

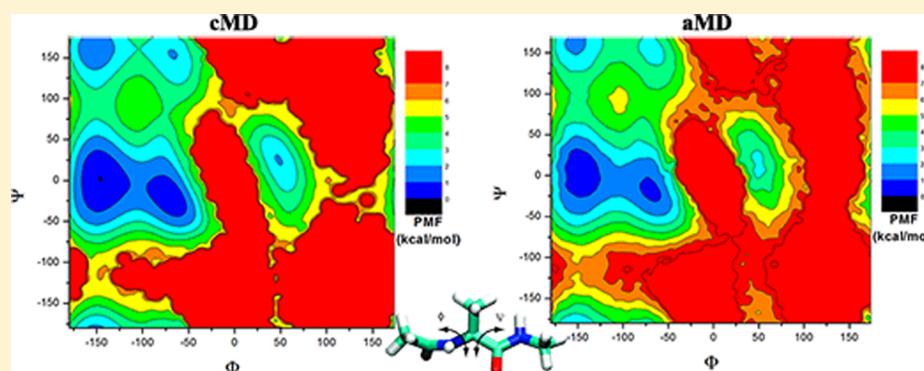
# Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation

Yinglong Miao,<sup>\*,†</sup> William Sinko,<sup>‡</sup> Levi Pierce,<sup>§</sup> Denis Bucher,<sup>§</sup> Ross C. Walker,<sup>§,||</sup> and J. Andrew McCammon<sup>†,‡,§</sup>

<sup>†</sup>Howard Hughes Medical Institute, <sup>‡</sup>Department of Pharmacology, <sup>§</sup>Department of Chemistry and Biochemistry and <sup>||</sup>San Diego Supercomputer Center, University of California at San Diego, La Jolla, California 92093, United States

<sup>§</sup>Schrodinger Inc., New York, New York 10036, United States

## Supporting Information



**ABSTRACT:** Accelerated molecular dynamics (aMD) simulations greatly improve the efficiency of conventional molecular dynamics (cMD) for sampling biomolecular conformations, but they require proper reweighting for free energy calculation. In this work, we systematically compare the accuracy of different reweighting algorithms including the exponential average, Maclaurin series, and cumulant expansion on three model systems: alanine dipeptide, chignolin, and Trp-cage. Exponential average reweighting can recover the original free energy profiles easily only when the distribution of the boost potential is narrow (e.g., the range  $\leq 20k_B T$ ) as found in dihedral-boost aMD simulation of alanine dipeptide. In dual-boost aMD simulations of the studied systems, exponential average generally leads to high energetic fluctuations, largely due to the fact that the Boltzmann reweighting factors are dominated by a very few high boost potential frames. In comparison, reweighting based on Maclaurin series expansion (equivalent to cumulant expansion on the first order) greatly suppresses the energetic noise but often gives incorrect energy minimum positions and significant errors at the energy barriers ( $\sim 2-3k_B T$ ). Finally, reweighting using cumulant expansion to the second order is able to recover the most accurate free energy profiles within statistical errors of  $\sim k_B T$ , particularly when the distribution of the boost potential exhibits low anharmonicity (i.e., near-Gaussian distribution), and should be of wide applicability. A toolkit of Python scripts for aMD reweighting “PyReweighting” is distributed free of charge at <http://mccammon.ucsd.edu/computing/AMDReweighting/>.

## INTRODUCTION

Accelerated molecular dynamics (aMD) is an enhanced sampling technique that works by flattening the molecular potential energy surface, often by adding a non-negative boost potential when the system potential is lower than a reference energy. The boost potential,  $\Delta V$  decreases the energy barriers and thus accelerates transitions between different low-energy states.<sup>1</sup> With this, aMD is able to sample distinct biomolecular conformations and rare barrier-crossing events that are not accessible to conventional molecular dynamics (cMD) simulations. AMD has been successfully applied to a number of biological systems<sup>2</sup> and hundreds-of-nanosecond aMD simulations have been shown to capture millisecond-time scale events in both globular and membrane proteins, such as

the bovine pancreatic trypsin inhibitor (BPTI)<sup>3</sup> and a G-protein coupled receptor (GPCR).<sup>4</sup>

In addition to enhanced conformational sampling, it is appealing to reweight aMD simulations to recover the original free energy landscapes of functional biomolecules. In theory, frames of aMD simulations can be reweighted by the Boltzmann factors of the corresponding boost potential (i.e.,  $e^{\Delta V/k_B T}$ ) and averaged over each bin of selected reaction coordinate(s) to obtain the canonical ensemble, an algorithm

**Special Issue:** Free Energy Calculations: Three Decades of Adventure in Chemistry and Biophysics

**Received:** February 5, 2014

**Published:** May 1, 2014



termed “exponential average” here. However, exponential reweighting is known to suffer from large statistical noise in practical calculations<sup>5</sup> because the Boltzmann reweighting factors are often dominated by a very few frames with high boost potential.

In an effort to avoid the energetic noise, scaled molecular dynamics (scaled MD) has been developed to flatten the biomolecular potential energy surface using a single scaling factor, enabling population-based reweighting and significantly improving the reweighted free energy profiles as demonstrated on alanine dipeptide and chignolin.<sup>5d</sup> While scaled MD provides a great advantage for simulation reweighting, it scales the system potential uniformly and does not allow selective acceleration of biomolecular conformational changes as does aMD, for example, applying boost potential to protein dihedrals only<sup>1a</sup> or subdomains of a protein.<sup>6</sup> Moreover, small numerical errors can also occur in the reweighting of scaled MD simulations and are the subject of an ongoing study.

AMD has also been combined with replica exchange (i.e., REXAMD) for optimal conformational sampling and free energy calculations of protein–ligand binding.<sup>7</sup> REXAMD provides an approach to calculate the free energy directly from the simulation replica with zero boost potential (i.e., unaccelerated).<sup>7c</sup> However, it has been shown that including statistics collected from accelerated runs can improve the final reweighted free energies.<sup>7b</sup> Moreover, REXAMD can be challenging in practice, as the optimal acceleration parameters for each replica are *a priori* not known. Thus, accurate reweighting of aMD simulations is still of key interest for free energy calculations. Such need also emerges in using aMD enhanced sampling simulations to build Markov state models.<sup>8</sup>

Importantly, aMD does not require any predefined reaction coordinate(s), unlike other biased free energy calculation methods, including the umbrella sampling,<sup>9</sup> metadynamics,<sup>10</sup> adaptive biasing force (ABF) calculations,<sup>11</sup> and orthogonal space sampling.<sup>12</sup> Thus, aMD can be advantageous for exploring the biomolecular conformational space without *a priori* knowledge or restraints and calculating the system free energy landscape through postsimulation analysis. However, the biasing potential applied in aMD simulations is on the order of tens to hundreds of kcal/mol, which is much greater in magnitude and wider in distribution than that of other biasing simulation methods that make use of predefined reaction coordinates (e.g., several kcal/mol). This leads to grand challenge for accurate reweighting of aMD simulations. A thorough analysis of the advantages and disadvantages of different algorithms for aMD reweighting is still lacking. Open questions include how to effectively suppress the energetic noise encountered in exponential reweighting, and what are the limiting parameters for accurate reweighting of aMD simulations in terms of system size and levels of acceleration. Note that the reweighting algorithms discussed in this study shall be applicable to all biasing simulation methods for recovering the original free energy profiles.

In previous aMD studies of proteins, Maclaurin series expansion was applied to approximate the exponential term  $e^{\Delta V/k_B T}$  and it was found less noisy than the exponential reweighting.<sup>3</sup> Furthermore, Shen and Hamelberg showed that cumulant expansion effectively smoothed the free energy profiles of alanine dipeptide obtained from aMD simulations.<sup>5a</sup> Cumulant expansion has also been widely used in free energy calculations of biomolecules in nonequilibrium simulations based on the Jarzynski’s equality,<sup>13</sup> for example, fast-growth

thermodynamic integration<sup>14</sup> and steered molecular dynamics.<sup>15</sup> Particularly, when the boost potential or its equivalent follows Gaussian distribution, the cumulant expansion to the second order is exact for calculating the average of exponentials. This has been known as “Gaussian fluctuation approximation” and applied to calculate electrostatic free energies of solutes in solution.<sup>16</sup>

Here, reweighting algorithms including the exponential average, Maclaurin series, and cumulant expansion are compared systematically for analyzing aMD simulations of three extensively studied biomolecular model systems, that is, alanine dipeptide, chignolin, and Trp-cage. The latter two are known to be fast-folding proteins.<sup>17</sup> The distributions of boost potential applied in the aMD simulations are analyzed in detail to address the limiting parameters for accurate aMD reweighting.

## METHODS

**Accelerated Molecular Dynamics.** Accelerated molecular dynamics (aMD) enhances the conformational sampling of biomolecules, often by adding a non-negative boost potential to the system when the system potential is lower than a reference energy:<sup>1,5b</sup>

$$\begin{aligned} V^*(r) &= V(r), & V(r) \geq E \\ V^*(r) &= V(r) + \Delta V(r), & V(r) < E \end{aligned} \quad (1)$$

where  $V(r)$  is the original potential,  $E$  is the reference energy, and  $V^*(r)$  is the modified potential. In the simplest form, the boost potential,  $\Delta V(r)$  is given by

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + E - V(r)} \quad (2)$$

where  $\alpha$  is the acceleration factor. As the acceleration factor  $\alpha$  decreases, the energy surface is flattened more and biomolecular transitions between the low-energy states are increased.

Two versions of aMD that provide different acceleration levels of biomolecules are termed “dihedral-boost”<sup>1a</sup> and “dual-boost”.<sup>1b</sup> In dihedral-boost aMD, boost potential is applied to all dihedrals in the system with input parameters ( $E_{\text{dihed}}$ ,  $\alpha_{\text{dihed}}$ ). In dual-boost aMD, a total boost potential is applied to all atoms in the system in addition to the dihedral boost (i.e.,  $E_{\text{dihed}}$ ,  $\alpha_{\text{dihed}}$ ;  $E_{\text{total}}$ ,  $\alpha_{\text{total}}$ ). Previous applications have suggested the following parameters for proper acceleration of conformational changes in globular proteins<sup>3</sup> and they are used in this study of alanine dipeptide, chignolin and Trp-cage

$$\begin{aligned} E_{\text{dihed}} &= V_{\text{dihed\_avg}} + 3.5N_{\text{res}}, & \alpha_{\text{dihed}} &= 3.5N_{\text{res}}/5 \\ E_{\text{total}} &= V_{\text{total\_avg}} + 0.175N_{\text{atoms}}, & \alpha_{\text{total}} &= 0.175N_{\text{atoms}} \end{aligned} \quad (3)$$

where  $N_{\text{res}}$  is number of protein residues,  $N_{\text{atoms}}$  is the total number of atoms, and  $V_{\text{dihed\_avg}}$  and  $V_{\text{total\_avg}}$  are the average dihedral and total potential energies calculated from short cMD simulations, respectively.

**Energetic Reweighting.** For an aMD simulation of a  $N$ -atom biomolecular system, the probability distribution along a selected reaction coordinate(s)  $A(r)$  is written as  $p^*(A)$ , where  $r$  denotes the atomic positions  $\{r_1, \dots, r_N\}$ . Given the boost potential  $\Delta V(r)$  of each frame,  $p^*(A)$  can be reweighted to recover the canonical ensemble distribution,  $p(A)$ , as

**Table 1.** Biomolecular Model Systems Simulated with Accelerated Molecular Dynamics (aMD) in the Present Study<sup>a</sup>

system	$N_{\text{res}}$	$N_{\text{atoms}}$	cMD	aMD	dihedral/dual-boost	$\Delta V_{\text{avg}}$ (kcal/mol)	$\sigma_{\Delta V}$ (kcal/mol)	$\delta_{\Delta V}$ (kcal/mol)
alanine dipeptide	2	1912	1000 ns	20 ns $\times$ 5	dihedral-boost	3.1	1.8	11.7
				20 ns $\times$ 5, 1000 ns	dual-boost	4.2	2.1	18.5
chignolin	10	6773	1000 ns $\times$ 2	300 ns $\times$ 3	dual-boost	8.5	3.2	31.9
Trp-cage	20	34 370	1000 ns $\times$ 2	500 ns $\times$ 4	dual-boost	38.8	7.0	70.8

<sup>a</sup> $N_{\text{res}}$  and  $N_{\text{atoms}}$  are the number of residues and system atoms, respectively.  $\Delta V_{\text{avg}}$ ,  $\sigma_{\Delta V}$ , and  $\delta_{\Delta V}$  are the average, standard deviation, and range (maximum–minimum) of the aMD boost potential,  $\Delta V$ .

$$p(A_j) = p^*(A_j) \frac{\langle e^{\beta \Delta V(r)} \rangle_j}{\sum_{j=1}^M \langle e^{\beta \Delta V(r)} \rangle_j}, \quad j = 1, \dots, M \quad (4)$$

where  $M$  is the number of bins,  $\beta = 1/k_B T$  and  $\langle e^{\beta \Delta V(r)} \rangle_j$  is the ensemble-averaged Boltzmann factor of  $\Delta V(r)$  for simulation frames found in the  $j^{\text{th}}$  bin. The above equation provides an “exponential average” algorithm for reweighting of aMD simulations. As the Boltzmann reweighting factors are often dominated by high boost potential frames, the distribution of  $\Delta V$  is characterized by its standard deviation  $\sigma_{\Delta V}$  and the range  $\delta_{\Delta V} = \Delta V_{\text{max}} - \Delta V_{\text{min}}$ . The reweighted potential of mean force (PMF) is calculated as  $F(A_j) = -(1/\beta) \ln p(A_j)$ .

Furthermore, the exponential term can be approximated by summation of the Maclaurin series of boost potential  $\Delta V(r)$  with the reweighting factor rewritten as

$$\langle e^{\beta \Delta V} \rangle = \sum_{k=0}^{\infty} \frac{\beta^k}{k!} \langle \Delta V^k \rangle \quad (5)$$

where the subscript  $j$  has been suppressed. The Maclaurin series expansion up to the 5th–10th order has been used in practice to reweight aMD trajectories.<sup>3</sup> The reweighted PMF profiles are typically less noisy than those obtained from exponential average reweighting.

Finally, the ensemble-averaged reweighting factor can also be approximated using cumulant expansion:<sup>14,18</sup>

$$\langle e^{\beta \Delta V} \rangle = \exp \left\{ \sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k \right\} \quad (6)$$

where the first three cumulants are given by

$$\begin{aligned} C_1 &= \langle \Delta V \rangle \\ C_2 &= \langle \Delta V^2 \rangle - \langle \Delta V \rangle^2 = \sigma_{\Delta V}^2 \\ C_3 &= \langle \Delta V^3 \rangle - 3\langle \Delta V^2 \rangle \langle \Delta V \rangle + 2\langle \Delta V \rangle^2 \end{aligned} \quad (7)$$

The free energy can then be derived from cumulant expansion as

$$F(A_j) = F^*(A_j) - \frac{1}{\beta} \sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k + F_c \quad (8)$$

where  $F^*(A_j)$  is the modified free energy surface sampled in the aMD simulation and the constant  $F_c = (1/\beta) \ln \sum_{j=1}^M \langle e^{\beta \Delta V(r)} \rangle_j$ .

In principle, reweighting using cumulant expansion is able to greatly suppress the energetic noise, since it collectively reweights many data points along the chosen reaction coordinate(s). Note that the Maclaurin series expansion is equivalent to cumulant expansion on the first order:

$$\langle e^{\beta \Delta V} \rangle = \sum_{k=0}^{\infty} \frac{\beta^k}{k!} \langle \Delta V^k \rangle = e^{\beta \langle \Delta V \rangle} \quad (9)$$

When the distribution of boost potential  $\Delta V$  within a bin satisfies the Gaussian distribution, cumulant expansion to the second order provides an accurate approximation of the exponential reweighting factor. In this study, we introduce a quantity called “anharmonicity”,<sup>19</sup>  $\gamma$  to characterize the distribution of aMD boost potential as follows. Given  $\Delta V$  for frames found in one bin of the selected reaction coordinate(s), the normalized probability distribution can be obtained  $p(\Delta V)$  and its entropy is calculated as

$$S_{\Delta V} = - \int_0^\infty p(\Delta V) \ln(p(\Delta V)) d\Delta V \quad (10)$$

Assuming  $\Delta V$  satisfies exactly Gaussian distribution, for which the boost potential achieves perfect sampling, its maximum entropy is given by

$$S_{\text{max}} = \frac{1}{2} \ln(2\pi e \sigma^2) \quad (11)$$

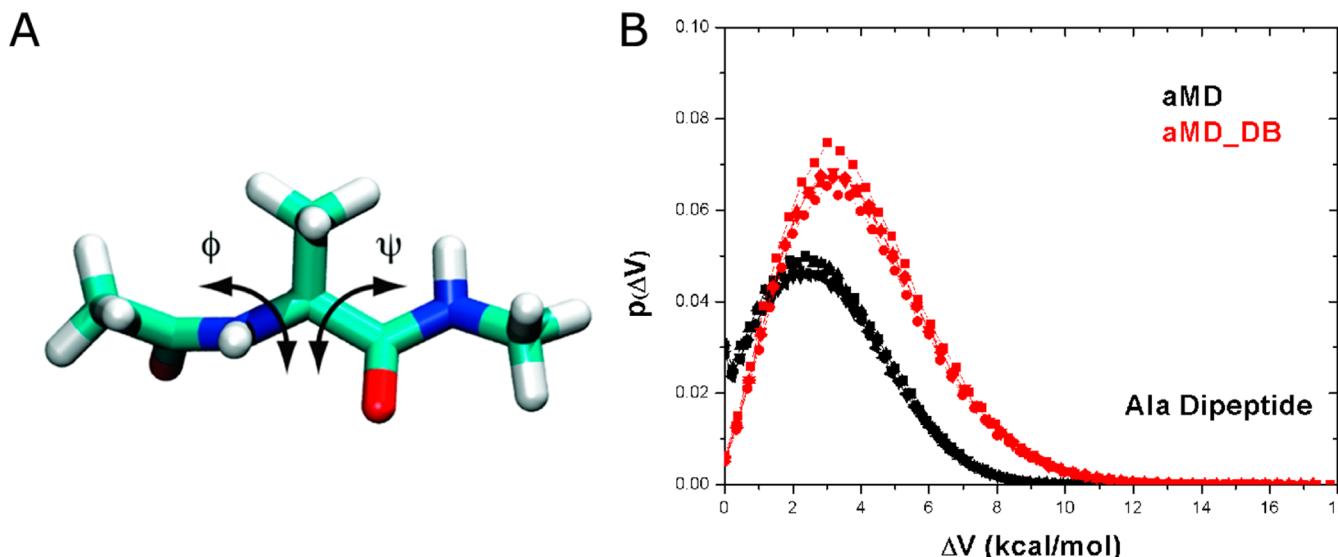
where  $\sigma$  is the standard deviation of the  $\Delta V$  distribution. Then, the distribution anharmonicity,  $\gamma$  of boost potential  $\Delta V$  is obtained as

$$\begin{aligned} \gamma &= S_{\text{max}} - S_{\Delta V} = \frac{1}{2} \ln(2\pi e \sigma^2) \\ &+ \int_0^\infty p(\Delta V) \ln(p(\Delta V)) d\Delta V \end{aligned} \quad (12)$$

When  $\gamma$  is zero, the boost potential follows ideal Gaussian distribution and cumulant expansion to the second order is expected to provide accurate reweighted free energy profiles. As  $\gamma$  increases, the  $\Delta V$  distribution becomes less harmonic and the reweighted PMF obtained from cumulant expansion to the second order deviates from the original free energy profile. Thus,  $\gamma$  serves as an effective indicator of aMD reweighting accuracy using cumulant expansion to the second order.

**Simulation Protocols.** All cMD and aMD simulations of alanine dipeptide, chignolin, and Trp-cage were performed using AMBER with the ff99SB force field on GPUs<sup>20</sup> using the SPFP precision model.<sup>21</sup> The simulated systems were built using the Xleap module of the AMBER package. Alanine dipeptide and chignolin were constructed as described previously.<sup>5d</sup> For Trp-cage, the amino acid sequence was obtained from the PDB code 2JOF<sup>22</sup> and an extended polypeptide was built as the starting structure. By solvating the structures in a TIP3P<sup>23</sup> water box that extends 8 to 10 Å from the solute surface, the alanine dipeptide system contained 630 water molecules, 2211 waters for chignolin, and 11 355 waters for Trp-cage. The total number of atoms in the three systems are 1912, 6773, and 34 370 for alanine dipeptide, chignolin, and Trp-cage, respectively (Table 1).

In AMBER simulations of the three systems, bonds containing hydrogen atoms were restrained with the SHAKE algorithm<sup>24</sup> and thus a 2 fs time step was used. Weak coupling to an external temperature and pressure bath was used to



**Figure 1.** (A) Scheme representation of backbone dihedrals  $\Phi$  and  $\Psi$  in alanine dipeptide. (B) The probability distribution of boost potential  $\Delta V$  obtained from five independent 20 ns dihedral-boost aMD (aMD) and dual-boost aMD (aMD\_DB) simulations.

control both temperature and pressure.<sup>25</sup> The electrostatic interactions were calculated using the PME (particle mesh Ewald summation)<sup>26</sup> with a cutoff of 8.0 Å for long-range interactions.

The three systems were initially minimized for 2000 steps using the conjugate gradient minimization algorithm and then the solvent was equilibrated for 50 ps in isothermal–isobaric (NPT) ensemble with the solute atoms fixed. Another minimization was performed with all atoms free and the systems were slowly heated to 300 K over 500 ps. Final system equilibration was achieved by a 200 ps isothermal–isovolumetric (NVT) and 400 ps NPT run to ensure that the water box of simulated systems had reached the appropriate density. Then, production simulations were performed in the NVT ensemble. The aMD acceleration parameters were calculated using the first 100 ns cMD simulation of each system.

For alanine dipeptide, one 1000 ns cMD simulation and five independent 20 ns dihedral-boost and dual-boost aMD simulations were obtained for calculating free energies. One of the five dual-boost aMD simulations was also extended to 1000 ns as a converged reference. For chignolin, two independent 1000 ns cMD simulations and three independent 300 ns dual-boost aMD simulations were compared. For Trp-cage, two independent 1000 ns cMD simulations and four independent 500 ns dual-boost aMD simulations were obtained for comparison. The simulation frames were saved every 0.2 ps for trajectory analysis. A summary of the simulations is listed in Table 1.

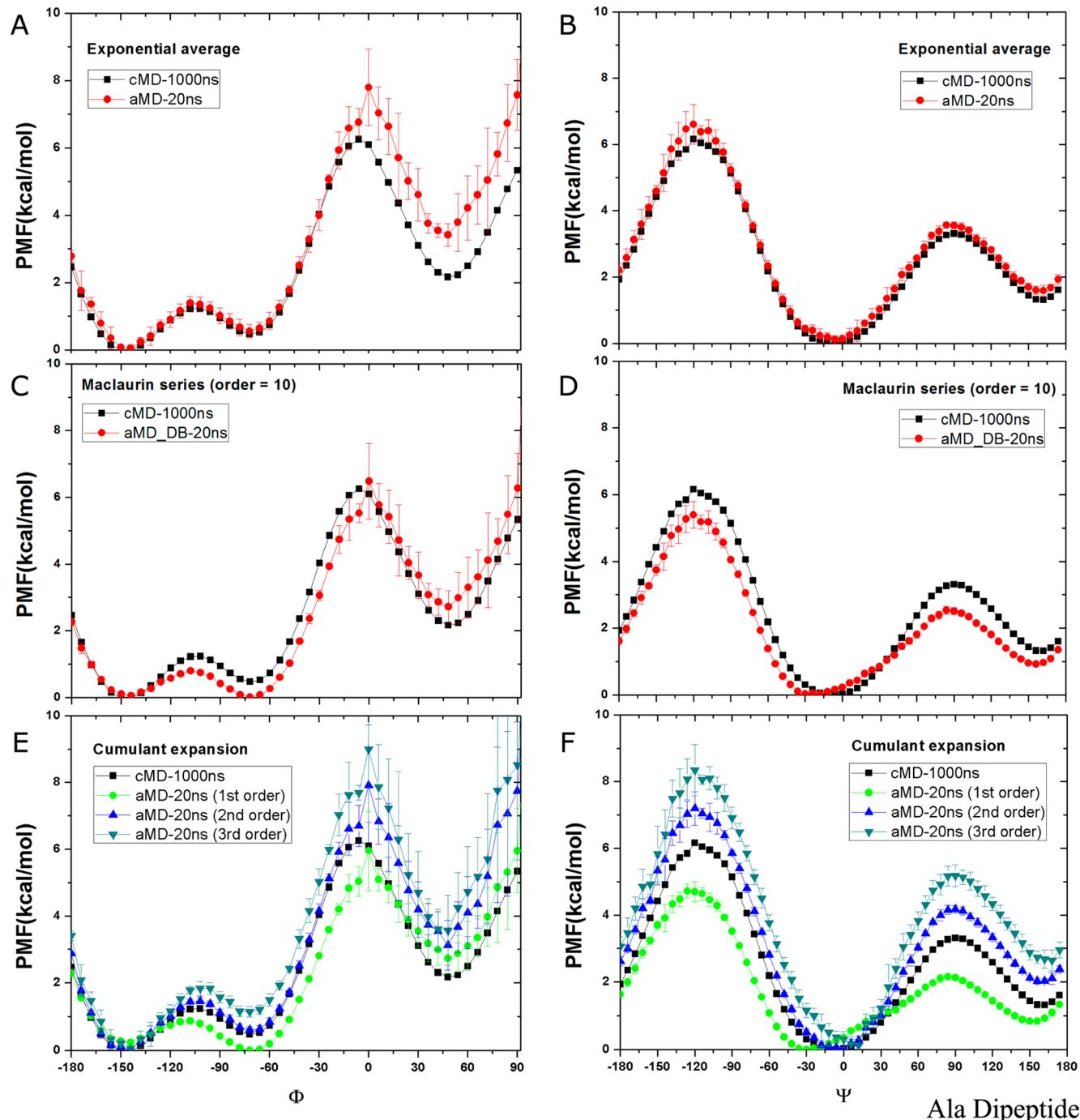
PMF profiles were calculated for backbone dihedrals  $\Phi$  and  $\Psi$  in alanine dipeptide,  $\Phi_4$  and  $\Psi_4$  for Pro4 in chignolin, and  $\Phi_{12}$  and  $\Psi_{12}$  for Pro12 in Trp-cage. Two-dimensional PMF profiles of  $(\Phi, \Psi)$  in alanine dipeptide were also analyzed. In case of multiple independent cMD/aMD simulations, PMF profiles obtained by averaging the independent runs were used. Calculations using five different bin sizes 3, 6, 9, 12, and 15 for backbone dihedrals in alanine dipeptide (Supporting Information Figure S1) showed that when the bin size is 3, the number of sampling points can be smaller than 10 with no converged free energy values for several bins in the energy barrier region at  $\Phi = 0^\circ$  (see Supporting Information Figure S1A). As the bin

size is increased to 6 or above, the free energy profiles exhibit slight variations. Furthermore, with increasing bin size the anharmonicity is found to decrease monotonically (Supporting Information Figures S1C and S1D), because of the fact that more data points with similar distribution width are collected in each bin. Therefore, a bin size of 6 is selected to balance between reducing the anharmonicity and increasing the bin resolution. When the number of simulation frames within a bin is lower than a certain limit (i.e., cutoff), the bin is not sufficiently sampled and thus excluded for reweighting. The cutoff can be determined by iteratively increasing it until the minimum position of the PMF profile does not change. The final cutoff was set as 10, 50, and 500 for reweighting of aMD simulations on alanine dipeptide, chignolin, and Trp-cage, respectively.

For Chignolin and Trp-cage, the root-mean-square deviation (RMSD) of the protein  $C_\alpha$  atoms relative to the experimentally determined folded structure were calculated from each cMD simulation and plotted in Supporting Information Figure S6. Folding of the two proteins were observed in all the presented simulations with RMSD decreased to  $\sim 2$  Å. This justifies that conformations of the two fast-folding proteins are well sampled. Furthermore, averaging of the two independent cMD runs for each system is presented. The error bars provide an estimate of the simulation accuracy. The RMSD plots for aMD simulations of both Chignolin and Trp-cage are obtained as well in the Supporting Information Figure S7, depicting more frequent folding of the two fast-folding proteins. AMBER trajectories were analyzed using the *ptraj* tool. A toolkit of Python scripts “*PyReweighting*” was used to reweight the aMD simulations to calculate both 1D and 2D free energy profiles.

## RESULTS AND DISCUSSION

Using exponential average, Maclaurin series and cumulant expansion, reweighting of aMD simulations yields PMF profiles of backbone dihedrals  $\Phi$  and  $\Psi$  in alanine dipeptide,  $\Phi_4$  and  $\Psi_4$  for Pro4 in chignolin, and  $\Phi_{12}$  and  $\Psi_{12}$  for Pro12 in Trp-cage. The two proline residues in chignolin and Trp-cage are located in the middle region of the protein chain and conformational changes in their backbone dihedrals have been suggested to



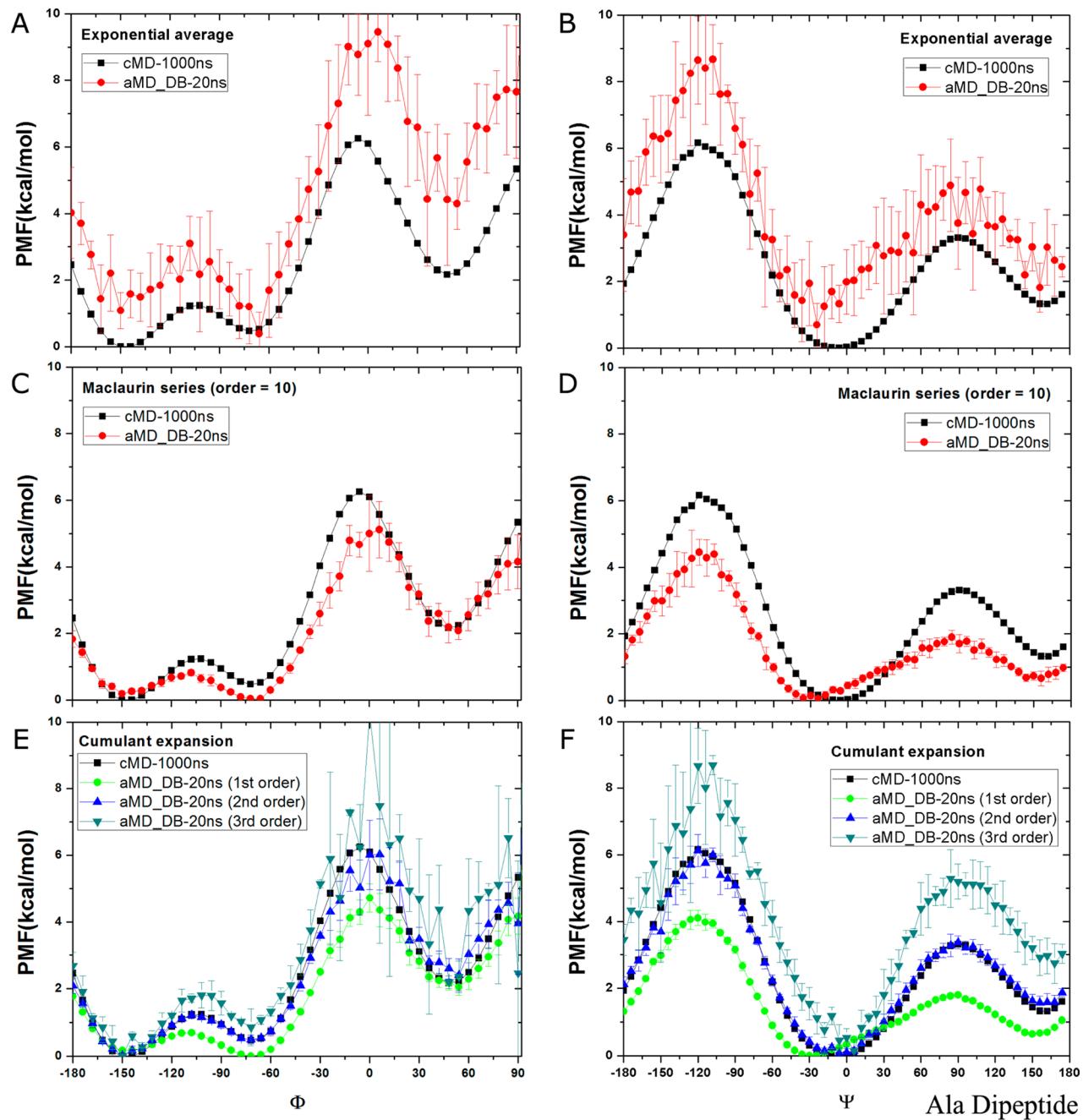
**Figure 2.** Free energy profiles of backbone dihedrals  $\Phi$  and  $\Psi$  in alanine dipeptide calculated by reweighting five independent 20 ns dihedral-boost aMD simulations using (A, B) exponential average, (C, D) Maclaurin series expansion to the 10th order, (E, F) cumulant expansion to the 1st, 2nd, and 3rd orders. The free energy profiles obtained from a 1000 ns cMD simulation are also plotted for comparison.

play a key role in the protein folding.<sup>17</sup> The reweighted PMF profiles of aMD simulations are compared with those of long-time scale cMD simulations to examine accuracy of the three different aMD reweighting algorithms.

**Alanine Dipeptide.** The PMF profiles of backbone dihedrals  $\Phi$  and  $\Psi$  in alanine dipeptide (Figure 1A) obtained by reweighting five independent 20 ns dihedral-boost and dual-boost aMD simulations are compared with those of a 1000 ns cMD simulation. In the dihedral-boost aMD simulations, the boost potential,  $\Delta V$  has an average of 3.1 kcal/mol, which is lower than that of the dual-boost aMD (4.2 kcal/mol).

Furthermore, the normalized distribution of  $\Delta V$  is not symmetric with the left wing truncated at 0 kcal/mol and right wing stretched to 11.7 kcal/mol (Figure 1B). Due to shift of  $\Delta V$  to higher values, the left wing of the  $\Delta V$  distribution is less truncated in dual-boost aMD simulations than in the dihedral-boost aMD. The standard deviation of  $\Delta V$  in dual-boost aMD is 2.1 kcal/mol, which is slightly higher than that of dihedral-boost aMD (1.8 kcal/mol).

In the dihedral-boost aMD simulations, the distribution of  $\Delta V$  is narrow with a range  $\delta_{\Delta V} \leq 12$  kcal/mol ( $20k_B T$ ). Exponential average reweighting is able to recover the original

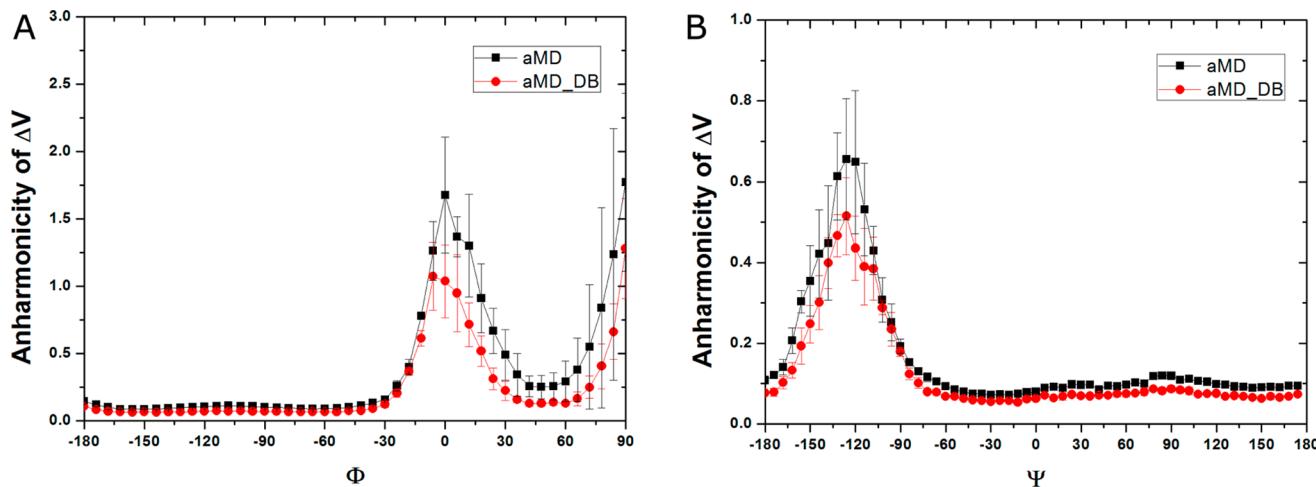


**Figure 3.** Free energy profiles of backbone dihedrals  $\Phi$  and  $\Psi$  in alanine dipeptide calculated by reweighting five independent 20 ns dual-boost aMD simulations using (A, B) exponential average, (C, D) Maclaurin series expansion to the 10th order, (E, F) cumulant expansion to the 1st, 2nd, and 3rd orders. The profiles are plotted similarly as in Figure 2.

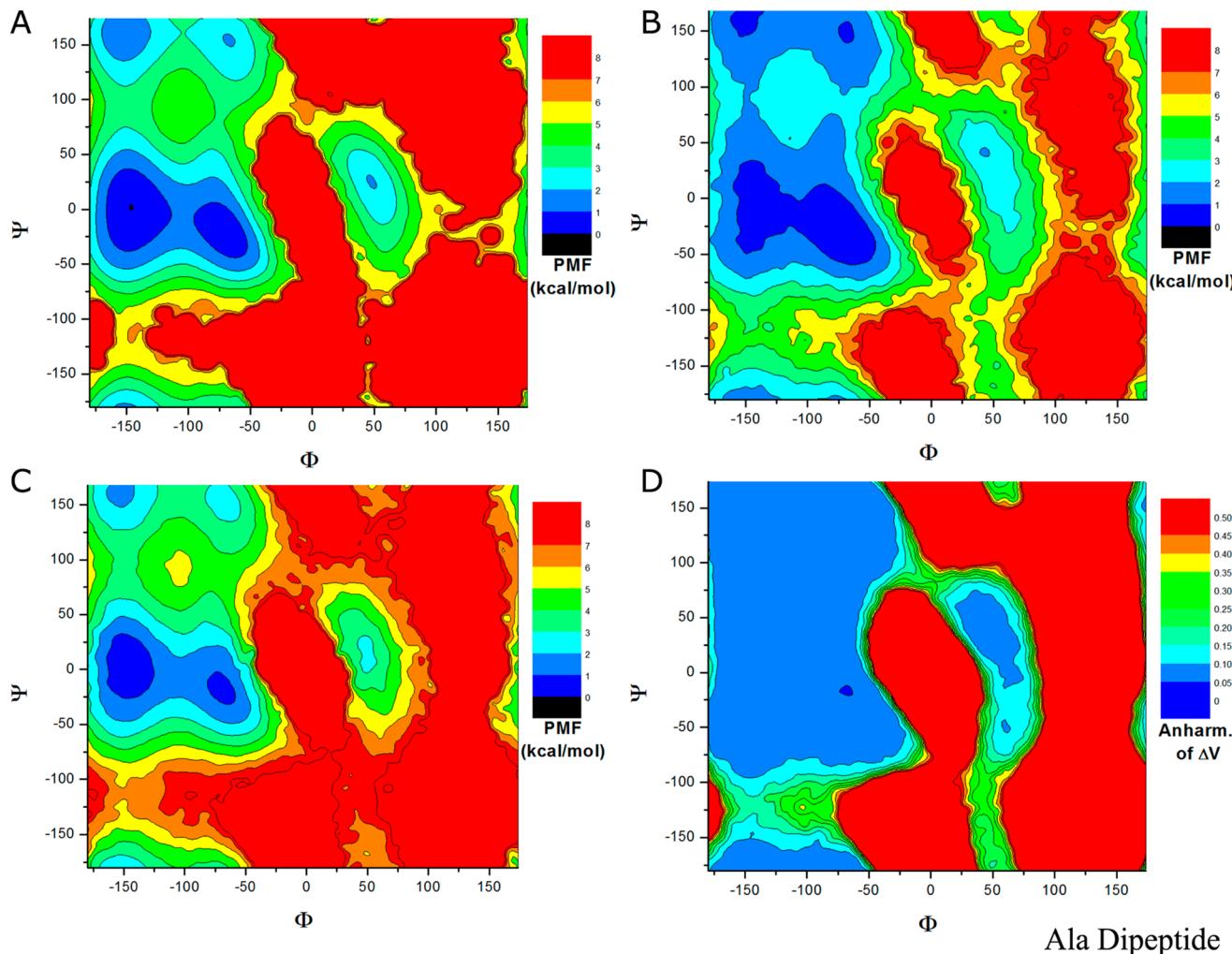
free energy profiles of  $\Phi$  (Figure 2A) and  $\Psi$  (Figure 2B) as obtained from 1000 ns cMD simulation, except that one local energy well of  $\Phi$  centered at  $48^\circ$  is increased by  $\sim 1$  kcal/mol, suggesting insufficient sampling of this region with dihedral-boost aMD. Using Maclaurin series expansion to the 10th order, the reweighted PMF profiles differ most in  $\Psi$  with position of the energy minimum shifted from  $0^\circ$  to  $24^\circ$  and the energy barriers lowered by  $\sim 1$  kcal/mol (Figure 2D), as well as in  $\Phi$  with the  $-72^\circ$  energy minimum lowered by  $\sim 0.5$  kcal/mol (Figure 2C). With cumulant expansion, reweighting on the first order that is equivalent to Maclaurin series expansion produces lower PMF with incorrect minimum positions, while expansion up to the third order overestimates the PMF values relative to

the energy minimum in both  $\Phi$  (Figure 2E) and  $\Psi$  (Figure 2F). In comparison, cumulant expansion to the second order provides the closest PMF in  $\Phi$  with two energy wells centered at  $-150^\circ$  and  $-72^\circ$  accurately reproduced, but PMF of the  $48^\circ$  local energy well exhibits significant increase similar to that of the exponential average reweighting (Figure 2E). Moreover, the second-order cumulant expansion overestimates energy barriers of  $\Psi$  by  $\sim 1$  kcal/mol at both  $-120^\circ$  and  $90^\circ$  as shown in Figure 2F.

In dual-boost aMD simulations, exponential average reweighting leads to significantly overestimated PMF with incorrect minimum positions and high fluctuations in both  $\Phi$  and  $\Psi$  (Figures 3A and B). This originates from the fact that



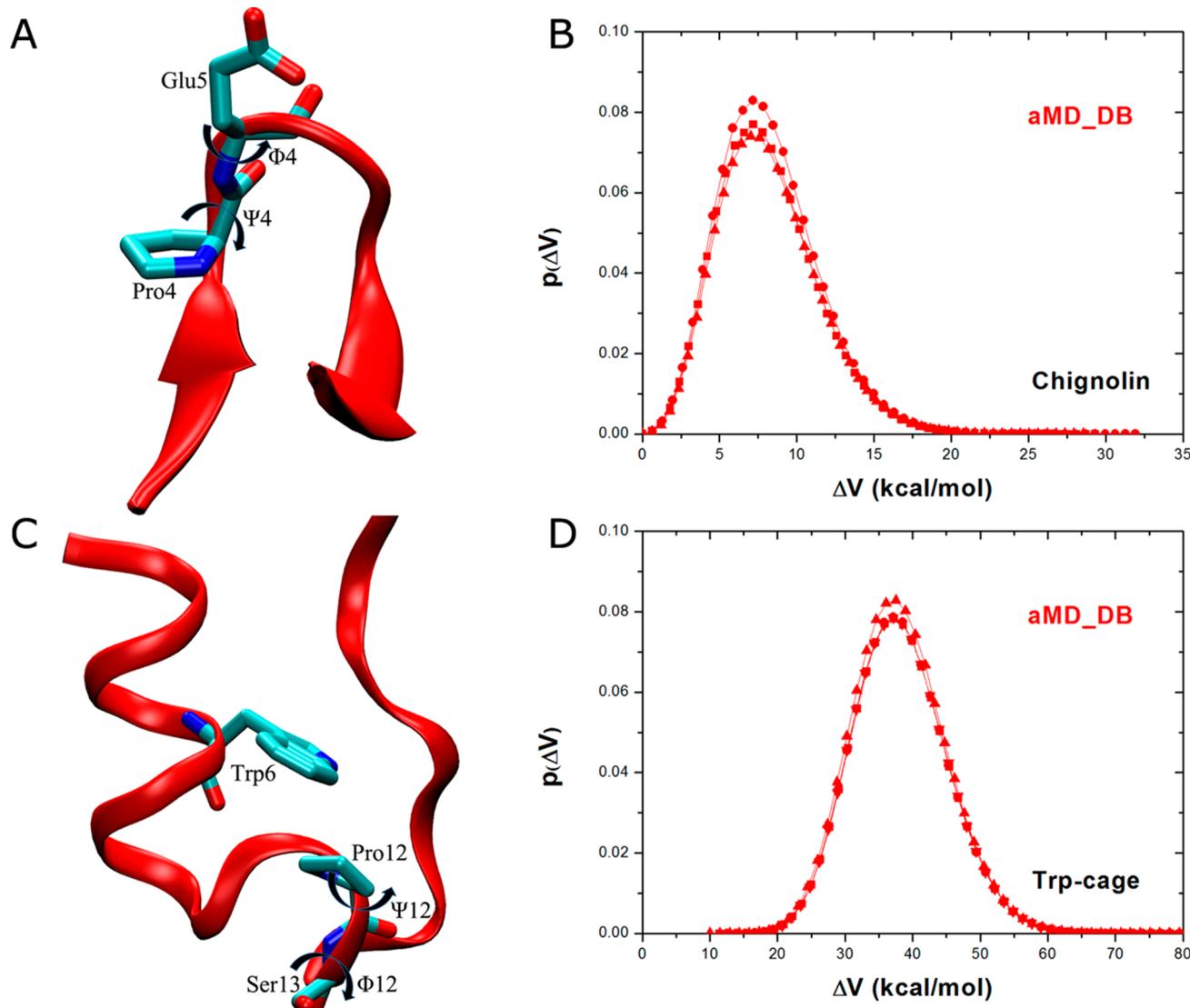
**Figure 4.** Distribution anharmonicity of boost potential  $\Delta V$  of frames found in each bin of (A)  $\Phi$  and (B)  $\Psi$  in the dihedral-boost aMD (aMD) and dual-boost aMD (aMD\_DB) simulations of alanine dipeptide.



**Figure 5.** Two-dimensional free energy profiles of backbone dihedrals ( $\Phi$ ,  $\Psi$ ) in alanine dipeptide calculated from (A) 1000 ns cMD simulation and reweighting of 1000 ns dual-boost aMD simulation using (B) Maclaurin series to the 10th order and (C) cumulant expansion to the 2nd order. (D) The distribution anharmonicity of  $\Delta V$  of frames found in each bin of ( $\Phi$ ,  $\Psi$ ) in the 1000 ns dual-boost aMD simulation.

the Boltzmann reweighting factors are dominated by a few high boost potential frames ( $\Delta V - \Delta V_{\min} = \sim 12-18$  kcal/mol as shown in Figure 1B). Using Maclaurin series expansion to the 10th order (cumulant expansion on the first order), the

reweighted PMF profiles are less noisy, but they do not reproduce the accurate energy minimum values in  $\Phi$  (Figure 3C) or the original minimum position in  $\Psi$  (Figure 3D), and the energy barriers are underestimated by  $\sim 1-2$  kcal/mol. In



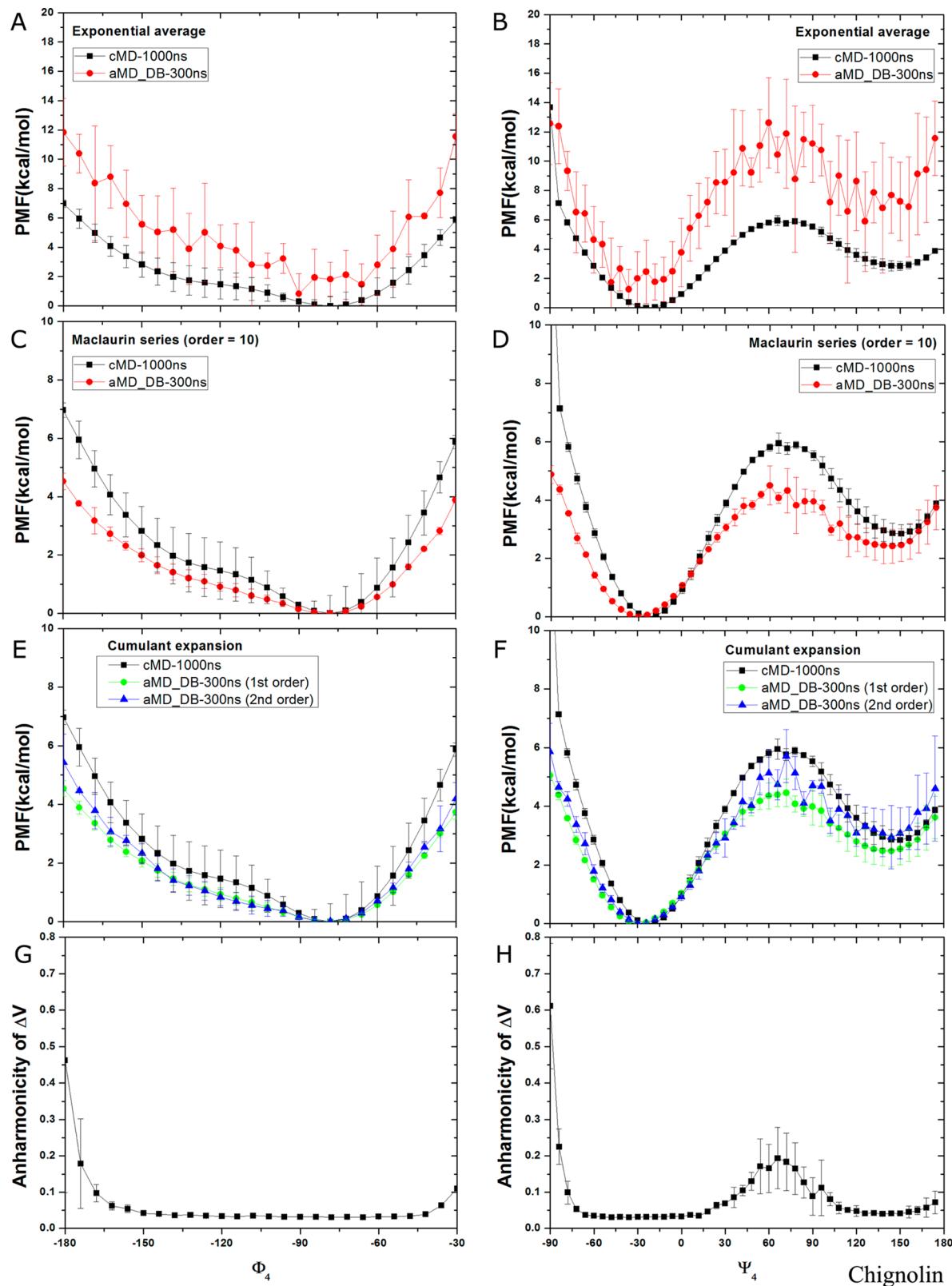
**Figure 6.** Chignolin: (A) scheme representation of backbone dihedrals  $\Phi_4$  and  $\Psi_4$  of the only proline residue (Pro4) and (B) the probability distribution of boost potential  $\Delta V$  obtained from three independent 200 ns dual-boost aMD simulations. Trp-cage: (C) scheme representation of backbone dihedrals  $\Phi_{12}$  and  $\Psi_{12}$  of Pro12 that is located in the middle of the protein chain and (D) the probability distribution of boost potential  $\Delta V$  from four independent 500 ns dual-boost aMD simulations.

contrast, while cumulant expansion to the third order still leads to deviated PMF profiles, expansion to the second order recovers the most accurate PMF profiles with negligible fluctuations that agree excellently with the original obtained from 1000 ns cMD simulation (Figures 3E and F). Reweighting of a converged 1000 ns dual-boost aMD simulation leads to closely similar results, as shown in Supporting Information Figure S3.

Next, we analyze the distribution anharmonicity,  $\gamma$  of  $\Delta V$  for aMD trajectory frames found in each 6°-bin of  $\Phi$  and  $\Psi$  (see details in Methods). In the region of 0°–90° for  $\Phi$ , the anharmonicity  $\gamma$  of the dihedral-boost aMD simulations is significantly higher than that of the dual-boost aMD (Figure 4A). This justifies the finding that less accurate PMF profiles are obtained from dihedral-boost aMD simulations using cumulant expansion to the second order than from the dual-boost aMD. Moreover, in the two regions of  $\Psi$  energy barriers at  $-120^\circ$  and  $90^\circ$  (especially the central barrier at  $-120^\circ$ ), dihedral-boost aMD exhibits marked higher anharmonicity

(Figure 4B), which appears as an indicator of incorrectly reweighted PMF in these regions as shown in Figure 2F. Overall,  $\gamma$  is close to zero in most regions of  $\Phi$  and  $\Psi$  except the energy barriers in the dual-boost aMD simulations.  $\Delta V$  satisfies near-Gaussian distribution in these regions, for which sufficient conformational sampling is achieved and highly accurate PMF is recovered through cumulant expansion to the second order.

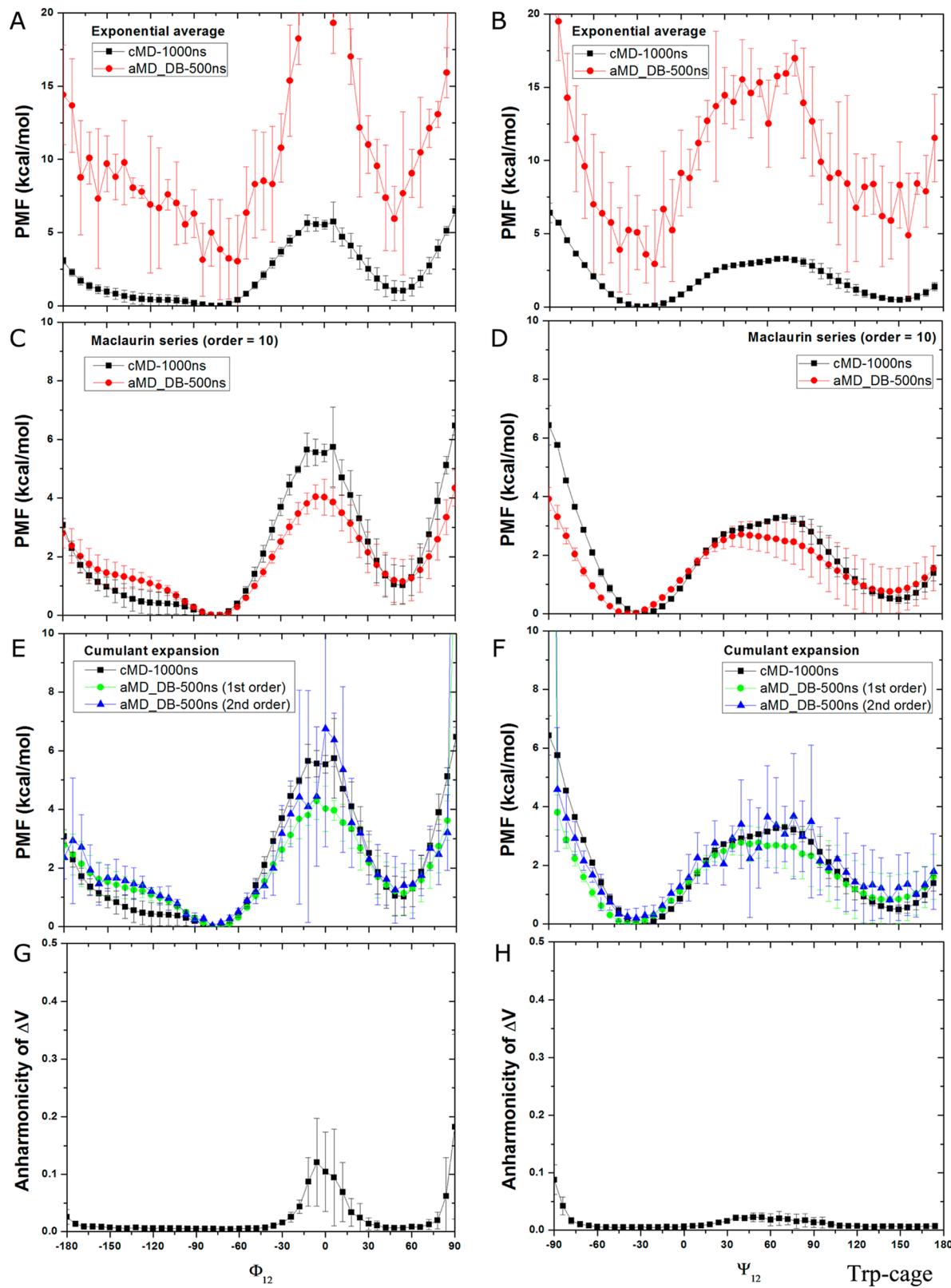
The two-dimensional PMF profiles of  $(\Phi, \Psi)$  in alanine dipeptide are presented in Figure 5. From the 1000 ns cMD simulation, the global energy minimum is identified at  $(-150^\circ, 0^\circ)$  and another four local minima at  $(-78^\circ, -6^\circ)$ ,  $(-150^\circ, 156^\circ)$ ,  $(-66^\circ, 150^\circ)$ , and  $(48^\circ, 30^\circ)$  (see Figure 5A). In comparison, the exponential average reweighting produces highly fluctuating PMF with large magnitudes of 20–40 kcal/mol in the 20 ns dual-boost aMD simulation (Supporting Information Figure S4A), even in the converged 1000 ns aMD simulation (Supporting Information Figure S4B). For reweighted PMF profiles using Maclaurin series and cumulant



**Figure 7.** Free energy profiles of backbone dihedrals  $\Phi_4$  and  $\Psi_4$  in chignolin calculated from reweighting three independent 300 ns dual-boost aMD simulations using (A, B) exponential average, (C, D) Maclaurin series expansion to the 10th order, (E, F) cumulant expansion to the 1st and 2nd orders. The free energy profiles obtained from two independent 1000 ns cMD simulation are also plotted for comparison. The distribution anharmonicity of  $\Delta V$  in the aMD simulations are plotted in (G)  $\Phi_4$  and (H)  $\Psi_4$ .

expansion, because the 20 ns dual-boost aMD simulation appears to lack converged sampling, as shown in Supporting Information Figure S5, results obtained from the 1000 ns aMD

simulation are presented in Figure 5. With Maclaurin series expansion to the 10th order, the reweighted PMF profile agrees reasonably well with the original in the overall shape, but it



**Figure 8.** Free energy profiles of backbone dihedrals  $\Phi_{12}$  and  $\Psi_{12}$  in Trp-cage calculated by reweighting four independent 500 ns dual-boost aMD simulations: (A, B) exponential average, (C, D) Maclaurin series expansion to the 10th order, (E, F) cumulant expansion to the 1st and 2nd orders. The free energy profiles obtained from two independent 1000 ns cMD simulation are also plotted for comparison. The distribution anharmonicity of  $\Delta V$  in the aMD simulations are plotted in (G)  $\Phi_{12}$  and (H)  $\Psi_{12}$ .

differs most in the relative PMF values and the global minimum is incorrectly predicted at  $(-72^\circ, -60^\circ)$  as shown in Figure S5B. The calculated root-mean-square error (RMSE) is 1.24 kcal/mol.

In contrast, using cumulant expansion to the second order, the global minimum is correctly reproduced at  $(-150^\circ, 0^\circ)$ , similarly for the other four local energy minima (Figure S5C).

The reweighted PMF profile is closely similar to the original with RMSE of 0.68 kcal/mol. Figure 5D plots the anharmonicity of  $\Delta V$  distribution in each bin of  $(\Phi, \Psi)$  in the 1000 ns dual-boost aMD simulations. Low anharmonicity values (typically  $<0.1$ ) are found in the sampled conformational space except the energy barrier regions, which again justifies the accurate aMD reweighting using cumulant expansion to the second order.

In summary, exponential average reweighting is accurate only when distribution of the boost potential is narrow with the range  $\delta_{\Delta V} \leq 20k_B T$  as found in dihedral-boost aMD simulation of alanine dipeptide. This is consistent with many previous aMD studies.<sup>1a,5a,27</sup> However, conformational changes of biomolecules can be limited by slow diffusion of the surrounding solvent molecules. Thus, dual-boost aMD has been developed to accelerate the dihedral transitions and slow diffusive motions concurrently.<sup>1b</sup> While dual-boost aMD provides acceleration at a higher level and greatly improves the conformational sampling,<sup>2b,4a</sup> boost potential of much wider distribution is applied in the simulations (see Figure 1B and Table 1). In this case, the exponential reweighting generally leads to high energetic noise because the Boltzmann reweighting factors are dominated by a few high boost potential frames. Although the Maclaurin series expansion greatly helps suppressing the energetic noise with the distribution range of the reweighting factors reduced by  $\sim 5$  orders of magnitude (see Supporting Information Figure S2A), it often misses reproducing the correct energy minimum positions and the PMF errors can be as large as  $\sim 2-3k_B T$ . The large noise of aMD reweighting factors can be suppressed using cumulant expansion as well (Supporting Information Figure S2B). When distribution of the boost potential exhibits low anharmonicity, reweighting using cumulant expansion to the second order is able to recover highly accurate free energy profiles within the statistical errors of  $\sim k_B T$ . Using cumulant expansion to the third order, the reweighted PMF profiles are rather deviated from the original even in the converged 1000 ns dual-boost aMD simulation (Supporting Information Figures S3E, F), largely due to propagation of small errors in the estimates of the third order cumulants.

**Chignolin and Trp-cage.** Apart from the alanine dipeptide, reweighting of aMD simulations is further analyzed on two fast-folding proteins (chignolin and Trp-cage) as compared with 1000 ns reference cMD simulations (see Methods). Particularly, free energy profiles are calculated for backbone dihedrals of Pro4 in chignolin ( $\Phi_4$  and  $\Psi_4$  in Figure 6A) and Pro12 in Trp-cage ( $\Phi_{12}$  and  $\Psi_{12}$  in Figure 6C). With increasing system size, higher boost potential  $\Delta V$  is applied in aMD simulations of chignolin and Trp-cage than in the alanine dipeptide simulations (Table 1). For chignolin, the  $\Delta V$  has an average of 8.5 kcal/mol with 3.2 kcal/mol standard deviation, exhibiting a wide distribution from 0 to 31.9 kcal/mol (Figure 6B). For Trp-cage, the  $\Delta V$  is further increased to an average of 38.8 kcal/mol with a broad distribution ( $\sigma_{\Delta V} = 7.0$  kcal/mol), ranging from 13.6 to 80.1 kcal/mol (Figure 6D).

In dual-boost aMD simulations of chignolin, reweighting using exponential average leads to high fluctuations in both  $\Phi_4$  and  $\Psi_4$ , as shown in Figure 7A and B. The energy wells are poorly reproduced and the energy barrier of  $\Psi_4$  at  $75^\circ$  is overestimated by  $\sim 6$  kcal/mol. Again, this originates from the broad distribution of  $\Delta V$ , for which the Boltzmann reweighting factors are dominated by the high boost potential frames with  $\Delta V = 20-32$  kcal/mol. Using Maclaurin series expansion to the

10th order, the reweighted PMF profiles are less noisy, but they differ most in the energy barrier region of  $\Psi_4$  near  $75^\circ$  by  $\sim 2$  kcal/mol (Figure 7D). With cumulant expansion, since the reweighted PMF profiles using expansion to the third order exhibit large energetic noise as seen in Figure 3E and F, they are not presented for chignolin, also for Trp-cage below. Cumulant expansion on the first order generates deviated PMF profiles, similar to the Maclaurin series expansion. In comparison, PMF profiles obtained from cumulant expansion to the second order exhibits better agreement with those of 1000 ns cMD simulations with errors of  $\sim 0.5$  kcal/mol (Figure 7E and F). High fluctuations are observed in the energy barrier region of  $\Psi_4$  centered at  $75^\circ$  (Figure 7F), which is strongly correlated with the increased anharmonicity of the  $\Delta V$  distribution found in the energy barrier region (Figure 7H). In the energy well regions of  $\Phi_4$  centered at  $-72^\circ$  and  $\Psi_4$  at  $-30^\circ$  and  $150^\circ$ , the anharmonicity of  $\Delta V$  distribution stays low at  $\sim 0.05$  (Figure 7G and H). This justifies the accurate PMF obtained in these regions through cumulant expansion to the second order.

For Trp-cage, high energetic noise is found in the reweighted PMF profiles obtained through exponential average (Figure 8A and B), similar to those of alanine dipeptide and chignolin. Based on Maclaurin series expansion to the 10th order, the reweighted PMF profiles are much less noisy, but the PMF values of  $\Phi_{12}$  in the energy barrier region  $-12^\circ-12^\circ$  differ most from the cMD simulations by  $\sim 2$  kcal/mol, and the energy minimum position of  $\Psi_{12}$  is shifted from  $-18^\circ$  to  $-36^\circ$  (Figure 8C and D). In comparison, cumulant expansion to the second order provides reweighted PMF profiles that are similar to the original as obtained from the 1000 ns cMD simulations within  $\sim k_B T$  statistical errors (Figure 8E and F). High fluctuations observed in the energy barrier regions are also strongly correlated with the increased anharmonicity of the  $\Delta V$  distribution as shown in Figure 8G and H. In summary, reweighting based on cumulant expansion to the second order provides the most accurate reweighted PMF profiles compared with the exponential average and Maclaurin series expansion.

## CONCLUSIONS

aMD is an efficient enhanced sampling technique that has been demonstrated to capture millisecond-time scale events in both globular and membrane proteins. Unlike many other biased sampling methods, aMD is able to explore the conformational space of biomolecules without the requirement of predefined reaction coordinate(s). However, accurate reweighting of aMD simulations has been a long-standing problem.<sup>5a</sup>

In this study, we systematically compared the currently used aMD reweighting algorithms, including the exponential average, Maclaurin series and cumulant expansion, with simulations on three model systems, that is, alanine dipeptide, chignolin and Trp-cage. In the dihedral-boost aMD simulations of alanine dipeptide, the boost potential exhibits a relatively narrow distribution with a range  $\delta_{\Delta V} \leq 12$  kcal/mol, for which exponential average is able to recover the original PMF profiles within errors of  $\sim 1$  kcal/mol. In the dual-boost aMD simulations of alanine dipeptide, chignolin and Trp-cage, larger boost potentials of broader distribution ( $\delta_{\Delta V} \geq 18$  kcal/mol) are applied. The exponential reweighting factors are dominated by a small number of simulation frames with the largest  $\Delta V$ , leading to highly noisy PMF profiles. Reweighting based on the Maclaurin series expansion that is equivalent to cumulant expansion on the first order greatly suppresses the energetic noise, but the energy minimum positions are often incorrectly

reproduced and the PMF values exhibit errors of  $\sim 1\text{--}2$  kcal/mol in the systems examined. In contrast, cumulant expansion to the second order normally recovers the most accurate PMF profiles (within  $\sim k_B T$  of the original), particularly when the distribution of  $\Delta V$  satisfies near-Gaussian distribution. In this context, anharmonicity of the  $\Delta V$  distribution,  $\gamma$  can serve as an indicator of the accuracy of aMD reweighting using cumulant expansion to the second order.

For simulations of larger systems (e.g., proteins with several hundred residues and hundreds-of-thousand to millions of atoms), significantly higher boost potential with broad distribution (e.g.,  $\delta_{\Delta V} \approx 100\text{--}200$  kcal/mol) often occurs with the current aMD scheme and accurate reweighting remains challenging.<sup>5c</sup> In future developments, the anharmonicity of the  $\Delta V$  distribution,  $\gamma$  could be used to propose new formula for aMD boost potential such that accurate reweighting can be achieved. Furthermore, the postprocessing of aMD trajectories could be optimized to minimize the anharmonicity of the boost potential distribution, for instance through the selection of a good set of multidimensional reaction coordinates. In general, we expect cumulant expansion to find great use in the reconstruction of accurate free energy profiles from aMD simulations.

## ■ ASSOCIATED CONTENT

### Supporting Information

Figures S1–S7 as described in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Email: yimiao@ucsd.edu.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Changsun Eun, Ferran Feixas, Donald Hamelberg, Hari Muddana, and Michael K. Gilson for valuable discussions. Computing time was provided on GPU nodes of the Triton Shared Computing Cluster (TSCC) at the San Diego Supercomputer Center (SDSC) and by the joint Memorial Sloan Kettering Cancer Center (MSKCC) and Walker Lab Exxact AMBER GPU Cluster. This work was supported by NSF (grants MCB1020765 and SI2-1148276), National Institutes of Health (NIH) (grant GM31749), Howard Hughes Medical Institute, Center for Theoretical Biological Physics (CTBP), NVIDIA Inc., and the National Biomedical Computation Resource (NBCR).

## ■ REFERENCES

- (1) (a) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919–11929. (b) Hamelberg, D.; de Oliveira, C. A. F.; McCammon, J. A. Sampling of Slow Diffusive Conformational Transitions with Accelerated Molecular Dynamics. *J. Chem. Phys.* **2007**, *127* (15), 155102.
- (2) (a) Wereszczynski, J.; McCammon, J. A. Nucleotide-Dependent Mechanism of Get3 As Elucidated from Free Energy Calculations. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (20), 7759–7764. (b) Gasper, P. M.; Fuglestad, B.; Komives, E. A.; Markwick, P. R. L.; McCammon, J. A. Allosteric Networks in Thrombin Distinguish Procoagulant vs Anticoagulant Activities. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (52), 21216–21222. (c) Markwick, P. R. L.; Pierce, L. C. T.; Goodin, D. B.; McCammon, J. A. Adaptive Accelerated Molecular Dynamics (Ad-AMD) Revealing the Molecular Plasticity of P450cam. *J. Phys. Chem. Lett.* **2011**, *2* (3), 158–164. (d) Bucher, D.; Grant, B. J.; Markwick, P. R.; McCammon, J. A. Accessing a Hidden Conformation of the Maltose Binding Protein Using Accelerated Molecular Dynamics. *PLoS Comput. Biol.* **2011**, *7* (4), e1002034. (e) Wang, Y.; Markwick, P. R. L.; de Oliveira, C. A. F.; McCammon, J. A. Enhanced Lipid Diffusion and Mixing in Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2011**, *7* (10), 3199–3207.
- (3) Pierce, L. C. T.; Salomon-Ferrer, R.; de Oliveira, C. A. F.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8* (9), 2997–3002.
- (4) (a) Miao, Y.; Nichols, S. E.; Gasper, P. M.; Metzger, V. T.; McCammon, J. A. Activation and Dynamic Network of the M2 Muscarinic Receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (27), 10982–10987. (b) Miao, Y.; Nichols, S. E.; McCammon, J. A. Mapping of Allosteric Druggable Sites in Activation-Associated Conformers of the M2Muscarinic Receptor. *Chem. Biol. Drug Des.* **2013**, *83* (2), 237–246.
- (5) (a) Shen, T. Y.; Hamelberg, D. A Statistical Analysis of the Precision of Reweighting-Based Simulations. *J. Chem. Phys.* **2008**, *129* (3), 034103. (b) Markwick, P. R. L.; McCammon, J. A. Studying functional dynamics in bio-molecules using accelerated molecular dynamics. *Phys. Chem. Chem. Phys.* **2011**, *13* (45), 20053–20065. (c) Miao, Y.; Nichols, S. E.; McCammon, J. A. Free Energy Landscape of G-Protein Coupled Receptors, Explored by Accelerated Molecular Dynamics. *Phys. Chem. Chem. Phys.* **2014**, *16* (14), 6398–6406. (d) Sinko, W.; Miao, Y.; de Oliveira, C. A. F.; McCammon, J. A. Population Based Reweighting of Scaled Molecular Dynamics. *J. Phys. Chem. B* **2013**, *117* (42), 12759–12768.
- (6) Wereszczynski, J.; McCammon, J. A. Using Selectively Applied Accelerated Molecular Dynamics to Enhance Free Energy Calculations. *J. Chem. Theory Comput.* **2010**, *6* (11), 3285–3292.
- (7) (a) Fajer, M.; Hamelberg, D.; McCammon, J. A. Replica-Exchange Accelerated Molecular Dynamics (REXAMD) Applied to Thermodynamic Integration. *J. Chem. Theory Comput.* **2008**, *4* (10), 1565–1569. (b) Fajer, M.; Swift, R. V.; McCammon, J. A. Using Multistate Free Energy Techniques to Improve the Efficiency of Replica Exchange Accelerated Molecular Dynamics. *J. Comput. Chem.* **2009**, *30* (11), 1719–1725. (c) Arrar, M.; de Oliveira, C. A. F.; Fajer, M.; Sinko, W.; McCammon, J. A. w-REXAMD: A Hamiltonian Replica Exchange Approach to Improve Free Energy Calculations for Systems with Kinetically Trapped Conformations. *J. Chem. Theory Comput.* **2013**, *9* (1), 18–23.
- (8) Bowman, G. R.; Huang, X.; Pande, V. S. Using Generalized Ensemble Simulations and Markov State Models to Identify Conformational States. *Methods* **2009**, *49* (2), 197–201.
- (9) (a) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199. (b) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules 0.1. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021.
- (10) Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry, and Material Science. *Rep. Prog. Phys.* **2008**, *71* (12), 126601.
- (11) (a) Darve, E.; Pohorille, A. Calculating Free Energies Using Average Force. *J. Chem. Phys.* **2001**, *115* (20), 9169–9183. (b) Darve, E.; Rodriguez-Gomez, D.; Pohorille, A. Adaptive Biasing Force Method for Scalar and Vector Free Energy Calculations. *J. Chem. Phys.* **2008**, *128* (14), 144120.
- (12) (a) Zheng, L. Q.; Yang, W. Practically Efficient and Robust Free Energy Calculations: Double-Integration Orthogonal Space Tempering. *J. Chem. Theory Comput.* **2012**, *8* (3), 810–823. (b) Zheng, L.; Chen, M.; Yang, W. Random Walk in Orthogonal Space to Achieve

Efficient Free-Energy Simulation of Complex Systems. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (51), 20227–32.

(13) (a) Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78* (14), 2690–2693. (b) Jarzynski, C. Equilibrium Free-Energy Differences from Nonequilibrium Measurements: A Master-Equation Approach. *Phys. Rev. E* **1997**, *56* (5), 5018–5035.

(14) Hummer, G. Fast-Growth Thermodynamic Integration: Error and Efficiency Analysis. *J. Chem. Phys.* **2001**, *114* (17), 7330–7337.

(15) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. Free Energy Calculation from Steered Molecular Dynamics Simulations Using Jarzynski's Equality. *J. Chem. Phys.* **2003**, *119* (6), 3559–3566.

(16) (a) Levy, R. M.; Belhadj, M.; Kitchen, D. B. Gaussian Fluctuation Formula for Electrostatic Free-Energy Changes in Solution. *J. Chem. Phys.* **1991**, *95* (5), 3627–3633. (b) Simonson, T.; Wong, C. F.; Brunger, A. T. Classical and quantum simulations of tryptophan in solution. *J. Phys. Chem. A* **1997**, *101* (10), 1935–1945.

(17) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334* (6055), 517–520.

(18) Eastwood, M. P.; Hardin, C.; Luthey-Schulten, Z.; Wolynes, P. G. Statistical Mechanical Refinement of Protein Structure Prediction Schemes: Cumulant Expansion Approach. *J. Chem. Phys.* **2002**, *117* (9), 4602–4615.

(19) Lange, O. F.; Grubmuller, H. Full Correlation Analysis of Conformational Protein Dynamics. *Proteins: Struct., Funct., Bioinf.* **2008**, *70* (4), 1294–1312.

(20) (a) D.A. Case, T. A. D., T.E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossvary, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and Kollman, P.A. AMBER 12; University of California: San Francisco, 2012; (b) Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8* (5), 1542–1555. (c) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9* (9), 3878–3888. (d) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3* (2), 198–210.

(21) Le Grand, S.; Gotz, A. W.; Walker, R. C. SPFP: Speed without Compromise—A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2013**, *184* (2), 374–380.

(22) Barua, B.; Lin, J. C.; Williams, V. D.; Kummler, P.; Neidigh, J. W.; Andersen, N. H. The Trp-cage: Optimizing the Stability of a Globular Miniprotein. *Protein Eng., Des. Sel.* **2008**, *21* (3), 171–185.

(23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.

(24) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341.

(25) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.

(26) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593.

(27) Wang, Y.; Harrison, C. B.; Schulten, K.; McCammon, J. A. Implementation of Accelerated Molecular Dynamics in NAMD. *Comput. Sci. Discovery* **2011**, *4* (1), 015002.