

# An Efficient Gaussian-Accelerated Molecular Dynamics (GaMD) Multilevel Enhanced Sampling Strategy: Application to Polarizable Force Fields Simulations of Large Biological Systems

Frédéric Célerse,<sup>▽</sup> Théo Jaffrelo Inizan,<sup>▽</sup> Louis Lagardère, Olivier Adjoua, Pierre Monmarché, Yinglong Miao, Etienne Derat, and Jean-Philip Piquemal\*



Cite This: *J. Chem. Theory Comput.* 2022, 18, 968–977



Read Online

ACCESS |



Metrics & More

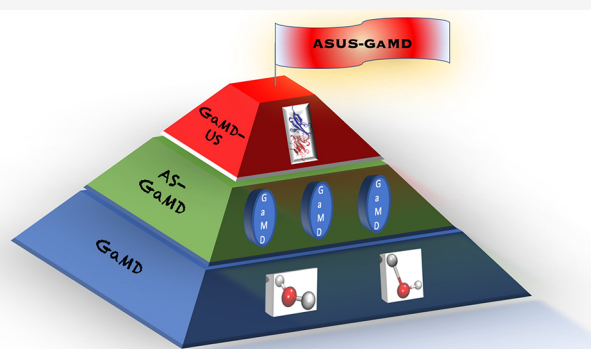


Article Recommendations



Supporting Information

**ABSTRACT:** We introduce a novel multilevel enhanced sampling strategy grounded on Gaussian-accelerated Molecular Dynamics (GaMD). First, we propose a GaMD multi-GPUs-accelerated implementation within the Tinker-HP molecular dynamics package. We introduce the new “dual-water” mode and its use with the flexible AMOEBA polarizable force field. By adding harmonic boosts to the water stretching and bonding terms, it accelerates the solvent–solute interactions while enabling speedups, thanks to the use of fast multiple–time step integrators. To further reduce the time-to-solution, we couple GaMD to Umbrella Sampling (US). The GaMD–US/dual-water approach is tested on the 1D Potential of Mean Force (PMF) of the solvated CD2–CD58 system (168 000 atoms), allowing the AMOEBA PMF to converge within 1 kcal/mol of the experimental value. Finally, Adaptive Sampling (AS) is added, enabling AS–GaMD capabilities but also the introduction of the new Adaptive Sampling–US–GaMD (ASUS–GaMD) scheme. The highly parallel ASUS–GaMD setup decreases time to convergence by, respectively, 10 and 20 times, compared to GaMD–US and US. Overall, beside the acceleration of PMF computations, Tinker-HP now allows for the simultaneous use of Adaptive Sampling and GaMD–“dual water” enhanced sampling approaches increasing the applicability of polarizable force fields to large-scale simulations of biological systems.



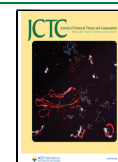
## INTRODUCTION

Understanding interactions within biomolecules is crucial for many topics such as drug discovery. Some structural modifications, sometimes undetected by experiment, can drastically change the nature of the physics ruling interacting complex systems. For this reason, predicting the long timescale conformational dynamics of proteins is a long-standing challenge within the conventional molecular dynamics (cMD) community.<sup>1–6</sup> It requires accurate models able to capture the true potential energy hyper-surface and long simulations to both access the large biological processes time-scale and satisfy the ergodicity principle.<sup>7</sup> Therefore, accelerating MD has been a central field of research in the last decades.<sup>8–11</sup> Beside these developments, several additional strategies have been pursued over the years to further accelerate the simulations. They include the extensive use of high-performance computing (HPC) resources<sup>4,12</sup> and the optimization of GPU-accelerated modeling platforms.<sup>13–15</sup> Alternatively, an intensive algorithmic work has been undertaken, introducing techniques such as multiple-time-step integrator schemes<sup>16,17</sup> or collective variables-driven MD methods.<sup>18,19</sup> The latter have been found useful in enhanced sampling and free-energy calculation.<sup>20–25</sup> Although such

methods are powerful, since they can estimate free energies of binding or the stability of secondary and quaternary structures of proteins,<sup>26,27</sup> the free-energy estimations can suffer from biases either generated by the initial choice of the collective variable (CV) or by the existence of multiple CV within the mechanism process (e.g., dual mechanisms).<sup>28</sup> For these reasons, collective variable–free methods have become increasingly popular.<sup>29</sup> Among them, the recent Gaussian-accelerated molecular dynamics (GaMD) has shown great promise, because of its high sampling acceleration, its user-friendly tunable parameters, and its minor additional computational cost.<sup>30</sup> GaMD accelerates conformational sampling by adding a harmonic boost to the potential energy. Coupled with the second-order cumulant expansion, GaMD allows us to

Received: October 12, 2021

Published: January 26, 2022



compute unbiased properties by using an accurate reweighting procedure through cumulant expansion to the second order.

Although new-generation many-body polarizable force field (PFFs) are more accurate in describing biomolecular interactions,<sup>31–34</sup> they are computationally more challenging than traditional approaches. Therefore, to overcome these limitations, here, we provide a novel general multilevel enhanced sampling strategy, which we apply to the PFF AMOEBA. To do so, we combine the Tinker–HP massively parallel multi-GPUs platform<sup>15</sup> together with a highly scalable GaMD implementation (*level 0*) and then additional enhanced sampling techniques based on recent developments of the field. As a first speedup, we propose an extension of the GaMD formalism with a new GaMD mode, enabling the use of flexible water models such as AMOEBA<sup>35,36</sup> and fast multiple-time-step integrators<sup>17</sup> (*level 1*). We then discuss the explicit coupling of such GaMD approach to Umbrella Sampling (US)<sup>37</sup> and Adaptive Sampling (AS)<sup>6</sup> techniques (*level 2*). To demonstrate their applicability to PFF, these physics-based hybrid enhanced sampling strategies are then applied to the Potential of Mean Force (PMF) study of a large biological complex CD2–CD58 interacting via salt bridges with the AMOEBA force field. Finally, we combine all of them together within the Adaptive Sampling–US–GaMD method (ASUS–GaMD) scheme (*level 3*).

## METHOD: INTRODUCING THE GaMD “DUAL WATER” MODE

GaMD is a potential-biasing method for unconstrained enhanced sampling without the need to set a predefined CV. It smooths the potential energy surface by adding a harmonic boost potential, as described in the seminal paper.<sup>11</sup> Its general framework makes it suitable for the development of hybrid schemes and variants, such as replica-exchange umbrella sampling GaMD (GaREUS),<sup>38</sup> ligand GaMD (LiGaMD),<sup>39</sup> and peptide GaMD (Pep-GaMD).<sup>40</sup>

If the system potential energy is lower than a threshold energy  $E$ , a harmonic potential energy boost is applied to smooth the potential energy surface. By denoting  $q \in \mathbb{R}^{3N}$  as the configurations when the system potential energy  $U(q)$  is lower than a threshold energy  $E$ , a boost, which is dependent on  $U(q)$  is added:

$$U'(q) = U(q) + \Delta U^{\text{GaMD}}(U(q)) \quad (1)$$

with  $\Delta U^{\text{GaMD}}(U(q))$  being the external harmonic potential boost:

$$\Delta U^{\text{GaMD}}(U(q)) = \begin{cases} 0 & U(q) \leq E \\ \frac{1}{2}k(E - U(q))^2 & U(q) < E \end{cases} \quad (2)$$

and  $k$  the harmonic force constant. The two adjustable GaMD parameters  $k$  and  $E$  are automatically determined following the original procedure described in ref<sup>30</sup>. The boost intensity can be managed through a user-specified upper limit labeled as  $\sigma_0$  (e.g.,  $10k_B T$ ) predefined before the simulation. To ensure accurate reweighting with the cumulant expansion the  $\Delta U^{\text{GaMD}}$  standard deviation,  $\sigma_{\Delta V}$ , should satisfy  $\sigma_{\Delta V} < \sigma_0$ .<sup>30,41,42</sup> GaMD provides different modes: the boost is either applied on the total potential (GaMD–pot), on the dihedral potential (GaMD–dih), or on both at the same time (GaMD–dual).<sup>43,44</sup> Recently, another mode was introduced: LiGaMD,

which adds the boost to a ligand nonbonded interactions,<sup>39</sup> accelerating the sampling of ligand–protein interactions. It is known that interactions involving water are essential for such systems and that protein stability processes are controlled by water–protein interactions.<sup>6,45,46</sup> To accelerate these interactions, one would like to use the GaMD–dual mode on the nonbonded interactions of water molecules. However, such a boost requires the evaluation of the complete nonbonded energies and, in the context of multistep integrators such as BAOAB–RESPA1,<sup>17</sup> where they are split between short range and long range, these are only available at the outer (large) time step. This type of integrators enables the use of larger time steps and thus a direct acceleration of MD. For example, the BAOAB–RESPA1 is based on a RESPA (Reference System Propagator Algorithm)<sup>16</sup> three-level splitting of forces (bonded, short-range nonbonded, and long-range nonbonded) within the Leimkuhler’s BAOAB discretization of Langevin dynamics.<sup>47</sup> It allows up to a 7-fold acceleration for polarizable point dipole molecular dynamics.<sup>17</sup> But the fluctuations of the associated bias are such that it must be evaluated at shorter timesteps, so that the entire procedure is not compatible with multistep integrators such as BAOAB–RESPA1. For similar reasons, the GaMD–dual mode with a bias applied to the complete potential energy is not compatible with even simple RESPA integrators in which the potential energy is split between bonded and nonbonded terms. Therefore, GaMD–dual mode becomes rapidly limited by the simulation time. To overcome this issue, we developed a new mode, GaMD–dual-water (denoted as “GaMD–dualwat”), which adds a boost to the protein dihedral potential energy term and the water stretching and bending terms, this time fully compatible with RESPA and RESPA1 like integrators, allowing the water molecule to be more flexible, thus favoring their conformational changes.

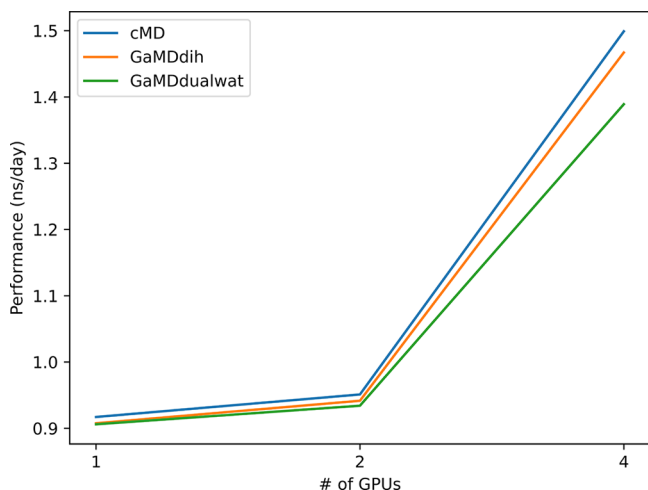
$$\begin{aligned} \Delta U^{\text{GaMD-dw}}(U(q)) \\ = \Delta U^{\text{dihedral}}_{\text{protein}}(U(q)) + \Delta U^{\text{stretch}}_{\text{water}}(U(q)) + \Delta U^{\text{bend}}_{\text{water}}(U(q)) \end{aligned} \quad (3)$$

This mode is enabled by the flexibility of the AMOEBA 03 water model<sup>35</sup> but is not compatible with rigid water models, such as TIP3P,<sup>48</sup> commonly used with the CHARMM and AMBER force field.<sup>49</sup> This framework allows one to further reduce the computational cost gap between PFFs and nPFFs. This new mode, in addition to the other GaMD–dih and GaMD–dual modes, is now available within the Tinker–HP software.<sup>12,15</sup> In the following, we first tested its GPU scalability and performance on the STMV system ( $\sim 1\,066\,624$  atoms), and its sampling efficiency is demonstrated on simulations of the alanine dipeptide and the CD2–CD58 complex. A technical appendix is present at the end of the manuscript and provides the formalism of the method and the associated debiasing equations.

## RESULTS AND DISCUSSION

**Level 0: Efficiency and GPU Scalability.** The GaMD implementation is such that only a small computational and communication (in parallel) overhead is added, compared to cMD. The GaMD–dih and GaMD–dualwat have been considered on the STMV system ( $1\,066\,624$  atoms) with the AMOEBA PFF and the 10 fs outer time-step HMR BAOAB–RESPA1 multiple-time-step integrator.<sup>17</sup> V100 GPUs from the national Jean Zay supercalculator have been used for all the

benchmark computations. Similar scalability studies have been performed on the Jean Zay multi-CPU (Figure S1 in the Supporting Information). The AMOEBA GPU simulations were performed on a single node, since the multinode extension of the AMOEBA PFF within the Tinker-HP package is still under development. On 1 and 2 GPUs (Figure 1), the GaMD data communications are negligible (1%). On 4



**Figure 1.** GaMD–dih and GaMD–dualwat scaling performance on 1/2/4 V100 GPUs (i.e., corresponding to a full node of the Jean Zay machine) on STMV (1 066 624 atoms) with the AMOEBA force field and the BAOAB–RESPA1 10 fs multiple-time-step integrator. The cMD reference, in blue, allows one to evaluate the GaMD impacts on the code communications.

GPUs, the communications are increasing and the performance decreases by 7%. Overall, the use of GaMD only slightly alters the performance. This high scalability opens the door to simulate at a high accuracy, large complex biomolecular systems with PFFs.

**Level 1: GaMD–dualwat with PFFs.** We compared GaMD–dih, GaMD–dual, and GaMD–dualwat sampling acceleration on the exploration of the relevant basins of the alanine dipeptide (e.g.,  $\alpha_r$ ,  $\alpha_L$ , and  $P_{II}$ ). The alanine dipeptide is solvated in a cubic 20 Å water box. We used the many-body AMOEBA18 PFF.<sup>50,51</sup> The system was minimized with a RMS of 1 kcal/mol and sampled within the NPT thermodynamic ensemble with the Bussi thermostat<sup>52</sup> and a MonteCarlo barostat<sup>53</sup> at 300 K and 1 atm. We used the Velocity Verlet integrator and a 1 fs time step.<sup>54</sup> The Smooth Particle Mesh Ewald (SPME) algorithm was employed to compute noncovalent interactions<sup>55</sup> with a real space cutoff equal to 7 Å and a van der Waals cutoff set to 9 Å. For AMOEBA, the convergence criteria for multipoles was set to  $10^{-5}$ . After short testing simulations, we found an optimal value of 3 kcal/mol for GaMD–dih and GaMD–dual  $\sigma_0$ , in accordance with ref 11, and 4 kcal/mol for GaMD–dualwat (see Figure S2 and Tables S1 and S2 in the Supporting Information). We ran three independent simulations of 60 ns for each mode. The different sampled basins are also compared to a 1  $\mu$ s cMD AMOEBA reference.

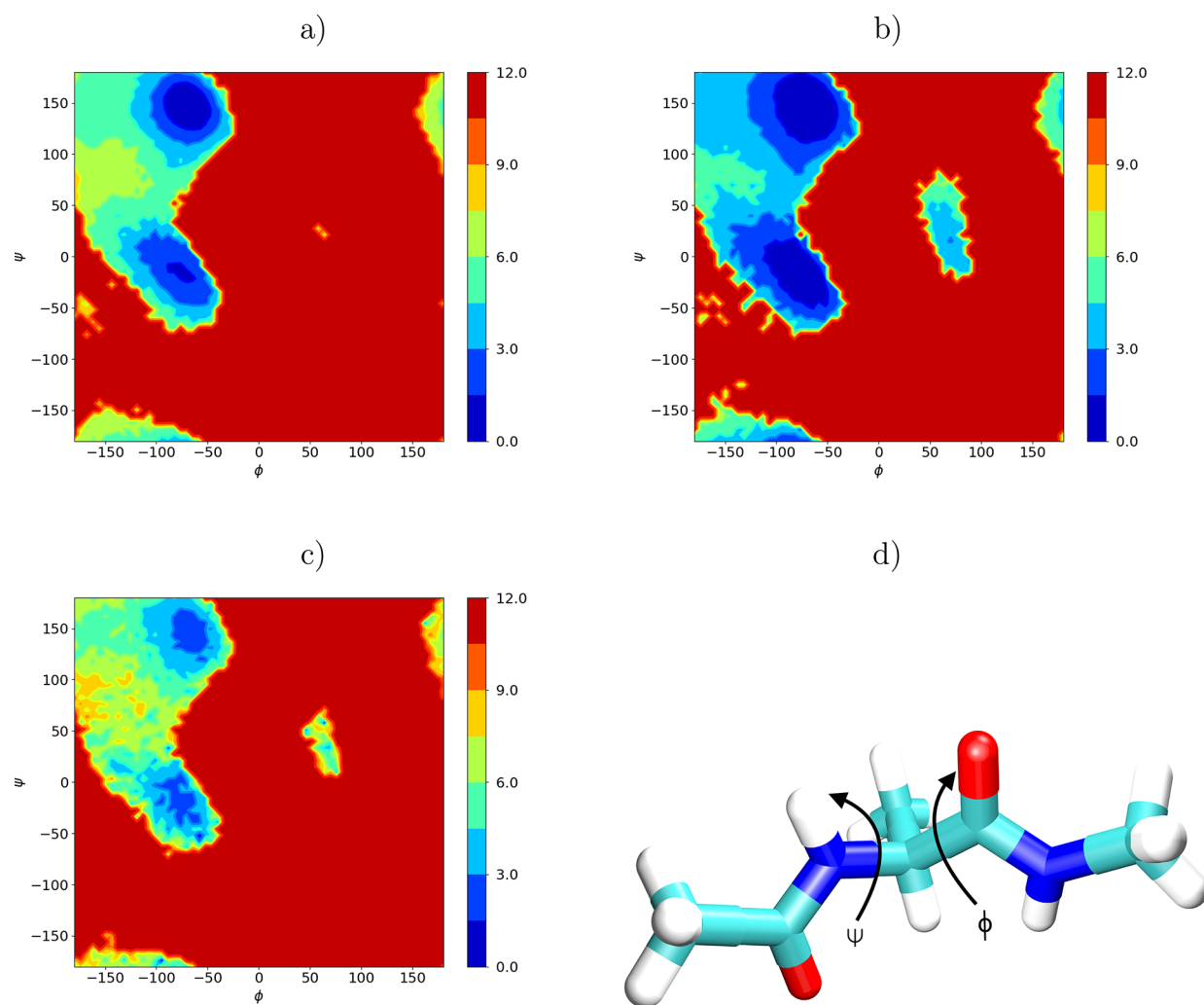
Reweighted, see the Technical Appendix, free-energy surfaces obtained from these simulations are depicted in Figure 2 and show that GaMD–dual captures the  $\alpha_r$  ( $50^\circ, 25^\circ$ ),  $\alpha_L$  ( $-75^\circ, -25^\circ$ ), and  $P_{II}$  ( $-75^\circ, 150^\circ$ ) basins well. These results are consistent with the 1  $\mu$ s cMD trajectory (see Figure

S3 in the Supporting Information) depicting these three basins. While the GaMD–dih mode captures the  $\alpha_r$  basin after 150 ns, the GaMD–dualwat captures it in 100 ns (see Figure S4 in the Supporting Information). We also observe a sampling acceleration between GaMD–dual and GaMD–dualwat compared to the ref 1 microsecond cMD. To characterize the GaMD boost harmonicity, its distribution anharmonicity ( $\gamma$ ) is calculated as in ref 30.  $\gamma$  serves as an indicator of the sampling convergence and reweighting procedure accuracy. Depicted in Figure S5 in the Supporting Information, GaMD–dih as well as GaMD–dual depicts high anharmonicity with, respectively, 0.252 and 0.016, compared to GaMD–dualwat with 0.0005. In addition, we see a steep anharmonicity convergence to less than  $10^{-3}$  for GaMD–dualwat while being relatively stable at  $2 \times 10^{-1}$  for GaMD–dih (see Figure S4). In comparison the anharmonicity is  $\sim 0.001$  with GaMD–dih and AMBER99SB. Therefore, PFFs increase the statistical noise and stress the importance of using low-anharmonicity GaMD modes. In that sense, GaMD–dualwat appears more suitable than GaMD–dual for PFFs simulations with an anharmonicity equal to 0.0005. As stated previously, another advantage of GaMD–dualwat is that it can be coupled to multiple-time-step procedures, such as BAOAB–RESPA1,<sup>17</sup> in contrast to the GaMD–dual mode, which remains limited to single-time-step integrators. Comparative results of GaMD–dualwat with both integrators can be found in Figure S6 in the Supporting Information. Its coupling with multiple-time-step procedures clearly compensates for the slightly lower sampling performance, compared with GaMD–dual. The sampling enhancement brought by the GaMD–dualwat can be partly related to how it affects the diffusion of water: in Table S3 of the Supporting Information, we report the self-diffusion coefficients of bulk water computed within a same setup (same size of box and same integrator) and observe that it is increased with the GaMD–dualwat mode, compared to the simple GaMD–dih one, favoring global conformational changes due to water reorganization. While the added sampling efficiency is already significant for the alanine dipeptide, we expect it to be larger on more complex and larger biological systems such as CD2CD58, where water reorganization plays a bigger role.

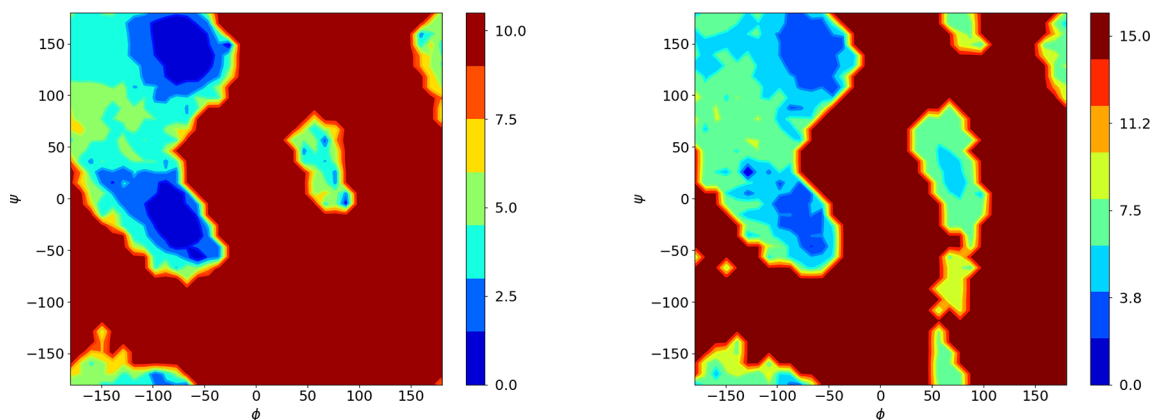
Combined with a highly parallel GPUs infrastructures and multiple-time-step integrators, the GaMD–dualwat should allow one to help reach very high-resolution conformational space of large molecular systems. In addition to the sampling acceleration it provides, the low associated anharmonicity drastically reduces the statistical noise associated with reweighting.

**Level 2: Accelerating Simulations with the Parallel AS–GaMD Scheme.** We further coupled our newly introduced GaMD mode to additional enhanced sampling strategies. Recently, we developed a new adaptive sampling (AS) technique, which was shown to allow massive sampling of the SARS–CoV–2 Main Protease conformational space.<sup>6</sup> We coupled these two methodologies together, yielding the AS–GaMD method. The principle is similar to the AS, the only modification being that each cMD at each iteration is now a GaMD simulation. The double bias coming from both AS and GaMD implies that a suitable and careful reweighting scheme must be introduced to reconstruct an unbiased free-energy surface. All mathematical tools for the reweighting scheme are provided in the Technical Appendix. We applied this methodology to the same system, the alanine dipeptide, using the same GaMD simulation protocol. At each iteration,





**Figure 2.** 2D PMF (in kcal/mol) of the alanine dipeptide obtained in AMOEBA for (a) GaMD–dih mode ( $3 \times 60$  ns), (b) GaMD–dual mode ( $3 \times 60$  ns), and (c) GaMD–dualwat mode ( $3 \times 60$  ns). (d) Alanine dipeptide representation with the corresponding  $\Phi$  and  $\Psi$  angles.

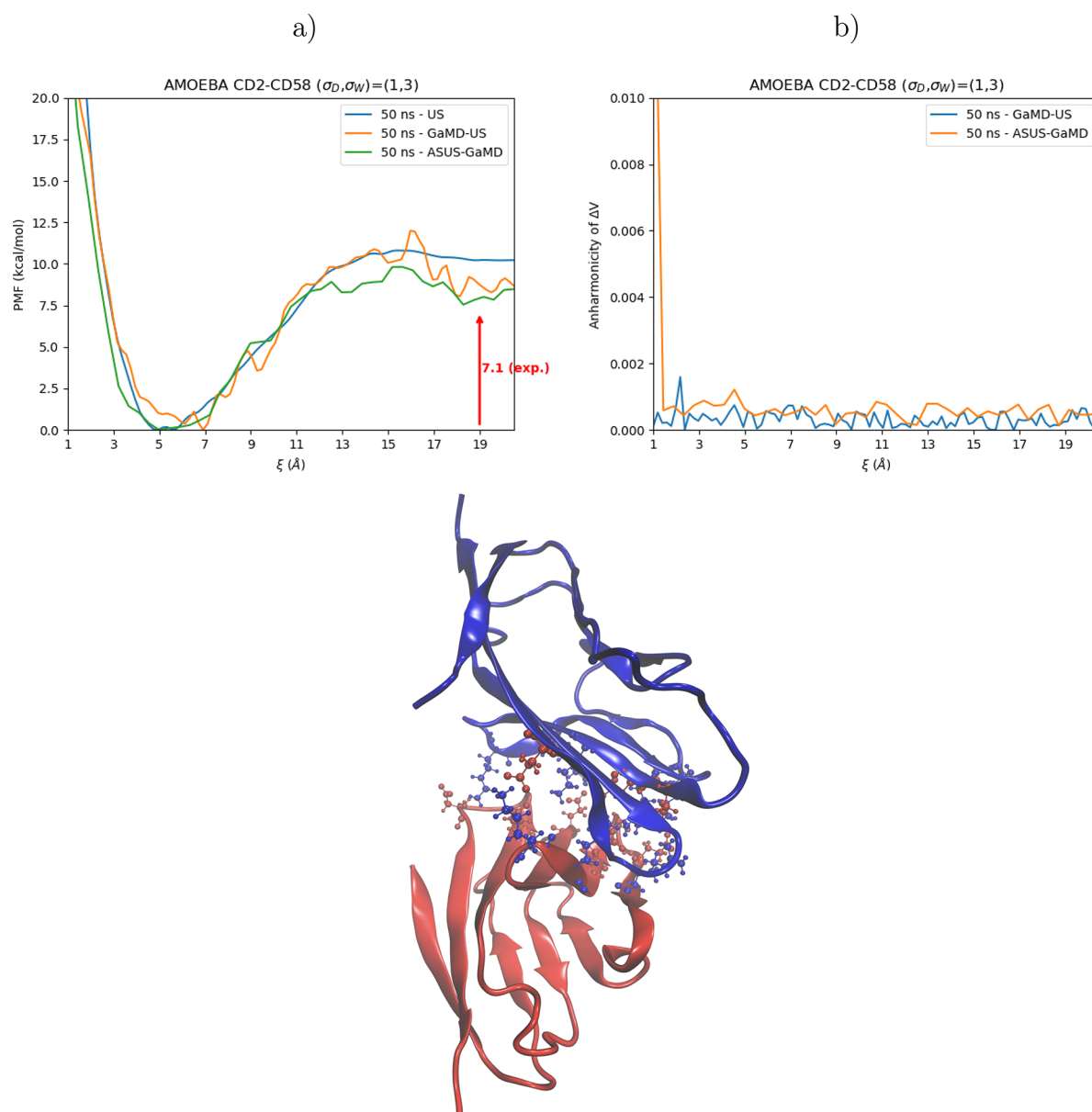


**Figure 3.** PMFs of (a) GaMD dualwat ( $3 \times 60 = 180$  ns) and (b) AS–GaMD dualwat ( $5 \times 25 = 125$  ns) simulations for the alanine dipeptide. For the GaMD dualwat simulations, we performed three independent simulations of 60 ns using 1 fs time step with the verlet integrator. The AS–GaMD dualwat simulations were performed using the BAOAB–RESPA1 10 fs multi-time-step integrator and 5 AS iterations of  $5 \times 5$  ns with a square term.

we projected the structures on the two main dihedral angle spaces. To push the limit of the AS–GaMD sampling capability, we combined a modified version of the AS selection scheme with the BAOAB–RESPA1 multi-time-step integrator. The probability law for selection of new structures was taken as

the inverse of the square of the probability density on the reduced space, which further amplifies the exploration of undiscovered region.

In Figure 3, we represented the 2D PMF obtained with both AS–GaMD/BAOAB–RESPA1 and GaMD/VERLET simula-



**Figure 4.** (c) PDB 1QA9 CD2CD58 representation with CD2 and CD58 subcomplexes represented respectively in blue and red, using the new ribbons representation. Residues at the interface considered in the COM distance between the two subcomplexes are represented in blue and red for respective basic and acid residues using the CPK representation. VMD software was employed to generate the structure. PMFs obtained with US, GaMD–US, and ASUS–GaMD are depicted in panel (a) and their respective anharmonicity in panel (b).

tions. For AS–GaMD/BAOAB–RESPA1, we performed 5 iterations of  $5 \times 5$  ns GaMD–dualwat simulations for a total simulation time of 125 ns. As in the previous section, the GaMD/VERLET is composed of three independent simulations of 60 ns (180 ns total). In 30% less simulation time and 5 times less computational time, thanks to the natural AS parallelism, the coupled AS–GaMD/multi-time-step integrator scheme greatly enhances the exploration of the free-energy surface. We observed that the  $\alpha_L$  region is already captured at the first iteration, i.e., with only 25 ns (see Figure S7 in the Supporting Information). In addition, other states, next to the  $\alpha_L$  region, are captured within tens of nanoseconds and are still not seen after the entire GaMD simulation. Therefore, this AS–GaMD/multi-time-step coupling can represent an important gain for the sampling of biomolecular systems.

**Level 3: Pushing the Limit of PMF Convergence with GaMD–US and ASUS–GaMD.** US has been widely used and is mathematically robust but it still suffers from several issues.<sup>56–58</sup> In addition to the choice of the CVs, it is also difficult to estimate the PMF convergence, since it is system-dependent. Good indicators to check if convergence is reached are the overlap between neighboring windows and the evolution of the PMF curve, as a function of the simulation time per window. To accelerate the sampling within each window, Oshima et al. recently combined GaMD with replica-exchange and US.<sup>38</sup> Here, we first only applied a GaMD boost in each US window in order to enhance the sampling in the orthogonal space.

To demonstrate the PMF convergence acceleration, we studied the dissociation of the salt bridges interface within the CD2CD58 complex. This system, made of several salt bridges

and hydrogen bondings interactions, was already studied by some of us.<sup>28</sup> Although it has been shown that PFFs allow a better description of the salt bridges interactions, their computational cost has long hindered the study of such a large system. Since the portability of Tinker-HP on multi-GPU and the global acceleration of the PFFs, reaching such system is now easily achievable. To start this study, we took the same CD2CD58 complex as in our previous work<sup>28</sup> but we solvated it in a waterbox of 100 Å × 100 Å × 100 Å. Counterion were added to neutralize the system. We used the AMOEBAIO18 PFF.<sup>50,51</sup> The system was minimized with a RMS of 1 kcal/mol in the NVT thermodynamic ensemble with the Bussi thermostat.<sup>52</sup> Temperature was set to 300 K while pressure was set to 1 atm. We used the multi-time-step BAOAB-RESPA1 with a 10 fs time step with the Hydrogen Mass Repartitioning scheme (HMR)<sup>17</sup> and Smooth Particle Mesh Ewald (SPME) algorithm to compute electrostatic and polarization interactions<sup>55</sup> with a real-space cutoff of 7 Å and a van der Waals cutoff of 9 Å. The convergence criteria for polarization was set to 10<sup>-5</sup>. Thirty nine (39) US windows were generated, ranging from 1 to 20 Å with a width of 0.5 Å between them. CV was chosen as the distance between the center of mass formed by the interfacial residues isolated by Bayas et al. on CD2 and CD58 (see Table 1 in ref 59). A spring constant of 10 kcal/(mol Å<sup>2</sup>) was employed to restrain the system along the chosen CV. Each window was run for 5 ns for equilibration and then for 50 ns. Histogram overlap as well as the PMF curve, as a function of the simulation time allocated per window, were employed to check the convergence of the simulations (see Figure S9 in the Supporting Information). The final US PMF shows a slow decrease of the free-energy barrier with the simulation time, suggesting a slow convergence to ~12.5 kcal/mol. Binding affinity was found to be, experimentally, ~7.1 ± 0.03 kcal/mol, suggesting that our simulations are not converged.<sup>59</sup> In order to improve sampling within each window, a new US procedure was performed, similar to the previous US protocol, but now with an additional GaMD–dualwat potential applied in each window. The GaMD parametrization protocol and reweighting procedure are described in the Technical Appendix and in Figure S8 and Table S4 in the Supporting Information). The optimized GaMD–dualwat parameter ( $\sigma_0$ ) values are equal to 1 and 3 kcal/mol for the dihedral and dual water modes, respectively. Figure 4 shows the difference between standard US and GaMD–US. The GaMD–US PMF and boost harmonicity converge at 40 ns per window (see Figures S10A–S10C in the Supporting Information). The predicted free-energy barrier is now within the 1 kcal/mol of the experiment. It shows that GaMD–dualwat, even without the presence of replica exchange, could considerably improve the PMF convergence of large systems. It also demonstrates that salt bridges and, more generally, protein–protein interactions are well-described with PFFs. Furthermore, as demonstrated in the work of Debiec et al.,<sup>60</sup> the improved accuracy of non-PFFs in describing these interactions requires the implicit incorporation of solvent polarization, underscoring the importance of polarization effects in these contexts.

To further push the sampling, we coupled together GaMD, AS, and US (ASUS–GaMD). We provide two reweighting schemes that either use modified Multistate Bennett Acceptance Ratio (MBAR) equations or the Rao–Blackwell estimator. The mathematical expressions are general and can be used with any weighted dynamics. Starting from an initial

US simulation (approximately equal to a few nanoseconds), each window is decomposed in several AS independent trajectories with an additional GaMD–dualwat potential boost (GaMD). Here, we ran 2 iterations of 5 × 5 ns GaMD–US per window. The PMF evolution can be found in Figure S11 in the Supporting Information, while the resulting PMF is depicted in Figure 4. We observe that ASUS–GaMD reaches GaMD–US in one iteration, showing the sampling acceleration impact provided by the AS part within ASUS–GaMD. Note that the rough aspect of the PMF obtained with the methods involving a GaMD bias comes from the debiasing of the boost potential, as can be seen in previous work involving US and GaMD.<sup>38</sup> Although a careful reweighting is needed for the different AS, GaMD and US layers, the overall ASUS–GaMD approach inherits the strong adaptive sampling advantages of being pleasantly parallelizable and considerably accelerates the PMF convergence.

## CONCLUSIONS

Combined with the use of modern GPUs, these sampling techniques allow one to drastically reduce time to solution in PFFs evaluation of PMFs. Although it is difficult to truly quantify the final acceleration (i.e., a PMF convergence remains partially system-dependent), one can see in Figure S12 in the Supporting Information that if we extrapolate the US convergence, ASUS–GaMD converges 1.4 times faster. Thanks to the native parallelism inherited from AS, the PMF evaluation can be done in one-fifth of the simulation time, yielding an acceleration of 7. If we consider that convergence was already reached with a 25 ns per window setup, this factor grows to 14. Thus, ASUS–GaMD that would have taken months can be reduced to days of computation. This work also allows one to invoke any variant of the combined approaches, thereby offering access to GPU-accelerated GaMD-adaptive sampling (AS–GaMD) simulations that will be helpful to further extend conformational space studies of proteins.<sup>6</sup> To conclude, these methodologies will contribute further to allow high-resolution sampling of large biological systems up to millions of atoms, using a polarizable force field.

## TECHNICAL APPENDIX

We use  $\xi(q)$  to denote the reaction coordinate along which we performed the US simulation and  $q$  is the configuration. Here, a configuration means the positions  $q \in \mathbb{R}^{3N}$  of all the atoms of the system. The imposed US bias potential is

$$U_j^{\text{US}}(q) = \mathcal{K}(\xi(q) - \xi_j)^2 \quad (\text{A1})$$

with  $\mathcal{K}$  being the force constant.

We combined the AS, US, and GaMD such that each US window  $j \in [1, \dots, M]$ ,  $\xi_1, \dots, \xi_M$ , is parallelized and accelerated by adaptive sampling replicas and GaMD boost potential:

$$U_j''(q) = U(q) + U^{\text{GaMD}}(q) + U_j^{\text{US}}(q) \quad (\text{A2})$$

We use  $(q_{j,n})_{n \in 1, \dots, N}$  to denote the  $N$  configurations generated by the AS replicas of US window  $j$  and  $(\omega_{j,n})_{n \in 1, \dots, N}$  to represent their respective AS weights. These normalized weights are defined as  $\omega_{j,n} = \frac{Nv_{j,n}}{\sum_{m=1}^N v_{j,m}}$  so that  $\sum_{n=1}^N \omega_{j,n} = N$  with  $v_{j,n}$  being the unnormalized AS weights. The canonical average of an observable  $\varphi$  is estimated by

$$\langle \varphi \rangle_j^n = \frac{\int \varphi(q) e^{-\beta U_j^n(q)} dq}{\int e^{-\beta U_j^n(q)} dq} \simeq \frac{\sum_{n=1}^N \varphi(q_{j,n}) \omega_{j,n}}{\sum_{n=1}^N \omega_{j,n}} = \frac{1}{N} \sum_{n=1}^N \varphi(q_{j,n}) \omega_{j,n} \quad (\text{A3})$$

In practice, to get a smooth reweighted PMF, the reaction coordinate  $\xi$  is discretized in  $K$  bins around values  $x_1, \dots, x_K$ . We want to estimate for each  $k \in [1, \dots, K]$  its free energy, up to an additive constant,

$$F(x_k) = -\frac{1}{\beta} \ln \mathbf{P}(\xi(q) \in \text{Bin}(x_k)) \quad (\text{A4})$$

where  $q$  is distributed according to the density probability law,  $\frac{e^{-\beta U}}{\int e^{-\beta U}}$ , i.e.,

$$F(x_k) = -\frac{1}{\beta} \ln \frac{\int 1_{\xi(q) \in \text{Bin}(x_k)} e^{-\beta U(q)} dq}{\int e^{-\beta U(q)} dq} = -\frac{1}{\beta} \ln \langle \varphi_k \rangle \quad (\text{A5})$$

with  $\varphi_k = 1_{\xi(q) \in \text{Bin}(x_k)}$ .

**First Step: GaMD with Cumulant Expansion.** We, first, remove the GaMD bias. Here, we want to find a relationship between  $\langle \varphi \rangle$  and  $\langle \varphi \rangle'$ , where the prime average represents the canonical average over the potential  $U' = U + U^{\text{GaMD}}$ . Starting from the canonical average, we notice

$$\begin{aligned} \langle \varphi \rangle &= \frac{\int \varphi(q) e^{-\beta U(q)} dq}{\int e^{-\beta U(q)} dq} \\ &= \frac{\int \varphi(q) e^{\beta U^{\text{GaMD}}(q)} e^{-\beta U'(q)} dq}{\int e^{\beta U^{\text{GaMD}}(q)} e^{-\beta U'(q)} dq} \\ &= \frac{\langle \varphi e^{\beta U^{\text{GaMD}}} \rangle'}{\langle e^{\beta U^{\text{GaMD}}} \rangle'} \end{aligned} \quad (\text{A6})$$

By applying this with  $\varphi = \varphi_k$

$$\begin{aligned} F(x_k) &= -\frac{1}{\beta} \ln \frac{\langle \varphi_k e^{\beta U^{\text{GaMD}}} \rangle'}{\langle e^{\beta U^{\text{GaMD}}} \rangle'} = -\frac{1}{\beta} \ln \langle \varphi_k e^{\beta U^{\text{GaMD}}} \rangle' + C \\ &= F'(x_k) - \frac{1}{\beta} \ln \frac{\langle \varphi_k e^{\beta U^{\text{GaMD}}} \rangle'}{\langle \varphi_k \rangle'} + C \end{aligned} \quad (\text{A7})$$

where  $C$  is a constant and  $F'(x_k)$  is the free energy,  $F'(x_k) = -\frac{1}{\beta} \ln \langle \varphi_k \rangle'$ . To reduce the estimator variance, we used the cumulant expansion to the second order,

$$\begin{aligned} \ln \frac{\langle \varphi_k e^{\beta U^{\text{GaMD}}} \rangle'}{\langle \varphi_k \rangle'} &\simeq \beta \frac{\langle \varphi_k U^{\text{GaMD}} \rangle'}{\langle \varphi_k \rangle'} \\ &\quad + \frac{\beta^2}{2} \left( \frac{\langle \varphi_k (U^{\text{GaMD}})^2 \rangle'}{\langle \varphi_k \rangle'} - \left( \frac{\langle \varphi_k U^{\text{GaMD}} \rangle'}{\langle \varphi_k \rangle'} \right)^2 \right) \end{aligned} \quad (\text{A8})$$

By combining with eq A7, the free energy is rewritten as

$$\begin{aligned} F(x_k) &\simeq -\frac{1}{\beta} \ln \langle \varphi_k \rangle' - \beta \frac{\langle \varphi_k U^{\text{GaMD}} \rangle'}{\langle \varphi_k \rangle'} \\ &\quad - \frac{\beta^2}{2} \left( \frac{\langle \varphi_k (U^{\text{GaMD}})^2 \rangle'}{\langle \varphi_k \rangle'} - \left( \frac{\langle \varphi_k U^{\text{GaMD}} \rangle'}{\langle \varphi_k \rangle'} \right)^2 \right) + C \end{aligned} \quad (\text{A9})$$

**Second Step: AS Modified MBAR.** Finally, we want to express  $\langle \varphi \rangle'$ , with respect to the AS weights in each US window  $j \in [1, \dots, M]$ . This can be done in two ways, using either the MBAR or the Rao–Blackwell estimator.

**Modified MBAR.** Let us define  $c'_j$  and  $F'_j$  as

$$c'_j = \int e^{-\beta U_j^n(q)} dq \quad F'_j = -\frac{1}{\beta} \ln c'_j \quad (\text{A10})$$

The prime comes from the use of the MBAR on the reference energy  $U'$  of the previous section. The starting point is to use the MBAR identity (eq 5 in ref 61) and notice

$$c'_j \langle e^{\beta U_j^n} \alpha_{i,j} \rangle'_i = \int e^{-\beta U_j^n(q)} e^{-\beta U_i^n(q)} \alpha_{i,j}(q) dq = c'_j \langle e^{\beta U_i^n} \alpha_{i,j} \rangle'_j \quad (\text{A11})$$

which holds for arbitrary functions  $q \rightarrow \alpha_{ij}(q)$  with  $i, j \in [1, \dots, M]$ . Notice that each window generated the same number of configurations  $N$ . The MBAR estimator has been proven to be optimal by using

$$\alpha_{i,j}(q) = \frac{1/c'_j}{\sum_{k=1}^M e^{-\beta U_k^n(q)}/c'_k} \quad (\text{A12})$$

and by summing over  $j$ :

$$c'_i \sum_{j=1}^M \left\langle \frac{e^{-\beta U_i^n}/c'_j}{\sum_{k=1}^M e^{-\beta U_k^n}/c'_k} \right\rangle'_i = \sum_{j=1}^M c'_j \left\langle \frac{e^{-\beta U_i^n}/c'_j}{\sum_{k=1}^M e^{-\beta U_k^n}/c'_k} \right\rangle'_j \quad (\text{A13})$$

We obtain a set of  $M$  equations for all  $i \in [1, \dots, M]$

$$c'_i = \sum_{j=1}^M \left\langle \frac{e^{-\beta U_i(q)}}{\sum_{k=1}^M e^{-\beta U_k(q)}/c'_k} \right\rangle'_j \quad (\text{A14})$$

Using eq A3 we obtain the estimators

$$\hat{c}'_i = \frac{1}{N} \sum_{j=1}^M \sum_{n=1}^N \frac{\omega_{j,n} e^{-\beta U_i(q_{j,n})}}{\sum_{k=1}^M e^{-\beta U_k(q_{j,n})}/\hat{c}'_k} \quad (\text{A15})$$

and finally with eq A10

$$F'_i = -\frac{1}{\beta} \ln \left( \frac{1}{N} \sum_{j=1}^M \sum_{n=1}^N \frac{\omega_{j,n} e^{-\beta U_i(q_{j,n})}}{\sum_{k=1}^M e^{\beta(F'_k - U_k(q_{j,n}))}} \right) \quad (\text{A16})$$

which must be solve self-consistently.

**Modified Rao–Blackwell Estimator.** Recently, Ding et al.<sup>62,63</sup> derived the MBAR equations using the Rao–Blackwell (RB) estimator. The RB theorem characterizes the transformation of a crude estimator into a better estimator that has smaller mean squared error, with respect to the dataset.

We wish to calculate the  $i \in [1, \dots, M]$  relative free energies  $F_i^*$  of  $M$  thermodynamic states sampled independently, with potential  $U_i$ . To compute the relative free energies, the system should be sampled according to the Boltzmann distribution. We note  $q_{i,n}$  with the  $n \in [1, \dots, N_i]$  configurations sampled from state  $i$ . To compute the relative free energies of the  $M$  thermodynamic states, the configurations  $q_{i,n}$  are combined and considered as samples from the generalized ensemble  $p_i(q) \propto e^{-\beta(U_i(q) + b_i)}$ , where  $b_i$  is an unknown biased energy. This biased energy was introduced<sup>62</sup> to adjust the relative weight of state  $i$  to be proportional to  $N_i$ , leading to



$$F_i = F_i^* + b_i = -\frac{1}{\beta} \ln \frac{N_i}{N} \quad (20)$$

where  $N = \sum_{i=1}^M N_i$ . From this equation, we can then use the RB estimator

$$\begin{aligned} F_i &= -\frac{1}{\beta} \ln p_i = -\frac{1}{\beta} \ln \frac{1}{N} \sum_{j=1}^M \sum_{n=1}^N p_j(q_{j,n}) \\ &= -\frac{1}{\beta} \ln \frac{1}{N} \sum_{j=1}^M \sum_{n=1}^N \frac{\omega_{j,n} e^{-\beta(U_i(q_{j,n}) + b_i)}}{\sum_{k=1}^M e^{-\beta(U_k(q_{j,n}) + b_i)}} \end{aligned} \quad (A18)$$

Combining with eq 20:

$$1 = \frac{1}{N} \sum_{j=1}^M \sum_{n=1}^N \frac{\omega_{j,n} e^{-\beta(U_i(q_{j,n}) + b_i)}}{\sum_{k=1}^M e^{-\beta(U_k(q_{j,n}) + b_i)}} \quad (A19)$$

Thus, the unbiased free energy  $F_i^*$  can be calculated using eq 20 after solving eq A19 for  $b_i$ . Equation A19 has major interests: (1) it is more stable, (2) it reduces the number of floating point operations, and (3) the problem is reduced to minimizing a convex function. Indeed, if we define

$$g_i(b_1, \dots, b_M) = \frac{1}{N} \sum_{j=1}^M \sum_{n=1}^N \frac{\omega_{j,n} e^{-\beta(U_i(q_{j,n}) + b_i)}}{\sum_{k=1}^M e^{-\beta(U_k(q_{j,n}) + b_i)}} - 1 \quad (A20)$$

then solving eq A19 is equivalent to finding the zeros of  $(g_1, \dots, g_M)$ . Moreover, we can remark that the function  $g_i = \nabla_{b_i} f$  where the function  $f$  is given by

$$f(b_1, \dots, b_M) = -\frac{1}{N} \sum_{j=1}^M \sum_{n=1}^N \omega_{j,n} \ln \left( \sum_{k=1}^M e^{-\beta(U_k(q_{j,n}) + b_k)} \right) - \sum_{j=1}^M b_j' \quad (A21)$$

which means solving eq A19 is equivalent to finding the critical points of  $f$ . Ding et al. have shown that  $f$  is convex, so the problem is reduced to minimizing this function which can be done with the L-BFGS method. The reweighting procedure, which uses part of the FastMBAR code, takes a few minutes on a single GPU. In this work we used the latter procedure, thanks to its GPU efficiency.

**Third Step: ASUS-GaMD Reweighting.** With either using the MBAR or the RB estimator procedure, we can extract the still-biased free energies. The final step is to derive an expression of  $\langle \varphi \rangle'$ , with respect to either  $\hat{c}_k'$  or  $F_k'$ . By setting  $c_0 = \int e^{-\beta U'(q)} dq$  and using (A3),

$$\begin{aligned} c_0 \langle \varphi \rangle' &= \int \varphi(q) e^{-\beta U'(q)} dq \\ &= \sum_{i=1}^M \frac{\int \varphi(q) e^{-\beta U'(q)} e^{-\beta U_i''(q)} / c_i' dq}{\sum_{j=1}^M e^{-\beta U_j''(q)} / c_j'} \\ &= \sum_{i=1}^M \left\langle \frac{\varphi}{\sum_{j=1}^M e^{-\beta U_j(q)} / c_j'} \right\rangle_i \\ &\simeq \frac{1}{N} \sum_{i=1}^M \sum_{n=1}^N \frac{\varphi(q_{i,n}) \omega_{i,n}}{\sum_{k=1}^M e^{-\beta U_k(q_{i,n})} / \hat{c}_k'} \end{aligned} \quad (A22)$$

in other words,

$$c_0 \langle \varphi \rangle' \simeq \frac{1}{NM} \sum_{i=1}^M \sum_{n=1}^N \varphi(q_{i,n}) r_{i,n} \quad (A23)$$

with  $r_{i,n}$  the weight of configuration  $q(i, n)$ :

$$r_{i,n} = \frac{M \omega_{i,n}}{\sum_{k=1}^M e^{-\beta U_k(q_{i,n})} / \hat{c}_k'} \quad (A24)$$

$c_0$  is unknown but is not dependent on  $k \in [[1, \dots, K]]$  so eq A9 can be rewritten as

$$\begin{aligned} F(x_k) &\simeq -\frac{1}{\beta} \ln \langle c_0 \varphi_k \rangle' - \beta \frac{c_0 \langle \varphi_k U^{\text{GaMD}} \rangle'}{c_0 \langle \varphi_k \rangle'} \\ &\quad - \frac{\beta^2}{2} \left( \frac{c_0 \langle \varphi_k (U^{\text{GaMD}})^2 \rangle'}{c_0 \langle \varphi_k \rangle'} - \left( \frac{c_0 \langle \varphi_k U^{\text{GaMD}} \rangle'}{c_0 \langle \varphi_k \rangle'} \right)^2 \right) + C' \end{aligned} \quad (A25)$$

with

$$\begin{aligned} c_0 \langle \varphi_k \rangle' &\simeq \frac{1}{NM} \sum_{i=1}^M \sum_{n=1}^N r_{i,n} \mathbf{1}_{\xi(q_{i,n}) \in \text{Bin}(x_k)} \\ c_0 \langle \varphi_k U^{\text{GaMD}} \rangle' &\simeq \frac{1}{NM} \sum_{i=1}^M \sum_{n=1}^N U^{\text{GaMD}}(q_{i,n}) r_{i,n} \mathbf{1}_{\xi(q_{i,n}) \in \text{Bin}(x_k)} \end{aligned} \quad (A26)$$

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.1c01024>.

CPUs scalability of GaMD (Figure S1); parametrization of GaMD parameters for the alanine dipeptide (Figure S2, Tables S1 and S2); free-energy surface of alanine dipeptide obtained from a long cMD simulation (Figure S3); the GaMD influence on the water diffusion property (Table S3); evolution of anharmonicity and basins's populations, as a function of the simulation time for the alanine dipeptide (Figure S4); boost distribution and anharmonicity from AMOEBA GaMD simulations on the alanine dipeptide (Figure S5); GaMD applied with Velocity Verlet and BAOAB-RESPA1 multi-time-step integrator (Figure S6); evolution of the AS-GaMD sampling, as a function of the simulation time (Figure S7); parametrization of GaMD parameters for the CD2-CD58 (Figure S8 and Table S4); U.S. convergence on CD2-CD58 (Figure S9); GaMD-US convergence on CD2-CD58 (Figure S10); ASUS-GaMD convergence on CD2-CD58 (Figure S11); US, GaMD-US, and ASUS-GaMD extrapolated convergence (Figure S12) (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

Jean-Philip Piquemal – LCT, UMR 7616 CNRS, Sorbonne Université, Paris 75005, France; The University of Texas at Austin, Department of Biomedical Engineering, Austin, Texas 78705, United States; Institut Universitaire de France, Paris 75005, France; [orcid.org/0000-0001-6615-9426](https://orcid.org/0000-0001-6615-9426); Email: [jean-philip.piquemal@sorbonne-universite.fr](mailto:jean-philip.piquemal@sorbonne-universite.fr)



## Authors

**Frédéric Célerse** – LCT, UMR 7616 CNRS, Sorbonne Université, Paris 75005, France; IPCM, UMR 8232 CNRS, Sorbonne Université, Paris 75005, France; Present Address: LCMD EPFL, CH-1015 Lausanne, Switzerland

**Théo Jaffrelot Inizan** – LCT, UMR 7616 CNRS, Sorbonne Université, Paris 75005, France

**Louis Lagardère** – LCT, UMR 7616 CNRS, Sorbonne Université, Paris 75005, France; IP2CT, FR 2622 CNRS, Sorbonne Université, Paris 75005, France

**Olivier Adjoua** – LCT, UMR 7616 CNRS, Sorbonne Université, Paris 75005, France

**Pierre Monmarché** – LCT, UMR 7616 CNRS, Sorbonne Université, Paris 75005, France; LJLL, UMR 7598 CNRS, Sorbonne Université, Paris 75005, France

**Yinglong Miao** – Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66045, United States; [orcid.org/0000-0003-3714-1395](https://orcid.org/0000-0003-3714-1395)

**Etienne Derat** – IPCM, UMR 8232 CNRS, Sorbonne Université, Paris 75005, France

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jctc.1c01024>

## Author Contributions

<sup>†</sup>These authors contributed equally to this work.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work has received funding from the European Research Council (ERC), under the European Union's Horizon 2020 Research and Innovation Program (Grant Agreement No. 810367), Project EMC2 (JPP). F.C. is thankful for funding from the French state funds managed by the CalSimLab LABEX and the ANR within the Investissements d'Avenir program (Reference No. ANR11-IDEX-0004-02) and support from the Direction Générale de l'Armement (DGA) Maîtrise NRBC of the French Ministry of Defense. Computations have been performed at GENCI (IDRIS, Orsay, France and TGCC, Bruyères le Chatel) on Grant No. A0070707671.

## REFERENCES

- (1) Caves, L. S.; Evanseck, J. D.; Karplus, M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.* **1998**, *7*, 649–666.
- (2) Schlitter, J.; Engels, M.; Krüger, P. Targeted molecular dynamics: a new approach for searching pathways of conformational transitions. *J. Mol. Graph.* **1994**, *12*, 84–89.
- (3) Markwick, P. R.; McCammon, J. A. Studying functional dynamics in bio-molecules using accelerated molecular dynamics. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20053–20065.
- (4) Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J. et al. Millisecond-scale molecular dynamics simulations on Anton. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis (SC '09)*, Portland, OR, November 2009; ACM Press, 2009; pp 1–11, Article No. 39.
- (5) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Shaw, D. E. Picosecond to Millisecond Structural Dynamics in Human Ubiquitin. *J. Phys. Chem. B* **2016**, *120*, 8313–8320.
- (6) Jaffrelot Inizan, T.; Célerse, F.; Adjoua, O.; El Ahdab, D.; Jolly, L.-H.; Liu, C.; Ren, P.; Montes, M.; Lagarde, N.; Lagardère, L.; Monmarché, P.; Piquemal, J.-P. High-resolution mining of the SARS-

CoV-2 main protease conformational space: supercomputer-driven unsupervised adaptive sampling. *Chem. Sci.* **2021**, *12*, 4889–4907.

(7) Cho, K.; Joannopoulos, J. Ergodicity and dynamical properties of constant-temperature molecular dynamics. *Phys. Rev. A* **1992**, *45*, 7089.

(8) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919–11929.

(9) Voter, A. F. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.* **1997**, *78*, 3908.

(10) Pierce, L. C. T.; Salomon-Ferrer, R.; de Oliveira, C. A.; McCammon, J. A.; Walker, R. C. Routine access to millisecond time scale events with accelerated molecular dynamics. *J. Chem. Theory. Comput.* **2012**, *8*, 2997–3002.

(11) Miao, Y.; Feixas, F.; Eun, C.; McCammon, J. A. Accelerated molecular dynamics simulations of protein folding. *J. Comput. Chem.* **2015**, *36*, 1536–1549.

(12) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P.; et al. Tinker-HP: A massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018**, *9*, 956–972.

(13) Götz, 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*, 1542–1555.

(14) Páll, S.; Zhmurov, A.; Bauer, P.; Abraham, M.; Lundborg, M.; Gray, A.; Hess, B.; Lindahl, E. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J. Chem. Phys.* **2020**, *153*, 134110.

(15) Adjoua, O.; Lagardère, L.; Jolly, L.-H.; Durocher, A.; Very, T.; Dupays, I.; Wang, Z.; Inizan, T. J.; Célerse, F.; Ren, P.; Ponder, J. W.; Piquemal, J.-P.; et al. Tinker-HP: Accelerating Molecular Dynamics Simulations of Large Complex Systems with Advanced Point Dipole Polarizable Force Fields Using GPUs and Multi-GPU Systems. *J. Chem. Theory. Comput.* **2021**, *17*, 2034–2053.

(16) Tuckerman, M. E.; Berne, B. J.; Rossi, A. Molecular dynamics algorithm for multiple time scales: Systems with disparate masses. *J. Chem. Phys.* **1991**, *94*, 1465–1469.

(17) Lagardère, L.; Aviat, F.; Piquemal, J.-P. Pushing the limits of multiple-time-step strategies for polarizable point dipole molecular dynamics. *J. Phys. Chem. Lett.* **2019**, *10*, 2593–2599.

(18) Fiorin, G.; Klein, M. L.; Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **2013**, *111*, 3345–3362.

(19) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

(20) Brotzakis, Z. F.; Limongelli, V.; Parrinello, M. Accelerating the calculation of protein–ligand binding free energy and residence times using dynamically optimized collective variables. *J. Chem. Theory. Comput.* **2019**, *15*, 743–750.

(21) Kabelka, I.; Brozek, R.; Vácha, R. Selecting Collective Variables and Free-Energy Methods for Peptide Translocation across Membranes. *J. Chem. Inf. Model.* **2021**, *61*, 819–830.

(22) Hovan, L.; Comitani, F.; Gervasio, F. L. Defining an optimal metric for the path collective variables. *J. Chem. Theory. Comput.* **2019**, *15*, 25–32.

(23) Bonati, L.; Rizzi, V.; Parrinello, M. Data-driven collective variables for enhanced sampling. *J. Phys. Chem. Lett.* **2020**, *11*, 2998–3004.

(24) Hashemian, B.; Millán, D.; Arroyo, M. Modeling and enhanced sampling of molecular systems with smooth and nonlinear data-driven collective variables. *J. Chem. Phys.* **2013**, *139*, 214101.

(25) Chen, P.-Y.; Tuckerman, M. E. Molecular dynamics based enhanced sampling of collective variables with very large time steps. *J. Chem. Phys.* **2018**, *148*, 024106.

- (26) Henin, J.; Fiorin, G.; Chipot, C.; Klein, M. L. Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *J. Chem. Theory. Comput.* **2010**, *6*, 35–47.
- (27) Gardner, J. M.; Abrams, C. F. Energetics of flap opening in HIV-1 protease: string method calculations. *J. Phys. Chem. B* **2019**, *123*, 9584–9591.
- (28) Célerse, F.; Lagardère, L.; Derat, E.; Piquemal, J.-P. Massively parallel implementation of Steered Molecular Dynamics in Tinker-HP: comparisons of polarizable and non-polarizable simulations of realistic systems. *J. Chem. Theory. Comput.* **2019**, *15*, 3694–3709.
- (29) Rodriguez, A.; d'Errico, M.; Facco, E.; Laio, A. Computing the free energy without collective variables. *J. Chem. Theory. Comput.* **2018**, *14*, 1206–1215.
- (30) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation. *J. Chem. Theory. Comput.* **2015**, *11*, 3584–3595.
- (31) Melcr, J.; Piquemal, J.-P. Accurate Biomolecular Simulations Account for Electronic Polarization. *Front. Mol. Biosci.* **2019**, *6*, 143.
- (32) Shi, Y.; Ren, P.; Schnieders, M.; Piquemal, J.-P. *Reviews in Computational Chemistry*, Vol. 28; John Wiley & Sons, Ltd., 2015; Chapter 2, pp 51–86; DOI: 10.1002/9781118889886.ch2.
- (33) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Ann. Rev. Biophys.* **2019**, *48*, 371–394.
- (34) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960–1986.
- (35) Ren, P.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **2003**, *107*, 5933–5947.
- (36) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T.; et al. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.
- (37) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (38) Oshima, H.; Re, S.; Sugita, Y. Replica-exchange umbrella sampling combined with gaussian accelerated molecular dynamics for free-energy calculation of biomolecules. *J. Chem. Theory. Comput.* **2019**, *15*, 5199–5208.
- (39) Miao, Y.; Bhattarai, A.; Wang, J. Ligand Gaussian accelerated molecular dynamics (LiGaMD): Characterization of ligand binding thermodynamics and kinetics. *J. Chem. Theory. Comput.* **2020**, *16*, 5526–5547.
- (40) Wang, J.; Miao, Y. Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. *J. Chem. Phys.* **2020**, *153*, 154109.
- (41) Miao, Y.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Theory, Implementation, and Applications. In *Annual Reports in Computational Chemistry*, Vol. 13; Elsevier, 2017; pp 231–278.
- (42) Wang, J.; Arantes, P. R.; Bhattarai, A.; Hsu, R. V.; Pawnikar, S.; Huang, Y.-m. M.; Palermo, G.; Miao, Y. Gaussian accelerated molecular dynamics: Principles and applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, *11*, e1521.
- (43) Wang, Y.-T.; Chan, Y.-H. Understanding the molecular basis of agonist/antagonist mechanism of human mu opioid receptor through gaussian accelerated molecular dynamics method. *Sci. Rep.* **2017**, *7*, 7828.
- (44) Palermo, G. Structure and dynamics of the CRISPR–Cas9 catalytic complex. *J. Chem. Inf. Model.* **2019**, *59*, 2394–2406.
- (45) de Courcy, B.; Piquemal, J.-P.; Garbay, C.; Gresh, N. Polarizable water molecules in ligand-macromolecule recognition. Impact on the relative affinities of competing pyrrolopyrimidine inhibitors for FAK kinase. *J. Am. Chem. Soc.* **2010**, *132*, 3312–3320.
- (46) El Ahdab, D.; Lagardère, L.; Inizan, T. J.; Célerse, F.; Liu, C.; Adjoua, O.; Jolly, L.-H.; Gresh, N.; Hobaika, Z.; Ren, P.; Maroun, R. G.; Piquemal, J.-P. Interfacial Water Many-Body Effects Drive Structural Dynamics and Allosteric Interactions in SARS-CoV-2 Main Protease Dimerization Interface. *J. Phys. Chem. Lett.* **2021**, *12*, 6218–6226.
- (47) Leimkuhler, B.; Matthews, C. Robust and efficient configurational molecular sampling via Langevin dynamics. *J. Chem. Phys.* **2013**, *138*, 174102.
- (48) Mark, P.; Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960.
- (49) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kucsera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M.; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (50) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable atomic multipole-based AMOEBA force field for proteins. *J. Chem. Theory. Comput.* **2013**, *9*, 4046–4063.
- (51) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA polarizable atomic multipole force field for nucleic acids. *J. Chem. Theory. Comput.* **2018**, *14*, 2084–2108.
- (52) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (53) Chow, K.-H.; Ferguson, D. M. Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. *Comput. Phys. Commun.* **1995**, *91*, 283–289.
- (54) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **1982**, *76*, 637–649.
- (55) 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*, 8577–8593.
- (56) You, W.; Tang, Z.; Chang, C.-e. A. Potential mean force from umbrella sampling simulations: What can we learn and what is missed? *J. Chem. Theory. Comput.* **2019**, *15*, 2433–2443.
- (57) Baştuğ, T.; Chen, P.-C.; Patra, S. M.; Kuyucak, S. Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling. *J. Chem. Phys.* **2008**, *128*, 155104.
- (58) Mills, M.; Andricioaei, I. An experimentally guided umbrella sampling protocol for biomolecules. *J. Chem. Phys.* **2008**, *129*, 114101.
- (59) Bayas, M. V.; Kearney, A.; Avramovic, A.; Van Der Merwe, P. A.; Leckband, D. E. Impact of salt bridges on the equilibrium binding and adhesion of human CD2 and CD58. *J. Biol. Chem.* **2007**, *282*, 5589–5596.
- (60) Debiec, K. T.; Gronenborn, A. M.; Chong, L. T. Evaluating the strength of salt bridges: a comparison of current biomolecular force fields. *J. Phys. Chem. B* **2014**, *118*, 6561–6569.
- (61) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.
- (62) Ding, X.; Vilseck, J. Z.; Hayes, R. L.; Brooks, C. L. Gibbs Sampler-Based Dynamics and Rao–Blackwell Estimator for Alchemical Free Energy Calculation. *J. Chem. Theory Comput.* **2017**, *13*, 2501–2510.
- (63) Ding, X.; Vilseck, J. Z.; Brooks, C. L. Fast Solver for Large Scale Multistate Bennett Acceptance Ratio Equations. *J. Chem. Theory. Comput.* **2019**, *15*, 799–802.