{v}_1^0$, the action of the propagator $e^{\delta t L_{\text{b}}^{[k]}}$ involves the following sequence

$$e^{\delta t L_{\text{b}}^{[k]}} \rightarrow \begin{cases} b = \frac{F_i^{[k]}}{\sqrt{Lm_i kT}} \\ v_i^* = v_i^0 \cosh(b \delta t) + \sqrt{\frac{LkT}{m_i}} \sinh(b \delta t) \\ s = \left[\frac{m_i (v_i^*)^2 + \frac{L}{L+1} Q_1 \sum_{j=1}^L (v_{1,i,j}^0)^2}{LkT} \right]^{-1/2} \\ v_i = s v_i^* \\ v_{1,i,j} = s v_{1,i,j}^0 \quad \text{for } j \in [1, L] \end{cases} \quad (26)$$

The effects of all other propagators present in Figure 2 are evaluated exactly as described in the original SIN(R) paper.²⁵ Therefore, for each degree of freedom i

$$e^{\delta t L_{\text{A}}} \rightarrow \{q_i = q_i^0 + v_i \delta t\} \quad (27a)$$

$$e^{\delta t L_{\text{IsoK}}} \rightarrow \begin{cases} v_{1,i,j}^* = v_{1,i,j}^0 e^{-v_{2,i,j} \delta t}, \quad \text{for } j \in [1, L] \\ s = \left[\frac{m_i (v_i^0)^2 + \frac{L}{L+1} Q_1 \sum_{j=1}^L (v_{1,i,j}^*)^2}{LkT} \right]^{-1/2} \\ v_i = s v_i^0 \\ v_{1,i,j} = s v_{1,i,j}^* \quad \text{for } j \in [1, L] \end{cases} \quad (27b)$$

$$e^{\delta t L_{\text{DOU}}} \rightarrow \begin{cases} z = e^{-\gamma \delta t} \\ \mu = \frac{Q_1 v_{1,i,j}^2 - kT}{\gamma Q_2} \\ v_{2,i,j} = v_{2,i,j}^0 z + \mu(1 - z) + \sqrt{\frac{kT}{Q_2}} (1 - z^2) \xi_{i,j} \end{cases} \quad (27c)$$

Note that the effect of $e^{\delta t L_{\text{Isok}}}$ is also expressed above in an update and renormalize form. In the case of the Ornstein–Uhlenbeck propagator, ξ_{ij} represents a random variable drawn independently from a standard normal distribution. Note, further, that the operator actions in eqs 26 and 27 are evaluated for a generic time step δt ; however, in applying these relations, one must follow the scheme in Figure 2 in order to determine the correct step length to be fed in for δt .

2.3. Solvation Free Energy Calculation. In this work, relative free energies are computed by independently simulating a system at several equilibrium states and then collectively analyzing configurations sampled from all of the simulations. All simulations are carried out using SIN(R) with constant numbers of molecules N , volume V , and heat-bath temperature T . To compute the free energies, we use the multistate Bennett acceptance ratio (MBAR) estimator,^{30,31} wherein the reduced free energy $f_i = F_i/kT$ of every state i in a set of m simulated states is estimated using the self-consistent solution of m equations given by³¹

$$\hat{f}_i = -\ln \sum_{k=1}^n \frac{e^{-u_i(\mathbf{q}_k)}}{\sum_{j=1}^m n_j e^{-u_j(\mathbf{q}_k) + \hat{f}_j}} \quad \text{for } i \in [1, m] \quad (28)$$

Here, n_j is the number of uncorrelated configurations sampled in state j , $n = \sum_{j=1}^m n_j$ is the size of the pooled sample obtained from simulating all states, and $u_i(\mathbf{q}_k)$ is the reduced potential, in state i , of the k th configuration in the pooled sample. This means that the probability density of configurations in state i is proportional to $e^{-u_i(\mathbf{q})}$. It is worth noting that a unique solution to the system in eq 28 can only be found if one of the free energies is set to a constant beforehand. It is common to set $\hat{f}_1 = 0$, meaning that we are actually estimating relative free energies relative to the first state. For the configurations to be uncorrelated, the sampling period in each simulation must be larger than the corresponding correlation time. Because this property is usually not known beforehand, in practice, we might need to eliminate chunks of correlated configurations in a postprocessing stage known as subsampling. This is done by estimating the correlation time by means of the autocorrelation function of some appropriate collective variable.⁴⁷ Here, in general, we use the reduced potential, $u_i(\mathbf{q})$, at the simulated state as such a variable. The pooled sample must cover a continuous region in phase space, meaning that there must be overlaps among the distributions sampled at the distinct states. Finally, we remark that MBAR also includes an estimator for the uncertainties of the computed free energies.^{30,31}

In the calculation of solvation free energies via alchemical decoupling of a solute molecule a from its surrounding solvent b , the reduced potential is

$$u_i(\mathbf{q}) = \frac{U_{ab}^{\text{alchem}}(\mathbf{q}, \lambda_i)}{kT} \quad (29)$$

where $U_{ab}^{\text{alchem}}(\mathbf{q}, \lambda)$ is the alchemical potential energy describing solute–solvent interactions, which are affected by the value of a coupling parameter λ . This parameter can vary continuously from 0 to 1 and must be defined so that $U_{ab}^{\text{alchem}}(\mathbf{q}, 0) = 0$ and $U_{ab}^{\text{alchem}}(\mathbf{q}, 1) = U_{ab}(\mathbf{q})$, where U_{ab} is the actual (or chemical, as opposed to alchemical) solute–solvent potential energy. In contrast, the solute–solute and solvent–solvent interaction energies, respectively, $U_{aa}(\mathbf{q})$ and $U_{bb}(\mathbf{q})$, are completely independent of λ . The set of states to be simulated is defined as a grid of m values of λ_i from 0 to 1. The number and locations

of the grid points must be chosen to provide the aforementioned distribution overlaps. The molar coupling free energy estimated from these simulations is

$$\Delta F_{\text{coup}} = kT(\hat{f}_{\lambda=1} - \hat{f}_{\lambda=0}) \quad (30)$$

In practice, molar solvation free energies are determined from liquid–gas-phase equilibrium experiments at specified temperatures and pressures. At low pressures, they are computed by⁴⁸

$$\Delta G_{\text{solv}} = kT \ln \frac{y_a P}{kT \rho_a^l} \quad (31)$$

where y_a is the solute mole fraction in the gas phase and $\rho_a^l = N_a/V$ is the solute concentration in the liquid phase, both measured at equilibrium. Due to this fact, liquid-phase simulations aimed at determining solvation free energies of model systems are usually carried out at constant-NPT conditions as opposed to the constant-NVT conditions simulated here. Through careful analysis, Shirts et al.⁴⁹ derived the equation

$$kT \ln \frac{y_a P}{kT \langle \rho_a^l \rangle_{\lambda=1}} = \Delta G_{\text{coup}} - kT \ln \frac{\langle V \rangle_{\lambda=0}}{\langle V \rangle_{\lambda=1}} \quad (32)$$

where the molar coupling free energy ΔG_{coup} and the ensemble averages $\langle \cdot \rangle$ are all computed from NPT simulations. Therefore, a comparison of ΔG_{coup} to an experimentally measured ΔG_{solv} requires a volume correction, which is often negligible.

Based on the procedure of Shirts et al.,⁴⁹ we can justify the use of ΔF_{coup} as a direct estimator for ΔG_{solv} . The chemical potential of the solute species a in the liquid phase is

$$\mu_a^l = \left(\frac{\partial F}{\partial N_a} \right)_{N_b, V, T} = -kT \ln \frac{Q(N_a)}{Q(N_a - 1)} \quad (33)$$

where Q is the canonical partition function, whose dependence on N_b , V , and T is kept implicit for simplicity, and $F = -kT \ln Q$ is the Helmholtz free energy. While N_a is the number of molecules of the solute species, the vector $\mathbf{N}_b$ represents the numbers of molecules of all species present in the solvent. All of these quantities refer to the liquid phase. The coupling free energy computed from the set of NVT simulations is

$$\Delta F_{\text{coup}} = -kT \ln \frac{Q(N_a, \lambda = 1)}{Q(N_a, \lambda = 0)} \quad (34)$$

where the partition functions refer to a system in which one of the solute molecules is distinct from the others due to its attachment to the coupling parameter. In the case of $\lambda = 1$, the only difference with the partition function of a solution with N_a indistinguishable solute molecules is the fact that it includes a factor $1/1!(N_a - 1)!$ instead of $1/N_a!$. In the case of $\lambda = 0$, we actually have two independent systems sharing the same simulation box: a solution with $N_a - 1$ solute molecules and a single, isolated solute molecule. Based on these facts, we can write

$$\Delta F_{\text{coup}} = -kT \ln \frac{N_a Q(N_a)}{Q(N_a - 1) \theta_a(T) V} \quad (35)$$

where $\theta_a(T)V$ is the ideal gas canonical partition function of a single solute molecule. Therefore, from eq 33,

$$\Delta F_{\text{coup}} = \mu_a^l - kT \ln \frac{\rho_a^l}{\theta_a(T)} \quad (36)$$

For an ideal gas phase at temperature T and pressure P_{gas} , the solute chemical potential is given by⁴⁹

$$\mu_a^g = kT \ln \frac{y_a P_{\text{gas}}}{kT\theta_a(T)} \quad (37)$$

Now, considering phase equilibrium between such a gas and the liquid with N_a solute molecules, it turns out that $\mu_a^g = \mu_a^l$ and $P_{\text{gas}} = \langle P \rangle_{\lambda=1}$. The latter equality stems from the fact that the average pressure, unlike the free energy, is insensitive to the mere tagging of a molecule. Therefore, we conclude that

$$\Delta F_{\text{coup}} = kT \ln \frac{y_a \langle P \rangle_{\lambda=1}}{kT\rho_a^l} \quad (38)$$

which is directly comparable to eq 31 as long as the average pressure of the NVT simulation with $\lambda = 1$ is close to the experimental value.

3. SIMULATION RESULTS

3.1. Common Simulation Details. In the simulations carried out here, unless stated otherwise, integration of the SIN(R) equations of motion is done exactly as described in Section 2.2. For this purpose, we implemented the method using the custom integrator functionality of OpenMM,⁵⁰ which is a very efficient MD code, especially when run on graphics processing units (GPUs). In all runs, the friction coefficient was set to $\gamma = 0.1 \text{ fs}^{-1}$ and the characteristic time scale chosen for the thermostats was $\tau = 10 \text{ fs}$. The inertial parameters of the thermostats are determined using $Q_1 = Q_2 = kT\tau^2$, as usual.^{19,41,42} These are the same parameter values employed in previous MD studies using the SIN(R) method.^{25,26} The temperature was set to 298.15 K in all simulations.

A truncated/smoothed Lennard-Jones potential is used to model the van der Waals (i.e., repulsion and dispersion) interactions. The standard potential $V_{\text{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]$ is multiplied by a switching function $f_s \left(\frac{r + \delta r - r_c}{\delta r} \right)$, where r_c is the cutoff distance, δr is a healing length (along which the potential smoothly decays to zero), and $f_s(z)$ is a fifth-order switching function given by

$$f_s(z) = \begin{cases} 1, & \text{if } z < 0 \\ 1 - 10z^3 + 15z^4 - 6z^5, & \text{if } 0 \leq z \leq 1 \\ 0, & \text{if } z > 1 \end{cases} \quad (39)$$

The interactions are evaluated using $r_c = 12 \text{ \AA}$ and $\delta r = 1 \text{ \AA}$, along with standard long-range corrections. Every simulation employed a cubic simulation box with a case-specific edge length L and subject to periodic boundary conditions. Electrostatic interactions are computed using the Ewald summation method, with the same cutoff distance r_c as was used for the real-space part. The Ewald damping parameter was set to $\alpha = \sqrt{-\ln(2\delta)}/r_c$, where $\delta = 5 \times 10^{-4}$, while the reciprocal-space part was solved by using the particle mesh Ewald (PME) method⁵¹ with $n_{\text{mesh}} \approx (2/3)\alpha L\delta^{-1/5}$.

To implement multiple time stepping, we split the interaction forces according to the so-called RESPA2 scheme.^{52,53} In the most general case, described as follows, such a splitting is done based on three time scales. The fastest forces include bonded interactions, namely, bond stretching, angle bending, and

torsion. A step size of 0.5 fs is used for integration at this fastest time scale. The middle time scale comprises the short-ranged components of the Lennard-Jones and Coulomb interactions, including scaled 1–4 interaction terms, with an internal cutoff distance $r_c^{\text{in}} = 8 \text{ \AA}$ and a smooth decay to zero starting from $5 \text{ \AA}$, yielding a healing length $\delta r^{\text{in}} = 3 \text{ \AA}$. In this case, a switching factor $f_s \left(\frac{r + \delta r^{\text{in}} - r_c^{\text{in}}}{\delta r^{\text{in}}} \right)$ multiplies the forces derived from the Lennard-Jones and Coulomb potentials, rather than the potentials themselves. The Coulomb potential included at this level includes short-range contributions from the full potential evaluated in real-space according to the so-called RESPA2 formulation.^{25,52,53} The step size chosen for the middle time scale is 3 fs, meaning that each step at this scale involves 6 substeps for the fastest forces. The slowest forces comprise the long-range components of nonbonded interactions, which include the reciprocal-space part of the electrostatic potential (with the proper subtraction of 1–2 and 1–3 Ewald exclusions, as well as short-range contributions already accounted for at the intermediate scale).^{52,53} From a computational standpoint, these are the most expensive force terms. The step size used for integration in this largest scale is the key parameter to be evaluated in the forthcoming sections. In specific cases where a two time scale scheme is employed, only the bonded forces are split from the nonbonded forces. When a single time step is used, mostly for benchmarking, no force splitting is employed. These various splittings were also implemented in OpenMM⁵⁰ using its custom nonbonded force functionality. It is worth mentioning that for the alchemical solute–solvent interactions (see Section 3.3 for details), the same truncation, smoothing, and time scale splitting as the solute–solute and solvent–solvent interactions.

In our experience, for the aforementioned splitting scheme to remain stable when very large time steps are employed, it is essential to apply the internal switching directly to the forces, as mentioned above. This observation seemingly contradicts an assertion by Morrone et al.⁵³ that switching on the force or on the potential leads to similarly smooth trajectories. This, in fact, remains true at smaller outer time steps (those employed with the slowest forces), but at the very large time steps employed in this study, applying the switching on the potential is not as effective as applying it directly on the forces. For a defined potential $V(r)$, the force that a particle j exerts on a particle i is computed by $\mathbf{F}_{ij} = -V'(r_{ij}) \frac{\mathbf{r}_{ij}}{r_{ij}}$, where $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and $r_{ij} = \|\mathbf{r}_{ij}\|$.

Applying the internal switching on the forces simply means that a modified force $\hat{\mathbf{F}}_{ij} = -f_s \left(\frac{r_{ij} - r_s^{\text{in}}}{\delta r^{\text{in}}} \right) V'(r_{ij}) \frac{\mathbf{r}_{ij}}{r_{ij}}$ will be used instead.

Because this force only appears in intermediate calculations, it is never actually necessary to evaluate its originating potential. Nevertheless, our particular implementation required us to feed such a potential into OpenMM,⁵⁰ and the details of its derivation are given in the Supporting Information.

For every system simulated in the canonical ensemble, the box volume was determined beforehand by carrying out a constant-NPT simulation at 298.15 K and 1 atm. For this, we employed a standard Langevin dynamics integrator⁵⁰ with $\Delta t = 1 \text{ fs}$ and a Monte Carlo barostat^{54,55} acting every 20 steps, with the maximum allowable volume change adjusted to yield an acceptance rate of roughly 50%. These simulations were carried out for 4 ns, and the average box volume sampled during the final half was taken over for the subsequent NVT simulations. In the case of solvation free energy calculations, this procedure is

applied only for the state in which the solute molecule is fully coupled to the solvent, as explained in Section 2.3.

3.2. Velocity Distribution in the Isokinetic Ensemble.

Our first test is a simple determination of the velocity distributions sampled during an equilibrium SIN(R) simulation as a function of the number of thermostats attached to each degree of freedom. Our goal is to confirm the theoretical predictions in Section 2.1. For this test, we simulated a water system with density $\rho = 0.998 \text{ g/cm}^3$, obtained by placing 500 molecules in a cubic box with $L_{\text{box}} = 24.653 \text{ \AA}$. The force field is the fully flexible SPC-Fw model.⁵⁶

As we seek only to reproduce the velocity distributions of the SIN(R) equations of motion with small time-discretization effects, we do not use RESPA in this particular study. Instead, we generate single time step trajectories with a time step size $\Delta t = 1 \text{ fs}$ and a traditional “side” integration scheme,⁴³ which is known to yield better velocity distributions than the middle scheme adopted throughout the rest of this paper. The total time of each trajectory is 5 ns, from which 10^4 configurations are sampled at regular intervals during the final 2 ns. This provides us with 1.5×10^7 velocity values, thus enabling us to build high-quality histograms. Reduced velocities $\hat{v}_i = \sqrt{\frac{m_i}{kT}} v_i$ are used so that oxygen and hydrogen atoms can be analyzed together. Four simulations are carried out at $T = 298.15 \text{ K}$, each one with a different number of thermostats L from 1 to 4. Normalized histograms constructed by uniformly dividing the interval $[-\sqrt{L}, \sqrt{L}]$ into 50 bins are shown in Figure 3. Comparison

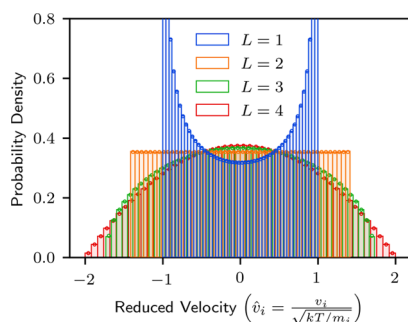


Figure 3. Reduced velocity histograms obtained for all atoms in a set of four SIN(R) simulations of SPC-Fw water⁵⁶ using different numbers L of thermostats attached to each degree of freedom. The circles represent theoretical predictions from eq 12. The equations of motion were solved using a single time scale and the standard side integration scheme.⁴³

with Figure 1 shows that the theoretical predictions are confirmed by the numerical results and, in particular, that the distribution only starts to approach a Boltzmann distribution for the larger values of L .

3.3. Hydration Free Energy of a Lennard-Jones Particle. In this section, we report a study on the hydration of a very simple solute, which is a single, electrically neutral Lennard-Jones particle. For the alchemical solute–solvent interactions, we employed two different strategies for incorporating the continuous coupling parameter λ such that the full Lennard-Jones potential is recovered when $\lambda = 1$ and no interaction occurs when $\lambda = 0$. The first strategy consists of scaling the solute–solvent interaction potential by a coupling function $g(\lambda)$ that increases monotonically from $g(0) = 0$ to $g(1) = 1$. In this case, for pairs of solute and solvent atoms, we have

$$V_{\text{scaling}}(r, \lambda) = 4g(\lambda)\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (40a)$$

where $g(\lambda) = \lambda^4 (5 - 4\lambda)$. The second coupling strategy employed here is the well-known softcore potential of Beutler et al.⁵⁷ given by

$$V_{\text{softcore}}(r, \lambda) = 4\lambda\epsilon \left[\left(\frac{\sigma^6}{r^6 + \frac{1-\lambda}{2}\sigma^6} \right)^2 - \frac{\sigma^6}{r^6 + \frac{1-\lambda}{2}\sigma^6} \right] \quad (41)$$

Both the scaled and softcore potentials are widely used alchemical pathways.^{1–12} Other approaches also exist that aim to combine features of both.^{58–63} We provide a few additional arguments on this subject in the Supporting Information, specifically regarding the use of V_{scaling} with different forms of the coupling function $g(\lambda)$.

To model the solute molecule, we use the parameters for methane taken from the TraPPE united-atom force field,⁶⁴ $\sigma = 3.73 \text{ \AA}$ and $\epsilon = 1.2305 \text{ kJ/mol}$. This model has been shown to reproduce experimentally determined properties of pure methane in a wide range of thermodynamic conditions.⁶⁵ For the solvent, we adopt the fully flexible SPC-Fw water model.⁵⁶ The experimental hydration free energy of methane is about $2.0 \pm 0.2 \text{ kcal/mol}$.^{1,48,66} The system contains 1 methane and 499 water molecules in a cubic box with edge length $L_{\text{box}} = 24.470 \text{ \AA}$, thus resulting in a density $\rho = 1.0206 \text{ g/cm}^3$. These values were determined according to the procedure described in Section 3.1. Within each of the two coupling strategies, simulations were carried out independently at 21 regularly spaced values of λ between 0 and 1. Each simulation consists of 4.5 ns of total simulation time, with the final 3.6 ns used for sampling configurations every 900 fs. For each coupling strategy, we repeat the procedure several times, each with a different size for the external time step of the SIN(R) simulation method. The outer time step (Δt) values are 1 fs, 3 fs, 6 fs, 9 fs, 15 fs, 30 fs, 45 fs, and 90 fs. A force splitting with only two time scales ($M = 2$) is considered for the two first cases. In all other cases, the three-scale splitting described in Section 3.1 is applied. The number of Nosé–Hoover thermostats per degree of freedom in the SIN(R) method is $L = 1$.

Statistical inefficiency is usually very high when considering a single solute molecule in a surrounding solvent. Using an integrated autocorrelation function approach,⁴⁷ we estimate the correlation time in the simulation at each state i based on the corresponding alchemical component of the potential energy, $U_{ab}^{\text{alchem}}(\mathbf{q}, \lambda_i)$. An exception is made for the first state, with $\lambda_1 = 0$, whose alchemical component is null by definition. In this case, we use $U_{ab}^{\text{alchem}}(\mathbf{q}, \lambda_2)$ instead. For each coupling strategy and each state i , the results obtained from simulations with different sizes of the outer time step were averaged and plotted against λ in Figure 4. It is interesting to observe the distinct behavior of the scaling and softcore approaches at mid-range values of λ , where the use of a softcore potential seems to make the alchemical energy more difficult to decorrelate.⁵⁹

When performing subsampling prior to MBAR calculations, we adopt, regardless of the employed Δt , the mean correlation times of Figure 4. By doing this instead of using the individual estimate for every simulation, comparable free energy values are obtained from equally sized samples. After subsampling, we end up with 27 998 configurations in the pooled sample obtained using scaling as the coupling strategy. This number is small

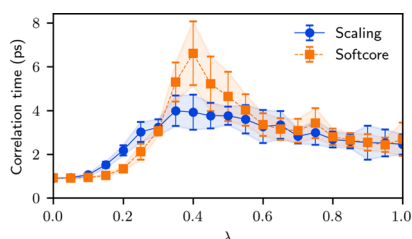


Figure 4. Mean correlation times estimated by means of the integrated autocorrelation function method⁴⁷ from the solute–solvent alchemical potential energy in methane–water simulations. Each error bar extends up and down to one standard deviation, determined from SIN(R) simulations executed with eight different outer time step sizes.

considering that the total production time for the combined set of simulations is 0.756 μ s. The corresponding number of configurations using the softcore potential is 25 963.

The free energy profiles and the final solvation free energies of methane in water obtained via MBAR are shown in Figure 5.

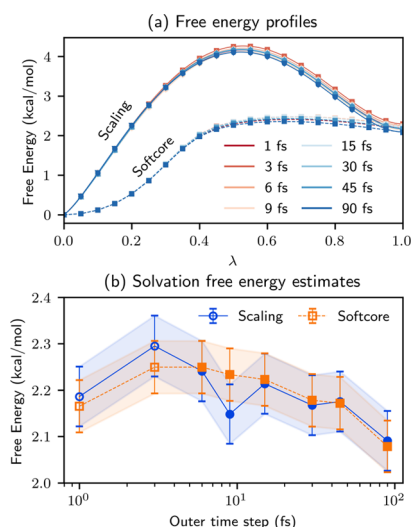


Figure 5. Solvation free energies computed for TraPPE-UA methane⁶⁴ in SPC-Fw water⁵⁶ using the middle-scheme SIN(R) method with one thermostat per degree of freedom. (a) Coupling free energy profiles computed via MBAR.^{30,31} (b) Estimated solvation free energy as a function of the outer time step size (log scale). The error bars and shaded areas denote a 95% confidence interval according to the MBAR uncertainty estimator. Empty symbols discriminate double from triple time scale RESPA integration. The inner time step size is always equal to 0.5 fs and the middle time step, when used, is always equal to 3 fs.

The symbols in each curve of Figure 5a correspond to the relative free energies computed for the 21 simulated states. The solid lines were obtained through reweighting while the dashed lines are simple linear interpolations. The distinction between the scaling and softcore profiles is clear. For scaling, a sharp increase in the profile occurs when λ departs from 0, and a relatively high peak of about 4.2 kcal/mol must be overcome before the fully coupled state is reached, whose free energy is approximately 2.2 kcal/mol above that of the uncoupled state.

Both the scaling and softcore approaches result in profiles that converge, within statistical sampling error, to the same final free energy. Due to time step discretization effects and the use of different outer time steps, a systematic component is expected to exist in these deviations.^{67,68} It is important to determine whether such a component can be overlooked in practice

(especially in the case of more complex molecules, as will be discussed in Section 2.2). For this, we can analyze the computed solvation free energy as a function of the time step size. The results are given in Figure 5b. For better visualization, a logarithmic scale is used to represent the outer time step sizes. Each error bar corresponds to a 95% confidence interval (1.96 times the mean standard error) obtained from the MBAR analysis. As is clear from the figure, both coupling schemes yield standard errors of similar magnitudes. Perceivable changes in free energy occur when Δt changes from 1 to 3 fs and when it changes from 45 to 90 fs. We conjecture that these are systematic deviations due, in the first case, to an increase in the time gap between nonbonded force updates and, in the second, to the large increase in the value of Δt . From a practical standpoint, however, the variations in free energy over all time steps employed are within typical experimental uncertainties. For instance, an uncertainty of 0.6 kcal/mol is the default value in the FreeSolv database,¹ and the particular value for methane is 0.2 kcal/mol. The maximum error in Figure 5b is well within this value. Therefore, the SIN(R) method is adequate for computing hydration free energies of single Lennard-Jones particles using outer time steps up to 45 fs or even 90 fs.

3.4. Hydration Free Energy of Polar Solutes. In this section, we evaluate the performance of the SIN(R) method with large time steps in the calculation of hydration free energies of polyatomic molecules. Four polar molecules of distinct sizes, shapes, and charge distributions are considered, namely, methanol, phenol, 1,4-dioxane, and dibenzo-1,4-dioxin, whose structures are shown in Figure 6. With respect to their bonded

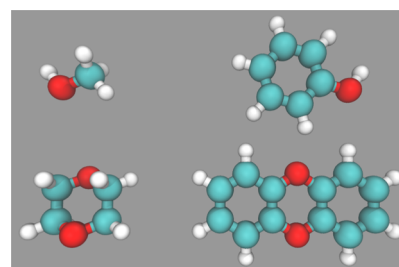


Figure 6. Structures of the polar molecules whose hydration free energies are computed in this work. Top-left, methanol; top-right, phenol; bottom-left, 1,4-dioxane; bottom-right, dibenzo-1,4-dioxin. The three-dimensional visualization was prepared using VMD⁷⁴ and Tachyon ray tracing.⁷⁵

and van der Waals interactions, the solute molecules are modeled according to the general AMBER force field (GAFF).⁶⁹ As is usual for this model, the partial atomic charges are computed by using the semi-empirical AM1-BCC method,⁷⁰ which is implemented in the Antechamber code.^{71,72} Once again, the fully flexible SPC-Fw water model⁵⁶ is used for the solvent. All other simulation details are exactly as described in Sections 3.1 and 3.3, including the numbers of molecules of each type and the procedures used to determine the box volume. The edge lengths that resulted from this procedure are 24.464 Å for methanol, 24.512 Å for phenol, 24.510 Å for 1,4-dioxane, and 24.560 Å for dibenzo-1,4-dioxin. Although such small boxes might appear susceptible to finite-size effects, this is not a particularly important issue as far as solvation free energies are concerned.⁷³

As usual, when electrostatic interactions play a role, we split the solvation free energy into two stages. First, the solute

coupling is carried out by completely disregarding the electrostatic interaction with the solvent. Only the Lennard-Jones contribution is considered, with its intensity modulated by a coupling parameter λ_{vdW} . We remark that the intramolecular interactions are always present, including the electrostatic components. Both the scaling and the softcore potential approaches are employed, as was done for the single particle in Section 3.3, and the calculation of ΔF_{vdW} is also done exactly as described in that section. In the second stage, with the Lennard-Jones interaction in its fully coupled state ($\lambda_{\text{vdW}} = 1$), we proceed with the coupling of the electrostatic contribution. This is done by a linear scaling such as

$$V_{\text{coulomb}}(r_{ij}, \lambda_{\text{coul}}) = \frac{\lambda_{\text{coul}}}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \quad (42)$$

where r_{ij} is the distance between a solute atom i and a solvent atom j with electric charges q_i and q_j , respectively. As already noted, the actual electrostatic calculation is performed via the PME algorithm.⁵¹ By simulating independent states with different values of $\lambda_{\text{coul}} \in [0, 1]$, we can finally apply MBAR to compute the total free energy difference ΔF_{coul} . Finally, the coupling free energy of a solute molecule is calculated as

$$\Delta F_{\text{coup}} = \Delta F_{\text{vdW}} + \Delta F_{\text{coul}} \quad (43)$$

To compute the van der Waals contribution to the solvation free energy, we employ the same sets of states as in Section 3.3. Once again, the total simulated time of each run is 4.5 ns, and the production time is 3.6 ns. All results in this section are obtained by using SIN(R) with a single thermostat per degree of freedom, that is, $L = 1$. We obtained free energy profiles via MBAR after carrying out, for every distinct size of the outer time step, 21 simulations using V_{scaling} and 21 simulations using V_{softcore} . The Supporting Information contains an analysis of these profiles, while here we focus on the total free energy difference computed in each case. Figure 7 depicts the van der Waals components of the coupling free energies obtained from the softcore potential simulations. Once again, increasing the outer step size causes small variations in the free energies obtained. The curves suggest systematic deviations, as expected due to the existence of discretization effects,^{67,68} but the magnitude of the uncertainties

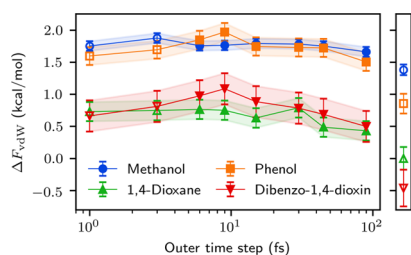


Figure 7. Estimated van der Waals contributions to the solvation free energies of different GAFF⁶⁹ solutes in SPC-Fw water.⁵⁶ The results are obtained via MBAR^{30,31} from independent simulations using the middle-scheme SIN(R) method with one thermostat per degree of freedom. The error bars and shaded areas denote 95% confidence intervals according to the MBAR uncertainty estimator. Empty symbols discriminate double from triple time scale RESPA integration. The inner time step size is always equal to 0.5 fs and the middle time step, when used, is always equal to 3 fs. Results shown in the narrow box on the right-hand side were obtained by switching off the long-range contribution of the nonbonded interactions.

and the presence of some possible outliers make it difficult to draw an unequivocal conclusion.

Clearly, the systems considered in Figure 7 are more challenging than that treated in Section 3.3 and would thus require longer production times to yield results with a similar degree of precision. Nevertheless, comparing uncertainties obtained for different systems after spending virtually the same amount of computational effort is a reasonable way of evaluating performance. In Figure 7, the magnitudes of the deviations and of the corresponding uncertainties are mutually consistent. In this sense, for estimating the van der Waals component of a coupling free energy, we argue that it is safe to employ outer time steps of tens of femtoseconds in the SIN(R) method without causing significant systematic deviations.

It is worth noting that if the force components assigned to the outer time scale had a negligible contribution to the free energy, then the SIN(R) method would play no role in keeping ΔF_{vdW} nearly independent of Δt . However, we can demonstrate that such force components are indispensable by switching them off completely and then analyzing the resulting free energies. This is shown in the narrow box on the right-hand side of Figure 7. These results were obtained by simply skipping the outermost operators of the RESPA integrator in eq 17. As one can see, they differ considerably from the free energies obtained when all interactions are included, thus confirming that the near constancy of ΔF_{vdW} is an actual achievement of the SIN(R) method.

As an additional test for the validity of the sampling algorithm,^{49,76} we compared the potential energy distributions obtained for dibenzo-1,4-dioxin in water at 298.15 and 306.15 K and checked how well the relation $\frac{\rho(U|\beta_2)}{\rho(U|\beta_1)} \propto e^{-(\beta_2 - \beta_1)U}$, where $\beta_i = \frac{1}{kT_i}$, is satisfied for different combinations of Δt and λ .

Details are given in the Supporting Information. As expected, the sampling quality diminishes somewhat as Δt approaches large values but is, nevertheless, satisfactory even for $\Delta t = 90$ fs.

Finally, we comment on the electrostatic component of the coupling free energy. Electrostatics are usually responsible for a significant shift in the total free energy value due to a decrease in the overall interaction potential between solute and solvent. However, as no cavity opening or major molecular rearrangements are necessary, it is relatively easy to switch the Coulomb interactions on and off once the van der Waals term is already switched on. Results that support this well-known aspect of computing solvation free energy are shown in the Supporting Information. Therefore, as the primary difficulty lies in the van der Waals contribution, our earlier conclusions regarding that part remain unchanged after the inclusion of the electrostatic contribution.

Before concluding this section, we comment briefly on gains in computational efficiency when such large time steps are used. In particular, the gains that can be achieved depend sensitively on the force decomposition employed, e.g., RESPA1 versus RESPA2,^{25,52,53} and general implementation details. In refs 25–27, we showed that factors of 10–20 could be gained in codes such as TINKER^{28,77} and PINY_MD⁷⁸ when large outer time steps are employed by ensuring that the short-range forces could be evaluated with an optimal speed relative to the long-range forces. As our purpose in this paper is to show that solvation free energy calculations could be performed accurately with large time steps, we have not focused on optimizing our OpenMM⁵⁰ implementation. In simulations of 512 flexible

water molecules carried out in a single GPU (model Nvidia Tesla V100-SXM2), we achieved a speed of 179 ns/day in a single time scale run with $\Delta t = 0.5$ fs, while the speed attained using 0.5/3/15 fs as the inner/middle/outer time steps was 645 ns/day. This corresponds to a 3.6 speed-up factor over an already impressively efficient simulation. On the other hand, increasing the outer step size up to 90 fs in the present OpenMM implementation resulted in a marginal gain, with a speed of 663 ns/day. Adding a solute molecule and its partially coupled interactions with the water does not change these numbers substantially. This illustrates the importance of optimizing the evaluation of bonded and short-range nonbonded forces when such large outer time steps are employed. Optimizing this part of the force evaluation in OpenMM will be the subject of future work.

3.5. Multiple Thermostats per Degree of Freedom.

Considering that the number L of thermostats per degree of freedom has a direct impact on the system evolution, the question naturally arises as to how beneficial it is to employ $L > 1$ in SIN(R) simulations. Some benefit is expected since eqs 13 and 14 show that, for a given temperature, larger values of L would impart higher speeds, on average, to the atoms with corresponding improvements in mixing and decorrelation. Clearly, increasing L to very large values increases the computational overhead with no clear gain in performance. Nevertheless, the effect of using large values of L is an interesting one from a numerical perspective and will be the subject of a future study.

To address the question of optimal choices of L , we repeated the simulations of Section 3.4 with four thermostats per degree of freedom ($L = 4$) instead of one. This number has been applied successfully in previous studies.^{25–27,29} In Figure 8, we show the

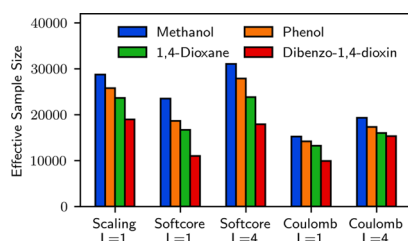


Figure 8. Sample sizes after subsampling are carried out by an integrated autocorrelation function method.⁴⁷ The original sizes are 84 000 for each Scaling and Softcore calculation and 44 000 for each Coulomb calculation.

size of the pooled sample obtained in every case after decorrelation is performed. We start by noting that, as anticipated in Section 3.4 for $L = 1$, the softcore-based calculations result in longer correlation times and, consequently, smaller sample sizes than their scaling-based counterparts. Despite this, the former calculations are able to produce results with smaller uncertainties due to their more effective exploration of the continuous phase-space regions that connect the uncoupled and fully coupled states. Another observation is that the solutes simulated possess similar degrees of difficulty in all types of calculations. Interestingly, we observe, in Figure 8, that using $L = 4$ always leads to better sampling than using $L = 1$, as expected. Average improvements of 46.8 and 30.9% in the numbers of uncorrelated configurations are achieved in the calculations of the van der Waals and electrostatic contributions, respectively.

Figure 9 shows the final solvation free energies obtained using MBAR on the combined samples. The fluctuations observed in

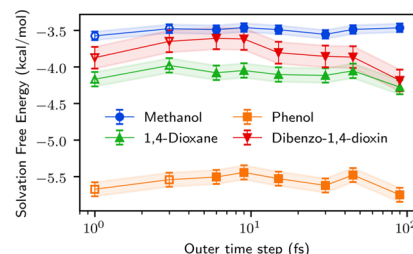


Figure 9. Final hydration free energy results. For each solute and outer time step size, MBAR analysis was carried out using the decorrelated samples obtained from SIN(R) simulations with both $L = 1$ and 4, combined. Error bars were determined by applying uncertainty propagation to eq 43.

the computed values are reasonable, particularly for the smallest solutes and for time step sizes up to 45 fs. In units of kcal/mol, the difference between the maximum and the minimum free energy estimates obtained with all Δt values up to 90 fs is 0.11 for methanol, 0.31 for phenol, 0.30 for 1,4-dioxane, and 0.57 for dibenzo-1,4-dioxin. If we consider $\Delta t \leq 45$ fs only, these numbers are 0.11, 0.23, 0.19, and 0.26, respectively. In both cases, these are within the standard 1 kcal/mol for chemical accuracy.

4. CONCLUSIONS

In this paper, we have tested a protocol for the computation of hydration free energies via multiple time scale molecular dynamics (MTS-MD) in which very large time steps can be used without resonance artifacts. Our results for organic solutes with various molecular sizes and charge distributions exhibit an acceptable variability when simulations with different parameters were applied to the same system condition. The innermost time step size is 0.5 fs in all cases. If bonded and nonbonded forces are split into two distinct time scales, perceivable but small shifts occur when the outer time step increases from 1 to 3 fs. By further splitting the nonbonded forces into short- and long-range contributions, thus adding one more time scale, and using 3 fs as the size of intermediate steps, we were able to employ outer steps ranging from 6 to 90 fs without observing substantial deviations in the computed free energies. The largest observed variations remained below 0.3 kcal/mol with time steps increasing up to 45 fs and below 0.6 kcal/mol if they were extended up to 90 fs. The free energies were also only marginally sensitive to the type of alchemical transformation applied for the van der Waals interactions; however, using a softcore potential proved to be generally better in terms of the resulting uncertainties than doing a simple scaling of the Lennard-Jones potential.

In addition to the solvation free energy calculations, we carried out a theoretical analysis of the isokinetic ensemble distribution of atomic velocities and its dependence on the number L of thermostats attached to each degree of freedom. This analysis allowed us to anticipate that having $L > 1$ is beneficial for the decorrelation of configurations sampled along a simulation, which we subsequently confirmed through statistical analysis.

After proving the principle that solvation free energies can be accurately computed via MTS-MD with very large time steps, an interesting follow-up study would be a performance comparison

between static λ protocols, such as the one employed here, and λ -dynamics approaches such as the adiabatic free energy dynamics (AFED) method. In addition, optimization of the calculation of bonded and short-range nonbonded forces in our implementation will be an important next step now that the proof of concept has been established.

One final comment on the incorporation of MTS into the framework of emerging machine learning (ML) potential models such as Gaussian process regression⁷⁹ and neural network^{80,81} potentials is in order. The promise of these methods is that when they are trained on high-quality *ab initio* data, they can then be used in lieu of these methods while retaining the accuracy but at a small fraction of the computational cost. However, these methods can still involve a high computational overhead compared to typical empirical force fields, and as a result, MTS methods have the ability to impact the use of these emerging models as well. For example, if a force field of reasonable quality exists for a system, then the training of the ML model could be focused on the difference between the *ab initio* and force field energies and forces, an approach known as Δ -learning. Not only would this require smaller training sets but also, after training, the force field could be used as the reference system in an MTS calculation with the ML model used as a correction applied with a large outer time step. In such a scheme, a long-range correction, either separately parameterized or extracted from the force field, could be applied along with the ML model or possibly using yet a third, larger step. These are questions that will also be explored in future work.

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.0c00698>.

Details of the simulation protocol, including modified potential for force-directed switching, alchemical coupling schemes for solute–solvent interactions, and ensemble validation and comparisons with experimental data (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Charles R. A. Abreu – Chemical Engineering Department, Escola de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21941-909, Brazil; orcid.org/0000-0002-6956-4045; Email: abreu@eq.ufrj.br

Mark E. Tuckerman – Department of Chemistry, New York University, New York, New York 10003, United States; Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, United States; NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China; orcid.org/0000-0003-2194-9955; Email: marktuckerman@nyu.edu

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.0c00698>

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

C.R.A.A. expresses gratitude to the Federal University of Rio de Janeiro for granting him a sabbatical year and to the New York University for hosting him as a visiting scholar. M.E.T.

acknowledges funding from the National Science Foundation through grants CHE-1565980 and CHE-1955381.

■ REFERENCES

- (1) 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.
- (2) Duarte Ramos Matos, G.; Kyu, D. Y.; Loeffler, H. H.; Chodera, J. D.; Shirts, M. R.; Mobley, D. L. Approaches for Calculating Solvation Free Energies and Enthalpies Demonstrated with an Update of the FreeSolv Database. *J. Chem. Eng. Data* **2017**, *62*, 1559–1569.
- (3) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150–160.
- (4) Mobley, D. L.; Klimovich, P. V. Perspective: Alchemical Free Energy Calculations for Drug Discovery. *J. Chem. Phys.* **2012**, *137*, No. 230901.
- (5) (a) 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. (b) Abel, R.; Wang, L.; Mobley, D. L.; Friesner, R. A. A Critical Review of Validation, Blind Testing, and Real-World Use of Alchemical Protein-Ligand Binding Free Energy Calculations. *Curr. Top. Med. Chem.* **2017**, *17*, 2577–2585.
- (6) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017**, *46*, 531–558.
- (7) Harger, M.; Li, D.; Wang, Z.; Dalby, K.; Lagardère, L.; Piquemal, J.-P.; Ponder, J.; Ren, P. Tinker-OpenMM: Absolute and Relative Alchemical Free Energies Using AMOEBA on GPUs. *J. Comput. Chem.* **2017**, *38*, 2047–2055.
- (8) Rizzi, A.; Murkli, S.; McNeill, J. N.; Yao, W.; Sullivan, M.; Gilson, M. K.; Chiu, M. W.; Isaacs, L.; Gibb, B. C.; Mobley, D. L.; Chodera, J. D. Overview of the SAMPL6 Host–Guest Binding Affinity Prediction Challenge. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 937–963.
- (9) Silveira, A.; Pereda, S.; Tavares, F.; Abreu, C. A Molecular Dynamics Study of the Solvation of Carbon Dioxide and Other Compounds in the Ionic Liquids [Emim][B(CN)₄] and [Emim][NTf₂]. *Fluid Phase Equilib.* **2019**, *491*, 1–11.
- (10) Duarte Ramos Matos, G.; Calabrò, G.; Mobley, D. L. Infinite Dilution Activity Coefficients as Constraints for Force Field Parameterization and Method Development. *J. Chem. Theory Comput.* **2019**, *15*, 3066–3074.
- (11) Radak, B. K.; Chipot, C.; Suh, D.; Jo, S.; Jiang, W.; Phillips, J. C.; Schulten, K.; Roux, B. Constant-pH Molecular Dynamics Simulations for Large Biomolecular Systems. *J. Chem. Theory Comput.* **2017**, *13*, 5933–5944.
- (12) 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.
- (13) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. New Approach to Monte Carlo Calculation of the Free Energy: Method of Expanded Ensembles. *J. Chem. Phys.* **1992**, *96*, 1776–1783.
- (14) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (15) Wu, P.; Hu, X.; Yang, W. λ -Metadynamics Approach To Compute Absolute Solvation Free Energy. *J. Phys. Chem. Lett.* **2011**, *2*, 2099–2103.
- (16) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci.* **2002**, *99*, 12562–12566.
- (17) Abrams, J. B.; Rosso, L.; Tuckerman, M. E. Efficient and Precise Solvation Free Energies via Alchemical Adiabatic Molecular Dynamics. *J. Chem. Phys.* **2006**, *125*, No. 074115.
- (18) Rosso, L.; Tuckerman, M. E. An Adiabatic Molecular Dynamics Method for the Calculation of Free Energy Profiles. *Mol. Simul.* **2002**, *28*, 91–112.
- (19) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible Multiple Time Scale Molecular Dynamics. *J. Chem. Phys.* **1992**, *97*, 1990–2001.

- (20) Schlick, T.; Mandziuk, M.; Skeel, R. D.; Srinivas, K. Nonlinear Resonance Artifacts in Molecular Dynamics Simulations. *J. Comput. Phys.* **1998**, *140*, 1–29.
- (21) Ma, Q.; Izaguirre, J. A.; Skeel, R. D. Verlet-I/R-RESPA/Impulse Is Limited by Nonlinear Instabilities. *SIAM J. Sci. Comput.* **2003**, *24*, 1951–1973.
- (22) Minary, P.; Martyna, G. J.; Tuckerman, M. E. Algorithms and Novel Applications Based on the Isokinetic Ensemble. I. Biophysical and Path Integral Molecular Dynamics. *J. Chem. Phys.* **2003**, *118*, 2510.
- (23) Minary, P.; Martyna, G. J.; Tuckerman, M. E. Algorithms and Novel Applications Based on the Isokinetic Ensemble. II. Ab Initio Molecular Dynamics. *J. Chem. Phys.* **2003**, *118*, 2527.
- (24) Minary, P.; Tuckerman, M. E.; Martyna, G. J. Long Time Molecular Dynamics for Enhanced Conformational Sampling in Biomolecular Systems. *Phys. Rev. Lett.* **2004**, *93*, No. 150201.
- (25) Leimkuhler, B.; Margul, D. T.; Tuckerman, M. E. Stochastic, Resonance-Free Multiple Time-Step Algorithm for Molecular Dynamics with Very Large Time Steps. *Mol. Phys.* **2013**, *111*, 3579–3594.
- (26) Margul, D. T.; Tuckerman, M. E. A Stochastic, Resonance-Free Multiple Time-Step Algorithm for Polarizable Models That Permits Very Large Time Steps. *J. Chem. Theory Comput.* **2016**, *12*, 2170–2180.
- (27) Chen, P.-Y.; Tuckerman, M. E. Molecular Dynamics Based Enhanced Sampling of Collective Variables with Very Large Time Steps. *J. Chem. Phys.* **2018**, *148*, No. 024106.
- (28) Albaugh, A.; Tuckerman, M. E.; Head-Gordon, T. Combining Iteration-Free Polarization with Large Time Step Stochastic-Isokinetic Integration. *J. Chem. Theory Comput.* **2019**, *15*, 2195–2205.
- (29) Zhang, Z.; Liu, X.; Yan, K.; Tuckerman, M. E.; Liu, J. Unified Efficient Thermostat Scheme for the Canonical Ensemble with Holonomic or Isokinetic Constraints via Molecular Dynamics. *J. Phys. Chem. A* **2019**, *123*, 6056–6079.
- (30) Kong, A.; McCullagh, P.; Meng, X.-L.; Nicolae, D.; Tan, Z. A Theory of Statistical Models for Monte Carlo Integration. *J. Royal Stat. Soc., Ser. B* **2003**, *65*, 585–604.
- (31) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, No. 124105.
- (32) Leimkuhler, B.; Matthews, C. *Molecular Dynamics With Deterministic and Stochastic Numerical Methods; Interdisciplinary Applied Mathematics*; Springer International Publishing: Cham, 2015; Vol. 39.
- (33) Tuckerman, M. E.; Mundy, C. J.; Martyna, G. J. On the Classical Statistical Mechanics of Non-Hamiltonian Systems. *Europhys. Lett.* **1999**, *45*, 149–155.
- (34) Tuckerman, M. E.; Liu, Y.; Ciccotti, G.; Martyna, G. J. Non-Hamiltonian Molecular Dynamics: Generalizing Hamiltonian Phase Space Principles to Non-Hamiltonian Systems. *J. Chem. Phys.* **2001**, *115*, 1678–1702.
- (35) Hörmander, L. *The Analysis of Linear Partial Differential Operators, Classics in Mathematics*, 2nd ed.; Springer: Berlin, New York, 2003.
- (36) Courant, R. *Differential and Integral Calculus*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1988; Vol. 2.
- (37) Martyna, G. J.; Klein, M. L.; Tuckerman, M. Nosé–Hoover Chains: The Canonical Ensemble via Continuous Dynamics. *J. Chem. Phys.* **1992**, *97*, 2635–2643.
- (38) Strang, G. On the Construction and Comparison of Difference Schemes. *SIAM J. Numer. Anal.* **1968**, *5*, 506–517.
- (39) Trotter, H. F. On the Product of Semi-Groups of Operators. *Proc. Am. Math. Soc.* **1959**, *10*, 545–551.
- (40) Suzuki, M. Generalized Trotter's Formula and Systematic Approximants of Exponential Operators and Inner Derivations with Applications to Many-Body Problems. *Commun. Math. Phys.* **1976**, *51*, 183–190.
- (41) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit Reversible Integrators for Extended Systems Dynamics. *Mol. Phys.* **1996**, *87*, 1117–1157.
- (42) Tuckerman, M. E. *Statistical Mechanics: Theory and Molecular Simulation; Oxford Graduate Texts*; Oxford University Press: Oxford, New York, 2010.
- (43) Zhang, Z.; Liu, X.; Chen, Z.; Zheng, H.; Yan, K.; Liu, J. A Unified Thermostat Scheme for Efficient Configurational Sampling for Classical/Quantum Canonical Ensembles via Molecular Dynamics. *J. Chem. Phys.* **2017**, *147*, No. 034109.
- (44) Leimkuhler, B.; Matthews, C. Rational Construction of Stochastic Numerical Methods for Molecular Sampling. *Appl. Math. Res. eXpress* **2012**, *14*, 17.
- (45) Leimkuhler, B.; Matthews, C. Robust and Efficient Configurational Molecular Sampling via Langevin Dynamics. *J. Chem. Phys.* **2013**, *138*, No. 174102.
- (46) Li, D.; Han, X.; Chai, Y.; Wang, C.; Zhang, Z.; Chen, Z.; Liu, J.; Shao, J. Stationary State Distribution and Efficiency Analysis of the Langevin Equation via Real or Virtual Dynamics. *J. Chem. Phys.* **2017**, *147*, No. 184104.
- (47) Chodera, J. D.; Swope, W. C.; Pitera, J. W.; Seok, C.; Dill, K. A. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *J. Chem. Theory Comput.* **2007**, *3*, 26–41.
- (48) Ben-Naim, A.; Marcus, Y. Solvation Thermodynamics of Nonionic Solutes. *J. Chem. Phys.* **1984**, *81*, 2016–2027.
- (49) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. Extremely Precise Free Energy Calculations of Amino Acid Side Chain Analogs: Comparison of Common Molecular Mechanics Force Fields for Proteins. *J. Chem. Phys.* **2003**, *119*, 5740–5761.
- (50) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLOS Comput. Biol.* **2017**, *13*, No. e1005659.
- (51) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (52) Zhou, R.; Harder, E.; Xu, H.; Berne, B. J. Efficient Multiple Time Step Method for Use with Ewald and Particle Mesh Ewald for Large Biomolecular Systems. *J. Chem. Phys.* **2001**, *115*, 2348–2358.
- (53) Morrone, J. A.; Zhou, R.; Berne, B. J. Molecular Dynamics with Multiple Time Scales: How to Avoid Pitfalls. *J. Chem. Theory Comput.* **2010**, *6*, 1798–1804.
- (54) Chow, K.-H.; Ferguson, D. M. Isothermal-Isobaric Molecular Dynamics Simulations with Monte Carlo Volume Sampling. *Comput. Phys. Commun.* **1995**, *91*, 283–289.
- (55) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular Dynamics Simulations of Water and Biomolecules with a Monte Carlo Constant Pressure Algorithm. *Chem. Phys. Lett.* **2004**, *384*, 288–294.
- (56) Wu, Y.; Tepper, H. L.; Voth, G. A. Flexible Simple Point-Charge Water Model with Improved Liquid-State Properties. *J. Chem. Phys.* **2006**, *124*, No. 024503.
- (57) 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.
- (58) Pham, T. T.; Shirts, M. R. Identifying Low Variance Pathways for Free Energy Calculations of Molecular Transformations in Solution Phase. *J. Chem. Phys.* **2011**, *135*, No. 034114.
- (59) 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*, No. 124120.
- (60) Buelens, F. P.; Grubmüller, H. Linear-Scaling Soft-Core Scheme for Alchemical Free Energy Calculations. *J. Comput. Chem.* **2012**, *33*, 25–33.
- (61) 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.
- (62) Naden, L. N.; Pham, T. T.; Shirts, M. R. Linear Basis Function Approach to Efficient Alchemical Free Energy Calculations. I. Removal of Uncharged Atomic Sites. *J. Chem. Theory Comput.* **2014**, *10*, 1128–1149.

- (63) Naden, L. N.; Shirts, M. R. Linear Basis Function Approach to Efficient Alchemical Free Energy Calculations. 2. Inserting and Deleting Particles with Coulombic Interactions. *J. Chem. Theory Comput.* **2015**, *11*, 2536–2549.
- (64) Siepmann, J. I.; Martin, M. G. Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B* **1998**, *102*, 2569–2577.
- (65) Aimoli, C. G.; Maginn, E. J.; Abreu, C. R. A. Force Field Comparison and Thermodynamic Property Calculation of Supercritical CO₂ and CH₄ Using Molecular Dynamics Simulations. *Fluid Phase Equilib.* **2014**, *368*, 80–90.
- (66) Abraham, M. H. Thermodynamics of Solution of Homologous Series of Solutes in Water. *J. Chem. Soc., Faraday Trans. 1* **1984**, *80*, 153.
- (67) Eastwood, M. P.; Stafford, K. A.; Lippert, R. A.; Jensen, MØ.; Maragakis, P.; Predescu, C.; Dror, R. O.; Shaw, D. E. Equipartition and the Calculation of Temperature in Biomolecular Simulations. *J. Chem. Theory Comput.* **2010**, *6*, 2045–2058.
- (68) Silveira, A. J.; Abreu, C. R. A. Refinement of Thermostated Molecular Dynamics Using Backward Error Analysis. *J. Chem. Phys.* **2019**, *150*, No. 114110.
- (69) 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.
- (70) (a) 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. (b) 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.
- (71) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
- (72) Case, D. A.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 2017*; University of California, 2017.
- (73) Parameswaran, S.; Mobley, D. L. Box Size Effects Are Negligible for Solvation Free Energies of Neutral Solutes. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 825–829.
- (74) Humphrey, W.; Dalke, A.; Schulten, K. VMD—Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (75) Stone, J. An Efficient Library for Parallel Ray Tracing and Animation. Ph.D. Thesis, University of Missouri-Rolla, 1998.
- (76) Merz, P. T.; Shirts, M. R. Testing for Physical Validity in Molecular Simulations. *PLoS One* **2018**, *13*, No. e0202764.
- (77) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.* **2018**, *14*, 5273–5289.
- (78) Tuckerman, M. E.; Yarne, D.; Samuelson, S. O.; Hughes, A. L.; Martyna, G. J. Exploiting Multiple Levels of Parallelism in Molecular Dynamics Based Calculations via Modern Techniques and Software Paradigms on Distributed Memory Computers. *Comput. Phys. Commun.* **2000**, *128*, 333–376.
- (79) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, No. 136403.
- (80) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, No. 146401.
- (81) Nguyen, T. T.; Székely, E.; Imbalzano, G.; Behler, J.; Csányi, G.; Ceriotti, M.; Götz, A. W.; Paesani, F. Comparison of Permutationally Invariant Polynomials, Neural Networks, and Gaussian Approximation Potentials in Representing Water Interactions through Many-Body Expansions. *J. Chem. Phys.* **2018**, *148*, No. 241725.