

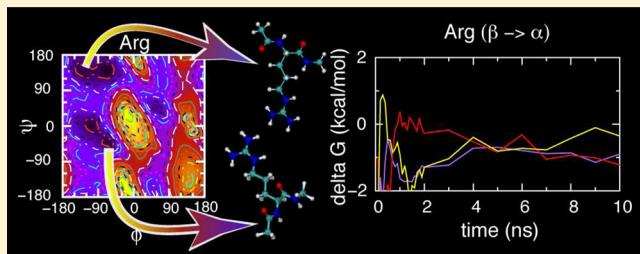
Peptide Backbone Sampling Convergence with the Adaptive Biasing Force Algorithm

Christina E. Faller, Kyle A. Reilly, Ronald D. Hills, Jr., and Olgun Guvench*

Department of Pharmaceutical Sciences, University of New England College of Pharmacy, 716 Stevens Avenue, Portland, Maine 04103, United States

Supporting Information

ABSTRACT: Complete Boltzmann sampling of reaction coordinates in biomolecular systems continues to be a challenge for unbiased molecular dynamics simulations. A growing number of methods have been developed for applying biases to biomolecular systems to enhance sampling while enabling recovery of the unbiased (Boltzmann) distribution of states. The adaptive biasing force (ABF) algorithm is one such method and works by canceling out the average force along the desired reaction coordinate(s) using an estimate of this force progressively accumulated during the simulation. Upon completion of the simulation, the potential of mean force, and therefore Boltzmann distribution of states, is obtained by integrating this average force. In an effort to characterize the expected performance in applications such as protein loop sampling, ABF was applied to the full ranges of the Ramachandran ϕ/ψ backbone dihedral reaction coordinates for dipeptides of the 20 amino acids using all-atom explicit-water molecular dynamics simulations. Approximately half of the dipeptides exhibited robust and rapid convergence of the potential of mean force as a function of ϕ/ψ in triplicate 50 ns simulations, while the remainder exhibited varying degrees of less complete convergence. The greatest difficulties in achieving converged ABF sampling were seen in the branched-side chain amino acids threonine and valine, as well as the special case of proline. Proline dipeptide sampling was further complicated by *trans*-to-*cis* peptide bond isomerization not observed in unbiased control molecular dynamics simulations. Overall, the ABF method was found to be a robust means of sampling the entire ϕ/ψ reaction coordinate for the 20 amino acids, including high free-energy regions typically inaccessible in standard molecular dynamics simulations.



INTRODUCTION

Molecular simulations are often used to compute a biomolecular system's free energy landscape as a function of one or more particular reaction coordinates (progress variables), providing insight into conformational heterogeneity and the mechanism of structural transitions.¹ Traditional, nonaccelerated simulations suffer from poor or incomplete sampling of the full reaction coordinate in complex systems, particularly in high-energy regions.^{2–6} If a high-energy region is a transition state, its presence will also hamper converged sampling of the low-energy regions that it separates. On time scales presently achievable by molecular dynamics (MD) simulations, reaction coordinates typically of interest in biopolymers involve landscape ruggedness sufficient to hamper exhaustive equilibrium sampling. Therefore, an important goal is to overcome the sampling problem of unbiased MD simulations with computationally efficient methods.

A variety of methods exist to accelerate local barrier crossings,^{7–9} a subset of which make it possible to determine the Boltzmann distribution of conformational states. Examples of the latter include umbrella sampling,¹⁰ metadynamics,^{11,12} use of constraint forces,^{13–15} and the adaptive biasing force (ABF) method.^{3–6} ABF aims for uniform sampling of reaction coordinates using an “on the fly” estimate of the forces needed

to overcome local free energy barriers encountered during the simulation.^{2–6,16,17} More precisely, the equilibrium force $\langle F \rangle$ is estimated by sampling the instantaneous force and then canceled by application of the adaptive bias $-\langle F \rangle$. Because the estimate relies on the previous history of sampling, the applied biasing force is “adaptive” and is continually updated to achieve the limit of uniform sampling along the reaction coordinate when the applied force is exactly the opposite of the true value of $\langle F \rangle$. Per the reversible work theorem,¹⁸ the potential of mean force (PMF) is then obtained by integrating the adaptive biasing force along the reaction coordinate, similar to constrained simulations where the constraint force is integrated to yield the PMF.¹⁹ Modifications have been made upon the original method developed by Darve and Pohorille³ to expand its use to multiple coordinates.² Analogous to metadynamics,^{11,12} applying the ABF methodology in multiple dimensions yields a multidimensional PMF.

The alanine dipeptide (Figure 1) is an often-used and well-characterized test case for new computational techniques, where backbone dihedral angles ϕ and ψ are used as system

Received: October 1, 2012

Revised: December 6, 2012

Published: December 10, 2012



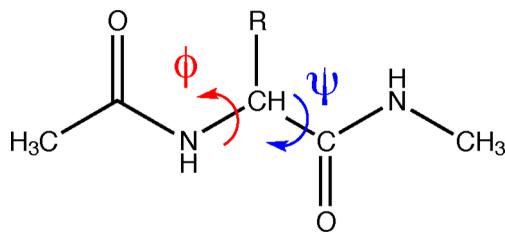


Figure 1. Amino acid “dipeptide” representation of the protein backbone with dihedral angles ϕ/ψ . The $-R$ group denotes the side chain.

descriptors. Alanine has previously been used to demonstrate the efficacy of ABF; using ranges of ϕ and ψ of $[180^\circ, 360^\circ]$ and $[-120^\circ, 180^\circ]$, respectively, Darve et al. rapidly achieved uniform sampling of the complete reaction coordinate, despite the presence of high free-energy regions that would be problematic for sampling in unbiased simulations.⁴ Alanine dipeptide has also served as an example to directly compare ABF and metadynamics,² as both techniques are intended to improve sampling of rare events, and hence should sample even the high-energy conformations of alanine dipeptide to generate a full Ramachandran map (ϕ and ψ of $[-180^\circ, +180^\circ]$ and $[-180^\circ, +180^\circ]$). In a comparison of these methods, ABF and metadynamics simulations yielded similar accuracy and results, but it was suggested that ABF might be a better approach in more “fragile” systems.² In a comparison of ABF and umbrella sampling,¹⁰ the reversible folding of deca-L-alanine was simulated under otherwise equivalent conditions; the ABF-generated PMF was comparable (in both qualitative and quantitative terms) to that of an umbrella sampling simulation 40 times longer.¹⁷ This implies that ABF may indeed be more efficient in complicated systems.

We investigated the free energy convergence of ABF on dipeptides of the 20 amino acids using backbone ϕ and ψ reaction coordinates. The effects of side chain branching on backbone sampling convergence were examined, as well as the role of peptide bond isomerization. Such studies are relevant to molecular simulations of proteins, since changes in just one or two backbone dihedral angles can result in biologically important structural changes such as flexible surface loop conformational switching.^{20,21}

METHODS

Preparation of Dipeptide Systems. Dipeptide systems were built in triplicate for the 20 amino acids, with three different ϕ/ψ conformations ($180^\circ/180^\circ$, “E” for extended; $-60^\circ/-60^\circ$, “HR” for right-handed helical; and $60^\circ/60^\circ$, “HL” for left-handed helical) and having side chain conformations based on the default internal coordinate values in the CHARMM force field topology files.²² Side chains were protonated as appropriate for neutral pH. Histidine dipeptides were built with the side chain imidazole singly protonated at either the delta (“Hsd”) or epsilon (“Hse”) nitrogen. Each dipeptide was centered in a box of $12 \times 12 \times 12 = 1728$ TIP3P²³ water molecules placed at evenly spaced grid points with a box size appropriate for the experimental density of water. Net system charges were balanced to 0 with the addition of either a single sodium or chloride ion as needed at the center of one of the faces of the box for dipeptides with charged side chains. Systems were constructed using the CHARMM software.^{24,25}

MD Simulations and Free Energy Calculations. Systems were represented using the CHARMM all-atom additive protein force field with the CMAP backbone potential^{22,26–29} and a modified TIP3P three-site rigid water molecule.^{23,30} Periodic boundary conditions³¹ were used for energy calculations with a cubic periodic unit. Bonds involving hydrogen were constrained to their equilibrium lengths using the SHAKE algorithm,³² and SETTLE,³³ a faster, noniterative algorithm, was used to maintain rigid water geometries. A time step of 0.002 ps was used with the BBK integrator³⁴ in combination with Langevin thermostating³⁵ and barostating.³⁶ Lennard-Jones interactions were smoothly truncated using an 8–10 Å switching distance.³⁷ Electrostatic interactions beyond the 10 Å cutoff were computed via the particle-mesh Ewald method using a ~1 Å grid spacing,³⁸ and the effect of long-range Lennard-Jones interactions on the system pressure was accounted for by an isotropic correction.³¹ Each system was energy minimized for 1000 steps using the conjugate-gradient algorithm^{39,40} and then heated to and equilibrated at 310 K and 1 atm for 40 ps by reassignment of velocities every 1000 steps, with harmonic positional restraints applied to dipeptide heavy atoms. The subsequent 50 ns production simulation replaced velocity reassignment with Langevin dynamics using a damping coefficient of 0.1 ps^{-1} , and ABF was in effect for the full 50 ns duration.

The full two-dimensional domain defined by ABF degrees of freedom ϕ and ψ was sampled to generate a complete Ramachandran map from -180 to $+180^\circ$ in both dimensions.⁴¹ Periodicity was allowed, enabling a system to transition directly from, e.g., $\phi = -179$ to $+179^\circ$ without having to cross $\phi = 0^\circ$. Hydrogen atoms attached to non-hydrogen atoms defining the ϕ/ψ angles, namely, the backbone amide hydrogens and the C_α hydrogen, were given increased masses of 12 amu. SHAKE bond constraints were not applied to bonds involving these hydrogen atoms in order to meet the ABF requirement that no constraints involve atoms used to define ABF degrees of freedom.¹⁷ Bin widths were 10° in each dimension. The adaptive biasing force was not applied fully to a given bin until n_{full} force samples had been collected for that bin. Specifically, until the number of force samples n equaled or exceeded n_{full} , the adaptive biasing force for a bin was scaled by a factor of f , where $f = 0$ for $n < n_{\text{min}}$ and $f = (n - n_{\text{min}})/(n_{\text{full}} - n_{\text{min}})$, with $n_{\text{full}} = 100$ and $n_{\text{min}} = n_{\text{full}}/2$. ABF history files were saved every 100 ps during the first 2 ns of the simulation and every 1 ns afterward for analysis. NAMD 2.8⁴² was used for minimization, MD, and ABF free energy sampling. Gradient and histogram data from the history files were used to generate 2D ϕ/ψ PMFs using the “abf_integrate” utility program, distributed as part of the NAMD package. “abf_integrate” was run with default settings, where a Monte Carlo simulation is performed in gradient space, and sampling of ϕ/ψ bins is biased by a memory-dependent potential (i.e., computed PMF) in the same spirit as metadynamics^{11,12} so as to eventually achieve uniform sampling when the forces due to the computed PMF cancel out the ABF gradients. Per the default settings, the integration is deemed converged when the RMSD between the ABF simulation gradients and the gradients from the “abf_integrate” computed PMF does not improve with additional Monte Carlo sampling. For the dipeptide systems, this typically occurred after 5–10 million MC iterations, and increasing the number of iterations by 10-fold for several select dipeptide data sets yielded no additional improvement in integration convergence.

RMSD values referenced in the text refer to the root-mean-square deviation (in kcal/mol) between pairs of PMFs, each generated from one of the three independent ABF simulations for a dipeptide. Pairs of PMFs were first least-squares aligned in the energy coordinate prior to RMSD calculation. For graphical display, PMFs were aligned along the energy coordinate such that the lowest free energy in each PMF is 0 kcal/mol.

RESULTS AND DISCUSSION

The PMFs obtained from three independent simulations of each dipeptide were used to gauge convergence as a function of

Table 1. Dipeptide Groupings Based on Final (50 ns) Intersimulation PMF RMSDs

group 0.5 ^a	group 1.0 ^a	group 1.5 ^a
Ala, Asp, Cys, Glu, Gly, His, Ser, Trp, Tyr	Arg, Asn, Gln, His, Ile, Leu, Lys, Met, Phe	Thr, Val

^aAll three intersimulation PMF RMSD values for a given dipeptide must fall below the indicated value in kcal/mol. Proline values were 1.3, 2.6, and 3.1 kcal/mol.

simulation time. Two metrics were used for comparison: (1) RMSD values between the three PMFs and (2) consistency of the free energy differences ΔG between different free-energy minima in ϕ/ψ space. Because RMSD measures convergence across the entire reaction coordinate space, it can be skewed by poor estimates for high free-energy (low probability) regions. Therefore, the ΔG metric is complementary in that it measures convergence in the low free-energy (high-probability) regions. For a given dipeptide, as convergence is achieved, the RMSD should approach zero and the ΔG values computed from each PMF should converge to a single consensus value; in practice, due to the stochastic nature of the “abf_integrate” program used to integrate the ABF gradients to generate a ϕ/ψ PMF, pairwise RMSDs for PMFs generated using the exact same alanine dipeptide ABF data are 0.1–0.2 kcal/mol, and such RMSD values thus represent a practical minimum in the limit of full convergence. As discussed below in detail, quantitative

convergence was observed for some dipeptides, qualitative convergence for others, and poor convergence for only a remaining few.

The dipeptides were clustered into three groups with increasing PMF RMSD cutoffs for more detailed analysis: RMSD < 0.5 kcal/mol (“group 0.5”), <1.0 kcal/mol (“group 1.0”), and <1.5 kcal/mol (“group 1.5”), where a given dipeptide was assigned to a particular group if all three pairwise intersimulation RMSDs met the cutoff at 50 ns. Group 0.5 and group 1.0 contained nine members each, while group 1.5 had two members (Table 1). Group 0.5 contains the two amino acids with the simplest side chains: Gly and Ala. Gly and Ala had both the quickest-converging and lowest final RMSDs, with values of 0.4 kcal/mol achieved at 5 ns and values of 0.2–0.3 kcal/mol at 50 ns. The other two amino acids that demonstrated quick RMSD convergence were Ser and Cys, with values of 0.4 and 0.5 kcal/mol achieved at 5 ns, respectively. Following Ala, Ser and Cys have the next simplest side chains, composed of two non-hydrogen atoms. At the other end of the spectrum was Pro, which fell outside of all three of the groups, as its intersimulation RMSDs were 3.1, 1.3, and 2.6 kcal/mol. Though not as problematic as Pro, Thr and Val, with two heavy atoms bound to their beta carbon, showed relatively poor convergence, placing them in group 1.5. Surprisingly, Ile, which like Thr and Val is also branched at the beta carbon, exhibited good convergence, with final intersimulation RMSDs ranging from 0.5 to 0.6 kcal/mol. Differences in side chain conformational behavior are discussed further below, and time courses for the intersimulation RMSDs are shown in Figure 2 and in Supporting Information Figure S1.

Computing the free energy changes for transition between the β , α , and α_L (left-handed α) regions of ϕ/ψ space showed trends consistent with the RMSD data. These three different conformational states were, for the purposes of this study, defined as different $10^\circ \times 10^\circ$ regions of the ϕ/ψ space. For the β conformation, this region was centered at $\phi/\psi = -135^\circ/135^\circ$, for α at $\phi/\psi = -65^\circ/-45^\circ$, and for α_L at $\phi/\psi = 65^\circ/45^\circ$. The β -to- α transition $\Delta G(\beta \rightarrow \alpha)$ was computed as $G(\alpha) - G(\beta)$ and the β to α_L transition $\Delta G(\beta \rightarrow \alpha_L)$ as $G(\alpha_L) - G(\beta)$.

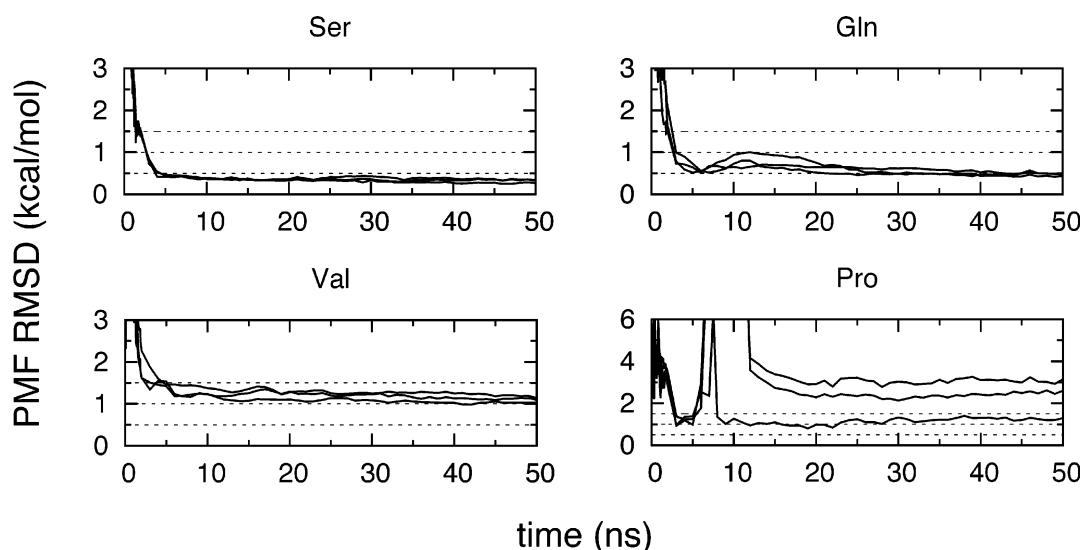


Figure 2. PMF heterogeneity in independent dipeptide simulations. The three intersimulation PMF RMSDs (kcal/mol) are shown as a function of simulation time. Note that the RMSD scale for Pro is 2× that of the other amino acids. Data for all 21 systems are presented in Supporting Information Figure S1.

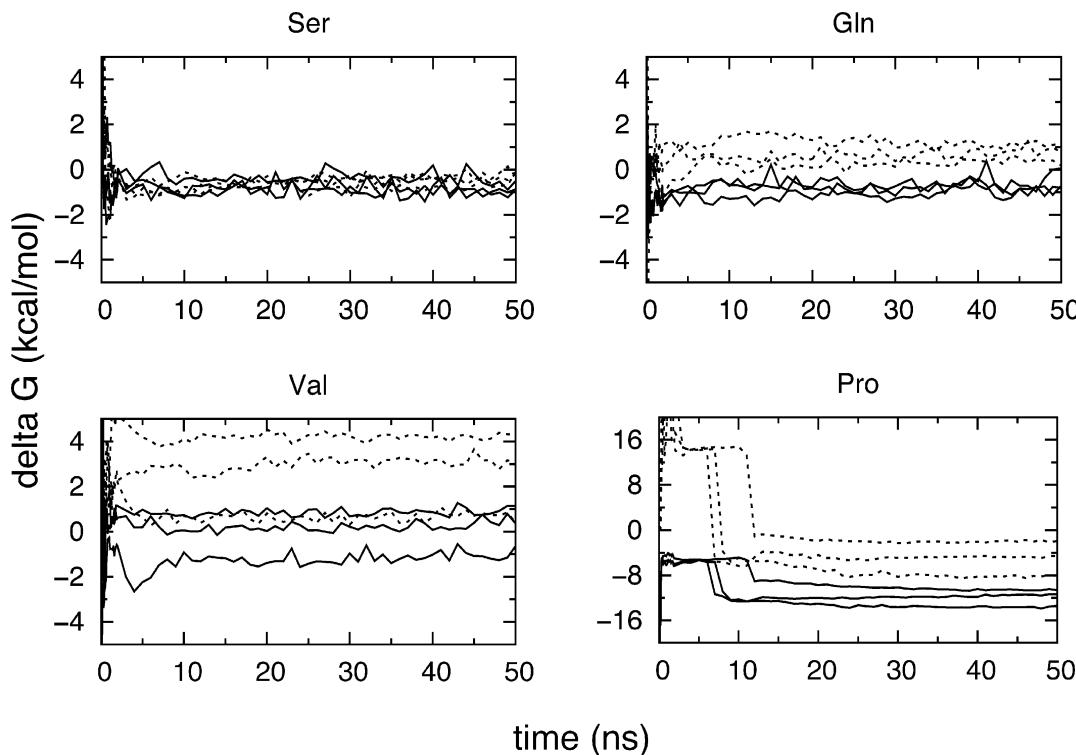


Figure 3. Intra-PMF ΔG 's for the $\beta \rightarrow \alpha$ (solid) and $\beta \rightarrow \alpha_L$ (dash) transitions in independent dipeptide simulations. Note that the ΔG scale for Pro is 4× that of the other amino acids. Data for all 21 systems are presented in Supporting Information Figure S2.

Table 2. Confidence Intervals for $\beta \rightarrow \alpha$ and $\beta \rightarrow \alpha_L$ Transition Free Energies^a

	$\langle \Delta G(\beta \rightarrow \alpha) \rangle$	95% CI	$\langle \Delta G(\beta \rightarrow \alpha_L) \rangle$	95% CI
Group 0.5				
Ala	-1.22	0.20	-0.29	0.66
Asp	-1.78	0.13	-0.55	1.01
Cys	-0.76	0.11	-0.87	1.72
Glu	-0.73	1.27	0.79	1.13
Gly	-2.41	0.14	-2.28	0.63
Hse	-0.38	0.55	-0.80	0.59
Ser	-0.75	0.56	-0.20	0.34
Trp	-0.72	0.74	-0.31	1.20
Tyr	-0.55	0.69	0.32	0.46
average		0.49		0.86
Group 1.0				
Arg	-0.48	0.73	0.50	0.43
Asn	-0.76	0.47	-0.46	0.59
Gln	-0.49	1.18	0.69	1.20
Hsd	-0.63	0.91	-0.33	1.39
Ile	-0.47	1.39	2.81	0.28
Leu	-1.23	1.21	0.19	2.19
Lys	-0.38	0.89	1.36	1.82
Met	-1.06	2.14	0.33	1.80
Phe	-0.72	1.84	0.21	2.25
average		1.19		1.33
Group 1.5				
Thr	-1.71	1.97	0.09	2.08
Val	0.29	2.22	2.60	4.94
average		2.10		3.51
Pro	-11.83	3.57	-4.90	7.78

^aAll data are from the final 50 ns timepoints.

Following these intra-PMF ΔG values as a function of time (Figure 3 and Supporting Information Figure S2), a similar trend is observed as in the inter-PMF RMSD data: most of the change occurs within the first 5 ns of the 50 ns ABF simulations. In the case of Val, which showed poor RMSD convergence, estimates for $\Delta G(\beta \rightarrow \alpha_L)$ reach stable values of 0, 3, and 4 kcal/mol by 5 ns and remain unchanged (and unconverged) for the next 45 ns. This is in contrast to quickly converging systems like Ser, for which nearly identical estimates of ΔG are reached within 5 ns. In aggregate, 95% confidence intervals at the 50 ns time point for $\Delta G(\beta \rightarrow \alpha)$ and $\Delta G(\beta \rightarrow \alpha_L)$ for dipeptides in group 0.5 are 0.49 and 0.86 kcal/mol, respectively (Table 2), showing the ability of ABF to achieve near-quantitative convergence. Error estimates were 1.19 and 1.33 kcal/mol for group 1.0, consistent with qualitative convergence, and 2.10 and 3.51 for group 1.5, consistent with poor convergence (Table 2). As mentioned previously, while RMSD measures convergence across the entire ϕ/ψ reaction coordinate including free-energy maxima, ΔG values measure convergence in the lowest free-energy, and hence thermodynamically most important, regions of the reaction coordinate. However, these separate measures of convergence are correlated, as demonstrated by increasing errors for ΔG estimates with increasing group RMSD values (Table 2).

Despite the differences in convergence properties and level of accuracy in free-energy estimates, comparing ϕ/ψ PMFs from each group demonstrates the ability of ABF to consistently sample the entire ϕ/ψ space and locate minima, maxima, and transition states. Figure 4 shows the 50 ns ϕ/ψ PMFs for Ser, Gln, Val, as well as Pro, which had the poorest ABF convergence based on RMSD and ΔG values. In the case of Ser dipeptide, there is quantitative agreement in the relative heights of the three minima in the β , α , and α_L regions, as well as the heights and locations of maxima (near $0^\circ/0^\circ$, $-15^\circ/-15^\circ$).

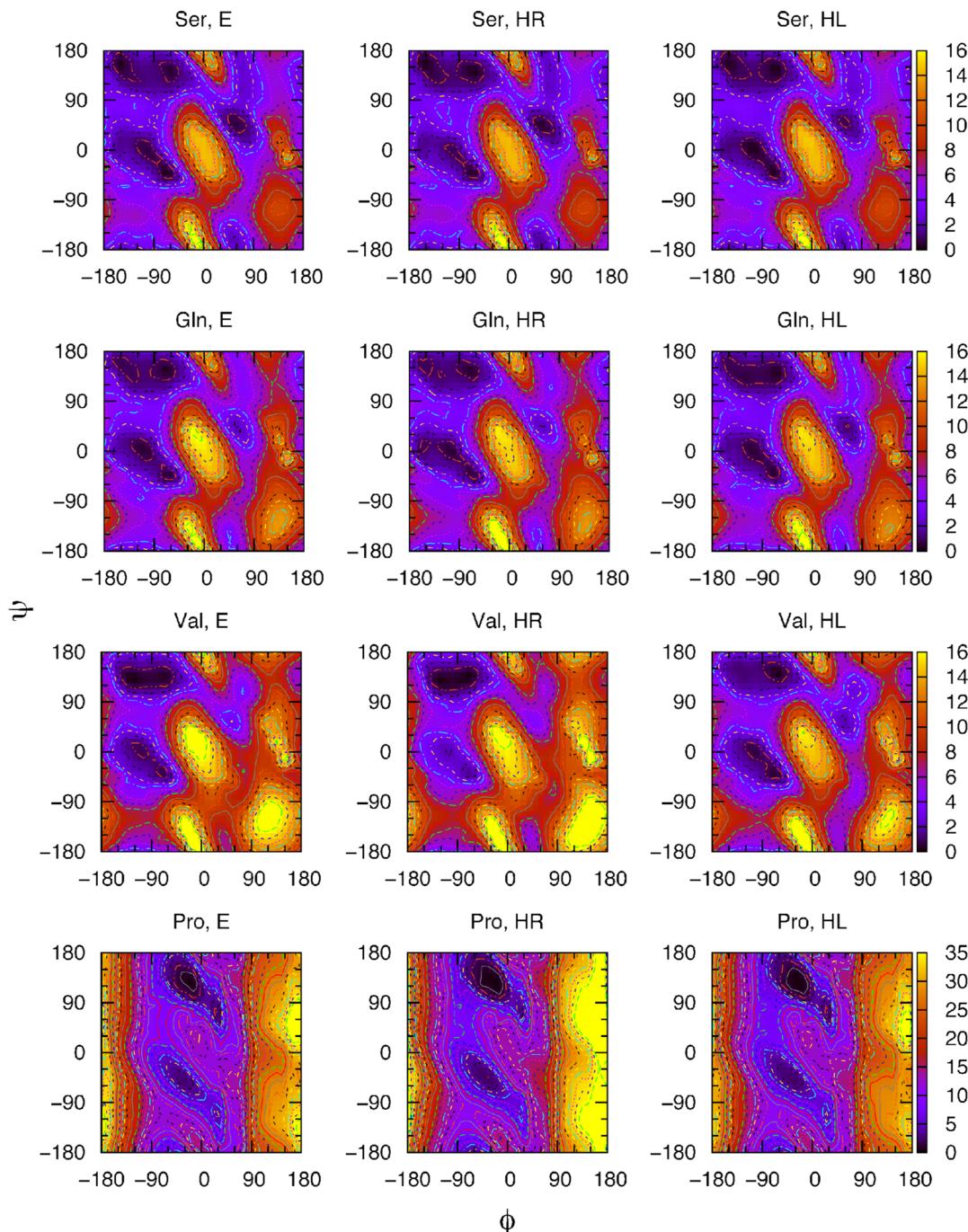


Figure 4. ϕ/ψ PMFs from three independent 50 ns ABF simulations of Ser (from group 0.5), Gln (group 1.0), Val (group 1.5), and Pro. “E” simulations were seeded with dipeptide conformations having $\phi/\psi = 180^\circ/180^\circ$ (extended), “HR” with $-60^\circ/-60^\circ$ (right-handed helical), and “HL” with $60^\circ/60^\circ$ (left-handed helical). Contours are drawn every kcal/mol. Note the difference in z-axis for Pro. Data for all 21 systems are presented in Supporting Information Figure S3.

165°, 135°/−115°, and 145°/0°) and saddle point regions between the minima. For Gln, there is qualitative agreement between the three simulations, but differences are observed in the α_L minimum and 0°/0° maximum. In contrast, for Val, which demonstrated poor convergence in RMSD and ΔG , agreement is lacking across the three simulations with regard to minima, maxima, and transition state free energies. Nevertheless, the locations of these regions are consistent across the Val simulations, and the entire reaction coordinate space is indeed sampled, demonstrating the utility of the ABF method. Finally, while the RMSD and ΔG time series for Pro suggest it

to be the most challenging dipeptide with regard to ABF sampling, there is some notable consistency between the three independent PMFs. Specifically, there are two major minima on the surface, one in the α region and the other near $-35^\circ/145^\circ$ in the poly proline region. The ΔG between these two regions is quantitatively consistent across the three simulations, and estimates of the transition state connecting them are within 1 kcal/mol. In the case of proline, it is the high free-energy regions of the reaction coordinate, located at $\phi < -120^\circ$ and $\phi > 60^\circ$ (Figure 4, “Pro”), which are inconsistently estimated, leading to large RMSD values (Figure 2). Additionally, because

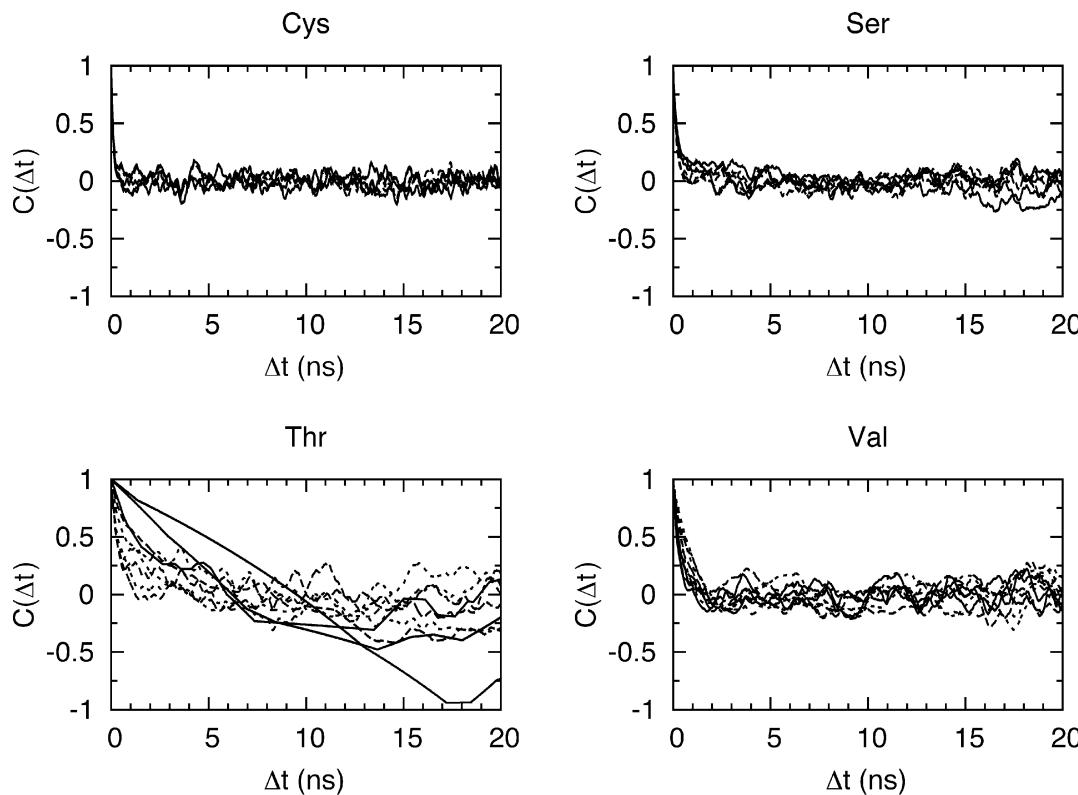


Figure 5. Time autocorrelation functions $C(\Delta t)$ for side chain dihedral χ_1 sampling in ABF dipeptide simulations. Data are shown for *trans* (solid), *gauche minus* (dash), and *gauche plus* (dot) states. Results for each state are shown only if its sampling exceeds a 15% threshold in two of the three simulations.

Table 3. Final (50 ns) ΔG and Inter-Simulation PMF RMSD Values (kcal/mol) for Val in Unrestrained and χ_1 -Restrained ABF Simulations

	unrestrained	restrained
$\Delta G(\beta \rightarrow \alpha)$	0.36, -0.65, 1.14	-1.39, -1.21, -1.66
$\Delta G(\beta \rightarrow \alpha_L)$	3.11, 0.40, 4.28	2.09, 2.50, 2.03
RMSD	1.03, 1.14, 1.10	0.42, 0.49, 0.48

the β and α_L conformations are in these high energy regions, the 50 ns estimates for $\Delta G(\beta \rightarrow \alpha)$ and $\Delta G(\beta \rightarrow \alpha_L)$ are inconsistent (Table 2).

We end the discussion on proline dipeptide by noting that, because the N and C_α atoms are connected in a 5-membered ring, the ϕ/ψ surface is in fact *not* periodic, unlike the other dipeptides. Rather, in the ϕ coordinate, rotation of the backbone dihedral angle will lead to build-up of ring strain due to twisting of the ring, and this strain may be released by *trans*-*cis* isomerization of the N-terminal peptide bond, leading to a new stable peptide backbone conformation with a different ϕ/ψ PMF.⁴³ This occurred in one of the three present simulations, namely, the “HR” simulation, and may be a contributing factor to the PMF from this simulation being conspicuously different than those from the “E” and “HL” simulations (Figure 4, “Pro HR”). The three Pro simulations were therefore repeated with half-harmonic restraining potentials in place for $\phi < -150^\circ$ and $150^\circ < \phi$; however, this still did not prevent isomerization, which was observed in both the “HR” and “HL” simulations. In contrast, all of the other amino acid dipeptide ABF simulations retained *trans* peptide bonds for their entirety. And, repeating the triplicate Pro simulations without ABF resulted in no isomerization,

though naturally in such unbiased simulations a very limited region of the ϕ component of the reaction coordinate is sampled. Such results suggest special care in applying ABF sampling to the peptide backbone of proline residues, for example, by limiting the extent of ϕ sampling to only the low free-energy regions (Figure 4, “Pro”).

One possible explanation for the dramatic side chain dependence of ABF convergence involves differences in the characteristic relaxation times associated with side chain conformational sampling. If the side chain conformation affects ϕ/ψ energetics and rotamer sampling times are on the order of or greater than the simulation length, then incomplete rotamer sampling will hinder PMF convergence. Time autocorrelation functions $C(\Delta t)$ ¹⁸ were analyzed for the χ_1 dihedral angle, defined as $N-C_\alpha-C_\beta-X_\gamma$, where X_γ is a non-hydrogen atom whose identity depends on the amino acid side chain. This dihedral is relevant, as it determines the location of X_γ relative to the backbone and its degree of interaction with the dipeptide. In contrast, C_β is fixed relative to the backbone and the X_δ position only exists in a subset of amino acids.

Relaxation times were compared for group 0.5 and group 1.5 dipeptides Cys and Ser and Thr and Val (Figure 5). $C(\Delta t)$ was computed by defining three χ_1 states: *trans*, $\chi < -120^\circ$ or $\chi \geq 120^\circ$; *gauche minus*, $-120^\circ \leq \chi < 0^\circ$; *gauche plus*, $0^\circ \leq \chi < 120^\circ$. $C(\Delta t)$ was computed for the memory of the three χ_1 states in each independent simulation. For example, for a *trans* $C(\Delta t)$ calculation, raw dihedral values were converted to 1's and 0's, with a raw value converted to 1 if it met the criteria for being in the *trans* state and 0 otherwise. $C(\Delta t)$ was then computed from the time series of 1's and 0's using the CORREL facility in CHARMM. Not surprisingly, $C(\Delta t)$ values for Cys and Ser

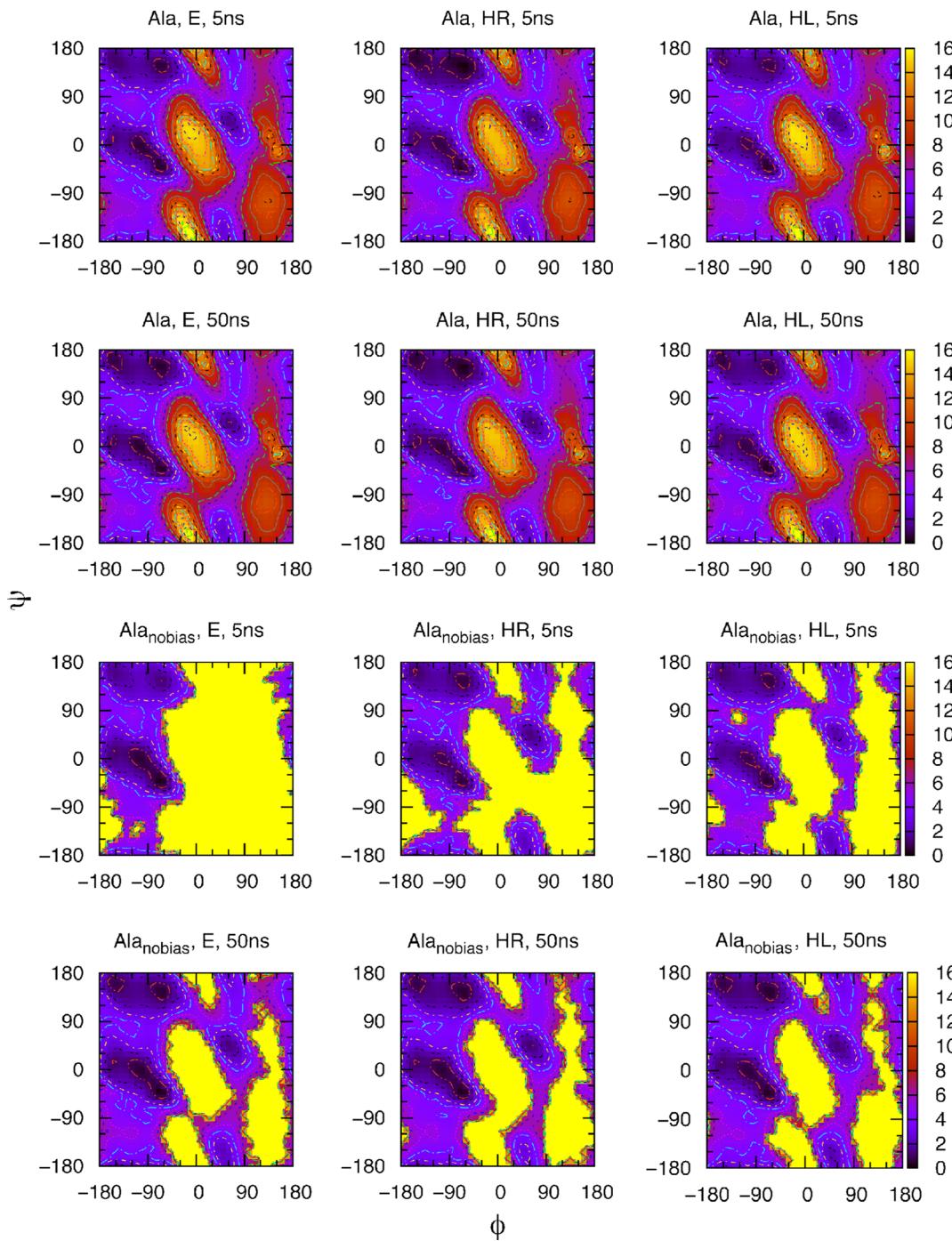


Figure 6. ϕ/ψ PMFs after 5 and 50 ns of sampling in independent Ala simulations using ABF (Ala) and standard MD ($\text{Ala}_{\text{no bias}}$). Contours are drawn every 1 kcal/mol. Bright yellow areas in the $\text{Ala}_{\text{no bias}}$ landscapes indicate unsampled regions.

decay rapidly to ~ 0 by $\Delta t < 3$ ns. Consistent with the hypothesis that side chain sampling affects ABF convergence, $C(\Delta t)$ for Thr shows slow decay. The $C(\Delta t)$ decay behavior for Val was intermediate between these two extremes.

Valine was further studied by repeating the ABF triplicate simulations with a biasing potential on χ_1 to restrain it to the *trans* state. The heating/equilibration protocol was unchanged, and the production ABF protocol only differed by the addition of a harmonic restraint to the $\text{N}-\text{C}_\alpha-\text{C}_\beta-\text{C}_{\gamma 1}$ dihedral of the form $k(\chi - \chi_0)^2/2$, where $\chi_0 = 180^\circ$ and $k = 0.050 \text{ kcal mol}^{-1} \text{ deg}^{-2}$. With this modification, Val switches from group 1.5 to group 0.5 and from inconsistent to consistent ΔG estimates

(Table 3), providing additional evidence that limitations in Val convergence are due to rotamer sampling. Therefore, if computational resources allow, it may be advisable to repeat a simulation several times with different starting conditions and compare intersimulation PMFs. Unfortunately, this is not always possible given that many biomolecular simulations tax the limits of available computational resources.

As a final measure of the efficacy of ABF for peptide ϕ/ψ sampling, the ABF data for alanine were compared with standard MD simulations performed in triplicate. PMFs from these latter simulations were generated by computing $G = -RT \ln p(\phi/\psi) + C$, where $p(\phi/\psi)$ is the number of MD samples in

a $10^\circ \times 10^\circ$ bin centered at ϕ/ψ and C is selected such that $G = 0$ at the global minimum. Figure 6 reveals the differences in sampling of minima, transition state regions, and maxima early in (5 ns) and at the end of (50 ns) the simulations. Two key differences between standard MD and ABF sampling are apparent: only ABF samples high free-energy regions of ϕ/ψ space, and only ABF demonstrates convergent sampling early in the simulations (Figure 6). Both standard MD and ABF of alanine give similar estimates at 50 ns for low-lying basins and their minimum energy transitions. Standard MD, however, fails to sample the less accessible, high-energy regions that separate conformational basins.

CONCLUSIONS

While ABF sampling in the 50 ns time regime investigated here was not sufficient to generate quantitative ϕ/ψ free energy estimates for all 20 amino acids, it was sufficient to generate such estimates for nearly half of them. And, ABF did sample the minima, maxima, and transition states across the entire span of ϕ/ψ values for all dipeptides. Backbone-rotamer coupling appears to explain the poor convergence of certain dipeptides. Rotational sampling of the Thr side chain is slow relative to the 50 ns time scale. Although sampling of the Val side chain is comparatively fast, removing side chain rotation converted the Val dipeptide system from poorly to well-converged. Finally, for the trivial alanine dipeptide, ABF sampling is clearly superior to unbiased MD with regard to generating converged estimates of $\Delta G(\phi/\psi)$ across the full ϕ/ψ reaction coordinate.

With regard to generalizing the results to MD studies on proteins, our own work has shown that in certain instances the overall conformation of a protein loop region is dependent upon the backbone conformation of a single key residue in the loop. In one instance, the peptide-binding cleft of an SH2 domain, formed by two loop regions, could be induced to go from a closed conformation to an open one capable of peptide binding solely by transition of a Tyr residue backbone in one of the loops from the extended region of ϕ/ψ space to the left-handed helical region.²⁰ In another instance, the transition from the extended region of ϕ/ψ space to the left-handed helical region for a Tyr residue in a loop forming the carbohydrate-binding cleft of the CD44 protein induced a loop conformational change allowing for formation of a direct contact between the side chain of an Arg residue in the loop and the carbohydrate ligand.²¹ This latter conformational change is thought to be a molecular basis for affinity-switching in CD44 binding to its carbohydrate ligand hyaluronan.⁴⁴ The ABF methodology can therefore be a very attractive solution for determining the thermodynamics associated with loop sampling in such instances. From the dipeptide data here, it is expected that ABF studies of loop conformation as a function of Ala, Asp, Cys, Glu, Gly, His, Ser, Trp, or Tyr backbone conformation ought to yield quantitative results on the 50 ns time scale. Results will be more qualitative for the other 11 amino acids, with Pro, Thr, and Val being especially challenging. Special attention should be paid to *trans-cis* peptide bond isomerization in proline, a phenomenon observed in the present ABF studies that does not occur in the absence of ABF sampling on the same time scale. Of course, in the case of loop sampling, in addition to side chain effects on backbone dynamics, other interactions involving additional degrees of freedom may also affect convergence times. On the basis of the present results, the most reliable way to assess convergence is using multiple simulations with different starting conditions. The present

results caution against overconfidence in measures of time-dependent behavior in single simulations as yardsticks for PMF convergence. The fact that ΔG between two states reaches a stable, seemingly time-invariant value is not a guarantee that it has converged. In conclusion, when combined with appropriate measures of convergence, the ABF methodology provides a useful means of free energy measurement for conformations of the protein backbone.

ASSOCIATED CONTENT

S Supporting Information

Figures of PMF heterogeneity plots, intra-PMF ΔG s of transition for all systems, and ϕ/ψ PMFs from three independent 50 ns ABF simulations for all dipeptide systems. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: oguvench@une.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

O.G. thanks University of New England for startup funding and Garlic Jones for helpful discussions. This work was supported by National Institutes of Health grant R15 GM099022 (O.G. and R.D.H) and by the National Science Foundation through XSEDE resources provided by the National Institute for Computational Sciences (grant TG-MCB120007; O.G.).

REFERENCES

- (1) Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2010**, *31*, 1569.
- (2) Hénin, J.; Fiorin, G.; Chipot, C.; Klein, M. L. *J. Chem. Theory Comput.* **2010**, *6*, 35.
- (3) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169.
- (4) Darve, E.; Rodriguez-Gomez, D.; Pohorille, A. *J. Chem. Phys.* **2008**, *128*, 144120.
- (5) Darve, E.; Wilson, M. A.; Pohorille, A. *Mol. Simul.* **2002**, *28*, 113.
- (6) Rodriguez-Gomez, D.; Darve, E.; Pohorille, A. *J. Chem. Phys.* **2004**, *120*, 3563.
- (7) Bhatt, D.; Zuckerman, D. M. *J. Chem. Theory Comput.* **2011**, *7*, 2520.
- (8) Isralewitz, B.; Gao, M.; Schulter, K. *Curr. Opin. Struct. Biol.* **2001**, *11*, 224.
- (9) Schlitter, J.; Engels, M.; Krüger, P. *J. Mol. Graphics* **1994**, *12*, 84.
- (10) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.
- (11) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562.
- (12) Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. *Curr. Opin. Struct. Biol.* **2010**, *20*, 148.
- (13) Carter, E. A.; Ciccotti, G.; Hynes, J. T.; Kapral, R. *Chem. Phys. Lett.* **1989**, *156*, 472.
- (14) den Otter, W. K. *J. Chem. Phys.* **2000**, *112*, 7283.
- (15) den Otter, W. K.; Briels, W. J. *J. Chem. Phys.* **1998**, *109*, 4139.
- (16) Chipot, C.; Henin, J. *J. Chem. Phys.* **2005**, *123*, 244906.
- (17) Henin, J.; Chipot, C. *J. Chem. Phys.* **2004**, *121*, 2904.
- (18) Chandler, D. In *Introduction to Modern Statistical Mechanics*; Oxford University Press: New York, 1987; p 201.
- (19) Darve, E. In *Numerical Methods for Calculating the Potential of Mean Force New Algorithms for Macromolecular Simulation*; Leimkuhler, B., Chipot, C., Elber, R., Laaksonen, A., Mark, A., Schlick, T., Schütte, C., Skeel, R., Eds.; Springer: Berlin, Heidelberg, 2006; Vol. 49; p 213.

- (20) Guvench, O.; Qu, C.-K.; MacKerell, A. D., Jr. *BMC Struct. Biol.* **2007**, *7*, 14.
- (21) Jamison, F. W., 2nd; Foster, T. J.; Barker, J. A.; Hills, R. D., Jr.; Guvench, O. *J. Mol. Biol.* **2011**, *406*, 631.
- (22) MacKerell, A. D., Jr.; Bashford, D.; Bellott, Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586.
- (23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (24) Brooks, B. R.; Brooks, C. L., 3rd; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. *J. Comput. Chem.* **2009**, *30*, 1545.
- (25) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
- (26) Guvench, O.; MacKerell, A. D., Jr. *Methods Mol. Biol.* **2008**, *443*, 63.
- (27) MacKerell, A. D., Jr.; Brooks, B.; Brooks, C. L., 3rd; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. CHARMM: The Energy Function and Its Parameterization. *Encyclopedia of Computational Chemistry*; John Wiley & Sons, Ltd: Chichester, 1998.
- (28) Mackerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd. *J. Comput. Chem.* **2004**, *25*, 1400.
- (29) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd. *J. Am. Chem. Soc.* **2004**, *126*, 698.
- (30) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. *J. Phys. Chem.* **1994**, *98*, 2198.
- (31) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, U.K., 1987.
- (32) Ryckaert, J. P.; Ciccotti, G.; Brenderson, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
- (33) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952.
- (34) Brünger, A.; Brooks, C. L., 3rd; Karplus, M. *Chem. Phys. Lett.* **1984**, *105*, 495.
- (35) Kubo, R.; Toda, M.; Hashitsume, N. *Statistical Physics II: Nonequilibrium Statistical Mechanics*, 2nd ed.; Springer: New York, 1991.
- (36) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613.
- (37) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667.
- (38) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (39) Hestenes, M. R.; Stiefel, E. *J. Res. Natl. Bur. Stand. (U.S.)* **1952**, *49*, 409.
- (40) Fletcher, R.; Reeves, C. M. *Comput. J.* **1964**, *7*, 149.
- (41) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95.
- (42) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781.
- (43) Fischer, S.; Dunbrack, R. L.; Karplus, M. *J. Am. Chem. Soc.* **1994**, *116*, 11931.
- (44) Banerji, S.; Wright, A. J.; Noble, M.; Mahoney, D. J.; Campbell, I. D.; Day, A. J.; Jackson, D. G. *Nat. Struct. Mol. Biol.* **2007**, *14*, 234.