JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

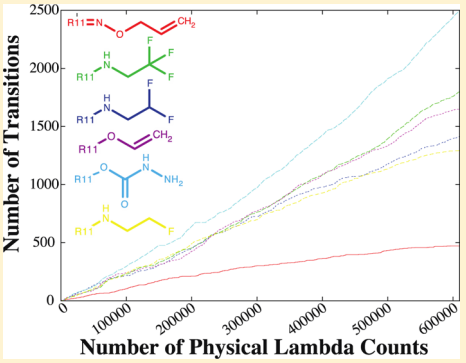
Biasing Potential Replica Exchange Multisite λ-Dynamics for Efficient Free Energy Calculations

Kira A. Armacost, Garrett B. Goh, and Charles L. Brooks, III*

Department of Chemistry, University of Michigan, 930 North University Avenue, Ann Arbor, Michigan 48109, United States

Supporting Information

ABSTRACT: Traditional free energy calculation methods are well-known for their drawbacks in scalability and speed in converging results particularly for calculations with large perturbations. In the present work, we report on the development of biasing potential replica exchange multisite λ-dynamics (BP-REX MSλD), which is a free energy method that is capable of performing simultaneous alchemical free energy transformations, including perturbations between flexible moieties. BP-REX MSλD and the original MSλD are applied to a series of symmetrical 2,5-benzoquinone derivatives covering a diverse chemical space and range of conformational flexibility. Improved λ-space sampling is observed for the BP-REX MSλD simulations, yielding a 2−5-fold increase in the number of transitions between substituents compared to traditional MSλD. We also demonstrate the efficacy of varying the value of *c*, the parameter that controls the ruggedness of the landscape mediating the sampling of λ-states, based on the flexibility of the fragment. Finally, we developed a protocol for maximizing the transition frequency between fragments. This protocol reduces the "kinetic barrier" for alchemically transforming fragments by grouping and ordering based on volume. These findings are applied to a challenging test set involving a series of geldanamycin-based inhibitors of heat shock protein 90 (Hsp90). Even though the perturbations span volume changes by as large as 60 Å^3, the values for the free energy change achieve an average unsigned error (AUE) of 1.5 kcal/mol relative to experimental K_d measurements with a reasonable correlation (R = 0.56). Our results suggest that the BP-REX MSλD algorithm is a highly efficient and scalable free energy method, which when utilized will enable routine calculations on the order of hundreds of compounds using only a few simulations.



INTRODUCTION

Free energy calculations comprise a set of simulation methods for determining the free energy differences between molecular systems. These can be translated into important thermodynamic properties, such as protein−ligand binding affinities, partition coefficients, hydration free energies, and protein stabilities.[1−5] Free energy calculations for protein−ligand binding affinities are of particular importance in the lead generation and optimization stages of the drug discovery process, as they can be faster, in principle, than the experimental exploration of chemical space.

Also known as alchemical transformations, free energy calculations utilize a thermodynamic cycle (Figure 1) to connect the free energy difference to a reference process. In the example of protein−ligand binding affinity illustrated below, the binding affinity of compound L_2 can be obtained from the alchemical free energy transformations of compound L_1 to L_2 in both the solvent (ΔG_{solv}) and protein (ΔG_{bound}) environment. Its binding affinity relative to L_1 can be calculated by the following expression:[3]

$$\Delta\Delta G^{bind}_{L_1\rightarrow L_2} = \Delta G_{bind}(L_2) - \Delta G_{bind}(L_1) = \Delta G_{bound} - \Delta G_{solv} \quad (1)$$

One major drawback of traditional free energy methods is that they require many simulations to compute the closely spaced λ values along the transformation pathway for a single compound pair. This results in significant computational cost when investigating a library of compounds. Thus, although this approach is useful, it is plagued with issues of computational feasibility, scalability, and speed.[4,6−11] Hardware advancements in the past few years, notably the development of GPU-accelerated MD code, have alleviated some of these issues. Free energy methods are also notorious for their relatively slow convergence, especially when poor overlap in phase space between λ-windows occurs. Consequently, some of the

Figure 1. Thermodynamic cycle for the calculation of the binding free energies of two ligands (L_1 and L_2) to a protein (R).

Received: October 7, 2014

DOI: 10.1021/ct500894k
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

algorithmic development that the Roux group has focused on is to integrate sampling methods, such as free energy perturbation (FEP) coupled with Hamiltonian replica exchange (FEP/H-REMD), to address this limitation. In recent work from their group, convergence was achieved for the binding affinity of *p*-xylene to the L99A mutant of the T4 lysozyme with extensive configurational sampling of the sides chains of lysozyme.[12,13] Further developments along these lines has been reported by Friesner,[14,15] Jorgensen,[16] and a team at Schrodinger[17] using FEP coupled to replica exchange with solute tempering (FEP/REST), which improves convergence by increasing the sampling of the conformational space around the ligand binding pocket. Alternatively, convergence can be enhanced by grouping ligands based on structural similarity, and such an approach has been used by the Mobley group in the Lead Optimization Mapper (LOMAP) program.[18]

While the above-mentioned strategies have advanced the capabilities of alchemical free energy methods, they do not directly address the fundamental problem concerning the scalability of free energy calculations. Multisite λ-dynamics (MSλD), an alternative free energy method developed by Knight et al.,[19,20] allows for the simultaneous evaluation of the free energy differences between multiple compounds in a single simulation. However, MSλD has only been validated on alchemical calculations that are relatively simple, where the changes in fragments modeled do not significantly perturb the local environment. In this manuscript, we report on the development of biasing potential replica exchange multisite λ-dynamics (BP-REX MSλD), which accelerates convergence and sampling by up to 5-fold and allows for the calculation of relative binding affinities of flexible moieties.

THEORY

The traditional free energy methods of thermodynamic integration and free energy perturbation compute the free energy of the transformation of one compound to another through a series of intermediate steps ($\lambda_0 \rightarrow \lambda_1$). The free energy is represented by[1,21]

$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\delta H}{\delta \lambda} \right\rangle_{\lambda} d\lambda \tag{2}$$

where the free energy difference is calculated from the ensemble average of the Hamiltonian (eq 3) of the system transforming between states A (λ=0) and B (λ=1), and λ represents the coupling parameter lying along the free energy transformation pathway. Here, $T(X,\{x\})$ represents the kinetic energy of the hybrid system, and the potential energy, $V(X, \{x\})$, is defined by eq 4.

$$H(X, \{x\}) = T(X, \{x\}) + V(X, \{x\}) \tag{3}$$

$$V(X, \{x\}) = V_{\text{env}}(X) + (1 - \lambda)V(X, x_1) + \lambda V(X, x_2) \tag{4}$$

X and x_i represent the atomic coordinates for the environment and ligand i, respectively, V_{env} represents the potential energy of the environment atoms only, and the interaction energy of the environment and ligand i is represented by (X, x_i).[3,19,22,23]

In the λ-dynamics approach, the λ parameter is converted to a dynamical variable, with fictitious mass (m_λ), and is propagated along with the Cartesian coordinates of the system.[3,19,22,23] The environment, which remains unchanged in the simulation, is composed of the atoms of the core ligand, the receptor, and the surrounding solvent. The remaining atoms that are unique to each fragment are treated as substituents attached to the core ligand. The potential energy function can be expressed as

$$V(X, \{x\}, \{\lambda\}) = V_{\text{env}}(X) + \sum_{i=1}^{N} \lambda_i (V(X, x_i) - F_i) \tag{5}$$

which has been expanded to include N ligands at a single site, with the constraints

$$0 \le \lambda_i \le 1 \tag{6}$$

$$\sum_{i=1}^{N} \lambda_i = 1 \tag{7}$$

Here, λ_i is the coupling parameter for ligand i, and F_i represents the biasing potential, which is also known as the fixed biasing potential in the context of λ-dynamics. The fixed biasing potential is an empirically predetermined variable used to increase the sampling of each $\lambda_i \approx 1$ state.[19] It allows for the calculation of ΔG values directly for an alchemical transformation of ligand i to ligand j. In the context of protein−ligand binding free energy calculations, if the fixed biasing potential is set to be ΔG_{solv} (left arm of the thermodynamic cycle as depicted in Figure 1), then the corresponding value calculated is $\Delta\Delta G_{\text{bound}}$ for an alchemical transformation from ligand i to ligand j bound in a protein. The free energy difference between two ligands with discrete biasing potentials (F_i and F_j) is thus calculated by

$$\Delta\Delta G_{j \rightarrow i} = -k_{\text{B}}T \ln \frac{P(\lambda_i = 1, \{\lambda_{m \neq i} = 0\})}{P(\lambda_j = 1, \{\lambda_{m \neq j} = 0\})} \tag{8}$$

where the free energy is governed by the probability that a state is dominated by ligand i or j, $P(\lambda_{i \text{ or } j} = 1, \{\lambda_{m \neq i \text{ or } j} = 0\})$.[3,19,22,23] Thus, the λ-dynamics approach allows for the sampling of multiple substituents in a single calculation, addressing the challenges associated with scalability and computational demand of traditional free energy methods.

Although λ-dynamics addresses the issue of scalability and computational demand, varying the functional groups at multiple sites on a common framework is of usual practice in many lead generation campaigns. Therefore, multisite λ-dynamics (MSλD) was developed to allow for multiple substituents at multiple sites to be considered simultaneously.[19,20] The potential energy is calculated via

$$V(X, \{x\}, \{\lambda\}) = V_{\text{env}}(X) + \sum_{S=1}^{M_{\text{sites}}} \sum_{i=1}^{N_S} \lambda_{S,i} (V(X, x_{S,i}) - F_{S,i}) + \sum_{S=1}^{M_{\text{sites}}-1} \sum_{i=1}^{N_S} \sum_{T=S+1}^{M_{\text{sites}}} \sum_{j=1}^{N_T} \lambda_{S,i}\lambda_{T,j} (V(x_{S,i}, x_{T,j})) \tag{9}$$

with similar holonomic constraints as before

$$0 \le \lambda_{\alpha,i} \le 1 \tag{10}$$

$$\sum_{i=1}^{N_\alpha} \lambda_{\alpha,i} = 1 \tag{11}$$

M_{sites} represents the total number of sites with multiple substituents, and N_S is the number of substituents on site S. The double summation signifies the interaction between the environment and all substituents at each site, and the third term

DOI: 10.1021/ct500894k
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

designates the interaction of the substituents at one site with the substituents at another site.[19] The substituents at a given site are independent of one another and do not interact with each other in the simulation. The free energy difference of the substituents at two sites can then be calculated by the ratio of populations for ligand A with fragment k at site 1 and fragment l at site 2 with the populations of ligand B with fragment i at site 1 and fragment j at site 2:

$$\Delta\Delta G_{1,i;2,j\rightarrow 1,k;2,l} = -k_B T \ln \frac{P(\lambda_{1,k} = 1, \lambda_{2,l} = 1)}{P(\lambda_{1,i} = 1, \lambda_{2,j} = 1)} \quad (12)$$

When λ approaches 1, a ligand is said to be dominant, and the threshold of $\lambda_{1,i} > 0.8$ was used to approximate physical states ($\lambda = 1$). λ values for N_α are governed by

$$\lambda_{\alpha,i} = \frac{e^{c \sin \theta_{\alpha,i}}}{\sum_{j=1}^{N_\alpha} e^{c \sin \theta_{\alpha,j}}} \quad (13)$$

which satisfies the necessary constraints (eqs 10 and 11). The coefficient (c) is given a value of 2.5 or 5.5 to further optimize the sampling of the end point state. Here, the values of θ have fictitious masses, m_θ, and are propagated along with the Cartesian coordinates of the system.[19,20]

Despite being a robust and accurate free energy method, sampling limitations under certain conditions can still arise in MSλD simulations. Under optimal conditions, substituents cannot vary by more than 2−3 kcal/mol from one another and an increased flexibility in the substituents decreases the effectiveness of sampling in λ-space.[19] In addition, λ values of substituents may become "trapped", which is a phenomenon in which a MSλD simulation predominantly samples a single substituent. This results in poor λ-sampling and is particularly exacerbated in situations where the protein environment perturbs the alchemical transformation (ΔG_{bound}) such that it is no longer within 2−3 kcal/mol from the reference (ΔG_{solv}). In the present work, these sampling issues are addressed with an implementation of the replica exchange algorithm in the biasing potential space.

Biasing Potential Replica Exchange MSλD. For the BP-REX MSλD method, simulations of multiple replicas with different biasing potential values (F_i) are performed and coupled via the following Metropolis criterion

$$P = \begin{cases} 1; & \text{if } \Delta \leq 0 \\ \exp(-\Delta); & \text{otherwise} \end{cases} \quad (14)$$

where

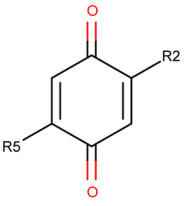
$$\Delta = \beta \begin{bmatrix} U^{F_{\alpha,i}}(\{\lambda_i\}; F'_{\alpha,i}) + U^{F_{\alpha,i}}(\{\lambda_i'\}; F_{\alpha,i}) \\ -U^{F_{\alpha,i}}(\{\lambda_i\}; F_{\alpha,i}) - U^{F_{\alpha,i}}(\{\lambda_i'\}; F'_{\alpha,i}) \end{bmatrix}$$

β represents $1/k_b T$, the first two terms are the potential energy terms for the two replicas after the exchange, and the last two terms are the potential energy values for the same two replicas before the exchange.[24,25] The central replica was assigned a fixed biasing potential as the calculated $\Delta\Delta G_{solv}$ value for each substituent. For the replicas immediately adjacent to the central replica, the fixed bias was incremented by ±1.4 kcal/mol for solvent (±1.6 kcal/mol for the bound systems) relative to the value assigned for the central bias. Replicas that are not immediately adjacent to the central replica were incremented by the same successive spacing (1.4 or 1.6 kcal/mol) from their neighboring replica. From a broader perspective, the BP-REX MSλD algorithm presented here is a generalized version of the pH-REX explicit solvent constant pH molecular dynamics method (CPHMD$^{MS\lambda D}$) developed by Brooks and co-workers,[24,25] which itself is a variation of Hamiltonian replica exchange.[26,27]

METHODS

Hybrid Ligand Setup. A series of symmetrical benzoquinone derivatives and the corresponding hybrid ligands were assigned parameters and partial charges consistent with the CHARMM General Force Field (CGenFF)[28−30] using our in house MATCH[31] automated parameterization tool. The benzoquinone derivatives selected in this study encompass a wide variety of substituents, including small, less flexible moieties and larger, more flexible moieties (Table 1). These substituents were chosen to explore the limits of traditional MSλD and the newly developed BP-REX MSλD algorithm in efficiently sampling multiple ligands simultaneously.

Table 1. Benzoquinone Fragment Designations for Site 2 (R2) and Site 5 (R5)



Substituent Designation	Substituent
a	R — CH_3
b	R — Cl
c	R — $CH(CH_3)_2$
d	R — C≡N
e	R — C_6H_5
f	R — O — CH_3
g	R — O — CH_2
h	R — NH_2
j	R — NH — $CH_2CH_2N(CH_3)_2$
k	R — NH — $CH(CH_3)CH_2N(CH_3)_2$

Simulation Details. *Benzoquinone Derivatives.* Molecular dynamics simulations were performed through the BLOCK/MSλD module in the CHARMM macromolecular modeling program, developmental version 39a1.[32,33] The ligand derivatives were solvated in a periodic 30 Å^3 cubic box of 690 TIP3P[34] water molecules. Langevin dynamics was used to propagate the equations of motion in all MSλD simulations using the Leapfrog Verlet algorithm,[35] and the SHAKE algorithm[36] was used to constrain the heavy atom hydrogen bonds. A nonbonded cutoff range of 10−12 Å was used with an electrostatic force shifting function and a van der Waals switching function, and the nonbonded lists were generated using a 15 Å cutoff. A fictitious mass of 12 amu·Å^2 was assigned to θ (see eq 13), and λ values were saved every 10 steps. NOE

DOI: 10.1021/ct500894k
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

restraints were imposed on the first common heavy atom of the substituents to maintain a flat-bottomed potential to increase the transition rates between substituents. Variable biases (F^{variable}) were added to the potential energy equation, eq 8, with force constants (k) varying between 0 to 12 kcal/mol (the a-k simulation (Table 1) required k = 28 kcal/mol) for each $\lambda_{\alpha,i}$. The variable bias does not alter the relative free energy values of the end states and has the following functional form:

$$F_{\alpha,i}^{\text{variable}} = \begin{cases} k(\lambda_{\alpha,i} - 0.8)^2 \text{ kcal/mol}, & \text{if } \lambda_{\alpha,i} < 0.8 \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

Triplicate simulations for traditional MSλD and BP-REX MSλD for each pair of substituents were run with different seed values of θ_i, which were initially heated to 310 K in a Langevin heat bath for 10 ps, equilibrated for 20 ps, before a production run of 3 ns was performed.

The fixed biasing potential was calculated with a weighted histogram analysis method (WHAM)[37] for λ-dynamics, which iterates the biasing potential from an initial trial value until the free energy barrier is reduced and converged sampling is achieved.[22] Initial fixed biases were estimated from the GBMV[38,39] implicit solvent and vacuum energies for each ligand in traditional MSλD.[19] For our simulations, we scanned a range of fixed biases between −50 to 50 kcal/mol in 10 kcal/mol increments using short 100 ps simulations to identify the free energy value yielding greater than 50 transitions/ns in λ-space. This calculated value was used as our fixed biasing potential (ΔG_{solv}) for the solvent and bound simulations. In all simulations, two sites were substituted concurrently (R2 and R5) with two substituents at each site. The root substituent introduced was methyl, unless otherwise noted, which transitioned between the others listed in Table 1 and itself. We performed each simulation for two different values of the parameter c, 5.5 and 2.5 (eq 13). In BP-REX MSλD, exchange attempts were made at every 1 ps, unless otherwise stated.

Geldanamycin Derivatives. Relative binding free energy values were computed from the thermodynamic cycle involving mutation of one ligand to another in a solvated system and bound to the N-terminal ATP-binding pocket of Hsp90. The crystal structure of the N-terminal domain of Hsp90 (213 residues) bound to geldanamycin (GDM) was used (PDB ID 1YET)[40] and solvated with 12,707 TIP3P[41] water molecules with periodic boundary conditions. The CHARMM22 all-atom force field for proteins with CMAP correction[42] was utilized for the protein, and no counterions were added to the simulation as the protein has a net charge of zero.

Three sets of simulations were performed, which included transformations of 4−6 substituents on the 11-site of GDM. The free energy estimation, F_i, for the solvated systems was calculated similarly to the benzoquinone derivatives. The calculated ΔG_{solv} values were used as the central replica bias for the solvent and bound systems (Table S1). The simulations were set up with the same parameters as listed for the benzoquinone derivatives. A short heating and equilibration stage was performed, followed by 3 and 5 ns production runs for the solvated and bound system, respectively. Three identical replica exchange simulations (3 replicas for the solvent systems and 9 replicas for the bound systems), with different seed numbers were performed as described above. The free energy values were averaged over all runs, where the central replica data was used to compute $\Delta\Delta G_{\text{bind}}$ values for the systems. Our simulations were all performed with the methoxy substituent at the C-17 position of GDM (Figure 6).

Analysis. Molecular dynamics simulations were performed for the solvated ligands (benzoquinone with substituents at the R2 and R5 site) for 1 ns of heating, 1 ns of equilibration, and 10 ns of dynamics to compute radial distribution functions (RDF). The RDF's were analyzed for the interaction of the H and O of water with the first common heavy atom of the substituents (C, O, N). The results were compared by a Wilcoxon rank sum test (Mann−Whitney U test) that examines the similarity of data sets.[43,44] A p-value is output that describes the similarity of the data sets, where a threshold greater than 0.1 signifies there is not enough evidence to differentiate the two data sets. The Wilcoxon test was performed within the R package.[45]

The fingerprint method (aka Tanimoto coefficient) is a metric used in computer-aided drug design to screen a number of compounds for molecular similarity.[46,47] The Tanimoto coefficient test was performed with the fingerprint package in Open Babel,[48] using the FP2 fingerprint. This fingerprint is a path-based (Daylight-type) fingerprint that indexes linear fragments of up to 7 atoms.[48] Other fingerprints are also available, including the FP3, FP4, and MACCS, but the FP2 identified superstructures for the class of ligand derivatives studied proficiently. Compounds are classified as superstructures if they have a chemical similarity to one another based on the fingerprint used.

All trajectories were analyzed within CHARMM. The transition rates ($\tau_{\text{trans}}(\alpha)$) for the substituents at site α indicate the frequency at which the identity of the dominant substituent changes in the simulation. The populations of these states were used to estimate the free energy difference calculated by eq 12.

RESULTS AND DISCUSSION

Traditional MSλD and BP-REX MSλD were used to calculate solvation free energies for the transformations of a series of symmetrical benzoquinone derivatives (Table 1). The derivatives encompassed a wide variety of substituents, including alkyl, aryl, hetero groups, and heteroalkyl side chains. The substituents were intentionally selected to represent a range of fragment sizes and flexibilities to explore the robustness and determine the limits of MSλD-based free energy calculations.

Relative Free Energy Calculations. Interconversion between symmetrical benzoquinone derivatives in water are expected to yield overall free energy differences of 0 ($\Delta\Delta G$ = 0). Therefore, the expected $\Delta\Delta G$ = 0 can be used as a specific means to evaluate the accuracy of MSλD-based free energy sampling, as it is independent of the quality of parameters used. Similar to Figure 1, a free energy cycle for the interconversion of the methyl fragment to the chloromethyl fragment should give free energy values of 0 for the vertical arms, whereas the horizontal arms will vary depending on the substituent change (Figure S1). Initial inspection of the free energy values for the MSλD simulations (Table S2) gives an overall free energy value of 0 for all substituents, but about half of the data points have nonzero values, i.e. varying by ±0.1 kcal/mol. For BP-REX MSλD, most simulations gave a $\Delta\Delta G$ = 0, overcoming the minor convergence issues seen with the traditional MSλD simulations. Only one simulation, BP-REX MSλD (a-g), gave an overall nonzero free energy value of 0.1 kcal/mol. When the simulations were performed with a different seed number, the calculated $\Delta\Delta G$ values were within 2 decimal points of the values previously calculated, demonstrating their high precision.

DOI: 10.1021/ct500894k
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Flexibility Changes Between Substituents Affects the Number of Transitions. In earlier studies employing the traditional MSλD a lower number of transitions in λ-space was observed when transforming a molecular substituent from a hydrogen to a flexible methoxy moiety.[19] Flexible substituents are capable of exploring a larger range of conformations, and this is manifested as a reduced efficiency in λ-space sampling, i.e. sampling between the two substituents of interest is diminished. Here, we compare the efficiency of λ-sampling between traditional and BP-REX MSλD simulations by examining the number of transitions observed between two competing substituents at a given site of derivitization.

From Figure 2 it is clear that the largest number of transitions occurs for a-a, a-b, a-d, and a-h. For traditional MSλD, the a-a simulation (CH_3 to CH_3) was our control as it provides an upper limit on the number of transitions that can occur. Closely following this substituent change is the a-h simulation involving a CH_3 to NH_2 substituent change. This simulation yielded ~700 transitions/ns. Two other simulations, CH_3 to CH_2Cl (a-b) and CN (a-d) yielded ~500 and ~400 transitions/ns, respectively. The sampling efficiency reported here is comparable to the original MSλD results,[19,20] where substituents with similar conformational flexibility resulted in a large number of transitions. The corresponding BP-REX MSλD simulations improved upon the MSλD results, with an increase by approximately 20−100 transitions/ns. Most notably, for the simulations with a flexible second substituent (a-c, a-e, a-g, a-j) the sampling efficiency increase is more pronounced, ranging from a 2−5-fold improvement.

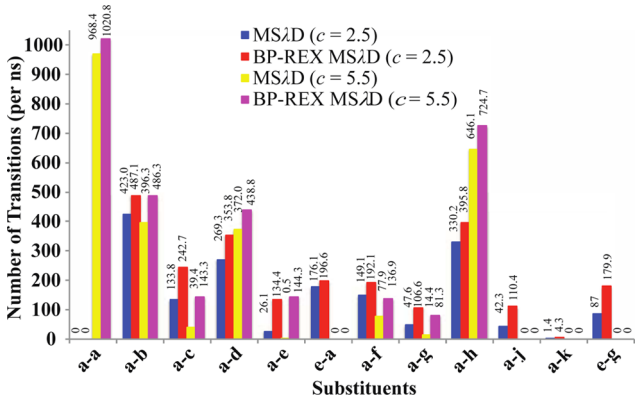


Figure 2. Number of transitions associated with the transformations of one substituent to another. The number of transitions between substituents is given per ns.

Varying the *c* Values. Prior work by Knight et al.[20] suggested that a coefficient (c) value of 5.5 is optimal for the functional form used in eq 13. This recommendation was generated primarily from simulations of very small perturbations between substituents, e.g. H to OH and OCH_3. A smaller c value is not suggested for these types of simulations.[20] However, when the changes in fragments are as large as the ones presented herein, our results indicate that a lower c value increases the number of transitions considerably albeit at a cost of increased sampling of intermediate λ states (see Discussion). Figure 3 provides a graphical representation of eq 13 with the two coefficient values of 5.5 and 2.5. For the high c value of 5.5, the energy landscape of θ-space is shaped in a manner that favors the sampling of physical end states. A steep slope between the extrema of 0 and 1 signifies that a minimal amount of time is spent sampling the intermediate states, which works well for small perturbations due to their strong overlap in phase space. For larger perturbations (i.e., involving flexible moieties), sampling the intermediate states is desirable as it allows for better overlap of phase space. Therefore, a smaller c value of 2.5 is utilized for these simulations, giving a more gradual slope between the extrema and allowing for a more staggered transition.

In our simulations of smaller perturbations, using a larger coefficient value of c = 5.5 for systems such as CH_3 to NH_2 and CN produces more transitions than a c = 2.5, on the order of 100−300 transitions/ns. Other simulations with larger perturbations, however, had more transitions when c = 2.5, including the transition from CH_3 to $CH(CH_3)_2$, OCH_2CHCH_2, and $NHCH_2CH_2N(CH_3)_2$. The lower coefficient value (c = 2.5) gave approximately a 2−5-fold increase in the number of transitions.

As points of reference, we performed simulations of a very simple transition, CH_3 to CH_3, and of a difficult transition, CH_3 to $NHCH(CH_3)CH(CH_3)N(CH_2CH_3)_2$. The CH_3 to CH_3 transition gave approximately 1000 transitions for the higher coefficient value (5.5), while the lower value (2.5) predominantly sampled an intermediate λ-value and consequently did not allow for a physical ligand to be present. The simulation involving CH_3 to $NHCH(CH_3)CH(CH_3)N(CH_2CH_3)_2$ was

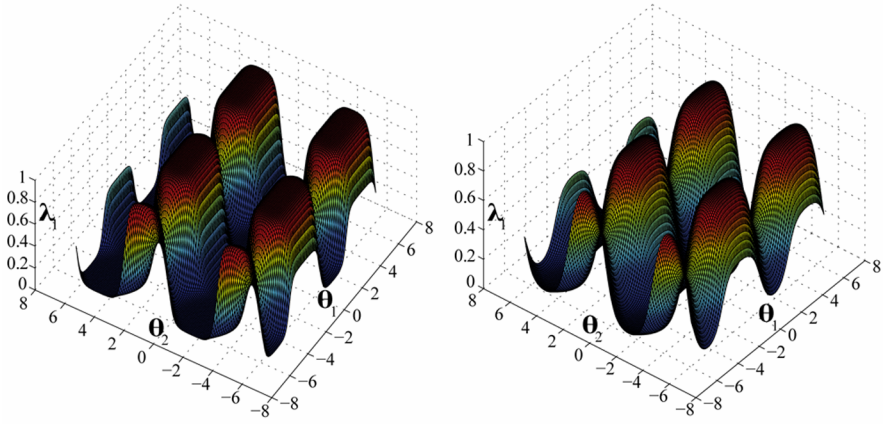


Figure 3. Representation of eq 13 that is used to satisfy the holonomic constraints. The left figure is the representation with c = 5.5, and the right figure represents c = 2.5.

DOI: 10.1021/ct500894k
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

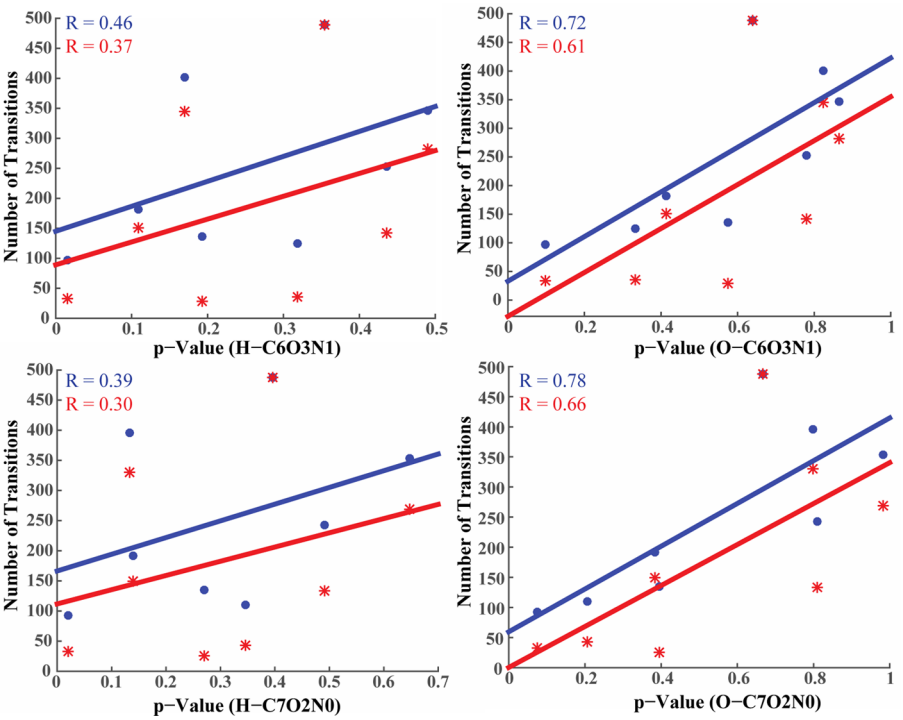


Figure 4. Correlation plots of the *p*-value of the radial distribution functions for H at the R2 position (top left), O at the R2 position (top right), H at the R5 position (bottom left), and O at the R5 position (bottom right). *R* values are given in each plot (red represents MSλD and blue represents BP-REX MSλD).

performed to determine the upper limits of the MSλD-based methods. A lower coefficient value gave a very limited number of transitions and did not achieve the target of 50 transitions/ns that is required to obtain reasonable convergence, while the higher c (5.5) value resulted in numerical instability. Although a lower coefficient value was effective for many of the flexible substituents, an extremely large perturbation cannot be surmounted even with the BP-REX MSλD algorithm.

Characterization of Substituents and Optimal Staging of Free Energy Calculations Based on a Variety of Metrics. To determine the practical upper limits and best practices for setting up BP-REX MSλD simulations, we examined further ways of grouping the substituents in the simulations. Sampling efficiency was determined by calculating λ-transition rates.

Water Radial Distribution Functions. Radial distribution functions (RDFs) were calculated to examine the interactions of the water molecules with the substituents. A Wilcoxon test was performed on the RDFs to compare the similarity of the RDFs across different substituents. The underlying hypothesis is that substituents with similar RDF profiles will occupy a similar region of conformation phase space and thus transition effectively between one another. Of the 8 simulations for the benzoquinone derivatives, only one *p*-value met the threshold (<0.1) to claim a substantial difference in the RDFs (substituent g compared to a in Table 1). Therefore, we constructed correlation plots of the *p*-value and the number of transitions between the end point states (Figure 4). Among the results, the O_{water}−C/O/$N_{substituent}$ RDFs correlated well with an *R* value of 0.61−0.66 for our BP-REX MSλD simulations, reinforcing our hypothesis that a strong overlap in conformation phase space is correlated with efficient λ-sampling. Although this metric performed well to predict sampling efficiency, extending this RDF calculation approach to more than a small library of compounds would substantially increase the total computational time. Therefore, we examined two "rapid" metrics for grouping substituents, the Tanimoto coefficient and volume occupancy of the ligands.

Tanimoto Coefficient (Chemical Similarities Using 2D Fingerprints). The Tanimoto coefficient was calculated to compare the similarity of each substituent to the CH_3 substituent. A series of fragments (CH_3, CH_2Cl, $CH(CH_3)_2$, CN, and phenyl) were designated as superstructures of one another (Table S3), and no superstructures in comparison to the methyl fragment were found for the O or N derivatives (Table 1, f-k). The FP2 fingerprint is used to distinguish between linear fragments of 7 atoms, giving superstructures for the fragments with carbon as the first heavy atom. Many of the superstructure groupings yielded a large number of transitions (Figure 2, a-b, a-c, a-d); however, the simulation between CH_3 and phenyl (a-e) had a lower number of transitions for the BP-REX MSλD method. Furthermore, comparing the superstructure groupings given with the number of transitions, the CH_3 to NH_2 (a-h) substituents gave the largest number of transitions but were not found to be superstructures. Therefore, while the Tanimoto coefficient may be a suitable metric to evaluate chemical similarities, its applicability to MSλD simulations, although leading to faster convergence with some groupings, did not show a significant discrimination in the groupings.

Examination of the Volume Occupancy. The simulation with the largest number of transitions, CH_3 to NH_2, motivated our choice to examine the volume of occupancy of each substituent (Table 2). Figure 5 demonstrates the correlation of volume occupied to the number of transitions in each simulation, where the change in volume was calculated as a difference in volumes between the two substituents. It is apparent that there is an inverse correlation between the

DOI: 10.1021/ct500894k
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 2. Volume Occupied by the Benzoquinone Derivatives in Solvent (Given in Å^3)

Substituent	Volume Occupied (Å^3)
R — CH_3	163.4
R — Cl	194.3
R — $CH(CH_3)_2$	237.2
R — C≡N	167.4
R — C_6H_5	315.5
R — O — CH_3	190.9
R — O — $CH_2CH{=}CH_2$	272.4
R — NH_2	145.8
R — NH — $CH_2CH_2N(CH_3)_2$	362.2
R — NH — $CH(CH_3)CH(CH_3)N(CH_3)_2$	523.6

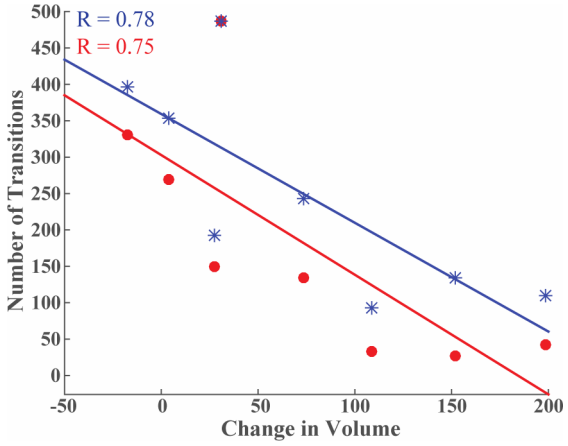


Figure 5. Correlation plot relating the change in volume of the substituents relative to the number of transitions per ns. The values of *R* are given in the plot (red represents MSλD and blue represents BP-REX MSλD).

volume and number of transitions, with an *R* value of 0.78 in our BP-REX MSλD calculations. These results are comparable with the RDF metric we tested when the water oxygen was used to represent the solvent species but with the added advantage that the volume calculations require a minimal computational expense.

The largest number of transitions occurred for the CH_3 to NH_2 simulation, followed by simulations with similarly occupied volumes (e.g., CH_3 to CN). To further demonstrate the effectiveness of optimal substituent grouping, we performed an additional simulation where the similarly occupied large volume substituents of phenyl (Ph) to OCH_2CHCH_2 (e-g) were grouped together. This simulation (e-g) gave an increased number of transitions (130−160 transitions/ns), compared to the initial CH_3 to Ph (a-e) and to OCH_2CHCH_2 (a-g) simulations. Therefore, our results suggest that grouping substituents by volume is an effective means to achieve efficient sampling in BP-REX MSλD.

Lastly, we looked at the effect of transitioning from a larger occupied volume substituent (Ph), the reference compound, to a smaller substituent (CH_3). In the Ph to CH_3 (e-a) simulation, the transitions increased by over 2-fold for the BP-REX MSλD simulation (312 ± 28.6 transitions/ns) compared to the CH_3 to Ph BP-REX MSλD simulation (135 ± 12.6 transitions/ns). The preparation of the structures involves constructing a water box around the initial ligand, where the only difference is the orientation and distribution of the water molecules around the larger Ph moiety, compared to the smaller CH_3 moiety. Therefore, it is not surprising that our results suggest that collapsing a substituent into a smaller volume (Ph to CH_3) and drawing in the nearby water molecules is kinetically more favorable than its counterpart (CH_3 to Ph), where the surrounding waters would need to be evacuated for such a growth to occur. While the $\Delta\Delta G$ values for both Ph to CH_3 and CH_3 to Ph simulations were effectively zero corresponding to the expected free energy difference, we observed that the magnitude of ΔG for each individual arm of the thermodynamic cycle for Ph to CH_3 and CH_3 to Ph simulations differed by 1.2 kcal/mol, suggesting that one of the simulations has yet to achieve global convergence (Table S2).

Geldanamycin Derivatives Bound to Hsp90. Thus, far, we have demonstrated the effectiveness of BP-REX MSλD on model benzoquinone compounds. In this section, we examine its performance on inhibitors of heat shock protein 90. Hsp90 is an ATP-dependent chaperone protein that is expressed in times of cellular distress. It utilizes a chaperone cycle for client proteins to bind (e.g., p53, Aha1) that promotes the growth and survival of malignant cancer cells.[49−51] Protein degradation occurs with inhibition of the ATP-binding pocket, causing the chaperone cycle to disassemble. One major class of inhibitors are geldanamycin (GDM) and its derivatives. The initial clinical trials of GDM bound to Hsp90 in cancer lines had toxicity and solubility concerns; therefore, derivatives of GDM have been developed, including mutations at the C-17, C-19, C-11, and C-7 positions (Figure 6 highlights the C-17 and C-11 positions).[52−56] Liu et al. examined a series of derivatives at the C-11 position of GDM, including ether, acyl, oxime, and amino substituents.[52] The crystal structures for the Hsp90-GDM and Hsp90-DMAG (dimethylaminoethylamino) complexes revealed a small binding pocket around the C-11 position, so the larger substituents at the C-11 position will not bind.[40,52,57]

Due to the large chemical space explored in GDM derivatives, it serves as a good test set for a series of challenging free energy perturbations. Using the best practices developed for performing BP-REX MSλD simulations, such as using a lower *c* for larger perturbations, grouping substituents based on similar volume occupancy, and ordering them by decreasing volume, we calculated the relative free energies of binding for 11-derivatives of GDM in Hsp90. Three sets of simulations were performed, where the small fragments (<685 Å^3), medium fragments (685−720 Å^3), and large fragments (720−782 Å^3) were grouped and simulated separately. This categorization allowed for 4−6 substituents per simulation. The volume

DOI: 10.1021/ct500894k
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

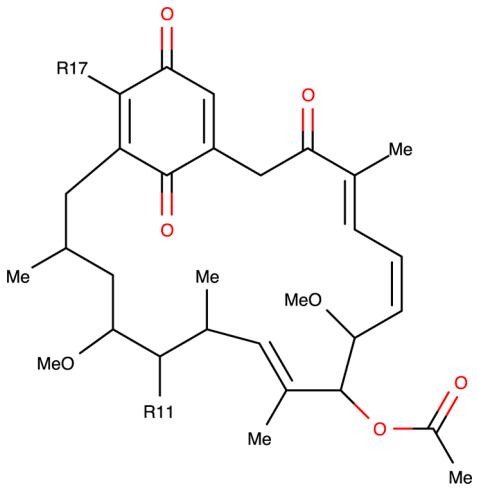


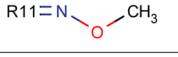
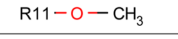
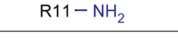
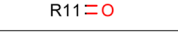
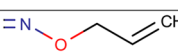
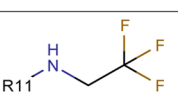
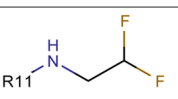
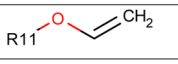
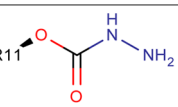
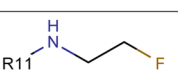
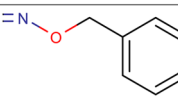
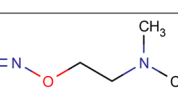
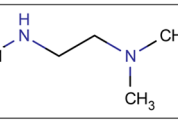
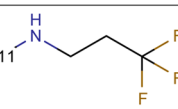
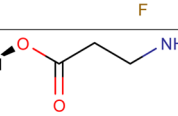
Figure 6. Structure for geldanamycin with the C-11 and C-17 sites of derivitization denoted as R11 and R17, respectively.

calculations, grouping pattern, and experimental free energy values (calculated from K_d)[52] are summarized in Table 3.

First, we analyzed the simulations by examining λ-space sampling efficiency, which is represented by the number of λ transitions between fragments. Analogous to the original MSλD approach,[19] we aimed for at least 50 transitions/ns for each substituent to yield adequate statistical precision in the final free energy calculations. An average of 424.4/ns, 608.3/ns, and 401.5/ns transitions were observed for the small, medium, and large fragments when bound to Hsp90. However, using the average transition rate may obscure the individual details of each substituent. It is possible that λ-sampling may not be uniform across all substituents simulated, for example, a system which transitions between only 2 substituents (without sampling the other substituents present) can also report a high overall λ-sampling rate. Therefore, we performed further calculations on the number of transitions/ns for every substituent in the BP-REX MSλD simulation. As presented in Figure 7, although the sampling efficiency may not be completely uniform across all substituents, each substituent achieved at least ~100 transitions/ns in the bound configuration to Hsp90, which is 2 times higher than the threshold used in the original MSλD approach.[19]

Next, we analyzed the simulations for accuracy. The free energy of binding was calculated from eq 1 using the thermodynamic cycle depicted in Figure 1. ΔG_{solv} and ΔG_{bound} were directly computed from the simulations, and the calculated ΔG_{bind} values were derived using an arbitrarily chosen reference ΔG_{bind} value. For the small and medium fragment simulations, the reference ΔG_{bind} corresponds to substituent *b* (in each simulation), which has accurate K_d values reported, and substituent *a* was used as the reference in the large fragment simulations. A positive correlation was observed between the experimental and calculated ΔG_{bind} values, with a Pearson correlation coefficient of 0.56 (Figure 8b). Our results are comparable to previous work by Knight et al., where calculations of the binding affinities of TIBO inhibitors against HIV-1 reverse transcriptase using the traditional MSλD approach yielded a Pearson correlation coefficient of 0.63.[19] From our calculations we find that the root-mean-square error (RMSE) is 3.1 kcal/mol, and the average unsigned error (AUE) is 2.4 kcal/mol; the calculated ΔG_{bind} for each substituent and

Table 3. Geldanamycin Derivatives Substituted at the 11 Position Grouped into Three Categories with Associated Volumes (Å^3) and the Experimental ΔG_{bind}[a]

Substituent	Designation	Volume (Å^3)	ΔG_{bind} (kcal/mol)
SMALL			
	a	682.6	-4.26*
	b	663.7	-7.53
	c	649.5	-5.11*
	d	643.1	-6.73
MEDIUM			
	a	719.8	-4.26*
	b	706.5	-6.24
	c	703.3	-6.44
	d	702.9	-6.44
	e	699.2	-6.10
	f	698.2	-4.68*
LARGE			
	a	781.0	-4.26*
	b	763.4	-4.26*
	c	748.9	-4.68*
	d	727.8	-5.67*
	e	722.9	-5.11*

[a]The * represents "nonbinding" free energy values estimated using the lower bound K_d values reported.

the corresponding experimental values are summarized in Figure S2.

An analysis of each substituent's calculated ΔG_{bind} (Figure S2) identified an outlier, the medium-a fragment, having a positive ΔG_{bind} value in our simulations. This suggests that this geldanamycin derivative is a nonbinder to Hsp90. Experimentally, the K_d value was reported to be >100 μM, which is an *estimate* based on the concentration range tested.[52] Due to the bulky nature of the allyl group, and the sp^3 to sp^2 hybridization change at the C-11 position of GDM for the medium-a fragment, significant steric clashes with the binding pocket are expected and consequently a compound based off this fragment would not be anticipated to bind well. Such an

DOI: 10.1021/ct500894k
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

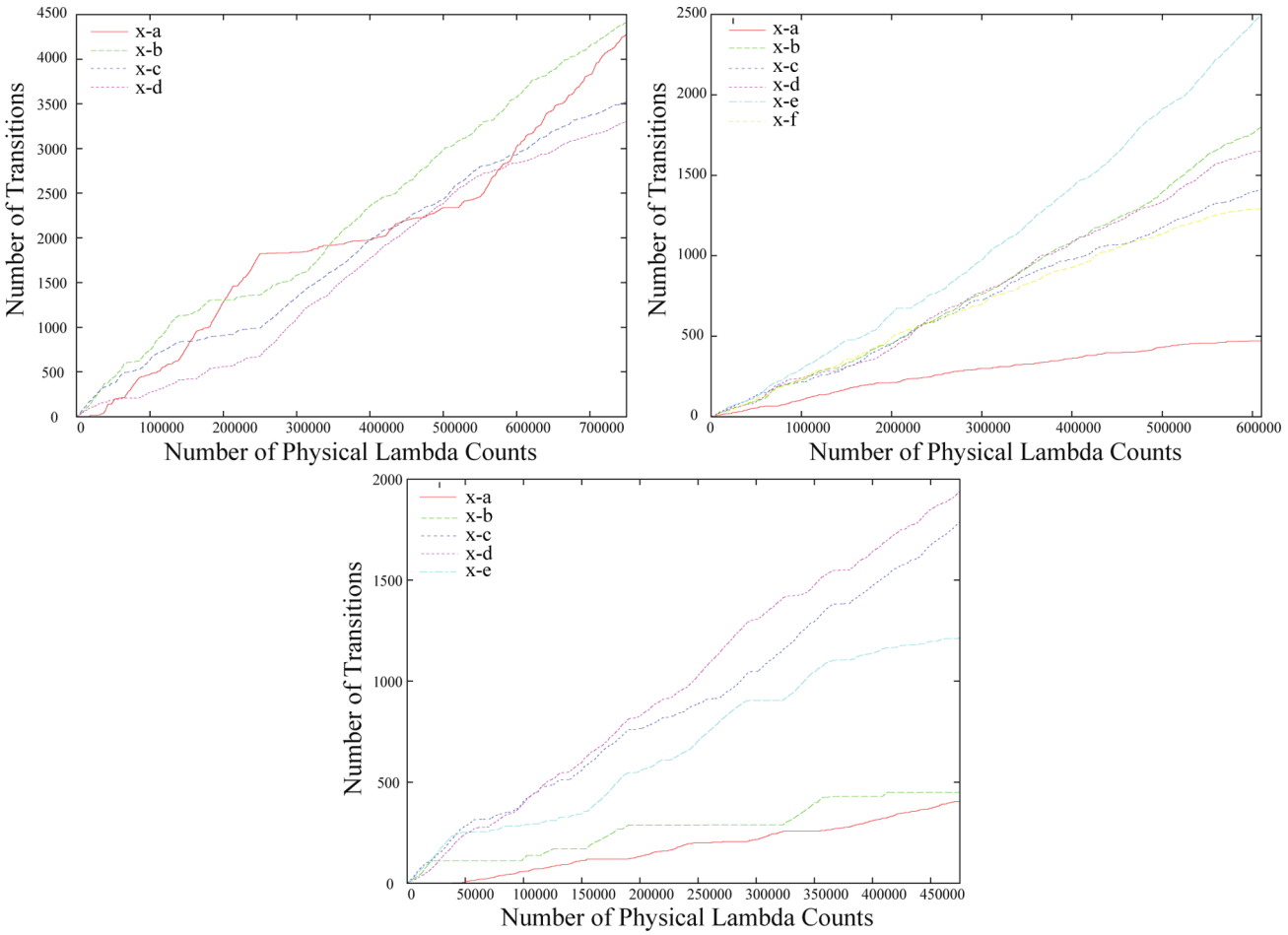


Figure 7. Number of transitions occurring between substituent x (any of the substituents in the given simulation) to substituent a through d, a through f, or a through e for small, medium, and large, respectively. The left plot represents the small fragments, right represents the medium fragments, and bottom represents the large fragments.

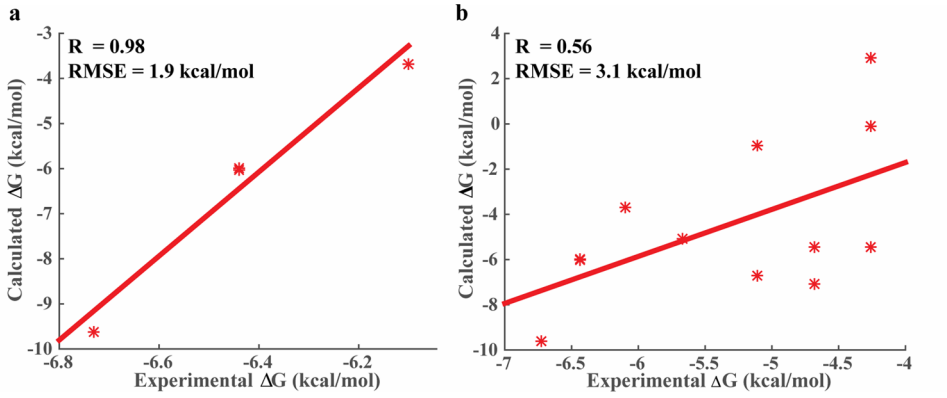


Figure 8. Experimental ΔG values relative to calculated ΔG values: (a) plot of the reliable correlated ΔG values and (b) plot of all correlated ΔG values. Pearson correlation coefficient values are given.

expectation is consistent with numerous experimental observations that larger substituents at the C-11 position do not bind.[40,52,57] Therefore, we suggest the lack of accuracy in the experimental data may explain the large difference between computed and measured ΔG_{bind} values. When the medium-a fragment is considered an outlier, the RMSE falls to 2.3 kcal/mol, and the AUE to 2.0 kcal/mol, which is a significant improvement. Additionally, we note that the order of binding affinities is, in fact, correctly recapitulated.

Lastly, from the experimental K_d values reported, it is apparent that many of the fragments only exhibit weak binding to Hsp90 (indicated by an * in Table 3). It should also be emphasized that for many of the weak binding compounds the experimental K_d values were reported as lower limit measurements (denoted as striped bars in Figure S2). There is a possibility that these ligands could be weaker binders than the reported threshold. Therefore, we performed a final analysis where the K_d estimates were filtered out and only the "reliable"

DOI: 10.1021/ct500894k
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

K_d measurements (solid black lines in Figure S2) were used. The computed Pearson correlation coefficient for this data was found to be 0.98 (Figure 8a). Using this "reliable" data set also yielded a slight improvement in accuracy: RMSE of 1.9 kcal/mol and an AUE of 1.5 kcal/mol, which is comparable to the AUE of 0.77−1.14 kcal/mol reported by Knight et al.[19] The RMSE of our results is close to the ∼2 kcal/mol RMSE threshold that Chodera and co-workers have reported to be adequate for using free energy calculations in the lead optimization stages of drug discovery.[6] This suggests that the free energy values obtained from BP-REX MSλD may be useful for enriching libraries that contain compounds with larger perturbations, which would be traditionally unassailable by conventional free energy methods.

CONCLUSION

In conclusion, we have improved the traditional multi-site λ-dynamics free energy method using a replica exchange algorithm. Biasing potential replica exchange MSλD (BP-REX MSλD) has demonstrated enhancements over traditional MSλD, notably in its ability to improve λ-sampling from 2−5-fold for flexible substituents. We also demonstrated the effectiveness of varying c values based on the flexibility of the fragment and developed a protocol for maximizing the transition between fragments, by grouping and ordering fragments based on volume occupancy.

These findings were further illustrated in our calculation of protein−ligand binding affinities for 11-derivatives of GDM-based inhibitors of Hsp90. Analysis of the individual λ-transition rates suggests that uniform sampling at greater than 50 transitions/ns was achieved for every substituent, even though the perturbations spanned volume changes by as much as 60 Å^3. The results from the central replica in our simulations were also in reasonable agreement with the experimental K_d values, giving a Pearson correlation coefficient of 0.56 and AUE of 2.4 kcal/mol (RMSE 3.1 kcal/mol) and a correlation coefficient of 0.98 and AUE of 1.5 kcal/mol (RMSE 1.9 kcal/mol) for derivatives that had reliably measured K_d values. Removal of an outlier value gave us even more accurate K_d values, yielding an RMSE of 2.3 kcal/mol. While we have only used the central replica for analysis of the binding free energies, it is also possible to utilize WHAM to combine the data from all replicas,[58] which may be helpful in dealing with systems that are difficult to converge. On the whole, the BP-REX MSλD algorithm is a highly efficient and scalable free energy method that addresses some of the currently unmet challenges in the field of free energy calculations. It has the potential to enable routine calculations on the order of hundreds of chemical variations of a lead compound using only a few simulations.

ASSOCIATED CONTENT

Supporting Information

The data for the benzoquinone systems and Hsp90-GDM derivative complexes. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: brookscl@umich.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by NIH GM037554 and the Howard Hughes Medical Institute (HHMI International Student Fellowship).

REFERENCES

(1) Bash, P. A.; Singh, U. C.; Langridge, R.; Kollman, P. A. *Science* **1987**, *236*, 564−568.
(2) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395−2417.
(3) Knight, J. L.; Brooks, C. L., III *J. Comput. Chem.* **2009**, *30*, 1692−1700.
(4) Shirts, M. R.; Mobley, D. L. In *Biomolecular Simulations: Methods and Protocols*; Monticelli, L., Salonen, E., Eds.; Springer Science +Business Media: New York, 2013; Vol. *924*, p 271−311.
(5) Hansen, N.; van Gunsteren, W. F. *J. Chem. Theory Comput.* **2014**, *10*, 2632−2647.
(6) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150−160.
(7) Jorgensen, W. L. *Science* **2004**, *303*, 1813−1818.
(8) Gilson, M. K.; Zhou, H.-X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21−42.
(9) Homeyer, N.; Gohlke, H. In *In Silico Drug Discovery and Design*; Unitec House: London, UK, 2013; p 50−63.
(10) Durrant, J. D.; McCammon, J. A. *BMC Biol.* **2011**, *9*, 71−79.
(11) Baron, R.; McCammon, J. A. *Annu. Rev. Phys. Chem.* **2013**, *64*, 151−175.
(12) Jiang, W.; Hodoscek, M.; Roux, B. *J. Chem. Theory Comput.* **2009**, *5*, 2583−2588.
(13) Jiang, W.; Roux, B. *J. Chem. Theory Comput.* **2010**, *6*, 2559−2565.
(14) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13749−13754.
(15) Wang, L.; Berne, B. J.; Friesner, R. A. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1937−1942.
(16) Cole, D. J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2014**, *10*, 565−571.
(17) Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Ling, T.; Abel, R. *J. Chem. Theory Comput.* **2013**, *9*, 1282−1293.
(18) Liu, S.; Wu, Y.; Lin, T.; Abel, R.; Redmann, J. P.; Summa, C. M.; Jaber, V. R.; Lim, N. M.; Mobley, D. L. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 755−770.
(19) Knight, J. L.; Brooks, C. L., III *J. Chem. Theory Comput.* **2011**, *7*, 2728−2739.
(20) Knight, J. L.; Brooks, C. L., III *J. Comput. Chem.* **2011**, *32*, 3423−3432.
(21) Mitchell, M. J.; McCammon, J. A. *J. Comput. Chem.* **1991**, *12*, 271−275.
(22) Guo, Z.; Brooks, C. L., III; Kong, X. *J. Phys. Chem. B* **1998**, *102*, 2032−2036.
(23) Kong, X.; Brooks, C. L., III *J. Chem. Phys.* **1996**, *105*, 2414−2423.
(24) Goh, G. B.; Hulberg, B. S.; Zhou, H.; Brooks, C. L., III *Proteins* **2014**, *82*, 1319−1331.
(25) Goh, G. B.; Knight, J. L.; Brooks, C. L., III *J. Phys. Chem. Lett.* **2013**, *4*, 760−766.
(26) Affentranger, R.; Tavernelli, I.; Iorio, E. E. D. *J. Chem. Theory Comput.* **2006**, *2*, 217−228.
(27) Faraldo-Gomez, J. D.; Kutluay, E.; Jogini, V.; Zhao, Y.; Heginbotham, L.; Roux, B. *J. Mol. Biol.* **2007**, *365*, 649−662.
(28) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. *J. Comput. Chem.* **2009**, *31*, 671−690.
(29) Vanommeslaeghe, K.; Mackerell, A. D., Jr. *J. Chem. Inf. Model.* **2012**, *52*, 3144−3154.
(30) Vanommeslaeghe, K.; Raman, E. P.; Mackerell, A. D., Jr. *J. Chem. Inf. Model.* **2012**, *52*, 3155−3168.

DOI: 10.1021/ct500894k
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(31) Yesselman, J. D.; Price, D. J.; Knight, J. L.; Brooks, C. L., III *J. Comput. Chem.* **2012**, *33*, 189−202.
(32) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(33) 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. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(35) Hockney, R. W. *Methods Comput. Phys.* **1970**, *9*, 136−211.
(36) Ryckaert, J. P.; Cicotti, G.; Berensden, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(37) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(38) Lee, M. S.; Feig, M.; Salsbury, F. R., Jr.; Brooks, C. L., III *J. Comput. Chem.* **2003**, *24*, 1348−1356.
(39) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III *J. Chem. Phys.* **2002**, *116*, 10606−10614.
(40) Stebbins, C. E.; Russo, A. A.; Schneider, C.; Rosen, N.; Hartl, F. U.; Pavletich, N. P. *Cell* **1997**, *89*, 239−250.
(41) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(42) Buck, M.; Bouguet-Bonnet, S.; Pastor, R. W.; Mackerell, A. D., Jr. *Biophys. J.* **2006**, *90*, L36−L38.
(43) Wilcoxon, F. *Biom. Bull.* **1945**, *1*, 80−83.
(44) Mann, H. B.; Whitney, D. R. *Ann. Math. Stat.* **1947**, *18*, 50−60.
(45) R Core Team; R Foundation for Statistical Computing: Vienna, Austria, 2013.
(46) Willett, P. *Drug Discovery Today* **2006**, *11*, 1046−1053.
(47) Willett, P. *Curr. Opin. Biotechnol.* **2000**, *11*, 85−88.
(48) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchinson, G. R. *J. Cheminf.* **2011**, *3*, 33.
(49) Walton-Diaz, A.; Khan, S.; Bourboulia, D.; Trepel, J. B.; Neckers, L.; Mollapour, M. *Future Med. Chem.* **2013**, *5*, 1059−1071.
(50) Miyata, Y.; Nakamoto, H.; Neckers, L. *Curr. Pharm. Des.* **2013**, *19*, 347−365.
(51) Trepel, J.; Mollapour, M.; Giaccone, G.; Neckers, L. *Nat. Rev. Cancer* **2010**, *10*, 537−549.
(52) Tian, Z.-Q.; Wang, Z.; MacMillan, K. S.; Zhou, Y.; Carreras, C. W.; Mueller, T.; Myles, D. C.; Liu, Y. *J. Med. Chem.* **2009**, *52*, 3265−3273.
(53) Tian, Z.-Q.; Liu, Y.; Zhang, D.; Wang, Z.; Dong, S. D.; Carreras, C. W.; Zhou, Y.; Rastelli, G.; Santi, D. V.; Myles, D. C. *Bioorg. Med. Chem.* **2004**, *12*, 5317−5329.
(54) Brazidec, J.-Y. L.; Kamal, A.; Busch, D.; Thao, L.; Zhang, L.; Timony, G.; Grecko, R.; Trent, K.; Lough, R.; Salazar, T.; Khan, S.; Burrows, F.; Boehm, M. F. *J. Med. Chem.* **2004**, *47*, 3865−3873.
(55) Kitson, R. R. A.; Moody, C. J. *Chem. Commun.* **2013**, *49*, 8441−8443.
(56) Kitson, R. R. A.; Chang, C.-H.; Xiong, R.; Williams, H. E. L.; Davis, A. L.; Lewis, W.; Dehn, D. L.; Siegel, D.; Roe, S. M.; Prodromou, C.; Ross, D.; Moody, C. J. *Nat. Chem.* **2013**, *5*, 307−314.
(57) Jez, J. M.; Chen, J. C.-H.; Rastelli, G.; Stoud, R. M.; Santi, D. V. *Chem. Biol.* **2003**, *10*, 361−368.
(58) Gallicchio, E.; Andrec, M.; Felts, A. K.; Levy, R. M. *J. Phys. Chem. B* **2005**, *109*, 6722−6731.

DOI: 10.1021/ct500894k
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX