ORIGINAL ARTICLES

Applying Efficient Implicit Nongeometric Constraints in Alchemical Free Energy Simulations

Jennifer L. Knight,[a,b] and Charles L. Brooks III*[a,b]

Several strategies have been developed for satisfying bond lengths, angle, and other geometric constraints in molecular dynamics simulations. Advanced variations of alchemical free energy perturbation simulations, however, also require nongeometric constraints. In our recently developed multisite λ-dynamics simulation method, the conventional λ parameters that are associated with the progress variables in alchemical transformations are treated as dynamic variables and are constrained such that: $0 \leq \lambda_i \leq 1$ and $\sum_{i=1}^{N} \lambda_i = 1$. Here, we present four functional forms of λ that implicitly satisfy these nongeometric constraints, whose values and forces are facile to compute and that yield stable simulations using a 2 fs integration timestep. Using model systems, we present the sampling characteristics of these functional forms and demonstrate the enhanced sampling profiles and improved convergence rates that are achieved by the functional form: $\lambda_i = \frac{e^{c \sin \theta_i}}{\sum_{j=1}^{N} e^{c \sin \theta_j}}$ that oscillates between $\lambda_i = 0$ and $\lambda_i = 1$ and has relatively steep transitions between these endpoints. © 2011 Wiley Periodicals, Inc. J Comput Chem 32: 3423–3432, 2011

Keywords: θ-dynamics · λ-dynamics · constraint

Introduction

Effectively including constraints in simulation methods is critical to achieve optimal sampling efficiency. These constraints limit the phase space that is explored so that sampling is focused in the regions of greatest interest. For example, in molecular dynamics (MD) simulations, bond length constraints are often used to eliminate the high-frequency motions that are associated with hydrogen atoms. These rapid oscillations do not significantly affect the longer timescale processes under investigation and yet require a small integration timestep to ensure the numerical stability of the simulations. Therefore, utilizing these hydrogen bond constraints allows for larger integration timestep and thus effectively generates longer trajectory lengths.

In unconstrained MD simulations where molecular models are represented in Cartesian coordinates, the equations of motion are described by a series of ordinary differential equations (ODEs). When rigid (holonomic) constraints are incorporated into these models, the equations of motion become significantly more complex. In the Lagrangian equations of motion, the forces of the constraints appear explicitly and the dependence of these forces on the positions and velocities of the centers of force is obtained from the corresponding set of constraint equations that contain undetermined Lagrange multipliers.[1] These equations can be solved to determine the constraint forces; however, because in MD simulations the equations of motion are solved approximately using finite difference methods, the constraints will gradually diverge from the target values.[2] In practice, this strategy for solving the equations of motion generally requires integration timesteps that are significantly smaller than the timescales of motion the constraints are seeking to eliminate, so is often impractical to implement.[2]

An alternative strategy for satisfying holonomic constraints is implemented by the family of SHAKE algorithms. In the SHAKE algorithm,[2] the equations of motion are solved in an unconstrained manner according to the ODEs to obtain an initial estimate of the new conformation and then the positions of the specific atoms are modified iteratively until all constraints are satisfied within a given tolerance level. Related algorithms also constrain the velocities (RATTLE[3]) and accelerations (WIGGLE[4]); other variants of the algorithm are specific for given topologies, for example, linear and ring systems (MILC SHAKE[5,6]) or semirigid molecules (Q-SHAKE[7], SETTLE[8]).

Another strategy for satisfying holonomic constraints in MD simulations may be described as using "implicit constraints," that is, using a functional form of the coordinate variables themselves to ensure that the constraints will be satisfied. This strategy has been adopted for modeling rigid molecules where Euler angles [9] or quaternions[10] are used to describe the rotational degrees of freedom of the system. This strategy has also been implemented in torsion-angle MD in which internal coordinates are used to define and sample atomic positions.[11–13] In this strategy, rigid units within a molecule are defined and atoms within these units remain fixed with respect to one another while the relative positions of the rigid units are sampled. Thus, the equations of motion are reduced to the usual ODEs and the geometric constraints that would

[a] J. L. Knight, C. L. Brooks III
Department of Chemistry, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109
E-mail: brookscl@umich.edu
[b] J. L. Knight, C. L. Brooks III
Department of Biophysics, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109

otherwise be required to ensure the appropriate rigidity of the system are satisfied at every timestep.

Nongeometric holonomic constraints have also been utilized in simulation methods and can be implemented using strategies that are analogous to those used to satisfy geometric constraints. λ-Dynamics simulations is an extension of alchemical free energy simulations in which the conventional $\{\lambda\}$ parameters that are associated with the progress variables in the chemical coordinates are treated as dynamic but constrained variables. In traditional free energy simulations in which one ligand is alchemically transformed into another, a nonphysical hybrid molecule is often constructed. In this case, atoms that are common to both ligands are represented once as a common core and are treated as "environment" atoms in the Hamiltonian. The atoms that are unique to each ligand are represented by individual noninteracting moieties that are attached to the common core. The corresponding hybrid potential energy function is defined by:

$$V(X, \{x\}) = V_{\text{env}}(X) + \lambda_1(X, x_1) + \lambda_2(X, x_2) \quad (1)$$

where X and x_i are the coordinates of environmental atoms and of those atoms which are unique to ligand i, respectively; V_{env} is the potential energy involving the environmental atoms only, $V(X, x_i)$ is the interaction energy between ligand i and the environment atoms and where the nongeometric constraints are defined by:

$$0 \leq \lambda_i \leq 1 \quad (2a)$$

$$\lambda_1 + \lambda_2 = 1 \quad (2b)$$

In λ-dynamics, the hybrid ligand is extended to N ligands where the hybrid potential energy function is defined by:

$$V(X, \{x\}) = V_{\text{env}}(X) + \sum_{i=1}^{N} \lambda_i(X, x_i) \quad (3)$$

where the nongeometric constraints are now extended to:

$$0 \leq \lambda_i \leq 1 \quad (4a)$$

$$\sum_{i=1}^{N} \lambda_i = 1 \quad (4b)$$

The hybrid Hamiltonian that governs the λ-dynamics simulations is defined by:

$$H_O(X, \{x\}, \{\lambda\}) = T_X + T_{\{x\}} + T_\lambda + V_{\text{env}}(X) + \sum_{i=1}^{N} \lambda_i V(X, x_i) \quad (5)$$

The original implementation of λ-dynamics in the chemistry at HARvard macromolecular mechanics (CHARMM) macromolecular modeling package[14,15] directly satisfied the constraints in Eq. (4). Specifically, the Lagrange multiplier method was used to determine the explicit constraint forces of λ and a subsequent renormalization of the λ positions and velocities was performed at every timestep to reduce the accumulation of small errors. However, due to the sensitivity of the total energy of the system to small changes in $\{\lambda\}$ at the λ endpoints, small integration timesteps are required to retain the stability of the numerical integrator for long trajectory lengths. In this study, we explore the use of implicit constraints in λ-dynamic simulations, namely implicitly satisfying the holonomic constraints on λ by judicious selection of a functional form of $\{\lambda\}$.

This strategy for implicitly satisfying nongeometric holonomic constraints has been implemented in contexts where only two related λs are being sampled simultaneously. Given the constraints listed in Eq. (2), this problem can be reduced to one-dimension with $\lambda_2 = 1 - \lambda_1$. For example, constant pH-MD[16–18] simulations adopt hybrid molecules and corresponding hybrid potential energy functions that are similar to those that are used in traditional alchemical free energy simulations. However, in constant pH-MD simulations, the "λ" is a dynamic variable that is a function of θ; θ is a volumeless particle with fictious mass and is propagated throughout the course of the simulation. The hybrid Hamiltonian that is used to govern the dynamics of the simulation is described by:

$$H_O(X, \{x\}, \{\lambda(\theta)\}) = T_X + T_{\{x\}} + T_\theta + V_{\text{env}}(X) + \lambda_1(\theta)V(X, x_1) + \lambda_2(\theta)V(X, x_2) \quad (6)$$

which have the same nongeometric constraints as are listed in Eq. (2). In this case, the coefficients $\{\lambda(\theta)\}$ describe the relative presence ($\lambda_1 = 0$; $\lambda_2 = 1$) or absence ($\lambda_1 = 1$; $\lambda_2 = 0$) of a hydrogen atom on a titratable amino acid and are defined by:

$$\lambda_1 = \sin^2 \theta \text{ and } \lambda_2 = 1 - \sin^2 \theta \quad (7)$$

Thus, by using this functional form for the λ values and sampling $\{\theta\}$ throughout the simulations, the nongeometric constraints are exactly satisfied at every timestep.

In another example, nongeometric implicit constraints are used in logistic regression for predicting event probabilities in which the logist or sigmoid function is defined by:

$$\lambda = \frac{1}{1 + e^\theta} \quad (8)$$

This function can be expressed in terms of two related variables:

$$\lambda_1 = \frac{e^\theta}{1 + e^\theta} \text{ and } \lambda_2 = \frac{1}{1 + e^\theta} \quad (9)$$

that implicitly satisfy the two constraints in Eq. (2). Variations of this function are applied in a wide variety of fields to model, for example, population growth,[19] nonlinearity in neural networks,[20] and sigmoidal behavior of dose–response curves.[21]

In constant pH-MD [Eq. (7)] and logistic regression [Eq. (9)], only two related variables are constrained and implicitly satisfy the nongeometric constraints in the simulations given the specific functional form of the variables themselves. However, λ-dynamics simulations require N λ variables to be sampled simultaneously and constrained. We are not aware of any implemented strategies in molecular simulations for defining N

related variables that implicitly satisfy the corresponding non-geometric constraints in Eq. (4). Here, we first explore the sampling profiles of eqs. (7) and (9) for defining two related λ variables and then we present two functional forms of $\{\lambda\}$ that implicitly enable nongeometric constraints for N related λs to be satisfied simultaneously. The first new functional form is based on the constant pH-MD formalism such that for a given set of λs:

$$\lambda_{\alpha,i}^{N\sin} = \frac{\sin^2 \theta_{\alpha,i}}{\sum\limits_{j=1}^{N} \sin^2 \theta_{\alpha,j}} \tag{10}$$

and the second new functional form is based on the logist function such that:

$$\lambda_{\alpha,i}^{N\exp} = \frac{e^{c\,\sin\,\theta_{\alpha,i}}}{\sum\limits_{j=1}^{N} e^{c\,\sin\,\theta_{\alpha,j}}} \tag{11}$$

In this study, we explore the sampling characteristics of these functional forms of $\{\lambda\}$ using our recently developed multisite λ-dynamics (MSλD) alchemical free energy simulations.[22] This simulation strategy is an extension of λ-dynamics in which $\{\lambda\}$ are treated as dynamic variables but with hybrid models that can have N distinct substituents at M sites on a common ligand framework.[23,24] For each functional form of $\{\lambda\}$, we evaluate the resulting relative free energy differences estimated by MSλD in vacuum and/or solvent environments for series of identical benzene, dihydrobenzene, or dimethoxy-benzene molecules to characterize its sampling behavior. The $\lambda^{N\exp}$ functional form clearly has the optimal sampling profile for these MSλD simulations; it enables facile transitions between $\lambda_i \approx 1$ and $\lambda_i \approx 0$, spends a significant amount of time sampling at the endpoints rather than the physically irrelevant intermediate values of $\{\lambda\}$, is easy to compute, and leads to numerically stable simulations.

Methods

Multisite λ-dynamics theory

In MSλD, we have extended the hybrid potential energy function to include multiple chemical modifications (i.e., substituents) at multiple sites on a common ligand core to be:

$$V(X, \{x\}, \{\lambda\}) = V_{\text{env}}(X) + \sum_{S=1}^{N_{\text{sites}}} \sum_{i=1}^{L_S} \lambda_{S,i}(V(X, x_{S,i}) - F_{S,i})$$
$$+ \sum_{S=1}^{N_{\text{sites}}-1} \sum_{i=1}^{L_S} \sum_{T=S+1}^{N_{\text{sites}}} \sum_{j=1}^{L_T} \lambda_{S,i}\lambda_{T,j}(V(x_{S,i}, x_{T,j})) \tag{12}$$

where the constraints are now given to be:

$$0 \leq \lambda_{\alpha,i} \leq 1 \text{ and } \sum_{i=1}^{L_\alpha} \lambda_{\alpha,i} = 1 \text{ for each site } \alpha, \tag{13}$$

where N_{sites} is the total number of sites which contain multiple substituents, L_S is the number of substituents at site S, and $F_{S,i}$

is a precalculated biasing potential that can enhance the sampling of each $\lambda_{S,i}$ state. The double summation in the second term of the hybrid potential accounts for the interactions between the environment and each substituent at each site in the system. The third term accounts for the interactions between each substituent and the substituents modeled on all other sites. Note that substituents at a given site do not "see" each other in these simulations.

A substituent at a given site is described to be "dominant" or "present" when its corresponding λ value approaches 1. A ligand is described to be dominant or present when the λ values associated with its constituent substituents are dominant at the same time. For systems with two sites, the relative free energies between two ligands is then computed by:

$$\Delta\Delta G_{1,i;2,j \to 1,k;2,l} = -k_B T \ln \frac{P(\lambda_{1,k} = 1, \lambda_{2,l} = 1)}{P(\lambda_{1,i} = 1, \lambda_{2,j} = 1)} \tag{14}$$

where $P(\lambda_{1,i} = 1;\{\lambda_{1,m \neq i} = 0\},\lambda_{2,j} = 1;\{\lambda_{2,m \neq j} = 0\})$ corresponds to the amount of time that substituent i is present at site 1 and substituent j is present at site 2, that is, when $\lambda_{1,i} \approx 1$ and $\lambda_{2,j} \approx 1$ concurrently during the λ-dynamics simulation. In practice, the amount of time $\lambda_{1,i} \approx 1$ is approximated by using a threshold, say $\lambda_{1,i} > 0.8$. MSλD has been implemented in the CHARMM macromolecular software package.[14,15]

Functional forms of λ

The functional forms of $\{\lambda\}$ that are assessed in this work are listed in Table 1 along with their corresponding partial derivatives with respect to θ. Lookup tables were used to efficiently approximate $\lambda^{N\exp}$.[25] Using this formalism for MSλD, it is the values of θ that have fictious masses, m_θ, and are propagated through the equations of motion, not the λ values directly. Thus, the extended Hamiltonian is:

$$H_o(X, \{x\}, \{\lambda(\theta)\}) = T_x + T_\theta + V(X, \{x\}, \{\lambda(\theta)\}) \tag{15}$$
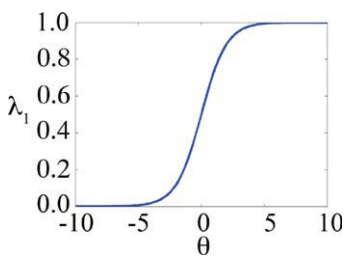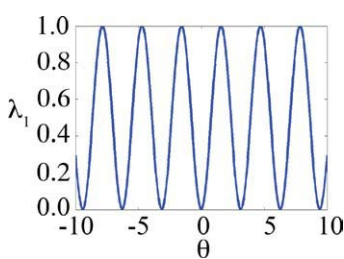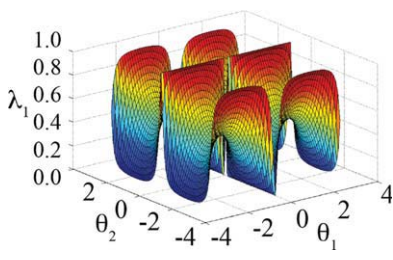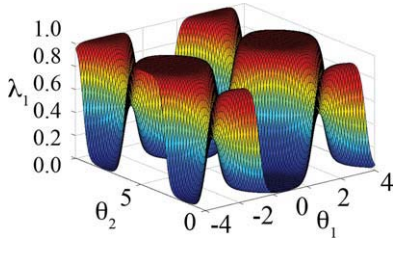
The Leapfrog Verlet algorithm is used to integrate the equations of motions and the forces on each θ are calculated by:

$$F_{\theta_{\alpha,i}} = -\frac{\partial V}{\partial \theta_{\alpha,i}} = -\frac{\partial \lambda_{\alpha,i}}{\partial \theta_{\alpha,i}}\left(V(X, x_{\alpha,i}) - F_{\alpha,i} + \sum_{T \neq \alpha}^{N_{\text{sites}}} \sum_{k=1}^{L_T} \lambda_{T,k}(V(x_{\alpha,i}, x_{T,k}))\right)$$
$$+ \sum_{j \neq i}^{L_\alpha} \frac{\partial \lambda_{\alpha,j}}{\partial \theta_{\alpha,i}}\left(V(X, x_{\alpha,j}) - F_{\alpha,j} + \sum_{T \neq \alpha}^{N_{\text{sites}}} \sum_{k=1}^{L_T} \lambda_{T,k}(V(x_{\alpha,j}, x_{T,k}))\right) \tag{16}$$

Model systems

Model hybrid ligands were constructed to represent multiple identical benzene, dihydroxybenzene, or dimethoxybenzene molecules. Each hybrid benzene molecule contained a single benzene ring with N distinct pairs of hydrogen and ipso carbon atoms at one or two sites on the common benzene ring (see Fig. 1A). As each C—H pair in the para-position interacts with each C—H pair at the ipso-position, the hybrid molecule with multiple substituents at two sites represents $N_{\text{site1}} \times N_{\text{site2}}$ distinct yet identical molecules. Similarly, each hybrid

Table 1. Summary of the functional forms for $\{\lambda\}$ that are used in this study where $\lambda_{\alpha,i}$ represents λ for the *i*th substituent at site α.

Functional form	$N_{\text{sub/site}}$	$\lambda(\theta)$	$\frac{\partial\lambda(\theta)}{\partial\theta}$	Schematic of $\lambda_{\alpha,1}$
$\lambda^{2\text{exp}}$	2	$\lambda_{\alpha,1} = \dfrac{e^{\theta_\alpha}}{1+e^{\theta_\alpha}};$ $\lambda_{\alpha,2} = \dfrac{1}{1+e^{\theta_\alpha}}$	$\dfrac{\partial\lambda_{\alpha,1}}{\partial\theta_\alpha} = \lambda_{\alpha,1}\lambda_{\alpha,2};$ $\dfrac{\partial\lambda_{\alpha,2}}{\partial\theta_\alpha} = -\lambda_{\alpha,1}\lambda_{\alpha,2}$	
$\lambda^{2\text{sin}}$	2	$\lambda_{\alpha,1} = \sin^2\theta_\alpha;$ $\lambda_{\alpha,2} = 1 - \sin^2\theta_\alpha$	$\dfrac{\partial\lambda_{\alpha,1}}{\partial\theta_\alpha} = 2\sin\theta_\alpha\cos\theta_\alpha;$ $\dfrac{\partial\lambda_{\alpha,2}}{\partial\theta_\alpha} = -2\sin\theta_\alpha\cos\theta_\alpha$	
$\lambda^{N\text{sin}}$	N	$\lambda_{\alpha,i} = \dfrac{\sin^2\theta_{\alpha,i}}{\sum\limits_{j=1}^{N}\sin^2\theta_{\alpha,j}}$	$\dfrac{\partial\lambda_{\alpha,i}}{\partial\theta_{\alpha,i}} = 2\cot\theta_{\alpha,i}\lambda_{\alpha,i}(1-\lambda_{\alpha,i});$ $\dfrac{\partial\lambda_{\alpha,j\neq i}}{\partial\theta_{\alpha,i}} = -2\cot\theta_{\alpha,i}(\lambda_{\alpha,i}\lambda_{\alpha,j})$	
$\lambda^{N\text{exp}}$	N	$\lambda_{\alpha,i} = \dfrac{e^{c\sin\theta_{\alpha,i}}}{\sum\limits_{j=1}^{N}e^{c\sin\theta_{\alpha,j}}}$	$\dfrac{\partial\lambda_{\alpha,i}}{\partial\theta_{\alpha,i}} = c\cos\theta_{\alpha,i}\lambda_{\alpha,i}(1-\lambda_{\alpha,i});$ $\dfrac{\partial\lambda_{\alpha,j\neq i}}{\partial\theta_{\alpha,i}} = -c\cos\theta_{\alpha,i}\lambda_{\alpha,i}\lambda_{\alpha,j}$	

dihydroxybenzene molecule consisted of a single benzene ring with N hydroxy groups and ipso carbon atoms at two sites on the common ring and each hybrid dimethoxybenzene molecule consisted of a single benzene ring with N methoxy groups and ipso carbon atoms at two sites on the common ring (see Fig. 1B). Parameters and partial charges for the model systems were assigned from the recently developed CHARMM general force field.[26] The hybrid molecules are identified in the text by the names "$N_{\text{site1substituent}} \times N_{\text{site2substituent}}$" where substituents "H," "OH," and "OCH$_3$" designate the hydrogen atoms, hydroxy and methoxy groups, respectively.

Simulation details

The Leapfrog Verlet algorithm was used to integrate the equations of motion and propagate the atomic coordinates, atomic velocities as well as the θ values and their velocities. For all simulations, a nonbonded cutoff of 15 Å was used with an electrostatic force shifting function and a van der Waals switching function between 10 and 12 Å. Hydrogen bonds were constrained using the SHAKE[27] algorithm and the integration timestep was 2 fs. Linear scaling by λ was applied to all energy terms except the bond and angle terms which were treated at full strength regardless of λ value to retain physically reasonable geometries. Each θ_i was assigned a fictious
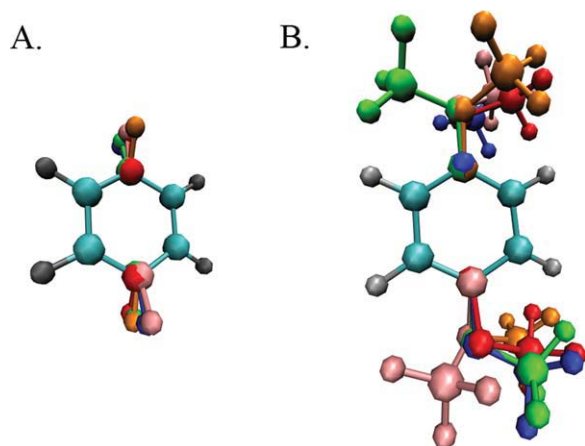
Figure 1. Schematic representation of three model systems used to assess the quality of the functional forms of λ in MSλD simulations. Hybrid molecules representing multiple identical benzene molecules by modeling distinct sets of hydrogen and corresponding ipso carbon atoms at A) sites 1 and 4 on a common benzene core. B) Hybrid molecule representing multiple dimethoxybenzene molecules at two sites on a common benzene core.

mass of 12 amu·Å^2 and λ values were saved every 10 steps. Solvent simulations were performed using 351 TIP3P[28] water molecules in a water box of 22 Å^3 with periodic boundary conditions. The temperature was maintained near 310 K by coupling to a Langevin heat bath using a frictional coefficient of 10 ps^{-1} for all atoms and 5 ps^{-1} for each θ_i. Production runs were 25 and 2 ns for vacuum and solvation simulations, respectively, and the threshold value for assigning $\lambda_{i,\alpha} \approx 1$ was $\lambda_{i,\alpha} \geq 0.8$ unless otherwise stated. Several values of c were assessed for the functional form λ^{Nexp} and results are reported for $c = 5.5$ unless otherwise specified. Ten simulations with different initial seed values for θ_i were performed for each combination of parameters and the resulting averages and standard deviations were reported. All simulations were performed on dual 2.66 GHz Intel Quad Core Xeon processors.

Model quality

All MSλD trajectories have been analyzed using new routines that we have implemented in CHARMM. The relative free energy difference for each pair of compounds (ij) that are represented in the hybrid molecules was estimated by averaging over results from 10 simulation trajectories. The average unsigned error (AUE), standard deviations (σ), and maximum errors that are reported in the tables and text represent the statistics compiled over all relative free energies that are estimated for the NP (i.e., $N(N-1)/2$) pairs of compounds in the hybrid molecule in the 10 simulation trajectories, for example:

$$\text{AUE} = \frac{1}{\text{NP}} \sum_{ij=1}^{\text{NP}} \left| \frac{1}{10} \sum_{k=1}^{10} \left(\Delta\Delta G_k(ij) \right) \right| \qquad (17)$$

where $\Delta\Delta G_k(ij)$ represents the relative free energy difference that is calculated between the ith and jth ligand from the kth trajectory.

Results

The functional forms of $\{\lambda\}$ that were explored in this study are summarized in Table 1 along with their corresponding first derivatives with respect to θ, which are required to calculate the forces on θ in Eq. (16). In each case, the constraints described in Eq. (2) are satisfied at every timestep. For the purposes of assessing the sampling properties of these functional forms in the MSλD free energy simulations, model compounds have been constructed that represent multiple identical molecules. These molecules are identical to one another in their structure and their force field parameters; thus, regardless of the environment, the relative free energy differences between any two molecules is exactly 0 kcal mol^{-1}. Therefore, any deviations in the simulation estimates from 0 kcal mol^{-1} can be understood as errors due to limitations in the MSλD sampling specifically. In the most simple simulation scenerio, multiple benzene compounds effectively "compete" with each other in vacuum. More flexible and thus more complicated cases are also considered with the dihydroxybenzene and then dimethoxybenzene hybrid models.

Implicit constraints for two variables: $\lambda^{2\text{exp}}$, $\lambda^{2\text{sin}}$

The first functional form of the implicit constraints that we consider is based on the logist function and the results for sampling the different model ligands using this functional form of $\{\lambda\}$ are summarized in Table 2. In this case, the functional form of $\{\lambda\}$ is:

$$\lambda_{\alpha,1}^{2\exp} = \frac{e^{\theta_\alpha}}{1 + e^{\theta_\alpha}} \text{ and } \lambda_{\alpha,2}^{2\exp} = \frac{1}{1 + e^{\theta_\alpha}} \qquad (18)$$

and this construct implicitly satisfies both constraints in Eq. (2). While this functional form never allows $\lambda_{\alpha,i} = 1$ and $\lambda_{\alpha,i} = 0$, exactly, the $\lambda_{\alpha,i}$ values approach sufficiently close to 1 and 0 to be of practical use. The implementation of this form of the constraints is very stable with timesteps up to 2 fs for trajectory lengths up to 25 ns. However, large average errors (0.6–1.3 kcal mol^{-1}), standard deviations (0.4–1.1 kcal mol^{-1}), and maximum errors (1.3–2.4 kcal mol^{-1}) are observed in the estimated relative free energies. The convergence of the ligand

Table 2. Quality of relative free energy estimates for 25 ns MSλD simulations using implicit constraints: $\lambda^{2\text{exp}}$ and $\lambda^{2\text{sin}}$ in vacuum.

Functional form	Hybrid ligand nSite1 $\times$ nSite4	Δt (fs)	τ_{trans} (ps^{-1})	$\Delta\Delta G$ (kcal mol^{-1}) AUE	σ	Max
$\lambda^{2\text{exp}}$	2H $\times$ 2H	2.0	0.02	1.194	1.114	2.366
	2OH $\times$ 2OH	2.0	0.02	1.310	0.876	2.001
	2OCH$_3$ $\times$ 2OCH$_3$	1.5	0.002	0.597	0.411	1.281
$\lambda^{2\text{sin}}$	2H $\times$ 2H	2.0	1.34	0.006	0.003	0.011
	2OH $\times$ 2OH	2.0	1.32	0.004	0.002	0.007
	2OCH$_3$ $\times$ 2OCH$_3$	0.5	0.90	0.010	0.005	0.017

The integration timestep is Δt and τ_{trans} is the average frequency of the change in the identity of the substituent with $\lambda \approx 1$ on each site. Statistics are averaged over the six pairs of compounds in the model hybrid ligands.

populations is quite slow due to the infrequent exchanges between $\lambda_{\alpha,1} = 1$ and $\lambda_{\alpha,2} = 1$. Essentially, once one benzene molecule for example becomes the dominant ligand, that is, one substituent at site 1 has its λ value assigned to 1 and one substituent at site 2 has its λ value assigned to 1, it is difficult to drive θ_1 and θ_2 into regimes where other combinations of substituents will have $\lambda = 1$ such that one of the other benzene molecule becomes the dominant ligand. Therefore, the relative free energy differences which are computed from the relative probabilities of each molecule being the dominant ligand in Eq. (13) is severely biased by the combination of substituents that first reach $\lambda = 1$ and thus first identify a dominant ligand.

The second functional form of $\{\lambda\}$ that we consider for implicitly satisfying these nongeometric λ constraints is defined by:

$$\lambda_{\alpha,1}^{2\,\sin} = \sin^2 \theta_\alpha \text{ and } \lambda_{\alpha,2}^{2\,\sin} = 1 - \sin^2 \theta_\alpha \qquad (19)$$

This formalism that is used for sampling two related λ values was originally implemented in CHARMM for constant pH-MD simulations by Lee and coworkers[16–18] and as a variant of λ-dynamics, termed "θ-dynamics", by Yang and coworkers (unpublished). This functional form of $\{\lambda\}$ also leads to stable simulations; although the integration timestep needed to be reduced to 0.5 fs to sample the more flexible methoxy moieties on the dimethoxybenzene hybrid model without compromising the numerical stability of the Verlet integrator. The results summarized in Table 2 demonstrate that the exchange rate between dominant substituents at each site is one to two orders of magnitude higher than the $\lambda^{2\exp}$ functional form and results in very low standard deviations of less than 0.002 kcal mol^{-1}. Similarly, low average and maximum errors of less than

Table 4. Fraction of θ-phase space in which a physically meaningful molecule is represented in a hybrid molecule with N substituents at a single site on a common ligand core, that is, when any substituent has $\lambda_i \approx 1$ defined by $\lambda_i >$ threshold.

Functional form	N	Threshold 0.80	0.90	0.95	0.99
$\lambda^{2\sin}$	2	0.588	0.412	0.284	0.114
$\lambda^{N\sin}$	2	0.416	0.274	0.186	0.076
	3	0.126	0.054	0.024	0.006
	4	0.030	0.008	0.003	0.0002
$\lambda^{N\exp}$	2	0.772	0.676	0.596	0.412
	3	0.660	0.540	0.447	0.270
	4	0.544	0.412	0.320	0.164

0.007 and 0.013 kcal mol^{-1}, respectively, are achieved for relative free energy differences between pairs of benzene, dihydroxybenzene, or dimethoxybenzene molecules in vacuum. This functional form implicitly encourages the change in the identity of the dominant substituents and thus exchanges in the identity of the dominant ligand throughout the simulation trajectories primarily due to its oscillating nature.

Implicit constraints for N variables: $\lambda^{N\sin}$, $\lambda^{N\exp}$

The first functional form of $\{\lambda\}$ that is generalized to N λs that we have examined is defined by:

$$\lambda_{\alpha,i}^{N\,\sin} = \frac{\sin^2 \theta_{\alpha,i}}{\sum\limits_{j=1}^{N} \sin^2 \theta_{\alpha,j}} \qquad (20)$$

Simulations were quite stable for this functional form, although the integration timestep also needed to be reduced to 0.5 fs to successfully sample the more flexible methoxy moieties on the dimethoxybenzene hybrid model. Simulation results based on sampling $\{\lambda\}$ with this functional form are summarized in Table 3. For sampling with any of the 2 × 2 hybrid ligands, that is, sampling two substituents at each site, the quality of the results is very high. However, these results are slightly degraded relative to those obtained from simulations using the $\lambda^{2\sin}$ functional form; average and maximum errors in the relative free energy estimates of less than 0.01 and 0.03 kcal mol^{-1}, respectively, are achieved for simulations based on the $\lambda^{N\sin}$ functional form while the corresponding errors are 0.007 and 0.013 kcal mol^{-1} for the $\lambda^{2\sin}$ functional form. For increasing numbers of substituents at each site, the transition rate decreases significantly and the overall quality of the relative free energy estimates tends to diminish. For hybrid ligands with four or more substituents at each of two sites, in

Table 3. Quality of relative free energy estimates for 1 250 000 steps MSλD simulations using implicit constraints: $\lambda^{N\sin}$ in vacuum.

Hybrid ligand nSite1 × nSite4	N	Δt (fs)	τ_{trans} (ps^{-1})	$\Delta\Delta G$ (kcal/mol) AUE	σ	Max
2H × 1H	2	2.0	0.87	0.0141		
3H × 1H	3	2.0	0.29	0.0072	0.0040	0.0108
4H × 1H	4	2.0	0.09	0.0035	0.0022	0.0064
5H × 1H	5	2.0	0.03	0.0201	0.0131	0.0476
2H × 2H	4	2.0	0.76	0.0150	0.0081	0.0278
3H × 3H	9	2.0	0.26	0.0490	0.0363	0.1217
4H × 4H	16	2.0	–	–	–	–
5H × 5H	25	2.0	–	–	–	–
2OH × 2OH	4	2.0	0.81	0.0051	0.0023	0.0090
3OH × 3OH	9	2.0	0.25	0.0349	0.0206	0.0814
4OH × 4OH	16	2.0	–	–	–	–
5OH × 5OH	25	2.0	–	–	–	–
2OCH$_3$ × 2OCH$_3$	4	0.5	0.80	0.0017	0.0009	0.0032
3OCH$_3$ × 3OCH$_3$	9	0.5	0.50	0.0193	0.0124	0.0494
4OCH$_3$ × 4OCH$_3$	16	0.5	0.22	0.1600	0.1102	0.4803
5OCH$_3$ × 5OCH$_3$	25	0.5	–	–	–	–

The integration timestep is Δt and τ_{trans} is the average frequency of the change in the identity of the substituent with $\lambda \approx 1$ on each site. Statistics are averaged over all $N(N-1)/2$ pairs of compounds in the model hybrid ligands. Rows with '–' indicate that not all ligands were sampled to be dominant in the course of the trajectory.

Table 5. Quality of relative free energy estimates for MSλD simulations using implicit constraints: λ^{Nexp}.

Hybrid ligand nSite1 $\times$ nSite4	N	Δt (fs)	τ_{trans} (ps^1)	$\Delta\Delta G_{vac}$ (kcal/mol)			Δt (fs)	τ_{trans} (ps^1)	$\Delta\Delta G_{solv}$ (kcal/mol)		
				AUE	Σ	Max			AUE	σ	Max
2H $\times$ 1H	2	2.0	1.10	0.0010			2.0	1.08	0.0000		
6H $\times$ 1H	6	2.0	0.65	0.0074	0.0050	0.0175	2.0	0.64	0.0088	0.0069	0.0213
10H $\times$ 1H	10	2.0	0.27	0.0077	0.0053	0.0229	2.0	0.27	0.0328	0.0211	0.0793
2H $\times$ 2H	4	2.0	1.08	0.0027	0.0019	0.0050	2.0	1.08	0.0098	0.0057	0.0177
5H $\times$ 5H	25	2.0	0.54	0.0326	0.0229	0.0923	2.0	0.54	0.0693	0.0529	0.2735
2OH $\times$ 2OH	4	2.0	1.01	0.0054	0.0033	0.0102	2.0	0.38	0.0068	0.0049	0.0133
5OH $\times$ 5OH	25	2.0	0.66	0.0322	0.0279	0.1171	2.0	0.71	0.0730	0.0524	0.2326
2OCH$_3$ $\times$ 2OCH$_3$	4	2.0	0.44	0.0046	0.0021	0.0083	1.5	0.15	0.0152	0.0080	0.0251
5OCH$_3$ $\times$ 5OCH$_3$	25	2.0	0.69	0.0116	0.0080	0.0383	1.5	0.29	0.0785	0.0548	0.2376

Vacuum simulations were run for 25 ns while solvent simulations were run for 3 ns (2.25 ns for the dimethoxybenzene simulations). The integration timestep is Δt and τ_{trans} is the average frequency of the change in the identity of the substituent with $\lambda \approx 1$ on each site. Statistics are averaged over all $N(N-1)/2$ pairs of compounds in the model hybrid ligands.

most cases, trajectory lengths of 25 ns were not even long enough to sample each ligand in the dominant state. This observation is due to the fact that as the number of substituents increases, the fraction of θ-phase space that is associated with a substituent having $\lambda_i \approx 1$ decreases. Table 4 summarizes the fraction of θ-phase space that is associated with a dominant substituent as a function of the threshold value for defining $\lambda_i \approx 1$. For the λ^{Nsin} functional form given a threshold value of 0.8, the fraction of θ-phase space that is dedicated to representing physical ligands reduces from 0.41 to 0.13 to 0.03 for hybrid ligands with two, three, and four substituents modeled, respectively. The schematic in Table 1 clearly shows that even for $N = 2$ a significant portion of θ-space yields intermediate rather than endpoint λ values.

We have explored another functional form of $\{\lambda\}$ that is generalized to N substituents at each site on a hybrid ligand and implicitly satisfies the nongeometric constraints in Eq. (4). This modified exponential is defined as:

$$\lambda_{\alpha,i}^{Nexp} = \frac{e^{5.5 \sin \theta_{\alpha,i}}}{\sum_{j=1}^{N} e^{5.5 \sin \theta_{\alpha,j}}} \tag{21}$$

Results from simulations based on this functional form are summarized in Table 5. This functional form was designed to combine the strengths of the previous functional forms: specifically, (i) an exponential term so that a single $\lambda_{\alpha,i}$ could approach 1 regardless of how many substituents were present, and (ii) a sin term as an argument of the exponential to encourage the oscillation of $\{\lambda\}$. Furthermore, we wanted a functional form that would have the same probability distribution for each $\lambda_{\alpha,i}$ so no further correction would be required to unbias the relative population counts in the analysis of the simulation trajectories.

Indeed, simulations based on this functional form are very stable; unlike the λ^{2sin} and λ^{Nsin} functional forms, the λ^{Nexp} functional form only required a small decrease in the integration timestep (to 1.5 fs) to successfully sample the more flexible methoxy moieties on the dimethoxybenzene hybrid model. This functional form yields frequent transitions among the dominant

substituents and leads to very high quality relative free energy estimates. For up to 10 substituents modeled on one site of the benzene core, the average and maximum errors are less than 0.008 and 0.025 kcal mol^{-1}, respectively, while the standard deviation is less than 0.006 kcal mol^{-1} in vacuum. Even the more flexible hybrid ligands representing 25 distinct dihydroxybenzene and dimethoxybenzene molecules have relative free energy estimates within 0.03 kcal mol^{-1} on average and at most have errors of 0.1 kcal mol^{-1}. The precision is also very good with standard deviations within 0.02 kcal mol^{-1}.

Simulations of each of the hybrid ligands were repeated in explicit solvent environments. In general, the transition rates for the benzene and dihydrobenzene models were similar in vacuum and solvent environments. By contrast, transition rates for the dimethoxybenzene hybrid ligands were systematically slower in solvent than in vacuum. Visual inspection of the trajectories confirmed that the methoxy groups explored a wide variety of conformations. Thus, the extra volume that is explored by the methoxy groups relative to the smaller substituents suggests that more substantial solvent rearrangements are required to sample each of the dimethoxybenzene ligands in the dominant state.[1]

Discussion

Simulation stability

From these simulation results based on model hybrid ligands, we have demonstrated that implicitly incorporating nongeometric constraints into the functional form of $\{\lambda\}$ yields relatively stable simulations with timesteps up to 2 fs in vacuum environments. Each of these functional forms and their corresponding forces in the MD simulations are relatively inexpensive to compute. To ensure that λ^{Nsin} would never be undefined, in the case when all $\sin \theta_i = 0$ a small offset could be applied to all θ_i; however, in practice, this event is so rare that this correction was not required even for simulations up to 25 ns.

With these functional forms of $\{\lambda\}$, the numerical stability of the equations of motion can become compromised when there is a small change in $\lambda_{\alpha,i}$ when $\lambda_{\alpha,i} \approx 0$. This situation
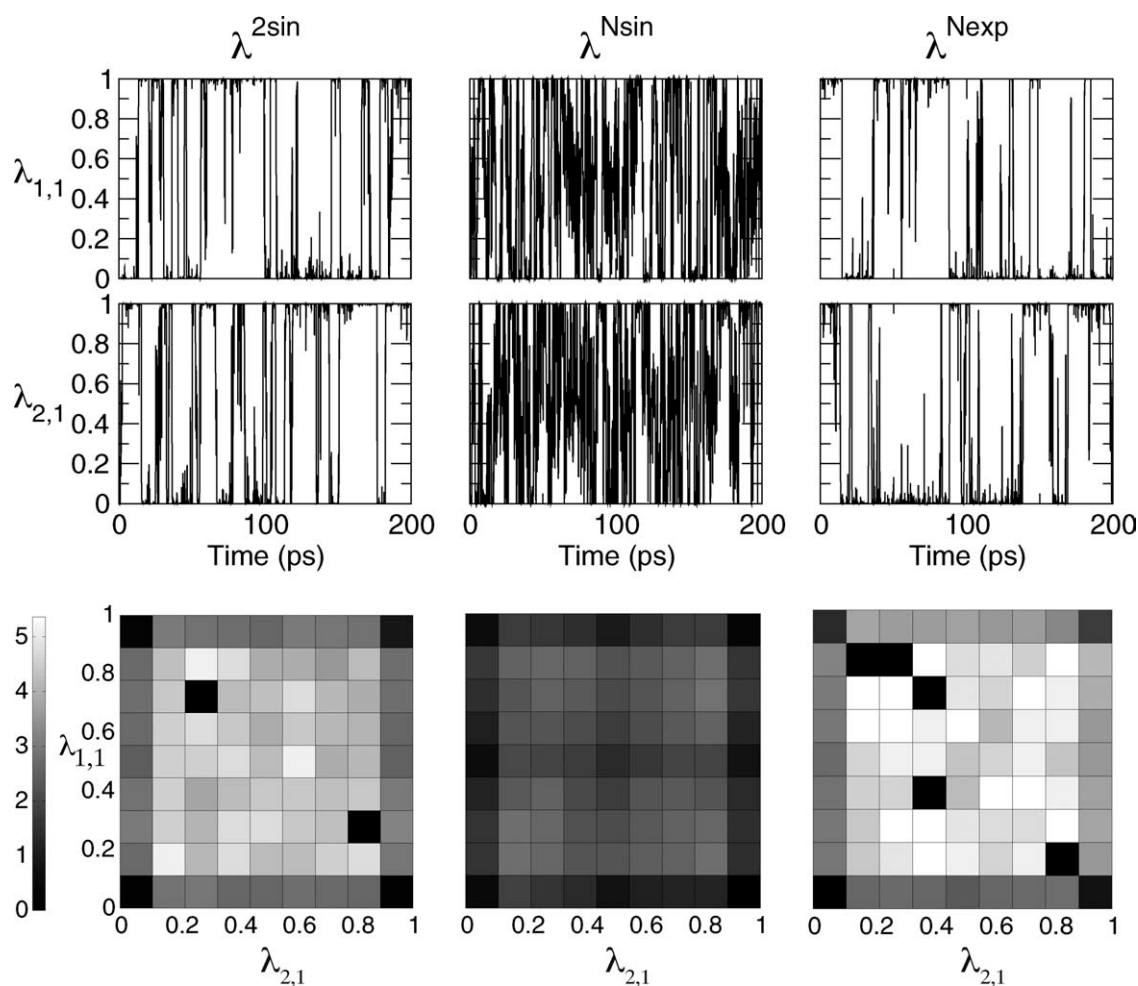
Figure 2. Representative data from explicit solvent 2OCH$_3$ × 2OCH$_3$ simulation trajectories based on λ^{2sin}, λ^{Nsin}, and λ^{Nexp}: λ-Value for first substituent on site 1 (top panel) and site 4 (middle panel); and the relative free energy free energy surface in kcal mol^{-1} calculated over 400 ps in the corresponding trajectories (bottom panel). Note the physically meaningful ligands coincide with the corners of each of the relative free energy surfaces.

arises when the substituent i on site α is in an energetically unfavorable conformation or, more frequently, is too close to an environment atom. When $\lambda_{\alpha,i}$ approaches 0, the contribution of this unfavorable interaction to the total energy of the system is very small. However, even a very small increase in the value of $\lambda_{\alpha,i}$ can contribute a significant amount of energy to the system and cause a spike in the energy and thus render the numerical solutions to the equations of motion unstable. In the original implementation of λ-dynamics,[23,29] due to small changes at the $\{\lambda\}$ endpoints introduced by the Lagrange multiplier method followed by the $\{\lambda\}$ renormalization at every timestep, the timestep often had to be decreased to 0.5 fs to retain numerical stability of the Verlet integrator. The λ^{2sin} and λ^{Nsin} functional forms require integration timesteps of 0.5 fs for modeling the flexible dimethoxybenzene compounds while the λ^{2exp}, λ^{2sin}, and λ^{Nsin} functional forms require integration timesteps of 0.5–1.0 fs for simulations when the hybrid compounds are modeled in explicit solvent environments (data not shown). By contrast, the λ^{Nexp} functional form is generally stable with an integration timestep of 1.5 fs in vacuum and in explicit solvent environments even for the flexible dimethoxy-

benzene molecules. This functional form is less sensitive than the other functional forms that we examined to slight changes at the $\{\lambda\}$ endpoints because the exact boundaries are:

$$\frac{e^{-c}}{e^c + (N-1)e^{-c}} \leq \lambda_{\alpha,i}^{N\exp} \leq \frac{e^c}{e^c + (N-1)e^{-c}} \tag{22}$$

where for $N = 5$ and $c = 5.5$ (as used in these simulations), the boundaries are $0.000016 < \lambda_{\alpha,i} < 0.99993$.

Leveraging the functional form of $\{\lambda\}$ to enhance sampling

In MSλD simulations, the efficiency of the simulations is directly related to the number of times that the identity of the substituent with $\lambda_{\alpha,i} \approx 1$ at each site changes, that is, the number of transitions, which leads to increased convergence of the relative free energy differences estimated by Eq. (13). The difference between the λ^{2exp} and λ^{2sin} functional forms in the "2 × 2" models clearly demonstrates the value of the λ^{2sin} functional form that oscillates in θ-space to improve the sampling of $\{\lambda\}$ itself, which in turn improves the likelihood of changing the identity of

the substituent with $\lambda_{\alpha,i} \approx 1$. Continually exerting a force that increases the magnitude of θ according to the λ^{2exp} functional form will perpetuate the "dominance" of the first substituent selected. By contrast, continually exerting a force that increases the magnitude of θ according to the λ^{2sin} functional form will eventually lead to an exchange in the substituent with $\lambda_{\alpha,i} \approx 1$.

While these oscillating functional forms do encourage more extensive sampling of λ-phase space, they do not render the simulations insensitive to the chemical identity of the compounds under investigation. Specifically, reasonable estimates of the biasing potentials in Eq. (12) must be used to effectively sample ligands whose relative free energies differ by more than 2–3 kcal mol^{-1}. We have demonstrated elsewhere[22] the effectiveness of MSλD sampling with the λ^{Nexp} functional form and good estimates of the biasing potentials to model relative hydration free energies of a series of benzene derivatives that range from 0 to 10 kcal mol^{-1}. In this study, MSλD simulations were performed based on a single hybrid ligand that contained two substituents modeled at one site and three substituents modeled at another site for a total of six distinct benzene derivatives. Reasonable biasing potentials $\{F_{\alpha,i}\}$ were determined that yielded sufficient transitions among the substituents at each site of the hybrid molecule and reproduced the hydration free energy estimates that were achieved by performing much more extensive traditional alchemical free energy perturbation simulations for pairs of these benzene derivatives.

Efficiency in MSλD simulations is also related to the fraction of time that a dominant ligand is present as compared with the time that several partial or intermediate and unphysical ligands are present. A functional form of $\{\lambda_i\}$ which is biased toward $\lambda_{\alpha,i} \approx 1$ and $\lambda_{\alpha,i} \approx 0$ will be more efficient than one in which intermediate $\lambda_{\alpha,i}$ values can dominate. Among the functional forms that we have examined in this study, the θ-phase space in λ^{Nexp} as compared with λ^{Nsin} is more biased towards λ values that are closer to 0 or 1. Furthermore, as seen from sample data from explicit solvent 2OCH$_3$ $\times$ 2OCH$_3$ simulation trajectories in Figure 2, simulations based on the λ^{Nexp} and λ^{2sin} functional forms spend a significant proportion of time at the physically meaning endpoints with λ values close to 0 or 1 as compared with λ^{Nsin} trajectories in which the endpoints are only weakly favored over intermediate λ values. Finally, the coefficient c in λ^{Nexp} can be tuned to describe the steepness of the switching between $\lambda_{\alpha,i} \approx 1$ and $\lambda_{\alpha,i} \approx 0$. We have identified a "sweet spot" coefficient of 5.5 that seems optimal for a broad range of these hybrid ligands and environments and should be robust for MSλD simulations regardless of the system. The relatively steep transition between $\lambda_{\alpha,i} \approx 1$ and $\lambda_{\alpha,i} \approx 0$ when $c = 5.5$ in λ^{Nexp} has several advantages in MSλD simulations. First, the quality of the results are relatively insensitive to the specific threshold value that is used to define $\lambda_{\alpha,i} \approx 1$. Second, there is little time in the simulations when intermediate $\lambda_{\alpha,i}$ values are explored and thus the simulation trajectories are predominantly sampling physical ligands which leads to faster simulation convergence. Finally, the simulations become much less sensitive than the λ^{Nsin} functional form to the number of substituents in the hybrid molecule. Coefficients of less than 5.5 do not sufficiently

approach the endpoints and spend a larger fraction of θ-space in intermediate λ values so were less efficient for these simulations. Coefficients of greater than 5.5 demonstrate increased transition rates in vacuum and thus increased convergence rates; however, the rates of change in $\{\lambda\}$ near the endpoints are too abrupt in solvent simulations to retain the stability in the numerical integrator (data not shown). A decrease in the integration timestep can alleviate this problem, but we have chosen instead to use $c = 5.5$ in all simulations in this study and recommend the use of this value across applications.

Enhancing simulation efficacy

Other MSλD parameters can be used to increase the efficacy of the simulations. For example, decreasing the m_θ parameters will tend to increase the mobility of the θ values and thus increase transition rates. Adding distance restraints that superimpose the ipso carbon atoms throughout the simulation ensure that substituents are in similar conformations to one another and increase the likelihood of the transitions. Finally, adding biases on the θ values that take effect only when $\lambda_{\alpha,i} < 0.8$ will also tend to increase the transition rates, but will also increase the amount of time spent at intermediate λ values (data not shown).

In addition, other advanced sampling methods could be used to further enhance simulation efficiency by reducing the effective barriers that are associated with environmental relaxation processes. For example, transitions between dominant substituents of a hybrid ligand can be hindered if the transition requires a conformational change in the solvent or protein side chain. This effective barrier to λ transitions may be exaggerated when substituents of significantly different charge distributions or volume are involved. In these cases, adopting strategies like temperature-accelerated MD [30,31] or λ-adiabatic free energy dynamics[32] in which the dynamics of the λ variables are adiabatically separated from the solvent dynamics may prove beneficial. Alternatively, the self-guided Langevin dynamics,[33,34] in which local average properties are used to enhance low-frequency conformational searching, or orthogonal space random walk,[35,36] in which the free-energy surface in both the λ-phase space and its generalized force space are flattened could enable those barriers directly related to λ and those related to the environment relaxation to be overcome more readily.

Model quality

This study has focused on the sampling characteristics of different functional forms of $\{\lambda\}$ in MSλD simulation trajectories using model hybrid ligands. Because we have been evaluating relative free energy differences between pairs of compounds that are identical to one another, we have isolated the contribution of the observed errors to errors in the sampling method itself. In most real applications, however, the quality of the results will have contributions from sampling errors and errors due to the force field parameters that define the modeled potential energy surface.

Even for the most flexible dimethyoxybenzene ligands in this study, the λ^{Nexp} functional form yields average and maximum unsigned errors of 0.01 and 0.04 kcal mol^{-1}, respectively, for the

10 and 25 ns vacuum trajectories and 0.08 and 0.24 kcal mol^{-1}, respectively, for the 10 and 2.25 ns solvent trajectories. The precision of the simulations is also very high with standard deviations within 0.06 kcal mol^{-1}. Other work is underway in our group to apply this method in both retrospective and prospective structure-based drug design applications and obtain better estimates of the combined modeling error. Because of the high quality and efficiency of this MSλD sampling method using the $\lambda^{N\exp}$ functional form of $\{\lambda\}$, MSλD simulations could be used as a method for optimizing new ligand force field parameters to reproduce available hydration free energy data or alternatively relative hydration free energies for series of functional groups that would be consistent for a given force field.

Conclusions

In this study, we have presented four different strategies for sampling the $\{\lambda\}$ variables in MD simulations based on alchemical free energy simulations. In these simulations, the dynamic variables $\{\lambda\}$ represent the coefficients that scale the interaction energies between the individual substituents and their environment. To satisfy the hybrid Hamiltonian that is used in the simulations, nongeometric constraints: $0 \leq \lambda_{\alpha,i} \leq 1$ and $\sum_{i=1}^{N_\alpha} \lambda_{\alpha,i} = 1$ for each site α, must be satisfied at every timestep. Four functional forms of $\{\lambda\}$ were evaluated for implicitly constraining either two or N λ parameters. The functional form of $\lambda_{\alpha,i}^{N\exp} = \frac{e^{c \sin \theta_{\alpha,i}}}{\sum_{j=1}^{N} e^{c \sin \theta_{\alpha,j}}}$ exhibits the ideal characteristics for our MSλD simulations. It implicitly satisfies the constraints and does not compromise the numerical stability of the simulations. It is oscillating in nature and so provides enhanced sampling of the λ_i values. It transits quickly between $\lambda_{\alpha,i} \approx 1$ and $\lambda_{\alpha,i} \approx 0$ such that (i) there is a significant fraction of θ-phase space in which a physical rather than unphysical ligand is present and (ii) it is relatively insensitive to the specific threshold that is used to define $\lambda_{\alpha,i} \approx 1$. Both the value of $\lambda_{\alpha,i}$ and the forces on $\lambda_{\alpha,i}$ are computationally inexpensive and each $\lambda_{\alpha,i}$ has same probability density function so no further bias or correction is required to account for differences in effective phase space volume sampled.

Acknowledgments

[1] M. P. Allen, D. J. Tildesley, Computer Simuation of Liquids; Oxford University Press: New York, NY, 1988.
[2] J. Ryckaert, G. Ciccotti, H. Berendsen, J Comput Phys 1977, 23, 327.
[3] H. Andersen, J Comput Phys 1983, 52, 24.
[4] S. Lee, K. Palmo, S. Krimm, J Comput Phys 2005, 210, 171.
[5] A. G. Bailey, C. P. Lowe, A. P. Sutton, J Comput Phys 2008, 227, 8949.
[6] A. G. Bailey, C. P. Lowe, J Comput Chem 2009, 30, 2485.
[7] T. Forester, W. Smith, J Comput Chem 1998, 19, 102.
[8] S. Miyamoto, P. Kollman, J Comput Chem 1992, 13, 952.
[9] A. Rahman, F. Stillinger, J Chem Phys 1971, 55, 3336.
[10] D. Evans, S. Murad, Mol Phys 1977, 34, 327.
[11] A. Jain, N. Vaidehi, G. Rodriguez, J Comput Phys 1993, 106, 258.
[12] A. Mazur, R. Abagyan, J Biomol Struct Dyn 1989, 6, 815.
[13] K. Gibson, H. Scheraga, J Comput Chem 1990, 11, 487.
[14] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, J Comp Chem 1983, 4, 187.
[15] B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, J Comput Chem 2009, 30, 1545.
[16] M. S. Lee, F. R. Salsbury, C. L. Brooks, III, Proteins 2004, 56, 738.
[17] J. Khandogin, C. L. Brooks, III, Biophys J 2005, 89, 141.
[18] J. Khandogin, C. L. Brooks, III, Biochemistry 2006, 45, 9363.
[19] P.-F. Verhulst, Correspondance mathématique et physique 1838, 10, 113.
[20] S. Samarasinghe, Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition; Taylor & Francis Group, LLC: Boca Raton, FL, 2007.
[21] R. L. Prentice, Biometrics 1976, 32, 761.
[22] J. L. Knight, C. L. Brooks, III, J Chem Theory Comput (in press).
[23] J. L. Knight, C. L. Brooks, III, J Comput Chem 2009, 30, 1692.
[24] X. Kong, C. L. Brooks, III, J Chem Phys 1996, 105, 2414.
[25] A. Yamamoto, Y. Kitamura, Y. Yamane, Ann Nucl Energy 2004, 31, 1027.
[26] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A. D. Mackerell, Jr., J Comput Chem 2010, 31, 671.
[27] W. F. van Gunsteren, H. J.C. Berendsen, Mol Phys 1977, 34, 1311.
[28] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, J Chem Phys 1983, 79, 926.
[29] Z. Guo, C. L. Brooks, III, X. Kong, J Phys Chem B 1998, 102, 2032.
[30] L. Maragliano, E. Vanden-Eijnden, Chem Phys Lett 2006, 426, 168.
[31] C. F. Abrams, E. Vanden-Eijnden, Proc Natl Acad Sci USA 2010, 107, 4961.
[32] J. B. Abrams, L. Rosso, M. E. Tuckerman, J Chem Phys, 2006, 125, 074115.
[33] X. Wu, B. R. Brooks, Chem Phys Lett 2003, 381, 512.
[34] X. Wu, B. R. Brooks, J Chem Phys 2011, 134, 134108.
[35] L. Zheng, M. Chen, W. Yang, Proc Natl Acad Sci USA 2008, 105, 20227.
[36] L. Zheng, M. Chen, W. Yang, J Chem Phys 2009, 130, 234105.