JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Multisite λ Dynamics for Simulated Structure−Activity Relationship Studies

Jennifer L. Knight and Charles L. Brooks, III*

Department of Chemistry & Department of Biophysics, University of Michigan, 930 N. University Ave., Ann Arbor, Michigan 48109, United States

ABSTRACT: Multisite λ dynamics (MSλD) is a new free energy simulation method that is based on λ dynamics. It has been developed to enable multiple substituents at multiple sites on a common ligand core to be modeled simultaneously and their free energies assessed. The efficacy of MSλD for estimating relative hydration free energies and relative binding affinties is demonstrated using three test systems. Model compounds representing multiple identical benzene, dihydroxybenzene, and dimethoxybenzene molecules show that total combined MSλD trajectory lengths of ∼1.5 ns are sufficient to reliably achieve relative hydration free energy estimates within 0.2 kcal/mol and are less sensitive to the number of trajectories that are used to generate these estimates for hybrid ligands that contain up to 10 substituents modeled at a single site or five substituents modeled at each of two sites. Relative hydration free energies among six benzene derivatives calculated from MSλD simulations are in very good agreement with those from alchemical free energy simulations (with average unsigned differences of 0.23 kcal/mol and R^2 = 0.991) and the experiment (with average unsigned errors of 1.8 kcal/mol and R^2 = 0.959). Estimates of the relative binding affinities among 14 inhibitors of HIV-1 reverse transcriptase obtained from MSλD simulations are in reasonable agreement with those from traditional free energy simulations and the experiment (average unsigned errors of 0.9 kcal/mol and R^2 = 0.402). For the same level of accuracy and precision, MSλD simulations are achieved ∼20−50 times faster than traditional free energy simulations and thus with reliable force field parameters can be used effectively to screen tens to hundreds of compounds in structure-based drug design applications.

1. INTRODUCTION

Free energy calculations are fundamental to obtaining accurate theoretical estimates of hydration free energies and protein−ligand binding affinities.[1−5] Traditionally, free energy differences are calculated from alchemical free energy simulations, which are analyzed by free energy perturbation, thermodynamic integration, or Bennett acceptance ratio methods.[6,7] These traditional alchemical simulations mutate one compound into another in a stepwise fashion using several unphysical intermediates to compute relative free energies or, alternatively, grow a compound from nothing to obtain an absolute free energy. In practice, however, due to the number of intermediates that must be investigated and the length of the simulations, these methods are generally too computationally intensive to be used routinely in structure-based drug design or systematic exploration of chemical modifications of a lead compound.

λ dynamics is an alternative free energy method in which the transformation coordinate between the end states, the parameter "λ", is treated as a dynamic variable in the simulations and is propagated in a manner that is analogous to changes in the atomic coordinates.[8,9] In this way, instead of performing simulations for fixed λ values, all of the intermediate states are explored in a single simulation. λ dynamics has been used to compute hydration free energies[9,10] and binding free energies[11,12] and to identify ligand binding modes.[13−15] λ-dynamics simulation methods have stimulated the development of other theoretical approaches for a variety of applications in which the λ parameter scales the potential energy and dynamically varies throughout the course of a simulation. For example, constant pHMD (CPHMD) simulations account for accurate protonation states of protein residues,[16−18] Abrams et al.'s[19] adiabatic free energy dynamics (AFED) generates free energy profiles along a reaction path, and Zheng et al.'s orthogonal space random walk method enhances free energy simulations[20] and conformational sampling.[21] Bitetti-Putzer et al.[22] uses λ as a self-regulating sampling variable to efficiently traverse high-energy barriers and to thoroughly explore low-energy basins, and Tivado-Rives et al.'s[23] Just Add Water (JAWS) strategy identifies positions of water molecules in binding sites of protein−ligand complexes. Still other simulation methods use dynamic λ variables in discretized λ space, such as simulated scaling,[24] FEP/REMD,[25] and BEDAM.[5,26]

In traditional alchemical free energy calculations, the hybrid potential energy function is defined by

$$V(X,\{x\}) = V_{\mathrm{env}}(X) + (1-\lambda)\,V(X,x_1) + \lambda V(X,x_2) \qquad (1)$$

where X and x_i are the atomic coordinates associated with the environment and ligand i, respectively, V_{env} is the potential energy involving the environment atoms only, and $V(X,x_i)$ is the interaction energy computed between ligand i and the environment. In this formalism, a hybrid ligand can be constructed in which all atoms that are common to both ligands are represented once, while all atoms that are unique to each ligand are represented explicitly. Atoms in the common ligand core are then treated as part of the environment, while atoms that are unique to each ligand are modeled as substituents on this common core. In λ dynamics, the hybrid molecule is extended

Received: June 28, 2011
Published: August 10, 2011

ACS Publications © 2011 American Chemical Society

dx.doi.org/10.1021/ct200444f | J. Chem. Theory Comput. 2011, 7, 2728–2739

to N ligands, and the corresponding hybrid potential energy function is constructed to be

$$V(X,\{x\},\{\lambda\}) = V_{\mathrm{env}}(X) + \sum_{i=1}^{N} \lambda_i (V(X,x_i) - F_i) \quad (2)$$

with the constraints

$$0 \le \lambda_i \le 1 \quad (3a)$$

$$\sum_{i=1}^{N} \lambda_i = 1 \quad (3b)$$

where λ_i is the coupling parameter associated with ligand i and F_i is a precalculated biasing potential that enhances the sampling of each $\lambda_i \approx 1$ state. The dynamics of the system are generated from the extended Hamiltonian:

$$H_o(X,\{x\},\{\lambda\}) = T_x + T_\lambda + V(X,\{x\},\{\lambda\}) \quad (4)$$

where T_x and T_λ are the kinetic energies of the atomic coordinates and λ variables, respectively (λ variables are treated as volumeless particles with mass m_λ). A ligand is defined to be "dominant" over other ligands when its corresponding λ value approaches 1. The difference in the changes in free energy in a given environment between ligands i and j is

$$\Delta\Delta G_{j\to i} = -k_B T \ln \frac{P(\lambda_i = 1, \{\lambda_{m\neq i} = 0\})}{P(\lambda_j = 1, \{\lambda_{m\neq j} = 0\})} \quad (5)$$

where $P(\lambda_i = 1,\{\lambda_{m\neq i} = 0\})$ corresponds to the amount of time ligand i has $\lambda_i = 1$ during the λ-dynamics simulation; in practice, the amount of time ligand i has $\lambda_i > 0.8$ during the simulation is counted. The relative binding affinities for each pair of ligands that are represented in the hybrid molecule can be estimated directly from simulations in the protein environment where the $\{F_i\}$ in eq 2 is assigned to be the relative free energy estimates for the solvent arm of the thermodynamic cycle. For a more detailed discussion of λ dynamics, the reader is referred to ref 9.

The original λ-dynamics methodology was implemented for modeling multiple substituents at a single site on a common ligand framework. By constrast, many experimental combinatorial chemistry approaches systematically vary substituents at multiple sites on a core compound. For example, structure−activity relationship studies often evaluate the efficacy of putative drug molecules that are chemical variants of a promising lead compound.

Here, we present a new version of λ dynamics that enables multiple substituents at multiple sites on a common ligand core to be modeled and demonstrate its usefulness for estimating series of relative hydration free energies and ligand binding affinities. With the ability to examine multiple substituents at multiple sites on a core molecule, this multisite λ-dynamics (MSλD) method has the potential to evaluate the relative free energies of many compounds simultaneously and further increase the efficiency of the λ-dynamics approach to free energy calculations. Furthermore, we have developed a strategy to implicitly satisfy the holonomic constraints on $\{\lambda\}$ that are defined in eq 3 and thus substantially improve the numerical stability of these simulations up to timesteps of 2 fs.[27]

First, using hybrid ligands that represent series of benzene, dihydroxybenzene, and dimethoxybenzene compounds, we demonstrate the robustness of MSλD for thorough sampling in vacuum and solvent environments. Second, we show that relative hydration free energies of six benzene derivatives that are estimated from MSλD compare very well with those obtained from traditional alchemical free energy simuations but at ∼1/20 the computational cost. Finally, we illustrate how estimates of the components of the relative binding affinities of 14 inhibitors of HIV-1 reverse transcriptase are computed about 50 times faster using MSλD relative to traditional free energy methods and yield comparable quality results. This study reveals the utility of MSλD as an effective sampling strategy in structure-based drug design to screen through on the order of tens to hundreds of variations of a lead compound in a reasonable amount of time.

2. METHODS

2.1. Multisite λ-Dynamics (MSλD) Theory. To enable multiple substituents at multiple sites to be sampled during λ dynamics, we have extended the hybrid potential energy function to be

$$V(X,\{x\},\{\lambda\}) = V_{\mathrm{env}}(X) + \sum_{S=1}^{M_{\mathrm{sites}}} \sum_{i=1}^{N_S} \lambda_{S,i}(V(X,x_{S,i}) - F_{S,i}) + \sum_{S=1}^{M_{\mathrm{sites}}-1} \sum_{i=1}^{N_S} \sum_{T=S+1}^{M_{\mathrm{sites}}} \sum_{j=1}^{N_T} \lambda_{S,i}\lambda_{T,j}(V(x_{S,i},x_{T,j})) \quad (6)$$

with the holonomic constraints:

$$0 \le \lambda_{\alpha,i} \le 1 \quad (7a)$$

$$\sum_{i=1}^{N_\alpha} \lambda_{\alpha,i} = 1 \quad (7b)$$

for each site α. M_{sites} is the total number of sites which contain multiple substituents, and N_S is the number of substituents at site S on the common ligand framework. 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 at all other sites. Note that substituents at a given site do not "see" each other in these simulations. In this case, a ligand is now 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 substituent sites, the relative free energies between two distinct compounds are estimated via

$$\Delta\Delta G_{1,i;2,j\to 1,k;2,l} = -k_B T \ln \frac{P(\lambda_{1,k} = 1, \lambda_{2,l} = 1)}{P(\lambda_{1,i} = 1, \lambda_{2,j} = 1)} \quad (8)$$

Instead of using the Lagrange multiplier method to satisfy the holonomic constraints in eq 7 explicitly, in this new implementation, we implicitly satisfy the constraints by defining λ's for the N substituents at site α to be

$$\lambda_{\alpha,i} = \frac{e^{5.5\sin\theta_{\alpha,i}}}{\sum_{j=1}^{N_\alpha} \exp(5.5\sin\theta_{\alpha,j})} \quad (9)$$

Using this formalism for MSλD, it is the values of θ that have fictitious masses, m_θ, and are propagated through the equations of motion, not the λ values directly. Thus, the extended Hamiltonian is defined as

$$H_o(X,\{x\},\{\lambda(\theta)\}) = T_x + T_\theta + V(X,\{x\},\{\lambda(\theta)\}) \quad (10)$$

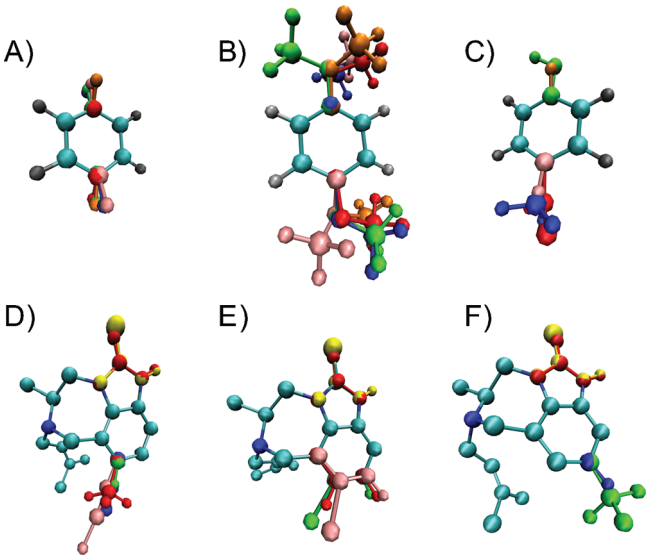

Figure 1. Schematic representation of three model systems that we have used to assess the quality of the MSλD implementation in CHARMM. Hybrid molecules representing multiple benzene molecules by modeling distinct sets of hydrogen and corresponding ipso carbon atoms at (A) site 1 and (B) sites 1 and 4 on a common benzene core. (C) A hybrid molecule representing six benzene derivatives modeled by three substituents at site 1 and two substituents on site 4 on a common benzene core. Hybrid TIBO molecules representing 14 inhibitors of HIV-1 reverse transcriptase containing both C═O and C═S variations at the X site and (D) variations involving nonhalides at the C-8 site or (E) halides at the C-8 site or (F) variations at the C-9 site.

Multisite λ dynamics has been implemented in the CHARMM macromolecular software package, version c36a4.[28,29]

2.2. Hybrid Ligands. Parameters and partial charges for the hybrid ligands were assigned from the recently developed CHARMM General Force Field (CGenFF).[30] Parameters and partial charge distributions for the TIBO compounds were optimized in our previous work[31] using our in-house parametrization tool MATCH[32] and quantum mechanical calculations and are included in that work's Supporting Information.

2.2.1. Model Compounds. 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 ring (see Figure 1A). Similarly, each hybrid dihydroxybenzene molecule consisted of a single benzene ring with N hydroxy groups and ipso carbon atoms at the *para* positions on the common ring, and each hybrid dimethoxybenzene molecule consisted of a single benzene ring with N methoxy groups and ipso carbon atoms at the *para* positions on the common ring (see Figure 1B). The hybrid molecules are identified in the text by the names "N_{site1} substituent × N_{site2} substituent" where substituents "h", "oh", and "och_3" designate the hydrogen atom, hydroxy, and methoxy moieties, respectively.

2.2.2. Benzene Derivatives. A hybrid ligand was constructed to represent six benzene derivatives: benzene, toluene, benzaldehyde, phenol, 4-methyl phenol, and 4-hydroxybenzaldehyde. At site 1, a hydrogen atom and methyl and aldehyde groups along with their corresponding ipso carbon atoms were modeled; at site 4, a hydrogen atom and hydoxyl group along with their corresponding ipso carbon atoms were modeled (see Figure 1C).

Table 1. Molecular Structures and the Corresponding Experimental IC_{50} and Binding Free Energies of the TIBO Analogues

compound	X	Y	IC_{50}[a] (μM)	ΔG_{bind}[b] (kcal/mol)	hybrid molecule
1	S	8-Br	0.0030	−12.09	E
2	S	8-CH_3	0.0136	−11.16	D
3	S	8-CCH	0.0296	−10.69	E
4	S	H	0.0440	−10.44	D, E, F
5	S	8-I	0.0474	−10.39	E
6	O	8-Br	0.0473	−10.39	E
7	S	8-CN	0.0563	−10.29	D
8	O	8-I	0.0880	−10.01	E
9	O	8-CCH	0.4376	−9.02	D
10	S	9-CF_3	0.4850	−8.96	F
11	O	8-CH_3	0.9890	−8.52	D
12	O	8-CN	1.1396	−8.43	D
13	O	H	3.1550	−7.81	D, E, F
14	O	9-CF_3	5.9190	−7.42	F

[a] Refs 34 and 35. [b] Calculated from $\Delta G_{bind}=RT \ln IC_{50}$ at 310 K.

Experimental hydration free energies were compiled from Cabini et al.;[33] the relative hydration free energies for all pairs of these compounds are listed in Table 2.

2.2.3. TIBO Compounds. Three hybrid TIBO molecules were constructed to represent a total of 14 unique inhibitors of HIV-1 reverse transcriptase. Figures 1D−F illustrate the hybrid ligands, and Table 1 summarizes the experimental binding free energies[34,35] of the TIBO compounds that were included in these calculations. Each hybrid molecule contained both C═O and C═S variations at the X site as well as variations involving nonhalides (Figure 1D) or halides (Figure 1E) at the C-8 site or variations at the C-9 site (Figure 1F). One TIBO pair (X═O→S, Y═H) is represented in each of the three hybrid molecules.

2.3. Simulation Details. In all MSλD simulations, the leapfrog Verlet algorithm was used to integrate the equations of motion and propagate the atomic coordinates and velocities as well as the θ values and their velocities. A nonbonded cutoff of 15 Å was used, and van der Waals switching and electrostatic shifting functions were implemented between 10 Å and 12 Å unless otherwise specified. Hydrogen bonds were restrained using the SHAKE[36] algorithm, and the time step was 2 fs unless otherwise specified. Each θ_i was assigned a fictious mass of 5 amu·Å^2, and λ values were saved every 10 steps. 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. The threshold value for assigning $\lambda_{\alpha,i} \approx 1$ was $\lambda_{\alpha,i} \geq 0.8$ unless otherwise specified. To enhance transition rates between substituents, restraint functions were employed to superimpose the ipso carbons on one another at each site, and variable biases ($F^{variable}$) were added to the hybrid potential energy function in eq 6 for each $\lambda_{\alpha,i}$:

$$F_{\alpha,i}^{variable} = \begin{cases} k(\lambda_{\alpha,i} - 0.8)^2 \text{kcal/mol}; & \text{if } \lambda_{\alpha,i} < 0.8 \\ 0; & \text{otherwise} \end{cases} \quad (11)$$

with force constants, k, assigned between 0 and 7 kcal/mol. The temperature was maintained near 310 K by coupling to a Langevin heat bath using a frictional coefficient of 10 ps^{-1} for all nonprotein atoms and 5 ps^{-1} for each θ_i. Ten independent simulations using

different initial seed values of θ_i were performed unless otherwise stated, and the resulting averages and standard deviations were reported. All calculations and analyses were performed using the new implementation of multisite λ dynamics in the BLOCK module in CHARMM version c36a4 on dual 2.66 GHz Intel Quad Core Xeon processors.

In all FEP/BAR calculations, for each pair of compounds, 23 λ windows (λ = 0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.75, 0.8, 0.85, 0.9, 0.925, 0.95, 0.975, 0.99, and 1.0) were used with soft-core potentials,[37] and the resulting trajectories were analyzed using the Bennett acceptance ratio (BAR) method.[38] Three independent series of simulations were performed for each pair, and the average and standard deviation over these three series are reported. All simuations were performed using the BLOCK module in CHARMM, and BAR analyses were performed using a modified version of pyMBAR.[39]

2.3.1. Model Compounds. Solvent simulations were performed using 351 TIP3P[40] water molecules in a water box of 22 Å^3 with periodic boundary conditions. Each θ_i was assigned a fictitious mass of 12 amu·Å^2. Heating and equilibration phases were 4 and 10 ps, respectively, and production runs were 25 and 3 ns for vacuum and solvation simulations, respectively, unless otherwise stated. A nonbonded cutoff of 15 Å was used, and van der Waals switching and electrostatic force shifting functions were implemented between 10 Å and 12 Å.

2.3.2. Benzene Derivatives. All relative hydration free energies were computed from the difference between relative free energy changes evaluated from solvated and vacuum simulations. Solvent simulations were performed using 351 TIP3P[40] water molecules in a water box of 22 Å^3 with periodic boundary conditions.

Only one set of MSλD simulations was required to compute the relative hydration free energies among all pairs of benzene derivatives. Relative free energies for each pair were averaged over results obtained from three independent trajectories. Initial biases $\{F_i\}$ for the six benzene derivatives were approximated in the solvent and vacuum simulations by the minimized GBMV[41,42] and vacuum energies for the individual ligands, respectively. Biases were optimized iteratively so that the final simulations in a given environment were projected to yield at least 50 transitions per site and at least 400 snapshots in which each unique compound was assigned to be "dominant", i.e., the substituent at site 1 and substituent at site 4 that corresponded to the compound had $\lambda_i \approx 1$. Heating and equilibration phases were 4 ps each. Production runs for the bias optimization stages were 200 and 20 ps for vacuum and solvent simulations, respectively, while the production runs for the final stage were 2 ns.

Nine series of traditional alchemical free energy simulations were performed for pairs of benzene derivatives that differed from one another at only one substituent site; from these simulations, relative hydration free energies for the remaining six pairs of compounds were inferred. Short heating and equilibration phases were performed, and production runs were 1 ns for both vacuum and solvent environments.

2.3.3. TIBO Compounds. Relative binding free energies were computed via thermodynamic cycles by performing MSλD simulations for subsets of TIBO compounds both in a solvent and while bound to the non-nucleoside reverse transcriptase inhibitor (NNRTI) binding pocket in HIV-1 RT. The same simulation parameters were used as was described in ref 31. For the solvation simulations, the hybrid molecules were solvated in a 20 Å^3 box of 244 TIP3P[40] water molecules, and periodic boundary conditions were employed. For the bound simulations, the PDB structure, 1TVR,[43] was truncated so that only residues within 20 Å of the crystallographic TIBO compound were retained, and the truncated protein−ligand system was solvated in a 20 Å sphere of 667 TIP3P[40] water molecules. Stochastic boundary conditions using a solvent boundary potential[44] of 22 Å with a 15 Å buffer region were employed. In addition to the bond and angle terms, the dihedral angle terms of the hybrid ligand were treated at full strength regardless of λ value.

Three sets of MSλD simulations were required to compute the relative binding free energies among the 14 TIBO compounds. Initial fixed biases $\{F_i\}$ in the solvent environment were estimated from the minimized energies of the individual ligands using the GBMV implicit solvent model[41,42] and were optimized iteratively using simulations of 200 ps in length such that the final simulations were projected to yield at least 50 transitions per site and 400 snapshots in which each unique ligand was assigned to be "dominant". Biases $\{F_i\}$ for the protein simulations were assigned from the MSλD-estimated $\Delta\Delta G_{solv}$ values. Final production runs were 2 and 1 ns for the solvated and bound simulations, respectively. Ligand populations from six 1 ns MSλD simulation trajectories were combined to compute relative free energies. The reported averages and standard deviations for MSλD simulations are calculated from three independent series of six MSλD trajectories.

Series of alchemical free energy simulations were performed for all 27 pairs of TIBO compounds, i.e., all pairs within the three hybrid molecules whose identities varied at only one site. Short heating and equilibration phases were performed, and production runs were 500 and 250 ps for solvent and bound environments, respectively.

2.4. Model Quality. All MSλD trajectories were analyzed using new routines that we have implemented in CHARMM. The fraction physical ligand (FPL) is a metric that describes the proportion of time that a full or physical ligand, as compared with only partial or unphysical ligands, is present during the course of a simulation. Transition rates ($\tau_{trans}(\alpha)$) among substituents on site α quantify the rate at which the identity of the "dominant" substituent changes throughout a simulation. The relative free energy difference for each pair of compounds (ij) in the hybrid molecules was estimated by averaging over results from multiple simulation trajectories. Alternatively, ligand populations from multiple trajectories were combined to compute the relative free energy difference by

$$\Delta\Delta G_k(ij) = -k_B T \ln \frac{\sum_{k=1}^{M} P(\lambda_j = 1)}{\sum_{k=1}^{M} P(\lambda_i = 1)} \quad (12)$$

3. RESULTS

A new version of λ dynamics, termed multisite λ dynamics (MSλD), has been implemented and is capable of simultaneously evaluating multiple chemical substituents at any number of sites on a common ligand core. The efficacy of this implementation of MSλD free energy simulations has been tested in three model systems. Figure 1 illustrates representative molecular structures of the alchemical hybrid molecule that are associated with each of these model systems. The first system includes a series of benzene, dihydroxybenzene, or dimethoxybenzene molecules in which multiple hydrogen atoms or hydroxy or methoxy groups

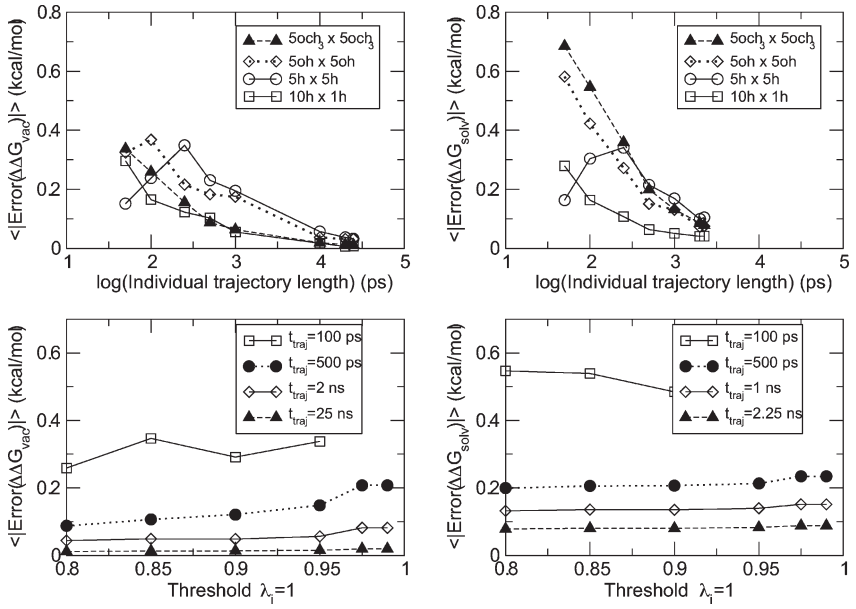

Figure 2. Sensitivity of MSλD simulations in estimating relative free energy differences based on hybrid ligands representing multiple, identical benzene (nh × nh) molecules, 25 dihydroxybenzene (5oh × 5oh) molecules, and 25 methoxybenzene ($5och_3 \times 5och_3$) molecules based on trajectory length in (A) vacuum and (B) solvent environments. Sensitivity of MSλD simulations for the 25 methoxybenzene ($5och_3 \times 5och_3$) molecules as a function of the threshold value used to assign $\lambda_{i,\alpha} \approx 1$ in (C) vacuum and (D) solvent environments. The reported unsigned errors were averaged over all ligand pairs whose individual relative free energies were obtained from data combined from 10 independent MSλD trajectories. Note: not all ligands were sampled as the "dominant" ligand in the 100 ps vacuum trajectories with threshold values $\lambda_{i,\alpha}$ >0.95 .

and their corresponding ipso carbon atoms are present at either site 1 or at site 1 and site 4 on a common benzene framework. The second system represents a series of six benzene derivatives: benzene, toluene, benzaldehyde, phenol, 4-methyl phenol, and 4-hydroxybenzaldehyde, for which experimental hydration free energies are available.[33] The third system represents a series of TIBO derivatives whose binding affinities to HIV-1 reverse transcriptase are known experimentally.[34,35] Since each substituent at each site interacts with each of the substituents at the other sites, a hybrid molecule with multiple substituents at two sites represents $N_{site1} \times N_{site2}$ distinct molecules where N_α is the number of substituents that are modeled at site α.

3.1. Relative Free Energies for Multiple Benzene, 1, 4-Dihydroxybenzene and 1,4-Dimethoxybenzene Molecules. In previous work,[27] we computed relative free energies in a vacuum and a solvent for pairs of identical benzene molecules that are represented in the model hybrid ligands. Given that each benzene molecule is assigned the same force field parameters, relative free energy differences of exactly 0 kcal/mol should theoretically be obtained in any environment; thus, any deviations in the simulation estimates from 0 kcal/mol can be understood as errors due to limitations in the MSλD sampling specifically. In addition, we considered the increasingly flexible and therefore more challenging systems of multiple, identical 1,4-dihydroxybenzene and 1,4-dimethoxybenzene compounds. For up to 10 substituents evaluated at a single site and five substituents evaluated at each of two sites, the average errors are on the order of 0.03 and 0.07 kcal/mol with maximum deviations of 0.12 and 0.28 kcal/mol for vacuum (25 ns) and solvent (2.25−3 ns) simulations, respectively.

Here, we characterize the sampling efficiency of MSλD in computing the free energy differences among these different hybrid ligands. First, the quality of results for the model compounds in vacuum environments is primarily dependent on the length of the simulation and is less affected by the complexity of the substituents which are being sampled, as demonstrated in Figure 2a. For each model system in a vacuum, the average unsigned error of the relative free energies estimated for all benzene, dihydroxybenzene, or dimethoxybenzene pairs that are averaged over 10 simulations each of length 250 ps are within 0.22 kcal/mol of the exact solution. The maximum observed errors for these short trajectory lengths are non-negligible, however, with deviations of up to 0.9 kcal/mol for any single ligand pair from simulations of the hybrid ligands representing 10 benzene, 25 benzene, 25 dihydroxybenzene, and 25 dimethoxybenzene molecules. However, within 10 ns (~8 CPU minutes on a single processor), the average unsigned error over all ligand pairs is 0.03 kcal/mol, and all ligand pairs have errors of less than 0.15 kcal/mol.

Benzene and dihydroxybenzene hybrid ligands experience similar transition rates among the "dominant" substituents at each site relative to one another and in both vacuum and explicit solvent environments. By contrast, transition rates for the dimethoxybenzene hybrid ligands are systematically slower than the benzene and dihydroxybenzene hybrid ligands and are systematically slower in a solvent than in a vacuum. This reduction in transition rates results in longer convergence times for simulations of the more flexible 1, 4-dimethoxybenzene relative to the corresponding benzene and dihydroxybenzene simulations, as shown in Figure 2b.

Finally, the quality of the relative free energy estimates is virtually insensitive to the specific threshold value that is used to define the presence of a "dominant" substituent, i.e., where $\lambda_{\alpha,i} \approx 1$ for the $\lambda_{\alpha,i}$ > threshold. While the actual ligand populations that are used in eq 8 to compute relative free energies depend on the threshold value, the ratio of the ligand populations is stable for threshold values between 0.8 and 0.95. For threshold values below 0.95, the accuracy of the relative free energies is primarily dependent on the length of the trajectory. Figure 2c,d

Table 2. Relative Hydration Free Energies ($\Delta\Delta G$) in kcal/mol for All Pairs of Six Benzene Derivatives Computed by MSλD and Alchemical Free Energy Perturbation Simulations Analyzed Using the Bennett Acceptance Ratio Method (FEP/BAR)[a]

		exptl	MSλD					FEP/BAR				
site1_site4		$\Delta\Delta G$[b]	$\Delta\Delta G_{vac}$		$\Delta\Delta G_{solv}$		$\Delta\Delta G$	$\Delta\Delta G_{vac}$		$\Delta\Delta G_{solv}$		$\Delta\Delta G$
react.	prod.		avg	σ	avg	σ	avg	avg	σ	avg	σ	avg
h_h	h_oh	−6.41	−10.58	0.07	−15.18	0.12	−4.60	−10.61	0.03	−15.15	0.04	−4.54
h_h	ch3_h	−0.02	−6.82	0.05	−7.29	0.13	−0.47	−6.76	0.03	−6.58	0.02	0.18
h_h	ch3_oh	*−5.88*	*−17.29*	*0.07*	*−22.34*	*0.14*	*−5.05*	*−17.24*	*0.05*	*−21.59*	*0.10*	*−4.34*
h_h	cho_h	−3.52	7.38	0.09	5.23	0.16	−2.14	7.39	0.01	5.78	0.04	−1.62
h_h	cho_oh	*−10.72*	*−2.41*	*0.20*	*−9.54*	*0.22*	*−7.12*	*−2.42*	*0.07*	*−8.77*	*0.08*	*−6.35*
h_oh	ch3_h	*6.39*	*3.76*	*0.08*	*7.89*	*0.13*	*4.12*	*3.85*	*0.06*	*8.56*	*0.07*	*4.72*
h_oh	ch3_oh	0.54	−6.71	0.04	−7.16	0.10	−0.45	−6.63	0.03	−6.44	0.06	0.20
h_oh	cho_h	*2.89*	*17.96*	*0.07*	*20.41*	*0.14*	*2.45*	*18.00*	*0.04*	*20.93*	*0.08*	*2.92*
h_oh	cho_oh	−4.30	8.17	0.16	5.64	0.19	−2.53	8.19	0.05	6.37	0.03	−1.81
ch3_h	ch3_oh	−5.85	−10.47	0.07	−15.04	0.09	−4.57	−10.51	0.03	−15.00	0.04	−4.50
ch3_h	cho_h	−3.50	14.20	0.10	12.52	0.10	−1.67	14.19	0.00	12.39	0.02	−1.81
ch3_h	cho_oh	*−10.69*	*4.41*	*0.19*	*−2.24*	*0.17*	*−6.65*	*4.24*	*0.06*	*−2.36*	*0.09*	*−6.61*
ch3_oh	cho_h	*2.35*	*24.67*	*0.05*	*27.57*	*0.09*	*2.90*	*24.80*	*0.03*	*27.54*	*0.07*	*2.73*
ch3_oh	cho_oh	−4.84	14.88	0.16	12.80	0.14	−2.08	14.85	0.03	12.79	0.05	−2.07
cho_h	cho_oh	−7.19	−9.79	0.16	−14.77	0.14	−4.98	−9.83	0.01	−14.56	0.02	−4.74

[a] Averages (avg) and standard deviations (σ) are calculated from three independent MSλD trajectories via eq 8 or a series of FEP/BAR trajectories. $\Delta\Delta G_{vac}$ and $\Delta\Delta G_{solv}$ represent the relative free energies estimated in the vacuum and solvent environments, respectively. FEP/BAR values in italics represent cases where the reactant and product differ at both sites, and thus $\Delta\Delta G$ estimates were obtained by combining results for the two FEP/BAR simulations evaluating changes at one site and then the other site. [b] Calculated from ref 33 at 310 K.

illustrate the quality of results in vacuum and solvent environments that are obtained for the dimethoxybenzene hybrid ligand which contained five substituents at each site as a function of the specific cutoff or threshold value that is used to define $\lambda_{\alpha,i} \approx 1$.

3.2. Relative Hydration Free Energies of Six Benzene Derivatives. Relative hydration free energies were computed for six benzene derivatives that differ from one another by the identity of the substituents in the para positions. To control for the influence of the force field parameters on the quality of the simulation results, MSλD results are first compared with those obtained from the Bennett acceptance ratio method analysis of traditional alchemical free energy simulations using the same ligand parameters (FEP/BAR; see Table 2). As shown in Figure 3a−c, the relative free energies estimated from MSλD and FEP/BAR agree very well with one another for each arm of the thermodynamic cycle (with R^2 values of 1.00 and 0.999 for the vacuum and solvent environments, respectively) and for the overall relative hydration free energies (with an R^2 value of 0.990 and slope of 1.0). The average unsigned difference between relative hydration free energies computed by FEP/BAR and MSλD for each pair of compounds is 0.29 kcal/mol with a maximum deviation of 0.56 kcal/mol.

The relative hydration free energies estimated from the MSλD simulations correlate well with those that were measured experimentally and have an average unsigned error (AUE) of 1.8 kcal/mol relative to experimental values and an R^2 of 0.959 (Figure 3d). The overall error, however, is dominated by contributions from pairs that include 4-hydroxybenzaldehyde; specifically, the five ligand pairs that include 4-hydroxybenzaldehyde have an AUE of 2.9 kcal/mol while the remaining 10 pairs have an AUE of 1.2 kcal/mol. The FEP/BAR results also yield an overall AUE of 1.8 kcal/mol with errors of 3.2 and 1.1 kcal/mol for pairs that include and exclude 4-hydroxybenzaldehyde, respectively.

We also explored the impact of averaging relative free energy changes obtained over multiple trajectories for a hybrid ligand. When multiple independent simulations initially explore different regions of phase space, the individual trajectories can be combined to give reasonable relative free energy estimates in a short amout of time. For example, ligand populations from 10 10-ps simulations were combined to give relative free energy estimates that are within 0.3 kcal/mol of the converged values, as shown in Figure 4a. However, the data from the explicit solvent simulation trajectory depicted in Figure 4b illustrate that longer trajectories are required to sufficiently explore the λ phase space. In fact, trajectory lengths of 100 ps are required for each of the 10 solvent simulations to sample all six distinct benzene derivatives. These results agree well with those that would be predicted from transition rates obtained in the previous section using model compounds. Using estimated transition rates for the model hybrid dimethoxybenzene compounds of 0.2−1.1 ps^{-1} that were reported in ref 27 and assuming that transitions are equally probable among each of the substituents at each site and that transitions at the two sites are not correlated, a minimum of 30 transitions or 30−150 ps would be required to sample each of the six distinct benzene derivatives five different times during a single trajectory. With total combined trajectory lengths of 1 ns, whether from a single simulation of 1 ns or 10 simulations of 100 ps, the relative free energy estimates are within 0.15−0.25 kcal/mol of the converged results.

3.3. Relative Binding Free Energies of 14 HIV-1 RT Inhibitors. In our previous study,[31] relative binding affinities for 44 pairs of TIBO compounds (among 21 unique TIBO compounds) were estimated from traditional alchemical free energy simulations analyzed by thermodynamic integration. In this current study, we are focusing on the ability of MSλD to estimate relative binding free energies when modeling different substituents at more than one site on a common ligand core simultaneously and so have only

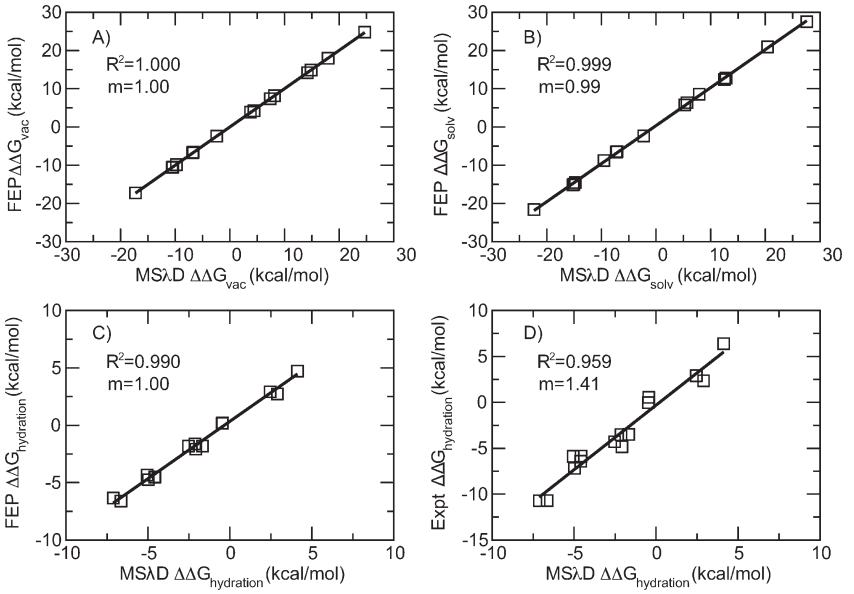

Figure 3. Correlation between MSλD-calculated and FEP/BAR-calculated relative free energies in (A) vacuum and (B) solvent environments for a series of six benzene derivatives and (C) the corresponding relative hydration free energies. (D) Correlation between MSλD-calculated and experimental relative hydration free energies.

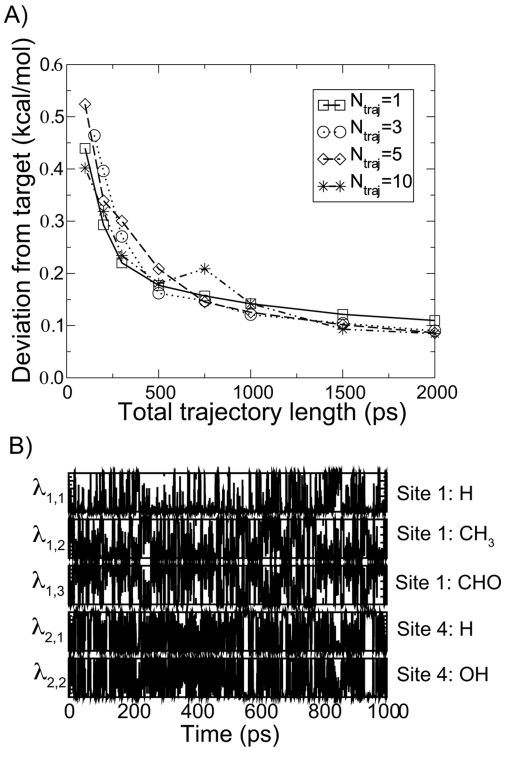

Figure 4. (A) Unsigned deviations for relative free energies in a solvent were averaged over N_{traj} simulations for all pairs in the hybrid ligand representing six benzene derivatives. The targets were defined for each pair by combining data from 10 independent trajectories each of 2 ns. (B) Representative data from an explicit solvent simulation trajectory for the hybrid benzene model.

included the 14 TIBO compounds for which experimental binding data are available for both X = O and X = S for a given substituent on site Y. In the previous study, we used different patch sizes (i.e., the number of atoms that were used to represent the substituents at each site) and scaled the dihedral angle terms by their corresponding $\{\lambda_i\}$; therefore, we repeated the free energy simulations so that we could compare each arm of the thermodynamic cycle with results obtained from the corresponding MSλD simulations. Free energy simulation trajectories were run for significantly longer than in our previous study, and soft-core potentials were used to ensure the convergence of the alchemical free energy simulations at the end points.

Table 3 and Figure 5 summarize the results for the relative free energies that have been computed for pairs of these TIBO compounds in solvent and protein environments by using either the new implementation of MSλD simulations or traditional alchemical free energy simulations analyzed by BAR. Transition rates at each site in the MSλD simulations are comparable to those observed for the model compounds and benzene derivatives modeled in a solvent and range from 0.02 to 0.9 ps^{-1}. The agreement between these two methods for estimates in the solvent environment is very good: the average unsigned difference and maximum difference are 0.55 and 1.30 kcal/mol, respectively. The R^2 is 0.998, and the slope of the regression line is 0.98. For the protein environment, the agreement in the relative free energies estimated with these two methods is also reasonably good: the average unsigned difference and maximum difference are 1.15 and 3.16 kcal/mol, respectively, while the R^2 and slope of the regression line are 0.987 and 0.96, respectively. In both solvent and protein environments, the difference between the relative free energies computed by FEP/BAR and MSλD tends to increase with the size differential of the pairs of compounds. For example, in solvent, the average unsigned difference in the relative free energies obtained from the two methods is 0.29, 0.67, and 0.96 kcal/mol for pairs of compounds whose substituents differ by 0, 1, and 2 heavy atoms, respectively. Similarly, in the protein environment, the average unsigned difference between the two methods is 0.73, 1.39, and 1.76 kcal/mol for

Table 3. Samping Characteristics of the Relative Free Energies for All Pairs of TIBO Compounds Estimated from Series of Alchemical Free Energy Simulations Analyzed by BAR (FEP/BAR) and from MSλD Simulations[a]

		FEP/BAR		MSλD				
hybrid ligand	pairs	$\langle\sigma\rangle$	FPL	τ_{trans}(X) (ps^{-1})	τ_{trans}(Y) (ps^{-1})	$\langle\sigma\rangle$	$\langle\Delta\rangle^b$	max Δ^c
				$\Delta\Delta G_{solv}$				
D	16	0.05	0.11	0.27	0.66	0.03	0.57	1.30
E	9	0.05	0.15	0.03	0.89	0.09	0.50	0.96
F	4	0.08	0.78	0.17	0.21	0.10	0.54	0.81
overall	*29*	*0.05*				*0.05*	*0.55*	*1.30*
				$\Delta\Delta G_{prot}$				
D	16	0.38	0.92	0.02	0.20	0.29	1.34	3.15
E	9	0.43	0.96	0.02	0.06	0.21	0.96	2.00
F	4	0.30	0.97	0.04	0.03	0.20	0.80	1.37
overall	*29*	*0.39*				*0.25*	*1.15*	*3.15*

[a] Averages and standard deviations (σ) are calculated from three independent series of six MSλD trajectories via eq 12 and three independent series of the FEP/BAR trajectories. $\Delta\Delta G_{solv}$ and $\Delta\Delta G_{prot}$ represent the relative free energies estimated in the solvent environment and binding pocket, respectively, and are reported in units of kcal/mol. The Fraction Physical Ligand (FPL) represents the fraction of time that a physical ligand is present during the simulation as opposed to several partial ligands. [b] $\langle|\Delta\Delta G^{MS\lambda D} - \Delta\Delta G^{FEP/BAR}|\rangle$. [c] Max $|\Delta\Delta G^{MS\lambda D} - \Delta\Delta G^{FEP/BAR}|$.

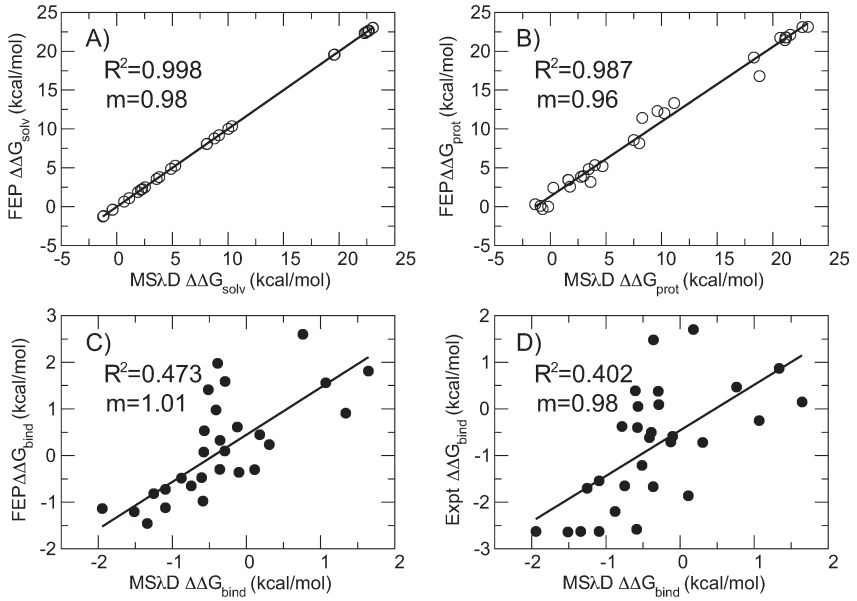

Figure 5. Correlation between MSλD-calculated and FEP/BAR-calculated relative free energies for 29 pairs of TIBO compounds that are inhibitors of HIV-1 reverse transcriptase in (A) solvent and (B) protein environments and (C) the corresponding relative binding affinities. (D) Correlation between MSλD-calculated and experimental relative binding free energies.

pairs of compounds whose substituents differ by 0, 1, and 2 heavy atoms, respectively.

As expected, in both methods, the uncertainty in the relative free energy estimates for the bound simulations is larger than that in the solvent environment. The average standard deviation of the free energy estimates from the three series of MSλD and FEP/BAR solvent simulations is 0.05 kcal/mol. By comparison, the average standard deviations for the protein simulations are 0.25 and 0.39 kcal/mol for the MSλD and FEP/BAR simulations, respectively. The average difference between the overall relative binding affinities for these pairs of TIBO inhibitors estimated by MSλD and FEP/BAR simulations is 0.7 kcal/mol, and the maximum difference is 2.4 kcal/mol. The range of predicted relative binding affinities for these pairs of TIBO inhibitors is quite small, so the R^2 of 0.473 between the relative binding affinities estimated from MSλD and FEP simulations is significantly lower than those obtained for relative free energy estimates in the solvent and bound environments.

Finally, the average unsigned error of the relative binding affinities for the 29 pairs of TIBO inhibitors is 0.9 and 1.3 kcal/mol for the MSλD and FEP/BAR simulations, respectively, and the largest errors are 2.0 and 2.6 kcal/mol, respectively (Table 4). The magnitudes of these errors are comparable to those that have been reported for FEP calculations of inhibitors bound to wild-type or mutant HIV-1 RT[45−49] and that may be considered reasonable for drug design applications. For example, in their recent paper, Chodera et al. state

Table 4. Average Unsigned Errors (AUE) and Maximum Errors (MaxE) in the Computed Relative Binding Affinities, $\Delta\Delta G_{bind}$ (in kcal/mol), for All Pairs of TIBO Compounds[a]

		FEP/BAR		MSλD	
hybrid ligand	pairs	AUE	MaxE	AUE	MaxE
D	16	1.42	2.62	0.77	1.97
E	9	1.23	2.38	1.05	1.99
F	4	1.00	1.17	1.14	1.84
overall	*29*	*1.30*	*2.62*	*0.91*	*1.97*

[a] Computed relative binding affinities are calculated from the differences in the $\Delta\Delta G_{prot}$ and $\Delta\Delta G_{solv}$ obtained from averaging three independent series of six MSλD trajectories via eq 12 and three independent series of the FEP/BAR trajectories.

Table 5. Percentage of Pairs of TIBO Compounds That Are Correctly Ranked with the Free Energy Methods and Force Field Parameters Used in This Work

$\Delta\Delta G^{expt}$ (kcal/mol)	# of pairs	FEP/BAR	MSλD
>2.0	6	100	100
>1.5	12	100	92
>1.0	14	93	93

that "statistical models of prediction-guided prioritization suggest that even moderate accuracy (RMS errors of 2 kcal/mol) could be sufficient to produce substantial efficiency gains in lead optimization campaigns."[4] The experimental relative binding affinities for these pairs of TIBO inhibitors are within 2.6 kcal/mol of each other, and thus the R^2 between MSλD-estimated and experimental relative binding affinities of 0.402 is quite encouraging. Perhaps a more interesting and relevant metric for the discriminatory ability of these free energy methods in structure-based drug design is the percentage of time that the better binder from a pair of compounds is correctly identified. Given that the experimental uncertainty associated with the binding affinity measurements are on the order of 0.5−1 kcal/mol[3], only pairs of compounds which have binding free energy differences above a certain threshold are examined. As demonstrated in Table 5, of the 14 pairs of TIBO compounds that have differences in their experimental binding affinities that are greater than 1 kcal/mol, FEP/BAR and MSλD correctly identify the more potent inhibitor 93% of the time. This percentage increases to 100 when the six pairs of compounds with experimental binding affinity differences of greater than 2 kcal/mol are considered.

4. DISCUSSION

4.1. Sampling Characteristics. Simulation results for the three model systems in this study clearly demonstrate the robustness of sampling using MSλD. The relative free energy differences between pairs of compounds that are estimated from MSλD agree well with theoretical values (in the case of the multiple identical benzene, dihydroxybenzene, and dimethoxybenzene compounds) and with traditional alchemical free energy methods (in the case of hydration free energies of the six benzene derivatives and binding free energies of the TIBO compounds). In MSλD simulations, the efficiency of the simulations is directly related to the proportion of time that a "dominant" ligand is represented relative to the partial, nonphysical ligands, and to the number of transitions or times that the identity of the "dominant" substituent at each site changes and more specifically, the number of times the identity of the "dominant" ligand changes. Thorough sampling enhances the convergence of the dominant ligands and thus the reliability of the relative free energies that are estimated from these populations using eq 8.

In the first model system, sampling convergence is most readily achieved for the smaller "h" and "oh" substituents since the substituents at each site sample very similar conformations at each time step. Given that each of these smaller substituents experiences comparable magnitudes of the unscaled interaction energies with the rest of environment at each time step, transitions between "dominant" substituents are relatively common. With the addition of the larger, more flexible methoxy moieties on the common benzene ring, however, the unscaled interaction energies between each substituent and the environment can be quite different from one another at a given point in time, which reduces the transition probabilities between dominant substituents for the dimethoxybenzene hybrid molecule relative to transition rates between the hybrid benzene and dihydroxybenzene molecules. This effect is magnified in the solvent simulations where systematically lower transition rates are observed. Visual inspection of the trajectories confirmed that the methoxy groups explore 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 1,4-dimethoxybenzene ligands in the "dominant" ligand state.

Transitioning between dominant ligands can be even more challenging in the context of a binding pocket where substituent sites interact with distinct parts of the pocket. In solvent environments, local solvent configurations surrounding two substituent sites that are remote from one another can rearrange relatively independently from each other. By contrast, fluctuations in different parts of a given protein binding pocket influence one another due to their backbone connectivity and specific intermolecular interactions and thus are correlated at longer time scales with one another than are solvent configurations. By constucting hybrid ligands that contain substituents of similar molecular volume or polarity to one another, the extent of the rearrangement of the environment that is required to enable one substituent to replace the "dominant" substituent is minimized; therefore, the probability of transitions among substituents on a hybrid ligand will tend to increase, and the ability of each of the dominant ligands to be sampled will improve. If clusters of substituents are selected such that at least one ligand in each hybrid molecule overlaps with that of another hybrid molecule, the potency of each represented ligand can be localized on a single relative scale with one another. With at least one experimental measurement, these calculated estimates are able to be placed on an absolute scale.

Other simulation parameters can be varied to enhance the efficacy of MSλD simulations. For example, decreasing m_θ will tend to increase the mobility of the θ values and thus increase transition rates, though, in our experience, values of m_θ on the order of 5−20 amu·Å^2 are usually reasonable. Alternatively, adding biasing potentials associated with the θ values that take effect only when $\lambda_{\alpha,i} < 0.8$ as in eq 11 or increasing the magnitude of the force constant, k, on these biasing potentials will also tend to increase the transition rates. Though, with increasingly large force constants, there will be a concomitant increase in the amount

of time spent at intermediate λ values (data not shown) and lower populations for the free energy differences required in eq 8.

4.2. Computational Expense. The most significant difference between MSλD and alchemical free energy simulations lies in their relative computational expense. Results estimated from MSλD simulations for the hybrid molecule representing six benzene derivatives were achieved ~250 times faster than those for the reported FEP/BAR calculations. However, the triplicate FEP/BAR calculations were more precise, and so a more appropriate comparison would be for trajectory lengths that achieve a similar level of precision. FEP/BAR simulations in which each λ window for each pair was sampled for ~40 ps yield an average precision in the relative free energy estimates of 0.15 kcal/mol and require ~20 times more computational resources than MSλD simulations of comparable precision.

Similarly, relative binding affinity estimates for the TIBO compound series for comparable levels of precision are achieved by MSλD simulatons about 50 times faster than the corresponding series of alchemical free energy simulations. The 1 ns MSλD trajectories for the bound environment were sufficiently long to ensure that each physical ligand was sampled in the dominant state in all but a few simulations; however, significantly shorter simulations were sufficient to identify the most favorable binders. Arguably, in prospective drug-design applications, it is primarily the best binders that are sought and not necessarily a quantitative ranking of each ligand. Therefore, these relative times represent a lower bound on the computational efficiency of MSλD over traditional alchemical free energy simulations that are restricted to sampling individual pairs of ligands.

4.3. MSλD Applications. Application of MSλD simulations to structure-based drug design strategies is promising given its computational expediency relative to traditional alchemical free energy simulations for exploring multiple variants at multiple sites on a common ligand framework. However, the quality of the relative hydration free energies or binding affinities is also dependent on the quality of the underlying force field parameters. In this study, reasonable overall agreement is observed between experimental relative free energies and those obtained from MSλD simulations using CGenFF parameters. The overall average unsigned error for the relative hydration free energies of the series of benzene derivatives is 1.8 kcal/mol, while the average unsigned error for the relative binding affinities for pairs of TIBO compounds is 0.9 kcal/mol. However, it was also clear that some ligands were poorly modeled relative to the others and suggestive of errors in the underlying force field parameters associated with the corresponding functional groups. Specifically, the quality of the 4-hydroxybenzaldehye parameters yields systematically underestimated relative hydration free energies (by 1.2−2.9 kcal/mol) compared with the rest of the benzene derivatives. 4-Hydroxybenzaldehye has the largest dipole moment of the ligands in this series, and its electronic structure is arguably the most sensitive to the local solvent environment. Thus, the comparably poor quality associated with its calculated relative hydration free energies is not unexpected, and it is possible that more sophisticated polarizable charge models may be needed to capture its true solvation properties.

Given the uncertainties in the sampling on the order of 1 kcal/mol combined with inconsistencies in the force field parameters, MSλD is not the silver bullet that will reliably identify the best binder out of a pool of very good binders. The strength of MSλD simulations is likely to be in the lead optimization stage where a compound has been identified that binds at micromolar concentrations to a given macromolecular target. Different functional groups at various sites on this lead compound can be systematically screened for their ability to improve the binding affinity relative to the lead compound.[11] Thus, these MSλD simulations could bridge the gap between high-throughput docking studies that survey libraries of hundreds of thousands of diverse compounds and the much more expensive alchemical free energy calculations that are usually performed on only a handful of chemical variants of the lead compound. MSλD also appears to be a reasonable and rapid method for validating and optimizing 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.

4.5. Functional Form of λ Values. In preparation for this study, we investigated various functional forms of the λ values that implicitly satisfy the holonomic constraints: $0 \leq \lambda_{\alpha,i} \leq 1$ and $\sum_{i=1}^{N} \lambda_{\alpha,i} = 1$. The functional form represented in eq 9 is used throughout this study and exhibits ideal characteristics for multisite λ dynamics simulations that mimick SAR strategies. First, it offers increased numerical stability at larger integration stepsizes over the original λ dynamics implementaion that used a simplified Lagrange multiplier method and renomalization at every time step.[27] Second, the periodicity of this functional form is oscillating in nature and so provides enhanced sampling of each of the $\lambda_i \approx 1$ states. Third, both the values of $\lambda_{\alpha,i}$ and the forces on $\lambda_{\alpha,i}$ are computationally inexpensive, and each $\lambda_{\alpha,i}$ has the same probability density function, so no further bias or correction is required to account for differences in effective phase space volume sampled. Finally, this functional form promotes rapid transitions 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 $\lambda_{\alpha,i} \approx 1$. The coefficient c in eq 7 can be tuned to describe the steepness of the switching between $\lambda_{\alpha,i} \approx 1$ and $\lambda_{\alpha,i} \approx 0$, and we have identified a "sweet spot" coefficient of 5.5 that seems optimal for MSλD simulations. Coefficients of less than 5.5 do not transition as quickly, so a larger fraction of θ space is restricted to intermediate λ values and 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 end points are too abrupt in solvent simulations to retain the stability in the numerical integrator. Shorter timesteps and/or soft-core potentials could alleviate this problem; however, the functional form in eq 9 implicitly protects the system by implicitly restraining the upper bound of λ to be

$$\lambda_{\alpha,i}^{\max} \leq \frac{e^{5.5}}{e^{5.5} + (N_\alpha - 1)e^{-5.5}} \tag{13}$$

or where $N = 5$ the boundaries are $0.000016 < \lambda_{\alpha,i} \leq 0.99993$ on site α.

5. CONCLUSION

In the present study, we have presented the multisite λ dynamics method and its application in three model systems, including computing relative hydration free energies for a series of benzene derivatives and relative binding affinities for a series of TIBO inhibitors of HIV-1 reverse transcriptase. Results from our model compounds of multiple identical benzene, dihydroxybenzene, and dimethoxybenzene molecules demonstrate the robustness of sampling in MSλD simulations by achieving relative free energy

differences within 0.2 kcal/mol of the theoretical values in both vacuum and solvent environments for combined trajectory lengths of 1.5 ns.

The relative free energies estimated for individual arms of the thermodynamic cycle for calculating the relative hydration free energies for the benzene derivatives and the relative binding affinities for the TIBO compounds were in very good agreement with those obtained from traditional alchemical free energy calculations with R^2 values above 0.987. The primary difference between relative free energies estimated using MSλD and traditional free energy methods was the computational expense in which MSλD simulations achieved the same level of accuracy and precision as the traditional calculations ~20−50 times more quickly.

Overall, these results compared well with experimental results with an AUE of 1.8 kcal/mol for the 15 pairs of hydration free energies and an AUE of 0.9 kcal/mol for the 29 pairs of binding affinities. These simulations also highlighted potential inconsistencies in the CGenFF where the pairs involving 4-hydroxybenzaldehye yielded systematically poorer relative hydration free energies than the rest of the pairs of benzene derivatives.

Systematically evaluating a series of compounds mimics a chemical optimization strategy in structure-based drug design in which various substituents are evaluated at specific sites on a promising new therapeutic lead compound. These examples provide proof of concept of both the accuracy of MSλD simulation results that can be obtained and the efficiency of this approach relative to traditional alchemical free energy calculations that rely on fixed λ values and are limited to pairs of compounds. As generalized force field parametrization strategies for drug-like molecules continue to mature and methods for constructing hybrid ligand molecules become more automated, sampling using MSλD simulations should be effective for routinely screening on the order of tens to a hundred variations of a lead compound.

AUTHOR INFORMATION

Corresponding Author

*E-mail: brookscl@umich.edu.

ACKNOWLEDGMENT

This work was supported by the National Institutes of Health through grant (GM037554).

REFERENCES

(1) Beveridge, D. L.; DiCapua, F. M. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–92.

(2) Foloppe, N.; Hubbard, R. Towards predictive ligand design with free-energy based computational methods? *Curr. Med. Chem.* **2006**, *13* (29), 3583–608.

(3) Gilson, M. K.; Zhou, H. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.

(4) 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* (2), 150–160.

(5) Gallicchio, E.; Levy, R. M. Advances in all atom sampling methods for modeling protein-ligand binding affinities. *Curr. Opin. Struct. Biol.* **2011**, *21* (2), 161–166.

(6) Zwanzig, R. W. High-temperature equation of state by a perturbation method. 1. nonpolar gases. *J. Chem. Phys.* **1954**, *22* (8), 1420–1426.

(7) Straatsma, T. P.; Berendsen, H. J. C. Free-energy of ionic hydration - analysis of a thermodynamic integration technique to evaluate free-energy differences by molecular-dynamics simulations. *J. Chem. Phys.* **1988**, *89* (9), 5876–5886.

(8) Knight, J. L.; Brooks, C. L., III. λ-Dynamics free energy simulation methods. *J. Comput. Chem.* **2009**, *30*, 1692–1700.

(9) Kong, X.; Brooks, C. L., III. λ-dynamics: a new approach to free energy calculations. *J. Chem. Phys.* **1996**, *105* (6), 2414–2423.

(10) Guo, Z.; Brooks, C. L., III; Kong, X. Efficient and flexible algorithm for free energy calculations using the l-dynamics approach. *J. Phys. Chem. B* **1998**, *102*, 2032–2036.

(11) Guo, Z.; Brooks, C. L., III. Rapid screening of binding affinities: application of the l-dynamics method to a trypsin-inhibitor system. *J. Am. Chem. Soc.* **1998**, *120*, 1920–1921.

(12) Guo, Z. Y.; Durkin, J.; Fischmann, T.; Ingram, R.; Prongay, A.; Zhang, R. M.; Madison, V. Application of the lambda-dynamics method to evaluate the relative binding free energies of inhibitors to HCV protease. *J. Med. Chem.* **2003**, *46* (25), 5360–5364.

(13) Damodaran, K. V.; Banba, S.; Brooks, C. L., III. Application of multiple topology lambda-dynamics to a host-guest system: beta-cyclodextrin with substituted benzenes. *J. Phys. Chem. B* **2001**, *105* (38), 9316–9322.

(14) Banba, S.; Brooks, C. L., III. Free energy screening of small ligands binding to an artificial protein cavity. *J. Chem. Phys.* **2000**, *113* (8), 3423–3433.

(15) Banba, S.; Guo, Z.; Brooks, C. L., III. Efficient sampling of ligand orientations and conformations in free energy calculations using the l-dynamics method. *J. Phys. Chem. B* **2000**, *104*, 6903–6910.

(16) Lee, M. S.; Salsbury, F. R.; Brooks, C. L., III. Constant-pH molecular dynamics using continuous titration coordinates. *Proteins* **2004**, *56* (4), 738–752.

(17) Khandogin, J.; Brooks, C. L., III. Constant pH molecular dynamics with proton tautomerism. *Biophys. J.* **2005**, *89* (1), 141–57.

(18) Khandogin, J.; Brooks, C. L., III. Toward the accurate first-principles prediction of ionization equilibria in proteins. *Biochemistry* **2006**, *45* (31), 9363–9373.

(19) Abrams, J. B.; Rosso, L.; Tuckerman, M. E. Efficient and precise solvation free energies via alchemical adiabatic molecular dynamics. *J. Chem. Phys.* **2006**, *125* (7), 074115–074126.

(20) Zheng, L.; Chen, M.; Yang, W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (51), 20227–20232.

(21) Zheng, L.; Chen, M.; Yang, W. Simultaneous escaping of explicit and hidden free energy barriers: Application of the orthogonal space random walk strategy in generalized ensemble based conformational sampling. *J. Chem. Phys.* **2009**, *130*, 234105–234114.

(22) Bitetti-Putzer, R.; Dinner, A.; Yang, W.; Karplus, M. Conformational sampling via a self-regulating effective energy surface. *J. Chem. Phys.* **2006**, *124* (17), 174901–174915.

(23) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of the water content in protein binding sites. *J. Phys. Chem. B* **2009**, *113* (40), 13337–13346.

(24) Li, H.; Fajer, M.; Yang, W. Simulated scaling method for localized enhanced sampling and simultaneous "alchemical" free energy simulations: A general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations. *J. Chem. Phys.* **2007**, *126* (2), 024106–024117.

(25) Jiang, W.; Hodoscek, M.; Roux, B. Computation of Absolute Hydration and Binding Free Energy with Free Energy Perturbation Distributed Replica-Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5* (10), 2583–2588.

(26) Gallicchio, E.; Lapelosa, M.; Levy, R. M. Binding Energy Distribution Analysis Method (BEDAM) for Estimation of Protein Ligand Binding Affinities. *J. Chem. Theory Comput.* **2010**, *6* (9), 2961–2977.

(27) Knight, J. L.; Brooks, C. L., III. Applying efficient implicit non-geometric constraints in free energy simulations. *J. Comput. Chem.* **2011**.

(28) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for

Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(29) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **2009**, *30* (10), 1545–614.

(30) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, *00*, 1–20.

(31) Knight, J. L.; Brooks, C. L., III. Validating CHARMM parameters and exploring charge distribution models in structure-based drug design. *J. Chem. Theory Comput.* **2009**, *5*, 1680–1691.

(32) Yesselman, J. D.; Price, D. J.; Knight, J. L.; Brooks, C. L., III. MATCH: An Atom-Typing Toolset for Molecular Mechanics Force Fields. (submitted) 2011.

(33) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. *J. Sol. Chem.* **1981**, *10* (8), 563–595.

(34) Ho, W.; Kukla, M. J.; Breslin, H. J.; Ludovici, D. W.; Grous, P. P.; Diamond, C. J.; Miranda, M.; Rodgers, J. D.; Ho, C. Y.; De Clercq, E.; Pauwels, R.; Andries, K.; Janssen, M. A. C.; Janssen, P. A. J. Synthesis and anti-HIV-1 activity of 4,5,6,7-tetrahydro-5-methylimidazo-[4,5,1-jk][1,4]benzodiazepin-2(1H)-one (TlBO) derivatives. 4. *J. Med. Chem.* **1995**, *38* (5), 794–802.

(35) Smith, R. H., Jr.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A.; Michejda, C. J.; Kroeger Smith, M. B. Prediction of binding affinities for TIBO inhibitors of HIV-1 reverse transcriptase using Monte Carlo simulations in a linear response method. *J. Med. Chem.* **1998**, *41* (26), 5272–86.

(36) van Gunsteren, W. F.; Berendsen, H. J. C. Algorithms for macromolecular dynamics and constrained dynamics. *Mol. Phys.* **1977**, *34*, 1311–1327.

(37) Zacharias, M.; Straatsma, T.; McCammon, J. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.* **1994**, *100* (12), 9025–9031.

(38) Bennett, C. H. Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. *J. Comput. Phys.* **1976**, *22* (2), 245–268.

(39) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129* (12), 124105–124114.

(40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(41) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III. New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* **2003**, *24* (11), 1348–56.

(42) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. Novel generalized Born methods. *J. Chem. Phys.* **2002**, *116* (24), 10606–10614.

(43) Das, K.; Ding, J.; Hsiou, Y.; Clark, A. D., Jr.; Moereels, H.; Koymans, L.; Andries, K.; Pauwels, R.; Janssen, P. A.; Boyer, P. L.; Clark, P.; Smith, R. H., Jr.; Kroeger Smith, M. B.; Michejda, C. J.; Hughes, S. H.; Arnold, E. Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. *J. Mol. Biol.* **1996**, *264* (5), 1085–100.

(44) Brooks, C. L., III; Brunger, A.; Karplus, M. Active site dynamics in protein molecules: a stochastic boundary molecular-dynamics approach. *Biopolymers* **1985**, *24* (5), 843–65.

(45) Rizzo, R. C.; Wang, D. P.; Tirado-Rives, J.; Jorgensen, W. L. Validation of a model for the complex of HIV-1 reverse transcriptase with sustiva through computation of resistance profiles. *J. Am. Chem. Soc.* **2000**, *122* (51), 12898–12900.

(46) Wang, D.; Rizzo, R.; Tirado-Rives, J.; Jorgensen, W. Antiviral drug design: computational analyses of the effects of the L100I mutation for HIV-RT on the binding of NNRTIs. *Bioorg. Med. Chem. Lett.* **2001**, *11* (21), 2799–802.

(47) Udier-Blagovic, M.; Tirado-Rives, J.; Jorgensen, W. Structural and energetic analyses of the effects of the K103N mutation of HIV-1 reverse transcriptase on efavirenz analogues. *J. Med. Chem.* **2004**, *47* (9), 2389–92.

(48) Smith, M.; Smith, R.; Jorgensen, W. Assault on resistance: the use of computational chemistry in the development of anti-HIV drugs. *Curr. Pharm. Des.* **2006**, *12* (15), 1843–56.

(49) Smith, M. B. K.; Rader, L. H.; Franklin, A. M.; Taylor, E. V.; Smith, K. D.; Smith, R. H.; Tirado-Rives, J.; Jorgensen, W. L. Energetic effects for observed and unobserved HIV-1 reverse transcriptase mutations of residues L100, V106, and Y181 in the presence of nevirapine and efavirenz. *Bioorg. Med. Chem. Lett.* **2008**, *18* (3), 969–972.