On the Generation of Quantum Configurational Ensembles Using Approximate Potentials

João Morado,[†] Paul N. Mortenson,[‡] J. Willem M. Nissink,[¶] Marcel L. Verdonk,[‡] Richard A. Ward,[¶] Jonathan W. Essex,[*,†] and Chris-Kriton Skylaris[*,†]

[†]*School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom*

[‡]*Astex Pharmaceuticals, 436 Cambridge Science Park, Milton Road, Cambridge CB4 0QA, United Kingdom*

[¶]*Medicinal Chemistry, Oncology R&D, AstraZeneca, Cambridge CB4 0WG, UK*

E-mail: j.w.essex@soton.ac.uk; c.skylaris@soton.ac.uk

Abstract

Conformational analysis is of paramount importance in drug design: it is crucial to determine pharmacological properties, understand molecular recognition processes, and characterize the conformations of ligands when unbound. Molecular Mechanics (MM) simulation methods, such as Monte Carlo (MC) and molecular dynamics (MD), are usually employed to generate ensembles of structures due to their ability to extensively sample the conformational space of molecules. The accuracy of these MM-based schemes strongly depends on the functional form of the force field (FF) and its parameterization, components that often hinder their performance. High-level methods, such as *ab initio* MD, provide reliable structural information but are still too computationally expensive to allow for extensive sampling. Therefore, to overcome these limitations, we present a multi-level MC method that is capable of generating quantum configurational ensembles while keeping the computational cost at a minimum.

We show that FF reparameterization is an efficient route to generate FFs that reproduce QM results more closely, which in turn can be used as low-cost models to achieve the gold standard QM accuracy. We demonstrate that the MC acceptance rate is strongly correlated with various phase space overlap measurements and that it constitutes a robust metric to evaluate the similarity between the MM and QM levels of theory. As more advanced applications, we present a self-parameterizing version of the algorithm, which combines sampling and FF parameterization in one scheme, and apply the methodology to generate the QM/MM distribution of a ligand in aqueous solution.

1 Introduction

The study of the conformational dynamics of molecules free in solution is essential for predicting molecular properties and to guide the rational development of new pharmaceutical compounds. The latter application is of utmost importance for the pharmaceutical industry since knowledge of the unbound state is vital to understand the fundamentals of molecular recognition.[1–5] Besides the displacement of water from protein binding sites,[6–8] one of the main phenomena that impacts binding affinity is the reorganization of the unbound state ligand upon binding to its target, a process that is influenced by the change in intramolecular energy of the ligand in adopting the bioactive conformer, as well as the associated loss of entropy.[4] Minimization of the free energy penalty associated with this structural change is vital to optimizing ligand potency, requiring knowledge of the physical interactions that control conformational preferences and methods for conformational analysis if a rational strategy is to be employed.[9] There is a wide range of experimental structural information of pharmaceutical compounds bound to their protein targets.[10,11] However, as it has been emphasized in various studies, the conformations of unbound compounds are still poorly characterized.[2,3,5,12] Therefore, the scientific community must put effort into developing tools that allow fast and reliable characterization of unbound molecular conformers as these can

potentially provide the so-called "missing link" in structure-based drug discovery.[2,12]

The most widely used experimental method to elucidate unbound conformational ensembles is NMR spectroscopy, which is often utilized in drug design to complement X-ray protein–ligand structural information.[9,12,13] Additionally, molecular mechanics (MM) simulation methods, such as molecular dynamics (MD) and Monte Carlo (MC), are also commonly employed to predict thermophysical molecular properties and generate structures for conformational analysis.[3,14] In particular, static properties, such as, *e.g.*, optical spectra, NMR spectra, and solvation free energies, can be determined from the relative populations of the free state conformers,[15] which are usually possible to estimate in MM-based simulations since, in many instances, these permit ergodic sampling.[16–21] Although these methods allow extensive sampling of the configurational space of molecules, the functional form used by the FF affects the sampling quality, and parameterization must be adequate to ensure accurate results. High-level simulation schemes, such as, *e.g.*, *ab initio* MD, have become the gold standard for simulation purposes as they provide reliable structural information at the quantum level, but are still too computationally expensive to allow achieving the time scales typically required for convergence of the simulations.[22–26] Hence, to attain extensive and reliable sampling, it is necessary to find a compromise between the efficiency of the MM-based methods and the accuracy of the quantum mechanics (QM) level of theory.

Several approaches have already been proposed to sample molecular conformations with QM accuracy at a nearly MM cost. In this context, Rosa *et al*[15] proposed a post-processing method wherein, through the use of conformational clustering and thermodynamic perturbation theory, it is possible to estimate the QM populations by correcting MM populations. Others have attempted to explore the conformational landscape of bioactive small molecules by using a combination of classical Hamiltonian replica exchange with high-level QM calculations.[27] In this work, we attempt to bridge the gap between the MM efficiency and the QM accuracy by presenting a methodology that is based on an *ab initio* MC algorithm. This approach enables recovery of the correct QM ensembles while keeping the computational

cost at a minimum. It is also capable of self-parameterizing FFs to a target level of theory in an iterative and on-the-fly fashion, a feature that can be applied whenever generation of high-quality FFs is required.

The method we propose in this paper uses a nested Markov chain Monte Carlo (nMC-MC) algorithm that combines sampling at the MM level with periodic switching attempts to the QM level. This nMC-MC algorithm works by firstly resorting to the hybrid Monte Carlo (hMC) scheme to rigorously generate configurations that belong to a target MM ensemble.[28–30] These are subsequently used as trial states for a second Markov chain, wherein they are accepted or rejected according to a correction step based on the difference between the MM and QM potentials.[31,32] In this way, it is possible to generate quantum configurational ensembles using approximate potentials (MM FFs). This multi-level *ab initio* MC algorithm has already been applied in various contexts, such as in the fitting of FF dihedral angles,[33] in a "stepping stone" approach for obtaining quantum free energies of hydration,[34] in a MC resampling approach for the calculation of hybrid classical/quantum free energies,[35] to improve the efficiency of Born models in MC simulations,[36] to model reactivity in small molecules,[37] and to enhance the conformational sampling of disordered regions of proteins.[38,39]

It is widely known that the key factor for convergence of multi-level approaches is ensuring a favorable overlap between the energy distributions of different levels of theory.[40–42] Otherwise, as FFs often predict conformations and energies that substantially deviate from the QM level, low acceptance rates are obtained when attempting to sample from the MM to the QM chain. Ultimately, the mismatch between the MM and QM descriptions becomes a bottleneck as it prevents a thorough exploration of the relevant QM potential energy surfaces (PES) and slows convergence of the sampling of the target quantum configurational distributions. Different strategies have already been proposed to improve the overlap between the probability distributions of the energies associated with different levels of theory.[40] One possibility is to artificially broaden the MM distribution by manipulating the thermodynamic variables (*e.g.*, pressure and temperature) characterizing the reference system[43–46] or using

Tsallis statistics.[47] It is also possible to introduce an intermediate level of theory (*e.g.*, a semi-empirical QM method) to bridge the gap between the MM and QM chains or employ an arbitrary number of intermediate potential energy layers with sufficient overlap between their probability distributions.[48] Another option is to increase the overlap between the distributions by improving the MM description so that it becomes more QM-like, which is the approach followed in this study. It can be achieved through either FF reparameterization (typically by force-matching[41,42,49]), by using machine-learning potentials,[50,51] or through fitting of *ad hoc* potentials.[43,52–54]

As application examples, we test the proposed methodology on a set of small organic molecules of increasing complexity which are representative fragments of molecules found in drug discovery programs. Specifically, we attempt to generate quantum configurational ensembles of aniline, acetanilide, biphenyl, diphenyl ether, and sulfanilamide. As a relevant druglike example, we investigate a fragment of cpd 26, which is the core of an efficacious low nM antagonist of the inhibitor of apoptosis proteins cIAP1 and XIAP.[55] Furthermore, as proof of principle, we use octahydrotetracene to demonstrate that the nMC-MC algorithm can be coupled with a reparameterization step, allowing for iterative optimization of the molecule's FF parameters using the on-the-fly QM-generated ensemble. This self-parameterization nMC-MC algorithm is similar in philosophy to the methods presented in some past applications,[50,52–54,56] though these studies did not use MM FFs or druglike molecules. Finally, as a more advanced application, we apply the nMC-MC algorithm to generate the QM/MM[57] distribution of aniline in aqueous solution.

This paper is structured as follows: we first present the basic theory underlying the proposed approach, *viz.* the hMC method, the nMC-MC algorithm, the FF reparameterization approach, the phase space overlap metrics, and the numerical experiments protocol. We then present applications of the algorithms to the previously mentioned test cases and conclude with final remarks.

2 Theory and Methods

2.1 Hybrid Monte Carlo

hMC is an exact sampling approach that combines the features of the MC and MD simulation methods in such a way that the trial steps of the MC algorithm are short MD runs. Therefore, hMC inherits the advantages of both algorithms, such as the tendency of MD to move the system towards regions of configuration space that are energetically favorable, and the possibility to relax the restriction on the size of the MD time step, dt, through the application of a MC step.[58] hMC also prevents the numerical instabilities that arise due to the finite time step size of MD algorithms.[59]

At the start of every iteration, hMC draws new velocities from the Maxwell-Boltzmann distribution at a chosen temperature T_K, a step which is performed using the Marsaglia polar method;[34,60] then, a short MD simulation in the microcanonical ensemble (NVE) is run using a sympletic integrator that preserves detailed balance[61] (*e.g.* the velocity-Verlet algorithm,[62,63] the integrator used in this study) during M steps; finally, the final configuration of the system is accepted or rejected according to a given acceptance criterion, which for the canonical ensemble (NVT) at temperature T_U reads[28]

$$\phi(\boldsymbol{q_i} \rightarrow \boldsymbol{q_f}) = \min\left\{1, \exp\left[-\beta_K \Delta K - \beta_U \Delta U^{MM}(\boldsymbol{q_i}, \boldsymbol{q_f})\right]\right\} \tag{1}$$

where $\Delta K = K_f - K_i$ is the difference of the kinetic energy between the final and initial states of the short MD run, $\Delta U^{MM} = U^{MM}(\boldsymbol{q_f}) - U^{MM}(\boldsymbol{q_i})$ is the difference between the potential energy of the system at configurations $\boldsymbol{q_f}$ and $\boldsymbol{q_i}$, and β_K and β_U are the thermodynamics betas corresponding to the temperatures T_K and T_U, respectively. The latter do not need have the same value, a feature that can be used as a means of increasing the conformational sampling efficiency by, for example, using high T_K values for the kinetic energy component. It is worth mentioning that M and dt are hyperparameters that have to be chosen properly in order to ensure sampling of uncorrelated snapshots while keeping the

wall time required for each MD manageable.[34] These hyperparameters may have an impact on the acceptance rates.[30,34,64]

The disadvantage of hMC is that its acceptance probability decays exponentially with the system size because the root-mean-square error in the energy increases with $N_a^{1/2}$,[30,65] where N_a is the number of atoms of the system. There have been many attempts to circumvent this bottleneck, the most widely studied being sampling from shadow Hamiltonians.[30,66] Nevertheless, owing to the relatively small size of the molecular systems covered in this study, this issue does not pose a problem for the current purposes. As will be next discussed, hMC can also be embedded in a nMC-MC algorithm, wherein the hMC moves are used as the trial steps of an *ab initio* MC algorithm.

2.2 Sampling From Approximate Potentials

As it was formalized by Gelb in his seminal work about sampling from approximate potentials, it is possible to create a nMC-MC simulation by coupling hMC with a correction step based on the difference between the MM and QM potentials.[31,32] The expression of this correction step reads

$$\theta(\boldsymbol{q_i} \rightarrow \boldsymbol{q_f}) = \min\left\{1, \exp\left[-\beta'_U \Delta U^{QM}(\boldsymbol{q_i}, \boldsymbol{q_f}) + \beta_U \Delta U^{MM}(\boldsymbol{q_i}, \boldsymbol{q_f})\right]\right\} \tag{2}$$

where $\Delta U^{MM} = U^{MM}(\boldsymbol{q_f}) - U^{MM}(\boldsymbol{q_i})$, $\Delta U^{QM} = U^{QM}(\boldsymbol{q_f}) - U^{QM}(\boldsymbol{q_i})$, and β_U and β'_U are the thermodynamic betas of the MM and QM ensembles, respectively. Note that, as in the hMC algorithm, the beta thermodynamic parameters also do not need to be the same in both Markov chains, something that can be exploited as a way of increasing the overlap between the MM and QM theory levels.[43,44]

The nMC-MC algorithm works by firstly generating a trial structure through the hMC algorithm, which is then attempted to be sampled into the QM level of theory by applying the acceptance criterion of equation (2). If the structure is accepted, the next hMC run starts

from this configuration; otherwise, the hMC run starts from the last accepted configuration. A detailed diagram of the workflow of this algorithm is shown in figure 1.

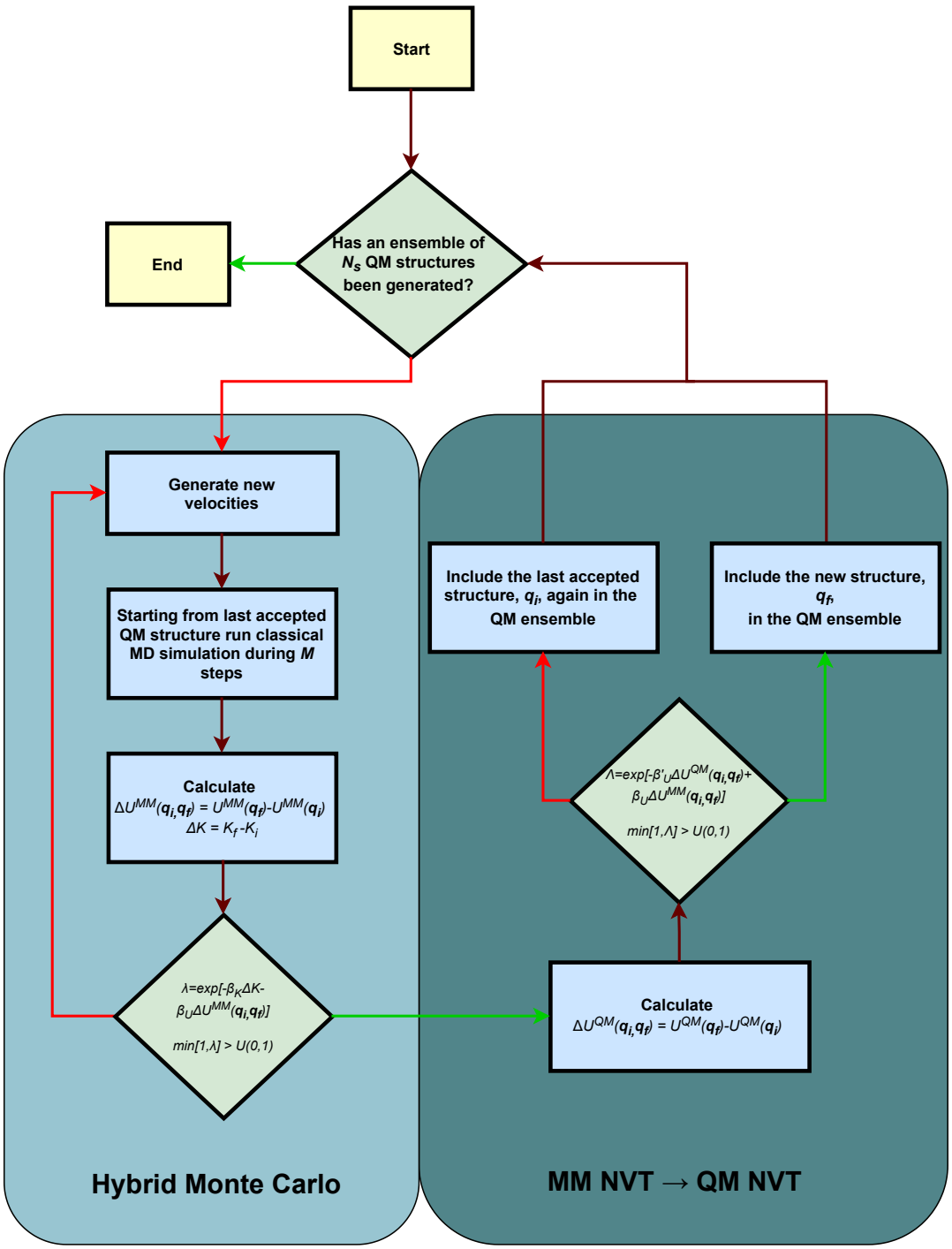

Figure 1: Diagram describing the workflow of nMC-MC algorithm as implemented in ParaMol.[67] The hMC part of the algorithm is used to generate an exact NVT ensemble (left), while the sampling from approximate potentials part is used as a switching step between the MM and QM levels of theory (right). $U(0,1)$ means a random number between 0 and 1 sampled from a uniform distribution, and the i and f subscripts refer to the initial and final states of a given iteration.

2.3 Force Field Reparameterization

The key requirement for convergence of the nMC-MC algorithm is ensuring a favorable distribution overlap between the MM FF and the QM level of theory. To fulfill this condition, we generate low-level models that are closer to the target high-level of theory than the general AMBER force field (GAFF)[68] energy function, in such a way that the overlap between the MM and QM energy difference distributions is improved. This is attained through reparameterization of GAFF-like MM FFs, a step that is performed to make the MM models more QM-like. To perform the optimization of the FF parameters, we resort to the methodologies implemented in ParaMol, wherein a FF is fitted to a target level of theory through minimization of the following objective function

$$X(\boldsymbol{p}) = X_F(\boldsymbol{p}) + X_U(\boldsymbol{p}) + \Theta(\boldsymbol{p}) \tag{3}$$

where $\boldsymbol{p}$ is the vector of parameters entering the optimization. X_F corresponds to the term of the objective function in which every component of the MM atomic forces is fitted to QM data and is given by

$$X_F(\boldsymbol{p}) = \frac{1}{3N_a} \sum_{i}^{N_s} \omega_i \sum_{j}^{N_a} \left[\Delta \boldsymbol{F}_{i,j}(\boldsymbol{p})^T \langle \boldsymbol{F}_{i,j}^{QM} \otimes \boldsymbol{F}_{i,j}^{QM} \rangle^{-1} \Delta \boldsymbol{F}_{i,j}(\boldsymbol{p}) \right] \tag{4}$$

where ω_i is the weight of the i-th conformation, $\Delta \boldsymbol{F}_{i,j} = \boldsymbol{F}_{i,j}^{MM}(\boldsymbol{p}) - \boldsymbol{F}_{i,j}^{QM}$, $\boldsymbol{F}_{ij}^{QM}$ and $\boldsymbol{F}_{ij}^{MM}$ are the QM and MM force vectors, respectively, of atom j in conformation i, N_s is the number of structures provided and N_a the number of atoms of the system. It is worthwhile mentioning that $\langle \boldsymbol{F}_{i,j}^{QM} \otimes \boldsymbol{F}_{i,j}^{QM} \rangle$ is the covariance of the atomic forces, herein used as a normalization factor so that the residuals of the forces are dimensionless and maximally of unit magnitude. Furthermore, X_E amounts for the fitting of energies to reference QM data, and the expression used for this term reads

$$X_U(\boldsymbol{p}) = \sum_i^{N_s} \omega_i \frac{\left(U_i^{MM}(\boldsymbol{p}) - U_i^{QM} - \langle \Delta U \rangle\right)^2}{\mathrm{Var}(U^{QM})} \tag{5}$$

where U_i^{QM} and U_i^{MM} are the QM and MM potential energies, $\mathrm{Var}(U^{QM})$ is the variance of the QM energies, here used as the normalization factor, and $\langle \Delta U \rangle = \frac{1}{N_s} \sum_i \left(U_i^{QM} - U_i^{MM}\right)$ is a term that brings the two distributions together by subtracting the average difference between the QM and MM potential energies from the energy residuals. Finally, $\Theta(\boldsymbol{p})$ is a regularization term that can be optionally included to prevent overfitting. In this study, we use a harmonic penalty function (L2 regularization) that assumes that the *prior* distribution of the m-th parameter, p_m, is a Gaussian centered at the initial guess p_m^0 with width γ_m. The expression of this harmonic regularization term reads

$$\Theta(\boldsymbol{p}) = \alpha \sum_m^{N_p} \frac{(p_m - p_m^0)^2}{\gamma_m^2} \tag{6}$$

where α is an adjustable scaling factor that controls the strength of the regularization.

2.4 Phase Space Overlap Metrics

As a means of establishing the similarity between the MM and QM levels of theory, we evaluate the phase space overlap using two different metrics. The first metric resorts to the idea that the phase space overlap can be calculated as the overlap of the total energy difference distributions between the two considered levels of theory.[42] These distributions are obtained by performing MD simulations using both the MM and the QM Hamiltonians, whereupon the differences $\Delta E_{MM}^{MM \to QM} = E_{MM}^{QM} - E_{MM}^{MM}$ and $\Delta E_{QM}^{QM \to MM} = E_{QM}^{MM} - E_{QM}^{QM}$ are evaluated for the trajectories obtained utilizing the MM and QM Hamiltonians, respectively, where the subscript indicates the level of theory used for sampling, and the superscript the level of theory used to evaluate the potential energy. The corresponding histograms of the calculated ΔEs are then approximated by assuming Gaussian-shaped distributions, in such a way that the energy difference distribution of the QM Hamiltonian is given by

$$\mathcal{N}_{QM}(\Delta E_{QM}^{QM\to MM}) = \sqrt{\frac{1}{2\pi\sigma_{QM}^2}}\exp\left[-\frac{\left(\Delta E_{QM}^{QM\to MM} - \langle \Delta E_{QM}^{QM\to MM}\rangle\right)^2}{\sigma_{QM}^2}\right] \tag{7}$$

where σ_{QM} is the standard deviation of the $\Delta E_{QM}^{QM\to MM}$ values. The Gaussian representation of the energy difference distribution of the MM Hamiltonian, $\mathcal{N}_{MM}$, can be written analogously. It is finally possible to measure the overlap, Ω, between the $\mathcal{N}_{QM}$ and $\mathcal{N}_{MM}$ distributions according to the following equation

$$\Omega = \frac{\langle \mathcal{N}_{QM}, \mathcal{N}_{MM}\rangle}{\max\left[\langle \mathcal{N}_{QM}, \mathcal{N}_{QM}\rangle, \langle \mathcal{N}_{MM}, \mathcal{N}_{MM}\rangle\right]} \tag{8}$$

where $\langle f, g\rangle = \int dx\ f(x)\cdot g(x)$ is the inner product or overlap integral between the f and g functions. The integration of this overlap integral is performed numerically using SciPy's integrate.quad function with default settings.[69]

We also use the descriptors of the phase space overlap between two states that were developed by Wu and Kofke.[42,70,71] In particular, we use a metric based on the overlap of total energy distributions that reads

$$\Sigma_{MM,QM} = 2\int_{-\infty}^{+\infty} dE_{QM}\ \rho_{QM}^{QM}(E_{QM}) \int_{-\infty}^{E_{QM}} dE'_{QM}\ \rho_{MM}^{QM}(E'_{QM}) \tag{9}$$

where ρ_A^A and ρ_B^A are the probability distributions of state A energies observed within simulations of states A and B, respectively. Similarly, the expression for $\Sigma_{QM,MM}$ can be written as

$$\Sigma_{QM,MM} = 2\int_{-\infty}^{+\infty} dE_{MM}\ \rho_{MM}^{MM}(E_{MM}) \int_{-\infty}^{U_{MM}} dE'_{MM}\ \rho_{QM}^{MM}(E'_{MM}) \tag{10}$$

The value of $\Sigma_{B,A}$ varies from 0 to 2 and indicates the offset of ρ_B^A relative to ρ_A^A. If ρ_B^A is centered left with respect to ρ_A^A, then $1 < \Sigma_{B,A} \leq 2$. Otherwise, if ρ_B^A is centered right with respect to ρ_A^A, then $0 \leq \Sigma_{B,A} < 1$. The integration of the double integrals of equations (9) and (10) is performed numerically using SciPy's integrate.dblquad function with default

settings.

2.5 Numerical Experiments Protocol

The numerical experiments presented in this study rely on using refined low-level models that attempt to approximate a high level of theory. Specifically, the high level of theory used throughout this work is the DFTB+[72,73] implementation of SCC-DFTB including the D3 dispersion correction[74] with Becke-Johnson damping.[75] This choice is based on the evidence that SCC-DFTB-D3 performs quite well in determining conformations of druglike molecules and respective energies,[76–78] as well as being computationally cheap, which allows for extensive testing of various compounds. The low-level models used to improve the overlap between the energy difference distributions of the low and high levels of theory are bespoke FFs, reparameterized to reproduce the SCC-DFTB-D3 level of theory. Specifically, the reparameterized FFs consist of refined versions of GAFF, for which the functional form reads[68]

$$U = \sum_{bonds} \frac{K_b}{2}(r - r_{eq})^2 + \sum_{angles} \frac{K_\theta}{2}(\theta - \theta_{eq})^2 + \sum_{dihedrals} V_n \left[1 + \cos\left(n\phi - \gamma_n\right)\right] + \sum_{i<j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{R_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i<j} \frac{q_i q_j}{4\pi\epsilon r_{ij}} \tag{11}$$

where r_{eq} and θ_{eq} are equilibrium structural parameters; K_b, K_θ and V_n are the bond, angle and dihedral force constants, respectively; n is the dihedral multiplicity, and γ_n is the dihedral phase; ϵ_{ij} is the well depth of the Leannard-Jones interaction between atoms i and j, σ_{ij} the distance at which said interaction vanishes, r_{ij} is the distance between atoms i and j, and q is the atomic charge. These refined MM models are designed systematically such that, for every molecule shown in figure 2, the following set of reparameterized FFs is generated

- B FF - bond force constants (K_b) and equilibrium values (r_{eq}) are optimized.
- BA FF - bond and angle force constants (K_θ) and equilibrium values (θ_{eq}) are optimized.
- BAT FF - bond, angle and dihedral force constants (K_b, K_θ, and V_n), bond and angle equilibrium values (r_{eq}, and θ_{eq}), and dihedral phase constants (γ_n) are optimized.
- BAT-LJ FF - in addition to the parameters optimized in the BAT FF, the σ and ϵ Lennard-Jones 12-6 parameters are also optimized.
- BAT-Q FF - in addition to the parameters optimized in the BAT FF, the atomic charges (q) are also optimized (under the constraint of total molecular charge conservation).
- The BAT-LJQ FF - in addition to the parameters optimized in the BAT FF, the σ and ϵ Lennard-Jones 12-6 parameters are also optimized, as well as the atomic charges (q) under the constraint of total molecular charge conservation.

The optimization of the parameters is performed using ParaMol with the SciPy's SLSQP optimizer.[79] All optimizations are deemed to be converged whenever the objective function between two successive iterations does not change by more than 10^{-6}, *i.e.*, $X_{n+1} - X_n < 10^{-6}$. The original GAFF parameters are used as the initial guess for the optimizations. These are obtained by initially parameterizing the druglike molecules using Antechamber packages, which are part of AmberTools.[80] AM1-BCC charges[81,82] are calculated after the geometry is optimized at the SCC-DFTB-D3 level of theory. The topology and coordinates files used as inputs to ParaMol are created using LEaP. All FF modifications given by the frcmod file created by parmchk2 are included. No atom-type symmetries are preserved during the reparameterization, which means that the presented results are close to the limits of accuracy that the GAFF functional form can achieve.

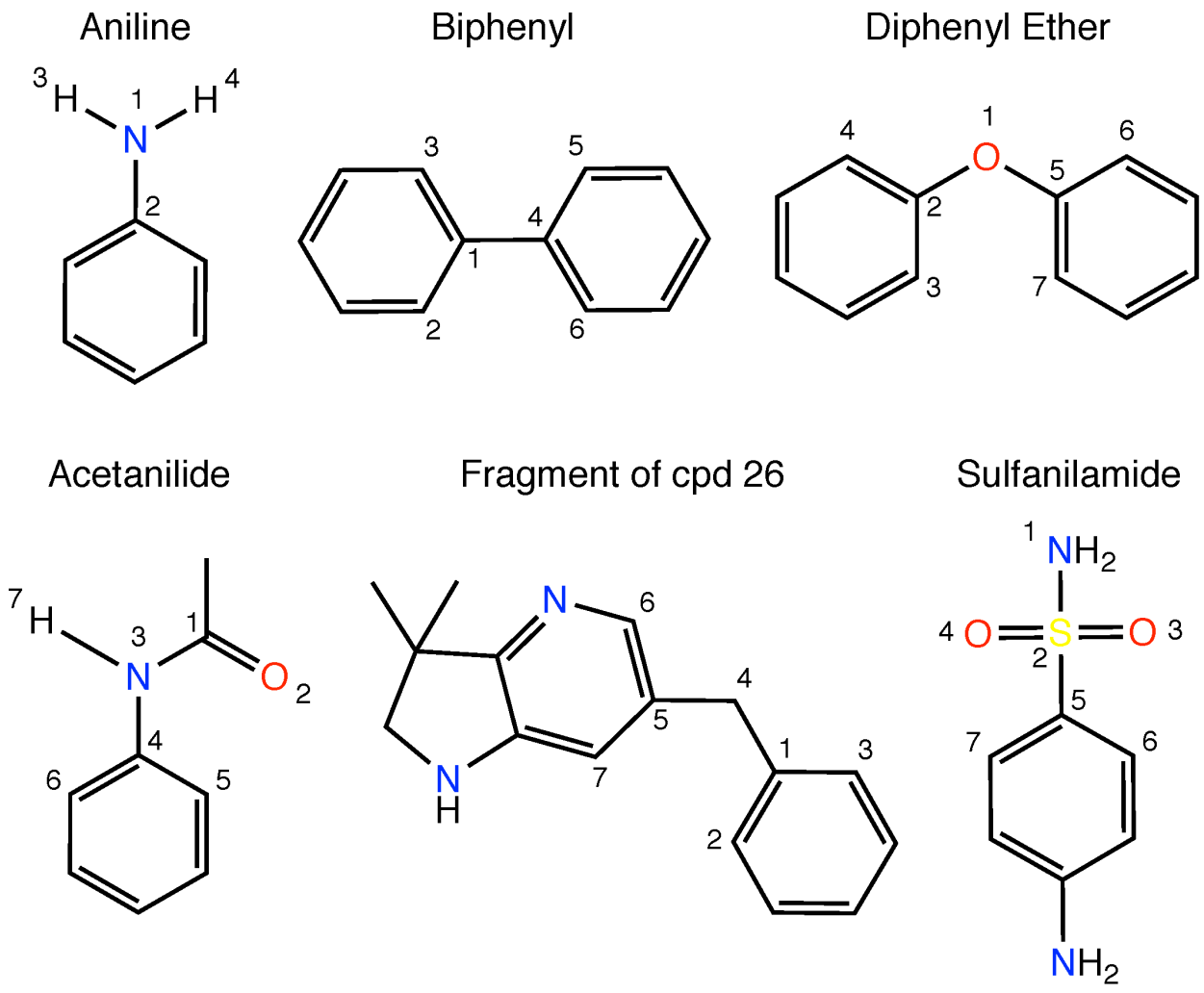

Figure 2: Molecular structures of the test molecules used in this paper.

The objective function minimized during the optimization procedure includes as targets both forces and energies, as shown in equation (3). We test reparameterizing the FFs using both the uniform and non-Boltzmann weighting (weighting temperature of 300 K) schemes available in ParaMol (see description of the weighting methods in the Supporting Information), as well as applying either no regularization or L2 regularization. Regarding the latter, the *prior* widths used throughout this study are those reported in the Supporting Information, and their choice is inspired by the values reported in refs 67 and 83. The value of the scaling factor used in regularization term is $\alpha = 1/N_p$. The training data sets consist of ensembles of configurations generated at the SCC-DFTB-D3 level of theory.

These are obtained by performing gas-phase Langevin dynamics (time step of 1 fs, and friction coefficient of 2 ps^{-1}) at a temperature of 500 K. We choose to simulate at a high temperature to ensure a thorough exploration of the SCC-DFTB-D3 conformational space. Snapshots of the simulations are collected every 1 ps, resulting in a final data set of 10000 configurations.

The nMC-MC calculations performed to estimate the acceptance rates use a time step of 1 (100 MD steps per hMC run). Velocities are sampled from the Maxwell-Boltzmann distribution at 300 K, which is also the temperature used for the MM and QM chains. No fine-tuning of these hyperparameters is attempted as these tend to be molecule-specific. A total of 4 independent samplers are run for each molecule, all starting from the same initial structure but using different random seeds.

3 Results and Discussion

3.1 nMC-MC Acceptance Rates

The most direct metric we can obtain from nMC-MC simulations are the acceptance rates. There are two of these: the hMC acceptance rate shown in equation (1), which gives information about the stability of the NVE MD runs and is useful to identify putative energy conservation issues; and the MM to QM switching step acceptance rate shown in equation (2), which gives information about the similarity between the MM and QM levels of theory. The latter is the focus of this work because, as will be discussed later, it is highly correlated with phase space overlap metrics. Therefore, it is a valuable metric of how close to the QM level of theory the MM FFs are sampling, *i.e*, it measures the MM $\rightarrow$ QM overlap. It should be noted, however, that this is a unidirectional relation since the acceptance rate does not give details about how close to the MM level of theory the QM Hamiltonian samples. Measuring the QM $\rightarrow$ MM overlap would require performing nMC-MC calculations using the QM Hamiltonian in the lower chain, which is computationally expensive and, in many

cases, unfeasible due to the requirement of performing *ab initio* MD. Hence, there is the possibility of predicting high acceptance rates that may misleadingly lead one to assume that a given MM FF is highly QM-like when, in reality, it only explores a subset of configurations included in the QM configurational distribution, leaving others unexplored. Nevertheless, as the phase space overlap metrics reveal, the MM configurational distributions at a given temperature are of a similar extent to, or broader, than their QM counterparts, supporting that the switching step acceptance rate is a robust metric of the similarity between the MM and QM levels of theory.

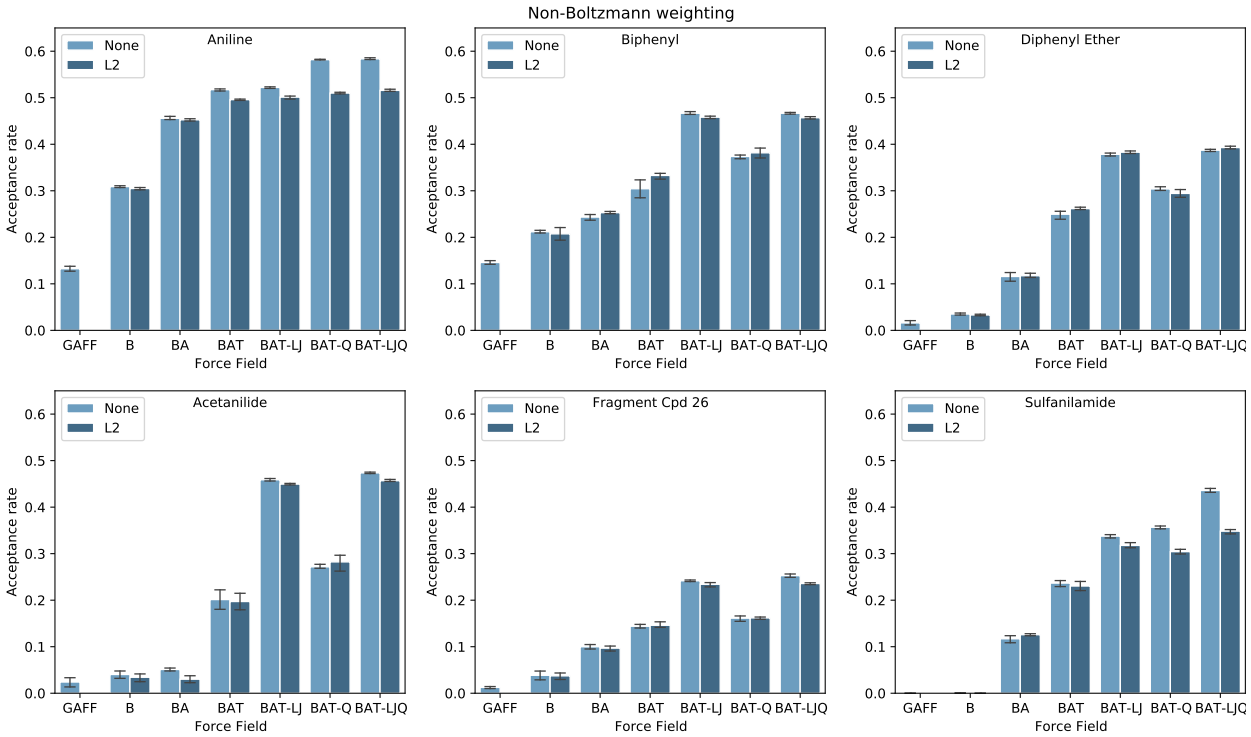

Figure 3: nMC-MC acceptance rates for the set of molecules represented in figure 2. The FFs used to calculate the acceptance rates are derived employing non-Boltzmann weighting with (dark blue) or without (light blue) L2 regularization. The training data set contains configurations sampled at 500 K. The errors bars correspond to the standard deviation of the results of 4 different nMC-MC samplers. Each sampler performs a total of 2×10^5 nMC-MC sweeps.

The acceptance rates obtained when using non-Boltzmann weighting in the reparameterizations are shown in figure 3. Through its analysis, it is possible to conclude that the variations observed in the acceptance rates are in line with what we expect from a systemat-

ical reparameterization of FFs. That is, for the non-regularized FFs, the more classes of FF parameters optimized, the higher the acceptance rates obtained, such that, in general, they follow the trend B < BA < BAT < BAT-Q < BAT-LJ < BAT-LJQ. The only disparity in this trend is observed for aniline and sulfanilamide, for which the BAT-Q FF performs better than the BAT-LJ, revealing that the optimization of charges is more important to accurate modeling of the aniline scaffold, which both molecules share, than the optimization of the Lennard-Jones parameters. It is worthwhile mentioning that the optimization of the non-bonded part of the FFs (charges, and 12-6 Lennard-Jones parameters) is here used as an *ad hoc* workaround to accelerate the convergence of the sampling of the target SCC-DFTB-D3 distribution, as well as to test the limits of the functional form of GAFF. However, it should be stressed that this approach is only viable for gas-phase situations like those presented in these examples because these parameters affect the intermolecular interactions. Therefore, the BAT-LJ, BAT-Q, BAT-LJQ parameters may have a limited applicability when applied to solutes in solution. Since solute-solvent interactions can influence the solute's configurational ensemble, and the nonbonded parameters influence the energetics of intermolecular interactions, a training data set including interactions with solvent needs to be considered if the Lennard-Jones parameters and partial atomic charges are to be optimized. Furthermore, the improvements seen in comparison to the original GAFF are substantial and indicate that the derived FFs increase their similarity with respect to the SCC-DFTB-D3 Hamiltonian. For all test cases, except aniline and biphenyl for which GAFF acceptance rates of *ca.* 13-15% are obtained, the GAFF acceptance rates are lower than 3%, being 0% for sulfanilamide, thus supporting the difficulty of MM FFs to correctly model sulfonamides.[84]

The properties of non-Boltzmann weighting support the previous observations because, during the FF optimization, this weighting scheme gives larger weights to conformations in which the MM energy is underestimated ($U^{MM} - U^{QM} < 0$) than to conformations in which the MM energy is overestimated ($U^{MM} - U^{QM} > 0$) with respect to the QM energy. Consequently, non-Boltzmann weighting mitigates the creation of spurious minima and drives

the errors towards high-energy regions, thus overestimating transition-state energies and underestimating fluctuations.[67,83] This is a desirable property because, if a perfect fit to the QM PES is impossible, it is preferable to have a mismatch that tends to make ΔU^{MM} greater than ΔU^{QM}, hence maximizing acceptance into the QM chain (see equation (2)) and, consequently, leading to the stable, systematic improvements that are observed.

Concerning the results obtained for the L2-regularized FFs, they follow similar trends to their non-regularized counterparts, except for some specific situations worth noting. For example, for the FFs wherein bonded parameters (B, BA, and BAT) are optimized, the acceptance rates of the regularized and non-regularized FFs are identical, presenting some variations that are not statistically significant. For the regularized FFs where parameters modeling the nonbonded interactions are also optimized (BAT-LJ, BAT-Q, and BAT-LJ), there is a noticeable decrease in the acceptance rates for some test cases such as, *e.g.*, aniline and sulfanilamide, in comparison to the non-regularized FFs. This may be entirely attributed to the *prior* widths that are used to constraint the charges and 12-6 Lennard-Jones parameters, which do not allow these parameters to stray too far away from their initial guesses and, as a result, lead to lower acceptance rates. For the remaining molecules, the regularized FFs perform equally or slightly better than the non-regularized ones, presenting differences that, in most cases, are also not statistically significant.

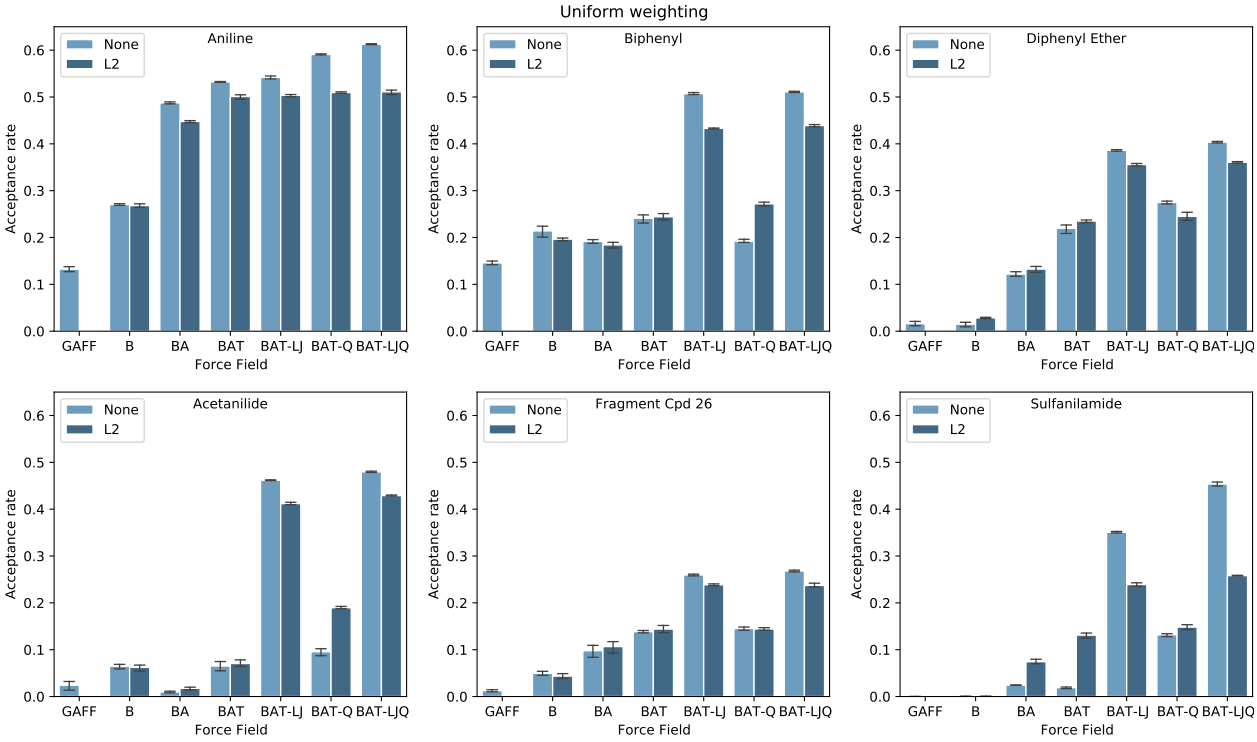

Figure 4: nMC-MC acceptance rates for the set of molecules represented in figure 2. The FFs used to calculate the acceptance rates are derived employing uniform weighting with (dark blue) or without (light blue) L2 regularization. The training data set contains configurations sampled at 500 K. The errors bars correspond to the standard deviation of the results of 4 different nMC-MC samplers. Each sampler performs a total of 2×10^5 nMC-MC sweeps.

A more interesting, yet somewhat more unpredictable, set of results are the acceptance rates obtained when employing uniform weighting in the reparameterizations (figure 4). Before analyzing the results obtained, it is important to discuss the consequences of the asymmetries that might be imposed on the MM PES by equally allowing positive and negative errors in the fittings, a feature of uniform weighting reported in previous studies.[67,83] To do this, let us consider the diagrams of figure 5, which show hypothetical MM PES fittings that can be obtained when employing uniform weighting. Assuming that the data set used in the reparameterizations comprises only the structures at configurations $\boldsymbol{q_i}$ and $\boldsymbol{q_f}$, all the represented uniform-weighted fittings have equal squared errors of the energy with respect to the QM PES, *viz.* $\sum_i \left(U_i^{QM} - U_i^{MM}\right)^2 = 2U^2$. Despite this, each case would lead to a different behavior if the corresponding FF were used in the nMC-

MC algorithm. Firstly, it is important to note that case A truly corresponds to a perfect fitting because the uniform-weighted MM PES can be superimposed with QM PES by a simple translation, as the difference between them is only a constant. Therefore, in what follows, we exclude this situation from the discussion. Furthermore, for case B, $\Delta\Delta U(\boldsymbol{q_i} \rightarrow \boldsymbol{q_f}) = \Delta U^{QM}(\boldsymbol{q_i} \rightarrow \boldsymbol{q_f}) - \Delta U^{MM}(\boldsymbol{q_i} \rightarrow \boldsymbol{q_f}) = -2U$ for uniform weighting, which leads to a fitting that maximizes acceptance into the QM chain. On the other hand, case C minimizes the likelihood of accepting structures into the QM chain since $\Delta\Delta U(\boldsymbol{q_i} \rightarrow \boldsymbol{q_f}) = 2U$. Hence, the uniform weighting scheme is prone to creating series of FFs that show unpredictable, non-systematic behavior since positive and negative $U^{MM} - U^{QM}$ differences are equally probable. This leads us to advocate for the use of the non-Boltzmann weighting if the aim is to use FFs in the nMC-MC algorithm because, as previously discussed, it tends to overestimate transition-state energies. As a result, non-Boltzmann-weighted fittings consistently show a skewing of the $U^{MM} - U^{QM}$ differences towards positive values, thus leading to the stable, predictable behavior observed between different FFs of the same series. Incidentally, non-Boltzmann weighting also proves to be superior to uniform weighting for general-purpose applications.[67,83] Note that the squared error of the energy with respect to the QM for the non-Boltzmann-weighted fitting is not equal to $2U^2$ as it is in the uniform-weighted fittings. Nevertheless, the point is to illustrate the possible asymmetries that uniform weighting may impose that can negatively impact the acceptance rates.

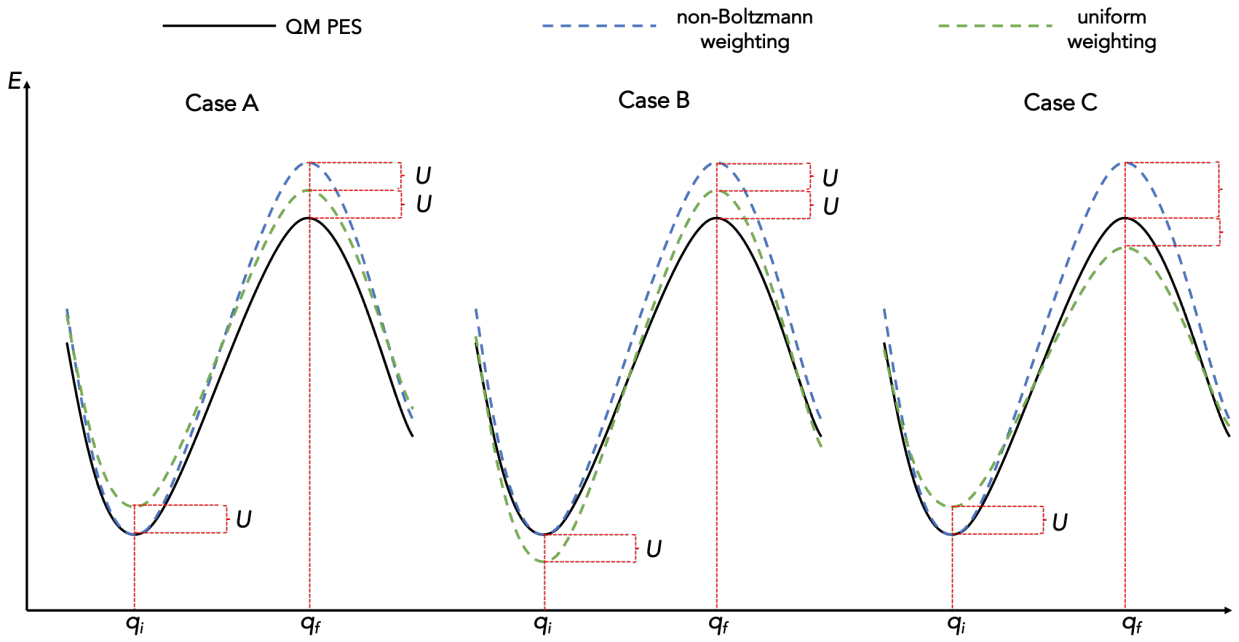

Figure 5: Diagram illustrating typical possible fittings that can be obtained when employing either the uniform and non-Boltzmann weighting schemes. All the represented uniform-weighted fittings have equal squared errors of the energy with respect to the QM PES, *viz.* $\sum_i \left(U_i^{QM} - U_i^{MM}\right)^2 = 2U^2$, but they behave differently when used in the nMC-MC algorithm.

Through the analysis of figure 4, it is possible to conclude that for the FFs that are optimized employing uniform weighting, the systematic improvement in the FFs only leads to systematically higher acceptance rates for aniline, diphenyl ether, and the fragment of cpd 26. This is observed for both the L2 and non-regularized versions of the FFs. On the other hand, for biphenyl and acetanilide, the BA FF results in lower acceptance rates than the B FF, and the BAT FF only performs slightly better than the latter. Furthermore, a set of results that is also unexpected are the acceptance rates for sulfanilamide, for which the non-regularized FFs perform equally for the BA and BAT FFs, even though the acceptance into the QM chain is very low (*ca. 2%*). Interestingly, higher acceptance rates are obtained in their L2-regularized counterparts, namely an increase from *ca.* 2% to *ca.* for 7-8% and 13% for the BA and BAT FFs, respectively. These unexpected results can be understood by inspecting the optimized angle force constants, K_θ, and angle equilibrium values, θ_{eq}, of

the non-regularized BA and BAT FFs of sulfanilamide (see figure 6). From these plots, it is clear that the non-regularized optimizations drive the parameters towards non-physical values such as close to zero or even negative K_θ, as well as θ_{eq} close to 0° or 360°, meaning that bent angles become practically linear. These artifacts, created by the optimization to minimize the objective function, ultimately have a strong impact on the acceptance rates due to poor dynamics and energy prediction of these FFs. These large, unphysical variations are lessened by applying L2 regularization, resulting in FFs that not only have optimized parameters with values in physically-sensible ranges but also are superior in terms of QM similarity in regards to their non-regularized analogues, thus leading to higher acceptance rates. Besides sulfanilamide, the only test case for which similar behavior is observed is acetanilide, in which the non-regularized uniform-weighted BA and BAT FFs also contain unphysical parameter values. This does not occur for any non-Boltzmann-weighted or L2-regularized FF.

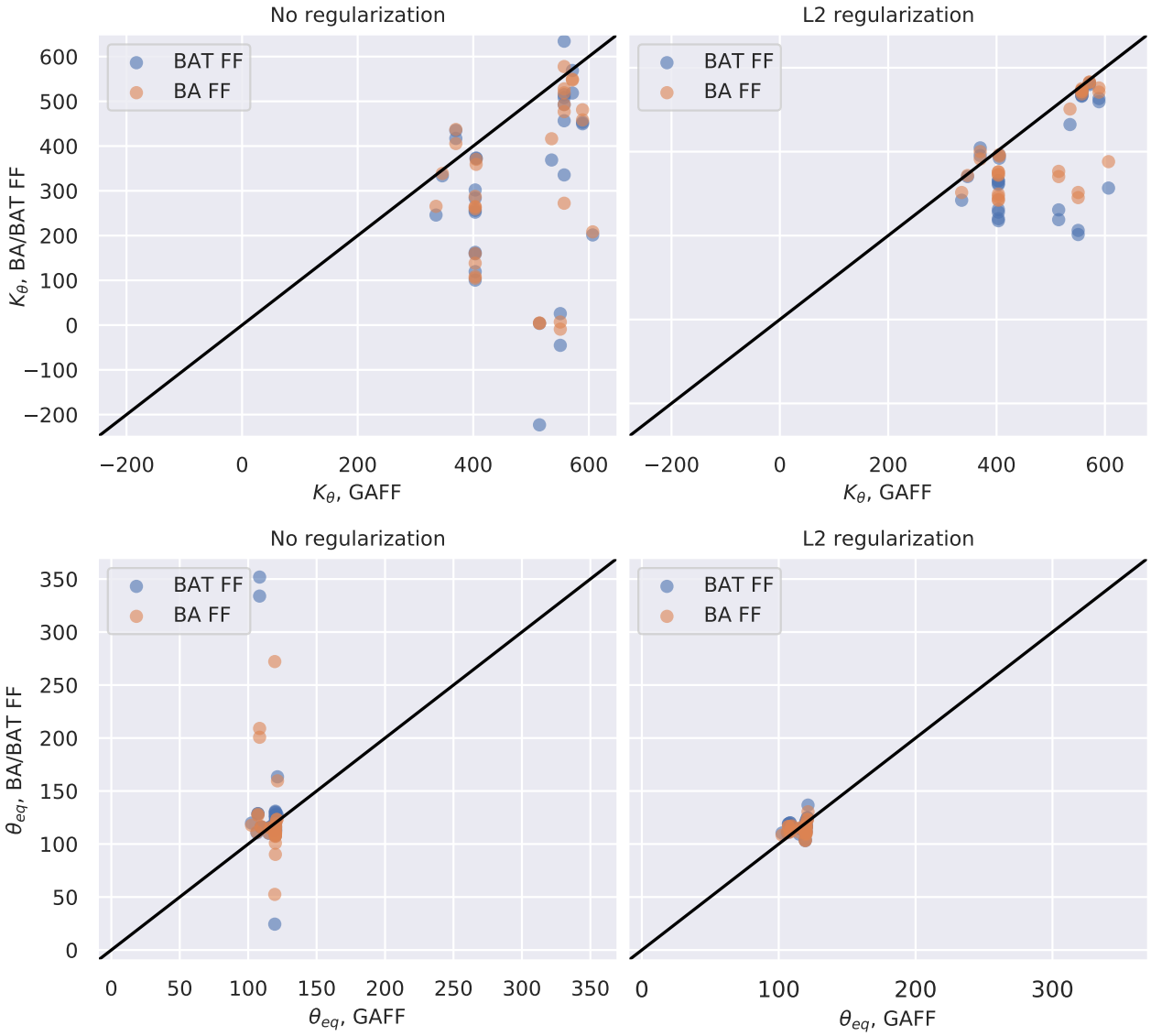

Figure 6: Sulfanilamide parameters before (GAFF; x axis) and after reparameterization (uniform-weighted BA/BAT FFs; y axis). The parameters represented are angle force constants (top panels) and angle equilibrium values (lower panels).

Two synergistic factors may contribute to the observed deviations from physically-sensible values in the non-regularized uniform-weighted FFs. First, some of the angle terms may have been used to compensate for deficiencies in other parts of the FF because when nonbonded parameters are concomitantly optimized these large deviations are not observed. In fact, as mentioned previously, the optimization of the nonbonded terms performed in this study is, in some sense, a workaround to make up for possible limitations in the FF functional

form. Second, the fact that the training data sets consist of NVT ensembles sampled at a temperature of 500 K also contributes to non-physical parameters. This can be observed by comparing the acceptance rates obtained for the training data sets generated at 500 K with the acceptance rates obtained for the training data sets generated at 300 K (figure 7; no regularization, uniform weighting). At the QM level, especially when sampling at very high temperatures, bonds and angles oscillate anharmonically, a phenomenon that may impact the FF reparameterization owing to limitations of the GAFF functional form that, by design, imposes harmonicity in these DOFs (see equation 11). The consequence is that the FF optimizers tend to pull the angle force constants towards lower values to generate wider potentials that better fit the sampled anharmonicities. Incidentally, this is the response observed in figure 6. Angles are particularly prone to straying away from physical-sensible values because they have the lowest force constants of the hard DOFs and, therefore, they may suffer more from using data sets containing configurations in the anharmonic regime.

Although the training data sets at 500 K do not lead to non-physical bond parameters, it is clear in figure 6 that, for some molecules, they lead to lower average acceptance rates in comparison to the FFs derived using data sets at 300 K. Interestingly, the differences in acceptance rates, which initially become apparent in the B or BA FF, are then propagated up to some degree into the FFs for which parameters of nonbonded terms are also optimized (BAT-LJ, BAT-Q, and BAT-LJQ). Optimization of nonbonded parameters either mitigates these contrasts or worsens them due to the presence of high-energy structures in the high-temperature ensemble that are unimportant to the low-temperature ensemble. Despite these observations, we still opt to use the high-temperature data sets for most of the reparameterizatons in this study, as they ensure a thorough exploration of the conformational space of the test cases being studied. Ideally, one would employ enhanced-sampling methods to have the best of both ensembles: the extensive conformational sampling of the high-temperature data set, and the harmonic behavior of the low-temperature data set.

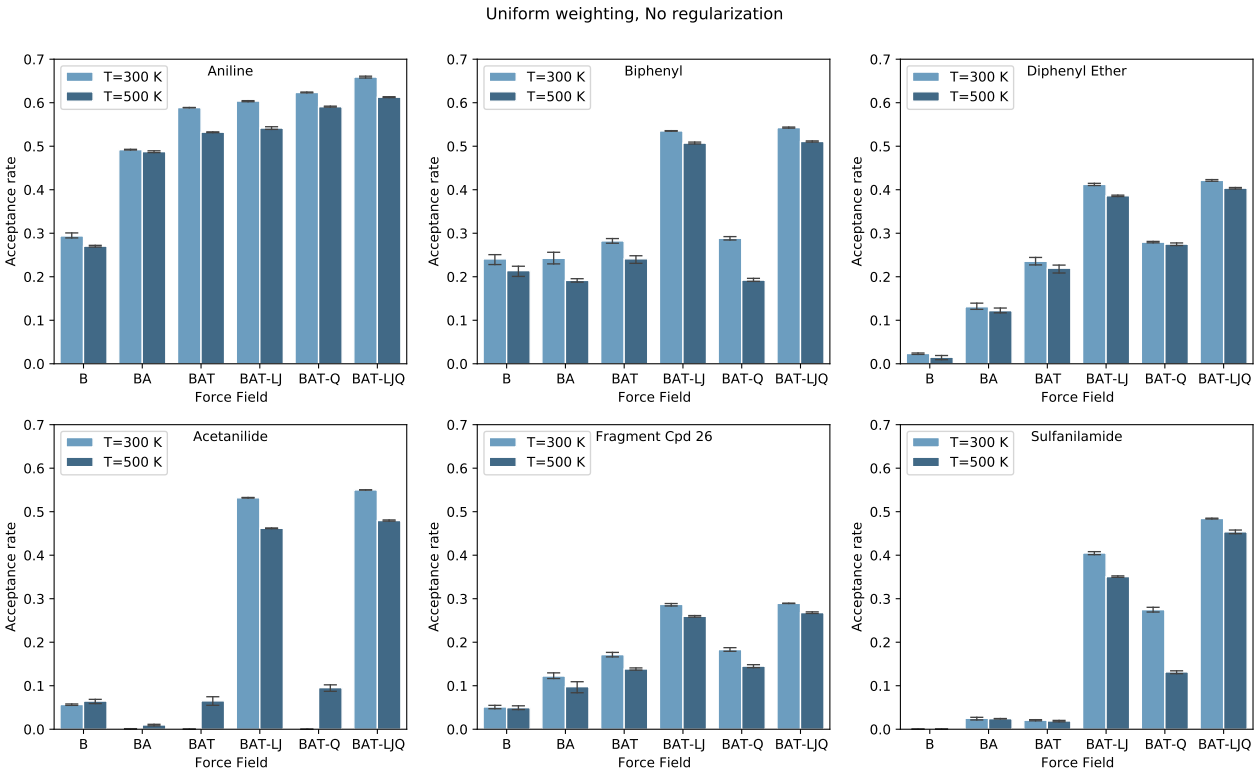

Figure 7: Comparison between the nMC-MC acceptance rates obtained for FFs reparameterized using data sets containing structure sampled at either 300 K or 500 K. The FFs used to calculate the acceptance rates are derived employing uniform weighting without any regularization. The error bars correspond to the standard deviation of the results of 4 different nMC-MC samplers. Each sampler performs a total of 2×10^5 nMC-MC sweeps.

Overall, FF reparameterization proves to be an efficient strategy to increase the acceptance rate of the switching step from the MM to the QM level of theory. This permits accelerating convergence of the sampling of the target QM configurational distribution, which would otherwise be impractical due to the very low acceptance rates that are obtained when using the original GAFF. The best acceptance rates are obtained for aniline (*ca.* 65%), whereas the molecule with lowest acceptance rate is the fragment of cpd 26. This is expected given that it is the largest and most complex molecule of the test set. It is expected that both molecular size and chemical complexity have an impact on the acceptance rates, the former because small differences between the MM and QM Hamiltonians sum up as the number of DOFs increases, and the latter due to the challenge that some functional groups pose to the functional form of GAFF. Furthermore, uniform-weighted FFs, especially without regular-

ization, are to be avoided as they are more prone to generate non-physical parameters to obtain the best possible fit. On the other hand, L2-regularized and non-Boltzmann-weighted FFs perform the best, especially given their stable behavior. Also noteworthy is that hard DOFs, such as bonds and angles, are crucial to reparameterize to increase the acceptance rates since they have large force constants. Therefore, small differences in their values lead to large changes in energy which may considerably impact the switching efficiency from the MM into the QM chain.

3.2 NH_2 Inversion in Aniline

As a first example of how nMC-MC allows recovering the exact quantum configurational distribution using an approximate FF, let us consider the inversion of the NH_2 group in aniline. It is well established, both experimentally[85,86] and theoretically,[87,88] that the primary amine of aniline has a pyramidal geometry, and that inter-conversion between two equally stable conformations occurs through nitrogen inversion. Nevertheless, although simple, this is a clear instance of a functional group for which GAFF fails to predict the correct conformational dynamics.

Through the analysis of the configurational distributions represented in figure 8, it can be seen that GAFF generates NVT configurational distributions at 300 K that differ substantially from those generated by SCC-DFTB-D3. Specifically, we can see that GAFF (lower panel) predicts that the NH_2 group assumes a trigonal planar geometry, hence failing to reproduce the inter-conversion between the two local minima, while SCC-DFTB-D3 (top panel) predicts the expected conformational behavior. Furthermore, the reparameterized BAT FF distribution is much closer to the SCC-DFTB-D3 distribution than the original GAFF, and the nMC-MC distribution successfully reproduces the SCC-DFTB-D3 distribution when sampling is performed using the BAT FF. The agreement obtained is excellent as there is negligible loss of accuracy.

It is noteworthy that the fast recovery of the target SCC-DFTB-D3 distribution through

the nMC-MC algorithm is only possible due to the increased acceptance rates that are achieved after reparameterization of the original FF. GAFF gave acceptance rates of *ca.* 12-13%, which, even though high in comparison with other test cases, are still much lower than the acceptance rates of *ca.* 49-50% that are obtained for the non-Boltzmann-weighted L2-regularized BAT FF. This high acceptance rate enables recovery of the target SCC-DFTB-D3 in only 2×10^5 nMC-MC sweeps (hMC runs of 100 steps with a 1 fs time step). No attempt is made to optimize the length of these calculations, as our main goal is to prove the implementation and principles of the methodology and not to optimize the protocol in itself.

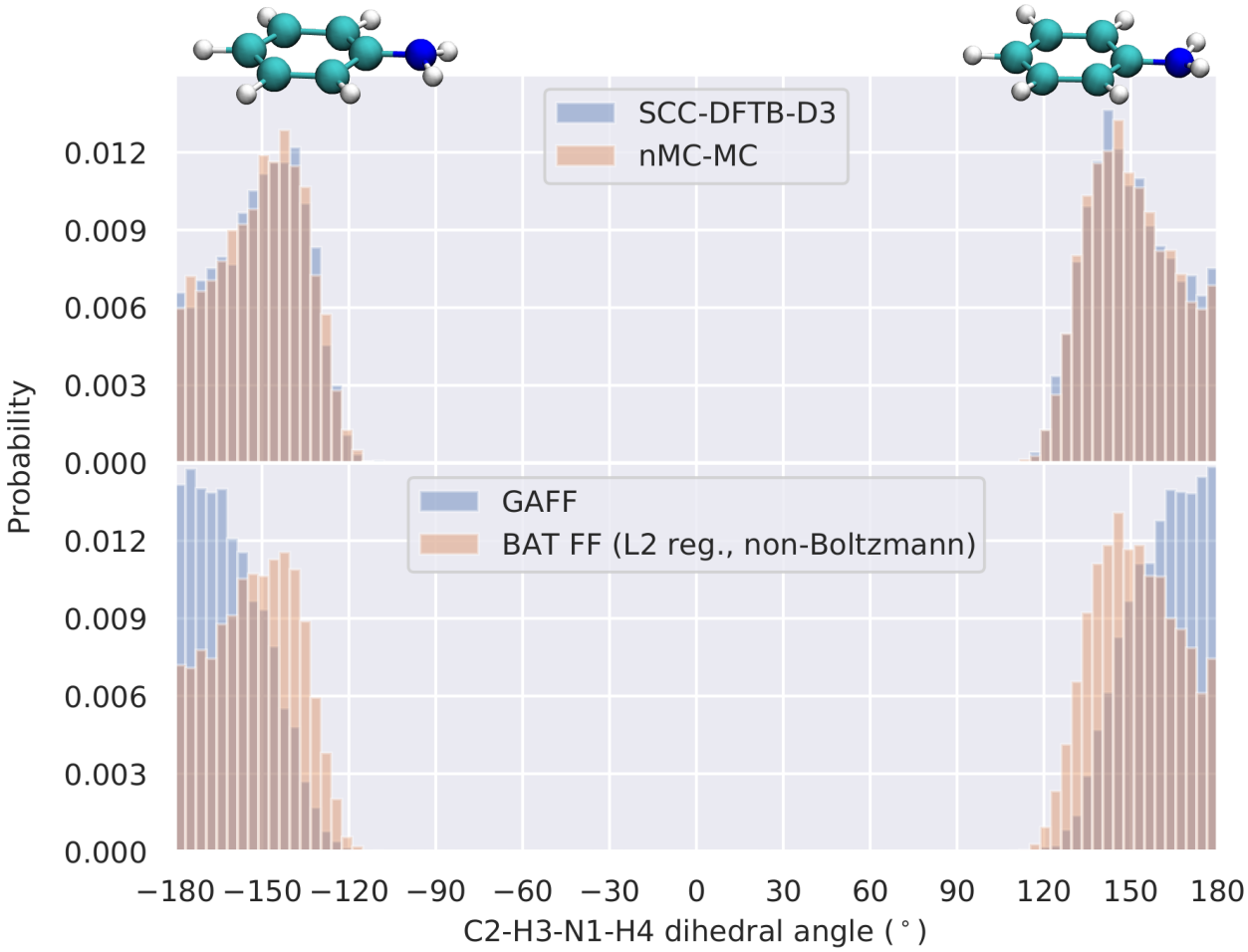

Figure 8: Top panel: Distribution of the C6-H9-N7-H8 improper dihedral of aniline as obtained in SCC-DFTB-D3 MD and nMC-MC simulations. Lower panel: Distribution of the C6-H9-N7-H8 improper dihedral of aniline as obtained in MD simulations using the original GAFF and the non-Boltzmann-weighted L2-regularized BAT FF. The SCC-DFTB-D3, GAFF, and BAT MD calculations were performed during 10 ns (snapshots collected every 1 ps), and the nMC-MC sampler performed a total of 2×10^5 MC sweeps. The simulations' temperature was 300 K.

3.3 Fragment of cpd 26

Let us now discuss the results obtained when applying the nMC-MC algorithm to the fragment of cpd 26 shown in figure 2, the largest molecule of our test set. Cpd 26 is a non-peptidic, orally bioavailable, and efficacious low nM antagonist of the inhibitor of apoptosis proteins cIAP1 and XIAP.[55] Therefore, this test case aims to mimic the application of the

presented methods to a recently designed and relevant druglike molecule.

By analyzing the SCC-DFTB-D3 configurational distribution represented in figure 9, we are able to identify 8 conformations for this molecule. Their molecular structures, which map to the red stars in the plot, are represented in figure 10. They arise essentially from the different relative positions that the phenyl group can assume relatively to the azaindoline ring. This conformational dynamics is in line with what is observed experimentally.[55] Surprisingly, although GAFF gives very low nMC-MC acceptance rates (close to 1%) and predicts incorrect relative abundances, it still fairly describes the global features of the configurational distribution. The BAT (L2-regularized, non-Boltzmann-weighted) optimization of GAFF leads to a much closer distribution to the target SCC-DFTB-D3, demonstrating the quality of this FF. This observation is further supported by the increase in acceptance rate to 23-24% that is obtained. Finally, as expected, the nMC-MC flawlessly reproduces the SCC-DFTB-D3 when using the BAT (L2-regularized, non-Boltzmann-weighted) FF in the low-level Markov chain, allowing convergence and thus recovery of the target distribution in 3×10^6 nMC-MC sweeps (hMC runs of 250 steps with a 1 fs time step). To accelerate sampling in the MM chain, a temperature of 400 K is used for the T_K and T_U values entering in equation (1), while the temperature of the target quantum NVT ensemble is kept at 300 K.

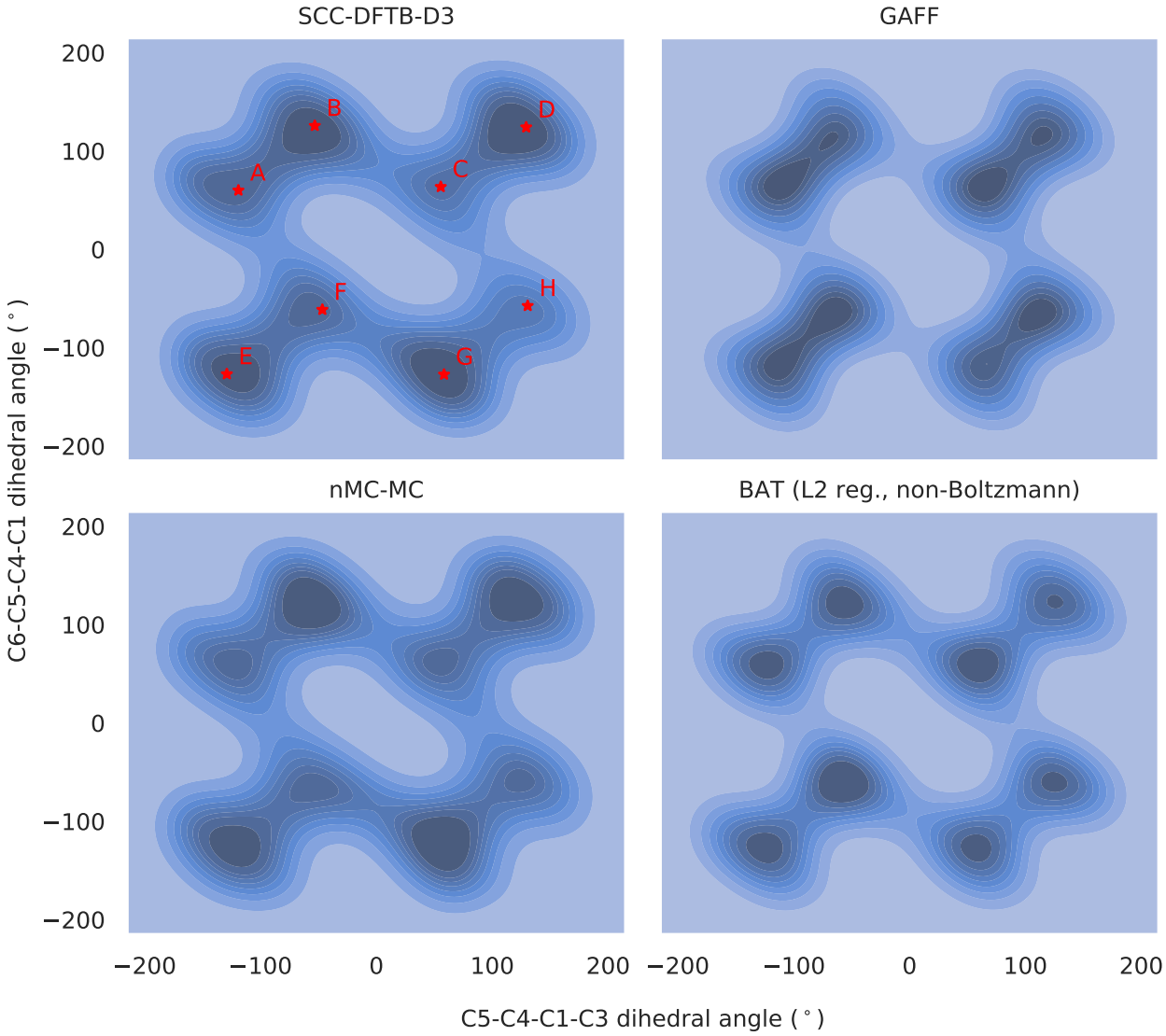

Figure 9: Configurational distributions of the C5-C4-C1-C3 *vs.* C6-C5-C4-C1 dihedrals for the fragment of cpd 26. The SCC-DFTB-D3 MD was simulated during 10 ns (snapshots collected every 1 ps), and the GAFF and BAT MD were simulated during 1 μs (snapshots collected every 100 ps). The nMC-MC sampler performed a total of 3×10^6 MC sweeps. The simulations' temperature was 300 K. The conformations identified on the top left plot are shown in figure 10.

Despite the success of the results obtained when applying the nMC-MC algorithm to the fragment of cpd 26, the main pitfall of this methodology becomes apparent in this test case. As mentioned before, the larger and more complex the molecule, the more difficult it is to reparameterize to the QM level of theory due to accumulations of errors that are unavoidable and usually related to FF functional form constraints. There are two possible solutions to

this bottleneck if a two-chain nMC-MC algorithm is to be kept. The simpler approach consists of artificially broadening the MM distribution by manipulating the thermodynamic variables of the MM chain. This would involve, *e.g.*, increasing the temperature of the MM chain such that its energy distribution becomes wider and overlaps to a greater extent with the QM energy distribution. This strategy was successfully applied in past studies,[43,44] though success is not guaranteed if mismatch between the energy distributions is too large. On the other hand, a more complex but perhaps more reliable method involves developing and employing more accurate low-level models. In this regard, there are different classes of FFs of increasing complexity that can be applied and are still computationally cheap in comparison with the QM calculations. Machine-learning potentials are also an attractive option, especially owing to their blindness to functional forms, which make them potentially more accurate than MM FFs.[89,90] Nevertheless, in principle, in a nMC-MC context they would have to be used alongside a high-level of theory similar to that they were trained to reproduce. Currently, we are performing further studies along these lines to address the performance of other low-level models in the nMC-MC algorithm. It is also possible to use intermediate levels of theory to bridge the gap between the low-level and high-level models. Unfortunately, this solution becomes computationally expensive, especially if hybrid energy models such as $\lambda_i U^{MM} + (1-\lambda_i) U^{QM}$, where λ_i controls the weight of each energy component in the i-th chain, are used, as they still require high-level calculations to be performed.

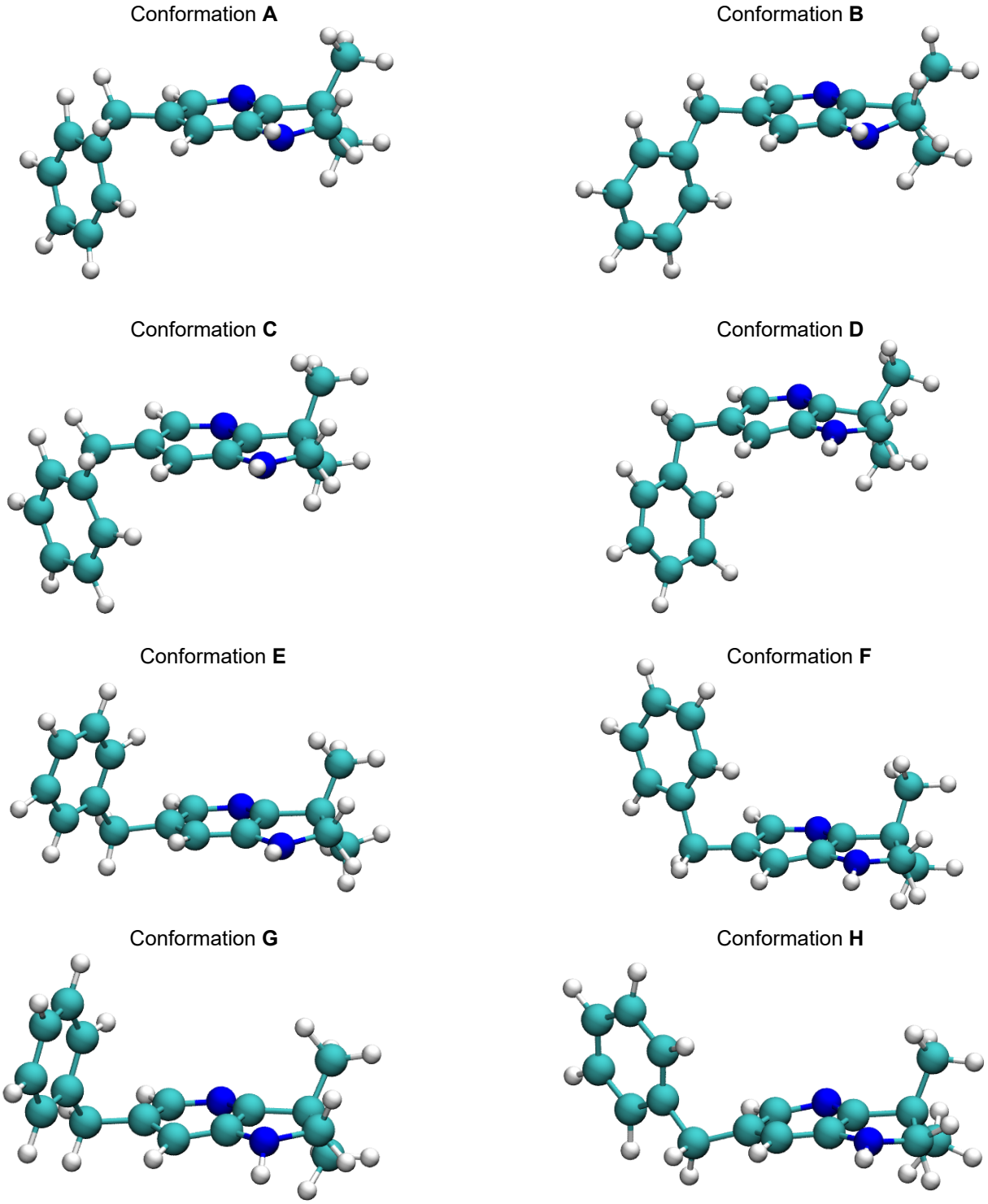

Figure 10: Main conformations of the fragment of cpd 26 identified in figure 9.

3.4 Analysis of Phase Space Overlap

To verify the variations in the acceptance rates observed when systematically reparameterizing the FFs, let us now turn our discussion to the evaluation of the phase space overlap between the ensembles generated using the MM FFs and the SCC-DFTB-D3 level of theory. In the following results, all the testing data sets contain 10000 configurations sampled from MM and QM MD simulations. These are performed with the same settings applied when generating the training data sets, except for the temperature which is here set to 300 K to make the results directly comparable to the nMC-MC acceptance rates. Furthermore, note that phase space is here employed as a synonym of configuration space as we only consider situations that compare total energy distributions at the same temperature, thus making the momentum coordinates irrelevant (see proof in the Supporting Information). Owing to this, even though these metrics were initially presented as depending on the total energy, in practical terms only the potential energy is used in their calculation.

The most direct and robust metric to measure the phase space overlap between ensembles obtained using different levels of theory is given by equation (8). It requires performing dynamics with the MM and QM Hamiltonians and, subsequently, evaluating the energies of each ensemble at both the MM and QM levels of theory. The potential energy difference between the QM and MM levels for structures sampled from a SCC-DFTB-D3 MD, $\Delta U_{QM}^{QM\rightarrow MM} = U_{QM}^{MM} - U_{QM}^{QM}$, and minus the potential energy difference between the MM FF and the QM level of theory for structures sampled using the MM FF, $-\Delta U_{MM}^{MM\rightarrow QM} = -\left(U_{MM}^{QM} - U_{MM}^{MM}\right)$, are then calculated, and the corresponding histograms determined. The resulting probability distributions are translated along the ΔU axis such that $\Delta U = 0$ is the midpoint between the two distribution means. Each histogram is fitted to a Gaussian function, and the overlap obtained between the Gaussians is evaluated numerically and used as an estimation of the phase space overlap. Any structure for which the absolute difference of its energy relative to the average energy of the respective distribution is larger than 100 kJ mol^{-1}, *i.e.* $|U - \langle U \rangle| > 100$, is removed from the data set.

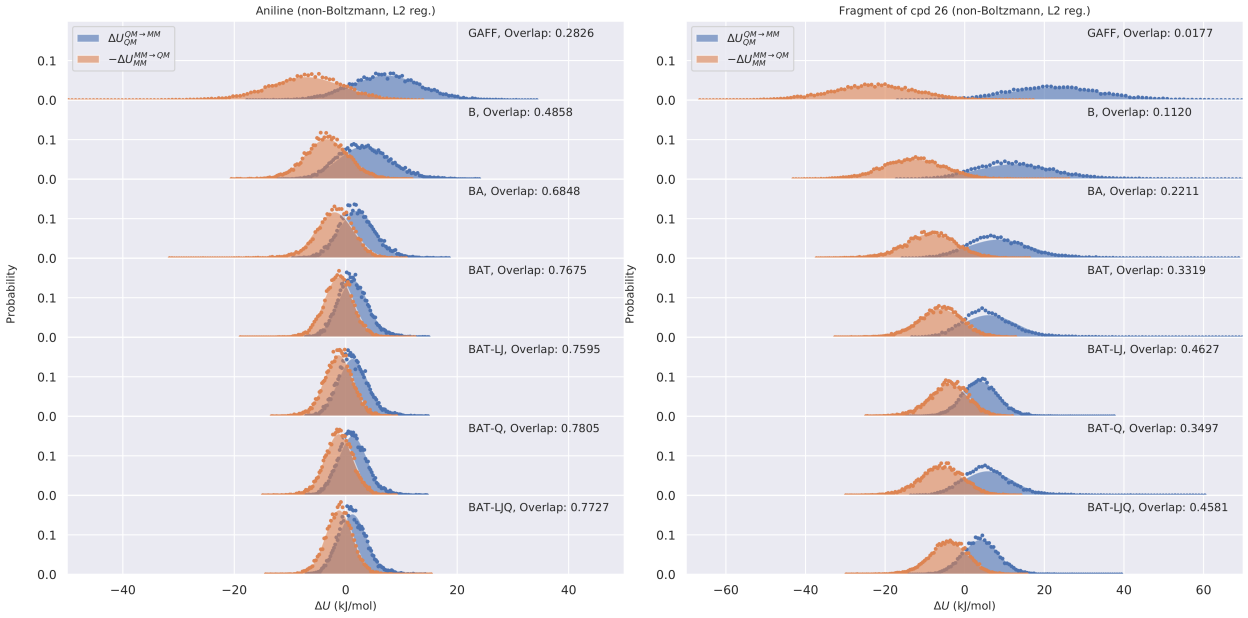

Figure 11: Energy difference histograms of MM→QM and QM→MM for aniline (left) and the fragment of cpd 26 (right). The distributions are translated along the ΔU axis such that $\Delta U = 0$ is the midpoint between the means of the two distributions.

The energy difference histograms and phase space overlaps for the two molecules in which we have primarily focused our discussion, *viz.* aniline and the fragment of cpd 26, are shown in figure 11 (non-Boltzmann weighting, L2 regularization). In these plots, it can be seen that the energy difference distributions are well approximated by Gaussian functions, and that the phase space overlap increases from the B FF to the BAT-LJQ FF, as observed in the acceptance rates of figure 3.

Using the nMC-MC switching step acceptance rate as a metric of the similarity between the MM and QM levels of theory necessarily requires a strong correlation between the acceptance rates of equation (2), θ, and the phase space overlap of equation (8), Ω. Hence, to assess the degree of correlation between both measurements, we compute the linear regressions of four sets of data: uniform-weighted data, non-Boltzmann-weighted data, non-Boltzmann-weighted L2-regularized data, and all data. From the results shown in figure 12, it is clear that there is a high degree of correlation between θ and Ω. Therefore, although these linear fittings are only an approximation of the true correlation between both measurements, we

consider that the observed correlations are close enough to linear behaviour to be considered as so. Due to this, a direct comparison between R^2 of different fittings is avoided. Nevertheless, it is worth point out that for the uniform-weighted data set, there is an outlier located at ($\theta \approx 0.20, \Omega \approx 0$), for which the nMC-MC acceptance rate is significant, but the phase space overlap is estimated to be practically 0. A close inspection reveals that this data point corresponds to the BAT-Q uniform-weighted L2-regularized FF of acetanilide. This observation is in line with results of a past publication, which reports situations wherein reparameterization of charges using uniform weighting also led to a decrease in the FF quality.[42] Interestingly, the nMC-MC algorithm still allows obtaining high acceptance rates for problematic FFs with non-physical parameters if the hMC runs are short enough to prevent the molecules from converting to the spuriously stable, non-physical geometries in which they get trapped in regular MD simulations. Overall, the results show that the nMC-MC acceptance rate is a robust metric of the phase space overlap and can be employed to evaluate the similarity between the levels of theory used in the nMC-MC low and high-level chains.

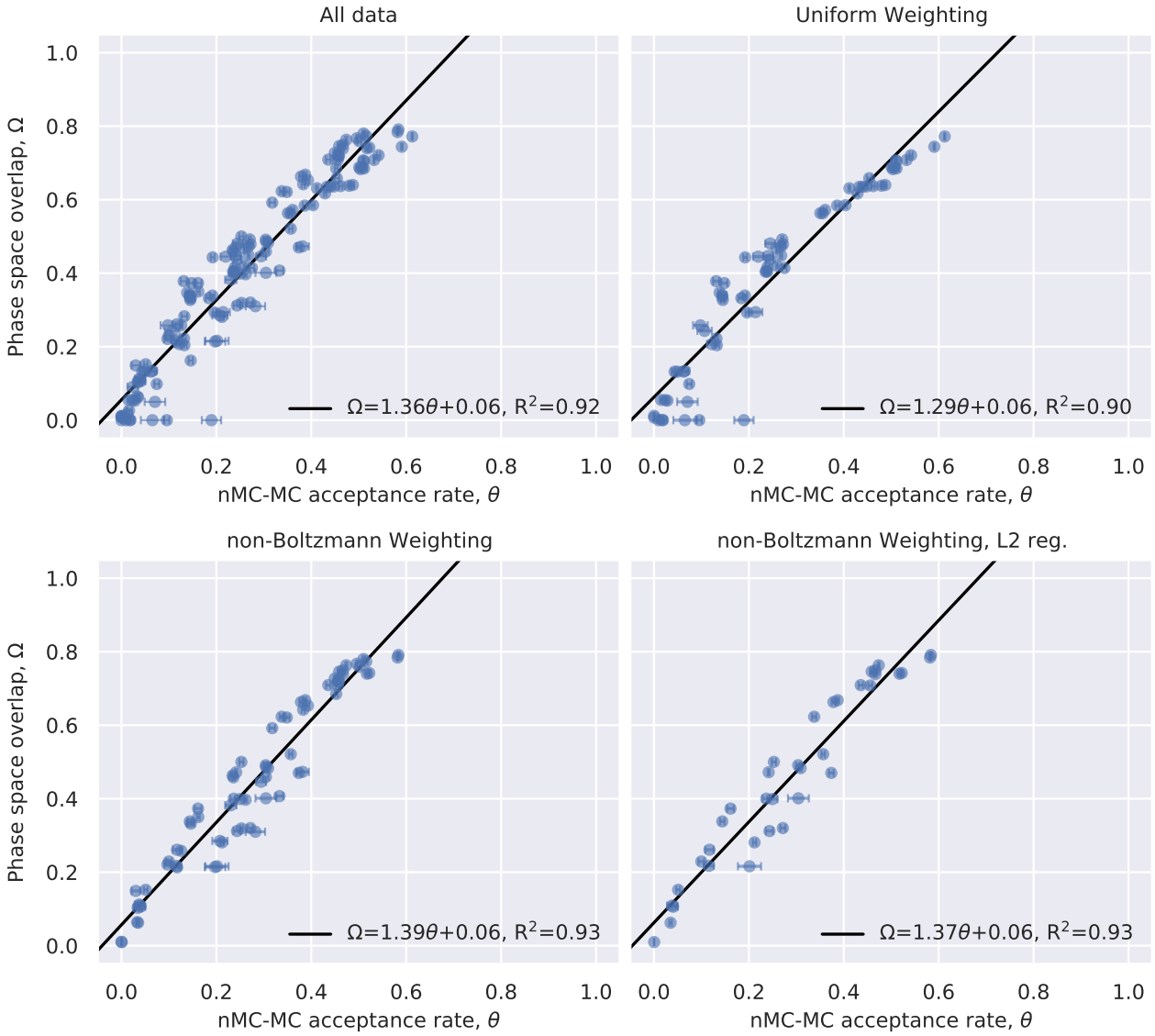

Figure 12: Correlation between the nMC-MC acceptance rate, θ, as given by equation (2), and the phase space overlap, Ω, as given by equation (8), for 4 different data sets: all data (top left), not-regularized data (top right), L2-regularized data (lower left), and non-Boltzmann-weighted L2-regularized data (lower right). The GAFF data points are included in the "all data" data set.

Before analyzing the results obtained for the Wu and Kofke metrics shown in equations (9) and (10), it is useful to explain their physical meaning. Since these metrics are a measure of the offset of an energy distribution with respect to another, they provide insights about how much the FF-sampled phase space lies inside the QM-sampled phase space. First, it is important to notice that when $0 \leq \Sigma_{QM,MM} < 1$, the probability distribution of the MM

energies observed for a simulation performed using the QM Hamiltonian, ρ_{QM}^{MM}, is centered right relative to the probability distribution of the MM energies observed for a simulation performed using the MM FF, ρ_{MM}^{MM}, meaning that the QM high-energy structures that lie above ρ_{MM}^{MM} are unimportant to the the MM FF, or that the MM FF low-energy structures that lie below ρ_{QM}^{MM} are undersampled by the QM Hamiltonian. On the other hand, when $1 < \Sigma_{QM,MM} \leq 2$, the probability distribution of the MM energies observed for a simulation performed using the QM Hamiltonian, ρ_{QM}^{MM}, is centered left relative to the probability distribution of the MM energies observed for a simulation performed using the MM FF, ρ_{MM}^{MM}, meaning that the QM Hamiltonian preferentially accesses a small set structures that are either not sampled by the MM FF or, if energetically favorable, are entropically disfavored, or that the MM FF samples high-energy structures that are unimportant to the QM Hamiltonian. Identical reasoning can be applied to $\Sigma_{MM,QM}$.

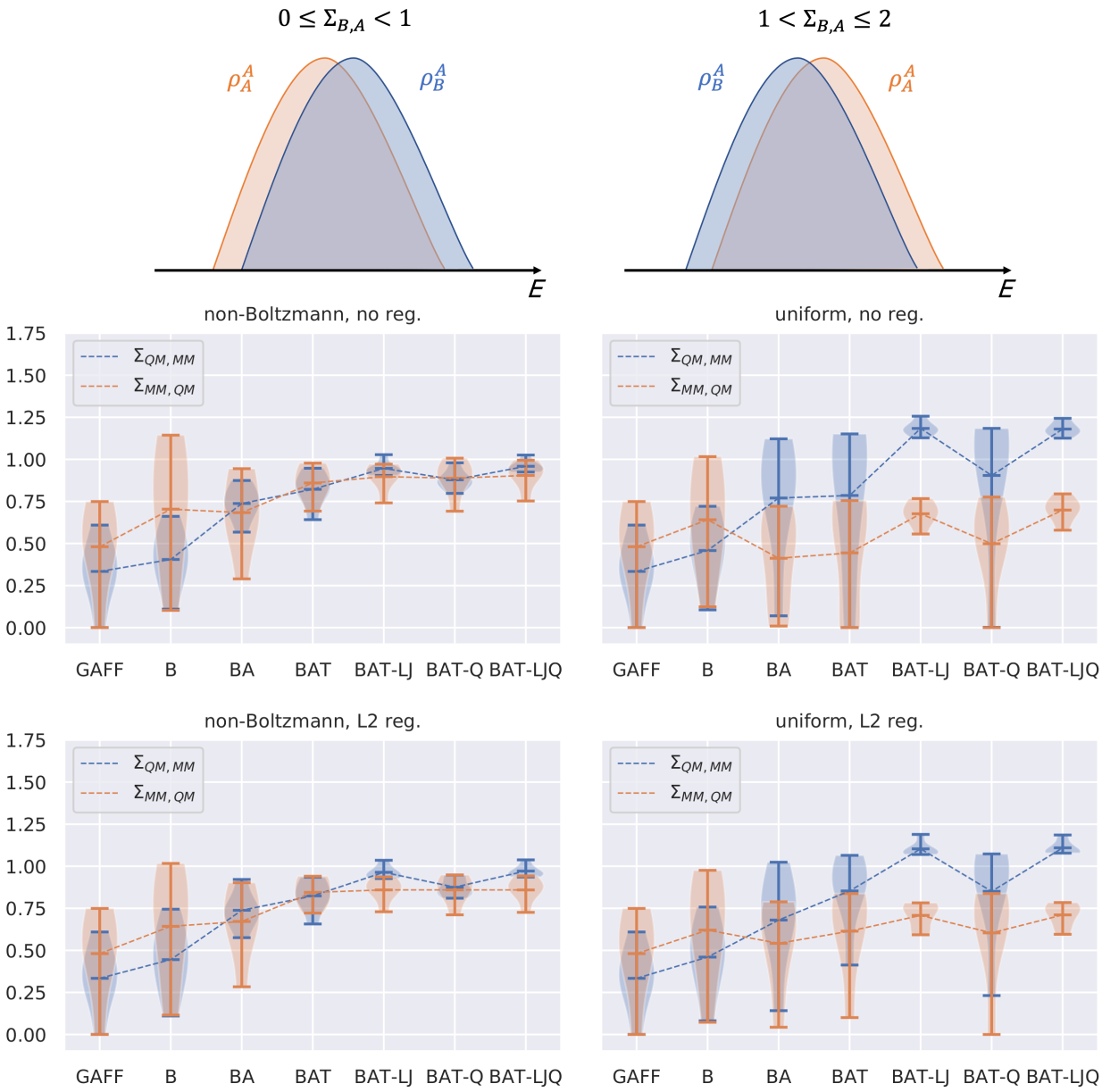

Figure 13: Violin plots showing the distribution of the Wu and Kofke overlap metrics between the MM and QM levels of theory, as given by equations (9) and (10), for all molecules represented in figure 2. The solid lines indicate the mean and the extrema of the distribution for each type of reparameterized FF, and the dashed lines connect their mean values. 4 data sets are represented: non-Boltzmann-weighted not-regularized data (top left), uniform-weighted not-regularized data (top right), non-Boltzmann-weighted L2-regularized data (lower left), uniform-weighted L2-regularized data (lower right).

From the results presented in figure 13, which shows the average behavior of $\Sigma_{QM,MM}$ and $\Sigma_{MM,QM}$ for all molecules in the test set, it can be seen that when the non-Boltzmann

weighting scheme is applied, $\Sigma_{QM,MM}$ starts with a value close to 0.4 (B FF), likely meaning that the energy of high-energy QM structures are overestimated in the MM FF or the MM FF is sampling spurious minima. Nevertheless, this mismatch is progressively diminished as more classes of FF parameters are reparameterized because $\Sigma_{QM,MM}$ is getting closer to 1, showing how well the MM FF can predict the energy of structures generated through a dynamics performed using the QM Hamiltonian. A similar trend is observed for $\Sigma_{MM,QM}$, though it reaches a plateau at $\Sigma_{MM,QM} \approx 0.8$ for the most-refined FFs (BAT, BAT-LJ, BAT-Q, and BAT-LJQ). This combination of $\Sigma_{QM,MM}$ and $\Sigma_{MM,QM}$ values indicates that systematic reparameterization of the FFs turns an overlap relation into a subset relation between the phase space distributions explored by the MM and QM levels of theory, supporting the idea that the MM distributions of the most-refined FFs are somewhat broader than their QM counterparts since the former sample high-energy structures that are unimportant to the QM FF.

An identical interpretation follows for the plots concerning the uniform-weighted FFs. Worthy of note are the BAT-LJ and BAT-LJQ FFs, for which $1 < \Sigma_{QM,MM} \leq 2$ and $0 \leq \Sigma_{MM,QM} < 1$. This is a case of special concern because it means that these FFs either sample spurious high-energy structures or undersample the QM minima. Finally, even though not observed in this study, $0 \leq \Sigma_{QM,MM} < 1$ and $1 < \Sigma_{MM,QM} \leq 2$ is a hugely undesirable case as it may indicate that a FF is either sampling spurious minima or overstabilizing the QM minima, situations that can potentially lead to trapping of MM simulations in overstabilized basins. Likewise, the case in which $1 < \Sigma_{QM,MM} \leq 2$ and $1 < \Sigma_{MM,QM} \leq 2$ also poses concerns as it often suggests that a FF is not only undersampling the true QM minima, but also sampling other spurious minimina.

3.5 Self-parameterizing nMC-MC

As proof of principle, we test the self-parameterizing methodology that iteratively couples the nMC-MC algorithm with a parameterization step. This algorithm allows on-the-fly deriva-

tion of bespoke FFs owing to its capability of performing sampling of relevant configurations and subsequent optimization of the FF parameters, all in one scheme. Specifically, we use the nMC-MC algorithm to sample QM configurations in such a way that a configuration belonging to the QM ensemble is added to the FF training data set for every 5 configurations accepted into this high-level ensemble. Configurations belonging to the MM ensemble or rejected configurations of any kind are discarded for reparameterization purposes. Despite this choice, in some situations they can be informative for the optimizations and, inclusively, accelerate the convergence of the self-parameterizing procedure.[50] Furthermore, every time 500 new QM structures are added to the existing training data set, a new reparameterization of the FF parameters is performed using the total training data. The temperature used for the Markov chains and the Maxwell-Boltzmann distribution is 300 K. The self-parameterizing nMC-MC procedure is deemed to be converged when the root-mean-square deviation (RMSD) of the FF parameters between two successive iterations is less than 10^{-4}.

The molecule used in this application is octahydrotetracene (see molecular structure in figure 14). This choice relies on the fact that this scaffold was previously identified as a challenging case for FFs, which struggle in reproducing its QM energies.[84] All bonded parameters are optimized in every optimization, such that the vector of optimizable parameters is given by $\boldsymbol{p} = \{\boldsymbol{K_b}, \boldsymbol{r_{eq}}, \boldsymbol{K_\theta}, \boldsymbol{\theta_{eq}}, \boldsymbol{V_n}, \boldsymbol{\gamma_n}\}$. The total number of optimizable parameters is 72. Using the nomenclature of the previous examples, this corresponds to generating a BAT-type FF. The objective function includes energy, force and regularization terms, as given by equations (4), (5) and (6). Uniform weighting is applied to weight the conformations. In contrast to the previous applications, AMBER atom-type symmetries are preserved, and the *prior* widths used in the regularization are estimated from the arithmetic mean for each class of parameters, a feature included in ParaMol.

The results obtained for the self-parameterizing nMC-MC calculation of octahydrotetracene are shown in figure 14. The algorithm takes in total 11 iterations of nMC-MC sampling and reparameterization to converge, resulting in a final training data set of 5500

QM structures. The convergence of the procedure can also be tracked and confirmed by looking at the components of the objective function as a function of the iteration number (see figure S1 in the SI). Moreover, the acceptance rate increased smoothly and monotonically, progressively stabilizing as the nMC-MC sweeps increase. Nonetheless, the behavior of the RMSD plot is somewhat more irregular, with sudden jumps that can be explained by the necessity of the optimization to adapt the parameters to a new set of configurations. Although this is not problematic for this example, in other applications large variations of the RMSD may lead to a premature ending of the self-parameterizing procedure, which is something that may be resolved by employing tighter thresholds.

To test the quality of the derived FF, we also generate a BAT-type FF using a training data set containing a total of 10000 QM structures. This FF is generated following the same philosophy applied in the previous examples, such that firstly we build the data set using Langevin dynamics at a temperature of 300 K, and afterward we optimize the FF. By comparing the blue and green lines of the top panel of figure 14, both approaches lead to identical acceptance rates of *ca.* 15-16%, strongly indicating the robustness of the self-parameterizing nMC-MC procedure. Overall, this self-parameterizing algorithm is quite appealing since, by combining sampling and parameterization in one scheme, it does not require *a priori* generation of a training data set of unknown size, therefore limiting the computational work to that which is strictly necessary.

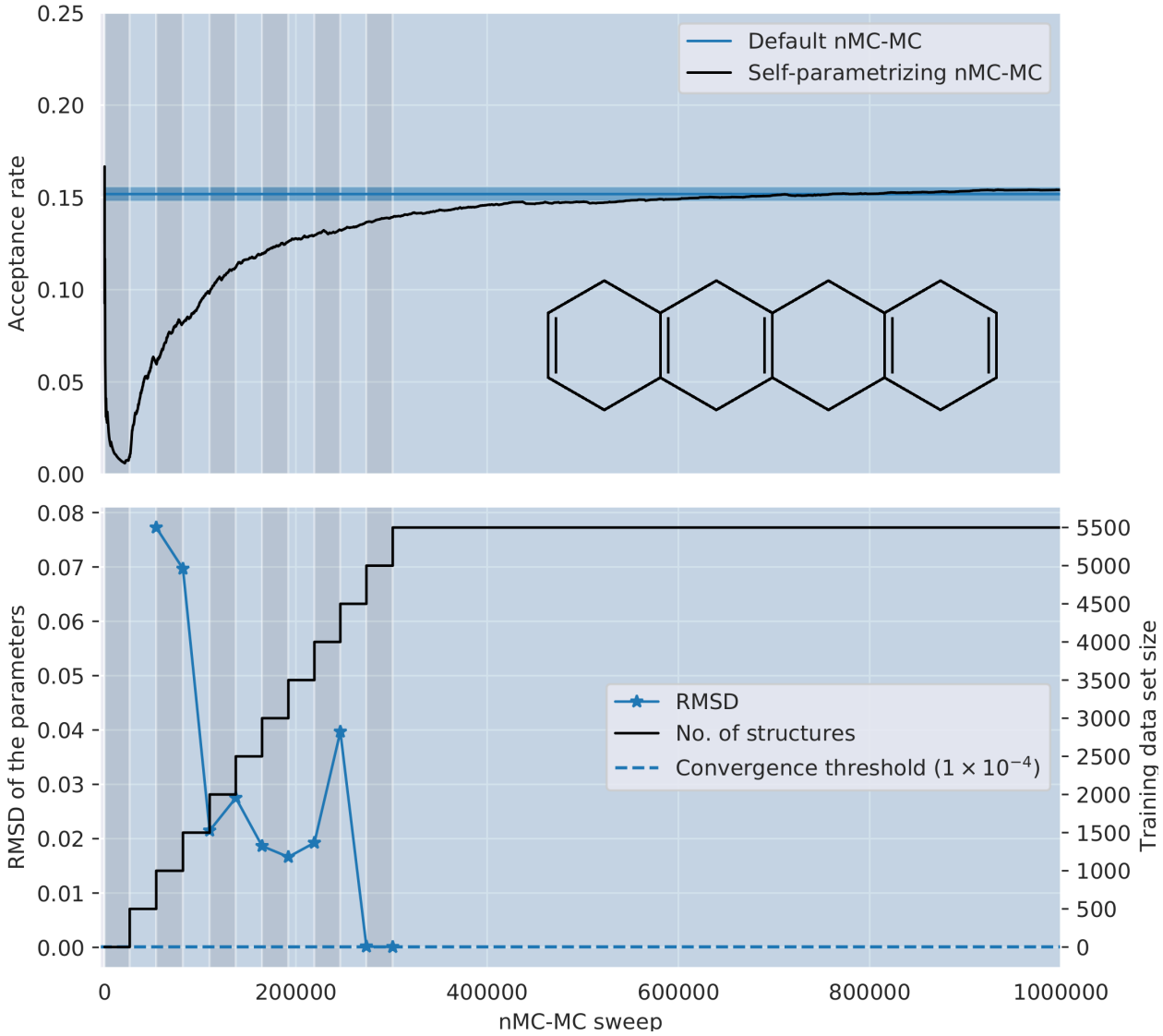

Figure 14: Top panel: acceptance rates of the self-parameterizing nMC-MC procedure as a function of the nMC-MC sweep for octahydrotetracene. The nMC-MC acceptance rate and standard deviation of the FF derived following the same philosophy applied for the test cases of figure 2 is also shown. The background shading indicates different iterations of the procedure. Bottom panel: plot of the RMSD of the FF parameters (left axis) and of the total number of structures in the training data set (right axis) as a function of the nMC-MC sweep.

3.6 nMC-MC Sampling into a QM/MM Hamiltonian

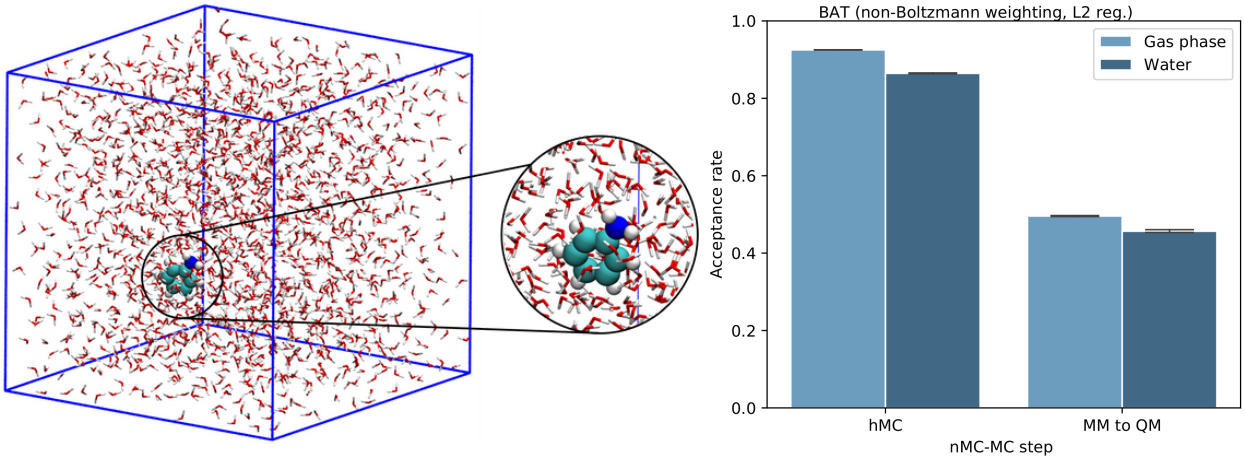

Figure 15: Left: Snapshot of the nMC-MC simulation of aniline in water. Right: Comparison of the hMC and switching step acceptance rates obtained for aniline in the gas phase and aqueous solution.

For a final test case, we immerse the aniline molecule in a TIP3P water box (the total system has 5327 atoms) and equilibrate the system in the NPT ensemble for 1 ns at 300 K using the Langevin integrator[91] (time step of 1 fs and friction coefficient of 2 ps^{-1}). The pressure of the system is maintained at 1 bar using the Monte Carlo barostat[92,93] implemented in OpenMM. Periodic boundary conditions are applied and long-range electrostatics are handled employing the Particle Mesh Ewald (PME) method.[94,95] The cutoff applied to all nonbonded interactions is 12 Å. The final configuration of the equilibration run, which has a box size of $37.38 \times 37.26 \times 38.55$ $Å^3$, is subsequently used as the starting point for a set of 4 NVT nMC-MC simulations in which the MM system is used in the low-level chain and a QM/MM model is employed in the high-level chain. The MM model used for aniline is the previously derived non-Boltzmann-weighted L2-regularized BAT FF.

In a system composed of a ligand in solution, the energy of the total MM system is given by

$$U^{MM}(\boldsymbol{q^s}, \boldsymbol{q^l}, \boldsymbol{q^{s-l}}) = U^{MM}_{sol}(\boldsymbol{q^s}) + U^{MM}_{lig}(\boldsymbol{q^l}) + U^{MM}_{lig-sol}(\boldsymbol{q^{s-l}}) \tag{12}$$

where U_{sol}^{MM} is the energy of the solvent (TIP3P waters), U_{lig}^{MM} the energy of the ligand (aniline), $U_{lig-sol}^{MM}$ is the ligand-solvent (aniline-TIP3P waters) interaction energy, and $\boldsymbol{q^s}$, $\boldsymbol{q^l}$ and $\boldsymbol{q^{s-l}}$ are the DOFs of the solvent, ligand, and the interaction between them. The QM/MM energy of a system in which only the ligand is included in the QM region and there are no covalent bonds between the ligand and solvent reads[96,97]

$$U^{QM/MM}(\boldsymbol{q^s}, \boldsymbol{q^l}, \boldsymbol{q^{s-l}}) = U_{sol}^{MM}(\boldsymbol{q^s}) + U_{lig}^{QM}(\boldsymbol{q^l}) + U_{lig-sol}^{MM}(\boldsymbol{q^{s-l}}) \tag{13}$$

where the only difference with respect to equation (12) is that now the potential energy of the ligand, U_{lig}, is calculated at the QM level. Note that the interaction between the MM and QM regions, $U_{lig-sol}^{MM}$, is still calculated at the MM level. For the present test case, the point charges used to calculate this interaction term do not change along with the simulation in the QM/MM Hamiltonian of equation (13). This corresponds to a mechanical embedding model with fixed-point charges in the QM region.[98] Consequently, by combining equations (12) and (13) we obtain that the $\Delta\Delta U$ that has to be introduced in the nMC-MC acceptance rate of equation (2) is given by

$$\begin{aligned} \Delta\Delta U(\boldsymbol{q_i}, \boldsymbol{q_f}) &= \left[U^{QM/MM}(\boldsymbol{q_f}) - U^{MM}(\boldsymbol{q_f})\right] - \left[U^{QM/MM}(\boldsymbol{q_i}) - U^{MM}(\boldsymbol{q_i})\right] \\ &= \left[U_{lig}^{QM}(\boldsymbol{q_f^l}) - U_{lig}^{MM}(\boldsymbol{q_f^l})\right] - \left[U_{lig}^{QM}(\boldsymbol{q_i^l}) - U_{lig}^{MM}(\boldsymbol{q_i^l})\right] \end{aligned} \tag{14}$$

from which we conclude that the switching step from the MM to the QM/MM Hamiltonian only requires the calculation of the energies of the ligand at the MM and QM levels. Equation (14) is employed in this example to sample the QM/MM distribution of aniline in a box of TIP3P waters.

The acceptance rates are shown in figure 15. As expected, the hMC acceptance rate decreases when going from gas phase to water solvent because of the increase in system size. There is also a small but significant decrease of *ca.* 4% in the switching step acceptance rate.

The successful application of the gas-phase-derived BAT FF is attributed to the fact that the conformational preferences of aniline in aqueous solution (see figure 16) are well captured by the nonbonded interactions. If this were not the case, a possible solution would involve reparameterizing GAFF using a training data set consisting of conformations extracted from explicit solution simulations. This is possible in ParaMol.

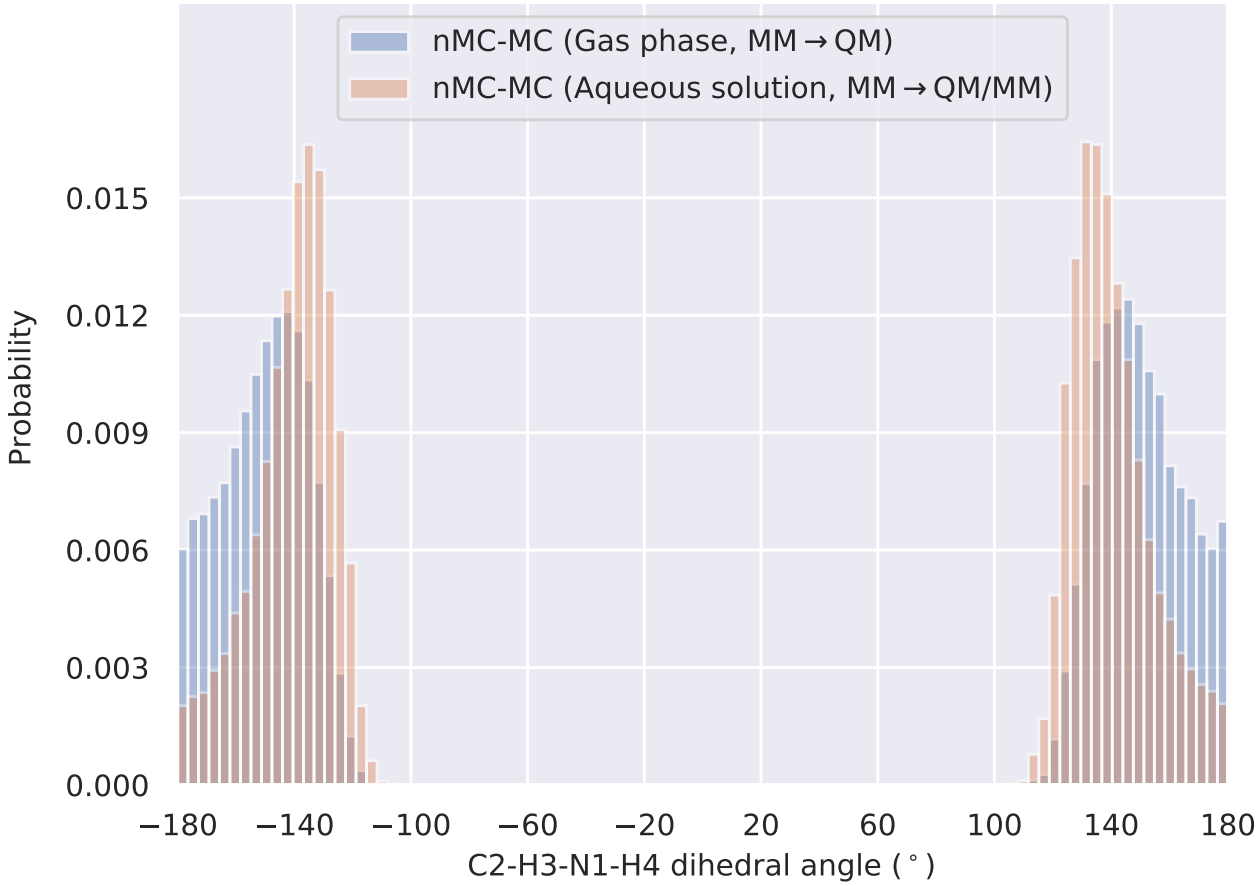

Figure 16: Comparison of the nMC-MC-sampled configurational distributions of aniline in gas phase and in aqueous solution.

In this application example, we use a fixed set of charges for the QM region, which not only simplifies the acceptance rate equation but also increases the similarity between the MM and the QM/MM models. Nevertheless, in QM/MM calculations it is also common to consider a QM region with varying partial charges, usually derived using a least-squares fitting to the QM electrostatic potential.[99–101] Although ParaMol provides the tools to parameterize nonbonded parameters in solution, from our experience, even in a mechanical embedding

context, in which the electrostatic coupling between the MM and QM regions is calculated at the MM level, a ligand with a varying set of MM partial charges is already a challenging case for the nMC-MC algorithm because the term $U_{lig-sol}^{MM'}(\boldsymbol{q}^{\boldsymbol{s-l}}) - U_{lig-sol}^{MM}(\boldsymbol{q}^{\boldsymbol{s-l}})$, where MM' is the FF with on-the-fly-fitted charges and MM the FF with original charges, does not vanish as it does in equation (14). Unfortunately, since energy is an extensive property, the larger the simulation box, the larger the mismatch between the QM/MM and MM levels of theory. Therefore, this QM/MM framework commonly leads to large energy differences in the ligand-solvent interaction energy that make the nMC-MC algorithm unviable for all but the simplest cases.

It is also important to stress that, besides conformational changes, there are important non-additive electrostatic effects that cannot be properly described by using a single set of charges. For example, the electrostatic embedding scheme, in which the electrostatic contribution from MM subsystem is included in the QM Hamiltonian, poses additional difficulties that worsen the mismatch between the MM and QM representations of the ligand due to polarization by the solvent. Hence, owing to the dynamic nature of the electrostatic cloud, these are typical applications for which fixed point-charges FFs are unsuitable, and where FF models that include descriptions of polarization and hyperconjugation, such as, *e.g.*, AMOEBA,[102,103] polarizable CHARMM[104,105] or fluc-q,[106–108] may prove their usefulness. The "electron spill-out" problem[109,110] is also a well-known pitfall of the electrostatic embedding scheme that may artificially distort the electron density of the QM region, thus increasing the mismatch between the energy of the QM/MM and MM models. So far, our research has set the mechanical-embedding fixed-point charge QM/MM model used above as the limiting case for successful sampling using the nMC-MC algorithm. In general, further complexity of the high-level models seems to be unsuitable to be reproduced by simple fixed point-charge FFs.

4 Conclusions

We present a multi-level procedure that allows estimation of the quantum configurational ensembles while keeping the computational cost at a minimum. This work is of paramount importance for conformational analysis because it combines the feasibility of computationally cheap methods, such as MM FFs, with the accuracy of the usually more expensive QM level of theory. The algorithms presented are implemented and made available in the ParaMol, free software that we develop and aims to ease the process of parameterization of MM FFs.[67] The code can be found at `https://github.com/JMorado/ParaMol`, and examples of how to use it are available through ParaMol's website `https://paramol.readthedocs.io`. We believe that the applications and results presented in this study may have implications in different areas of chemistry with biological relevance, especially for the drug design community.

The presented methodology involves coupling the hMC algorithm with a switching step between two Markov chains, the latter as formalized by Gelb. In the context of this work, the low-level Markov chain corresponds to a GAFF-like MM FF wherein sampling of configurations is performed, whereas the high-level Markov chain is the SCC-DFTB-D3 level of theory to which conformations are periodically attempted to be sampled. Owing to the low energetics similarity between GAFF and SCC-DFTB-D3, a straightforward application of the methodology leads to very slow convergence of the target configurational distributions due to low acceptance rates. Therefore, we resort to FF reparameterization as a means of ensuring sufficient overlap between the MM and QM levels of theory. We demonstrate this to be a successful strategy of generating more QM-like FFs and, consequently, increasing the nMC-MC switching step acceptance rates, thus accelerating the convergence of the sampling of the target quantum configurational distribution.

Overall, systematic reparameterization of FFs proves to be an efficient strategy to increase the acceptance rates of the switching step from the MM to the QM level of theory. The best acceptance rates are obtained for aniline (*ca.* 65%), whereas the molecule with the lowest possible acceptance rate is the fragment of cpd 26. This is expect since both molecular size

and chemical complexity have an impact on the acceptance rates. Moreover, we determine that the optimal reparameterization recipe involves employing non-Boltzmann weighting alongside L2 regularization. Uniform-weighted FFs, especially without regularization, are to be avoided as they easily sacrifice physicality in the FF parameters to obtain the best possible fit. These observations are in line with conclusions from a previous study.[67] The systematic parameterization also shows that hard DOFs, such as bonds and angles, are crucial to be reparameterized to increase the acceptance rates, mainly due to their large force constants. Reparameterization of charges with the uniform weighting scheme seems to be deleterious to the quality of the FFs. The acceptance rate data are supported by information obtained from various phase space overlap metrics. The latter reveal further insights into the features of the weighting methods and also lead us to suggest the switching step acceptance rates as a robust metric of phase space overlap.

We also present a self-parameterizing algorithm that combines sampling and FF parameterization in one scheme. This method does not require *a priori* generation of a training data set of unknown size, thus limiting the computational work to the strictly necessary. We illustrate its *modus operandi* and show that it gives identical results to the standard approach.

Finally, we also apply the nMC-MC algorithm to generate the QM/MM distribution of a ligand in an aqueous solution. We prove that, within a fixed-point charge mechanical embedding framework, the nMC-MC algorithm is a viable methodology that permits recovering the target QM/MM configurational ensemble. This application example also provides useful guidelines for future research efforts because it illustrates the limitations of using a GAFF-like MM FF as the low-level model. Since this FF type has fixed-point charges, it appears to be generally unsuitable for application in contexts involving varying solute charges, which may occur either due to conformational changes or polarization originating from electrostatic embedding. A possible solution for this bottleneck may involve resorting to polarizable FFs or machine-learning models.

Acknowledgement

The authors acknowledge the use of the IRIDIS High Performance Computing Facility and associated support services at the University of Southampton, as well as the UK Materials and Molecular Modelling Hub, which is partially funded by EPSRC (EP/P020194/1 & EP/T022213/1), for providing the computational resources used in the completion of this work. The authors also thank AstraZeneca for funding this study, and are grateful for the support from the EPSRC Centre for Doctoral Training, Theory and Modelling in Chemical Sciences under Grant EP/L015722/1.

Supporting Information Available

- *Prior* widths' values, description of the weighting methods, details of the QM and MM calculations, proof of independence of the phase space overlap metrics on the momentum coordinates, self-parameterizing nMC-MC convergence data, acceptance rates data, configurational distributions plots, nMC-MC acceptance rate *vs.* phase space correlation plots, and phase space overlap plots.
- Examples of input and output files associated with the results reported.

Conflict of Interest Disclosure

This research was partially funded by AstraZeneca.

References

(1) Tong, J.; Zhao, S. Large-Scale Analysis of Bioactive Ligand Conformational Strain Energy by Ab Initio Calculation. *J. Chem. Inf. Model.* **2021**, *61*, 1180–1192, PMID: 33630603.

(2) Chiarparin, E.; Packer, M. J.; Wilson, D. M. Experimental Free Ligand Conformations: A Missing Link in Structure-Based Drug Discovery. *Future Med. Chem.* **2019**, *11*, 79–82.

(3) Chen, I.-J.; Foloppe, N. Tackling the Conformational Sampling of Larger Flexible Compounds and Macrocycles in Pharmacology and Drug Discovery. *Bioorg. Med. Chem.* **2013**, *21*, 7898–7920.

(4) Blundell, C. D.; Nowak, T.; Watson, M. J. Measurement, Interpretation and Use of Free Ligand Solution Conformations in Drug Discovery. *Prog. Med. Chem.* **2016**, 45–147.

(5) Foloppe, N.; Chen, I.-J. Conformational Sampling and Energetics of Drug-Like Molecules. *Curr. Med. Chem.* **2009**, *16*, 3381–3413.

(6) Cui, D.; Zhang, B. W.; Matubayasi, N.; Levy, R. M. The Role of Interfacial Water in Protein-Ligand Binding: Insights from the Indirect Solvent Mediated PMF. *J. Chem. Theory Comput.* **2018**, *14*, 512–526.

(7) Li, Z.; Lazaridis, T. Water at Biomolecular Binding Interfaces. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573–581.

(8) Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for Molecular Recognition Exploiting Hydrophobic Enclosure in Protein–Ligand Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 808–813.

(9) Blundell, C. D.; Nowak, T.; Watson, M. J. In *Chapter Two - Measurement, Interpretation and Use of Free Ligand Solution Conformations in Drug Discovery*; Lawton, G., Witty, D. R., Eds.; Prog. Med. Chem.; Elsevier, 2016; Vol. 55; pp 45–147.

(10) Murakami, Y.; Omori, S.; Kinoshita, K. NLDB: A Database for 3D Protein–Ligand Interactions in Enzymatic Reactions. *J. Struct. Funct. Genomics* **2016**, *17*, 101–110.

(11) Berman, H. M. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(12) Balazs, A. Y. S.; Carbajo, R. J.; Davies, N. L.; Dong, Y.; Hird, A. W.; Johannes, J. W.; Lamb, M. L.; McCoull, W.; Raubo, P.; Robb, G. R.; Packer, M. J.; Chiarparin, E. Free Ligand 1D NMR Conformational Signatures To Enhance Structure Based Drug Design of a Mcl-1 Inhibitor (AZD5991) and Other Synthetic Macrocycles. *J. Med. Chem.* **2019**, *62*, 9418–9437.

(13) Li, Y.; Kang, C. Solution NMR Spectroscopy in Target-Based Drug Discovery. *Molecules* **2017**, *22*, 1399.

(14) Huggins, D. J.; Biggin, P. C.; Dämgen, M. A.; Essex, J. W.; Harris, S. A.; Henchman, R. H.; Khalid, S.; Kuzmanic, A.; Laughton, C. A.; Michel, J.; Mulholland, A. J.; Rosta, E.; Sansom, M. S. P.; van der Kamp, M. W. Biomolecular Simulations: From Dynamics and Mechanisms to Computational Assays of Biological Activity. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2019**, *9*, e1393.

(15) Rosa, M.; Micciarelli, M.; Laio, A.; Baroni, S. Sampling Molecular Conformers in Solution with Quantum Mechanical Accuracy at a Nearly Molecular-Mechanics Cost. *J. Chem. Theory Comput.* **2016**, *12*, 4385–4389.

(16) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.

(17) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(18) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431–9438.

(19) Swendsen, R. H.; Wang, J.-S. Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.

(20) Hukushima, K.; Nemoto, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608.

(21) Torrie, G.; Valleau, J. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

(22) Heenen, H. H.; Gauthier, J. A.; Kristoffersen, H. H.; Ludwig, T.; Chan, K. Solvation at Metal/Water Interfaces: An Ab Initio Molecular Dynamics Benchmark Of Common Computational Approaches. *J. Chem. Phys.* **2020**, *152*, 144703.

(23) Aminpour, M.; Montemagno, C.; Tuszynski, J. A. An Overview of Molecular Modeling for Drug Discovery with Specific Illustrative Examples of Applications. *Molecules* **2019**, *24*, 1693.

(24) Zhang, L.; Li, W.; Fang, T.; Li, S. Ab Initio Molecular Dynamics With Intramolecular Noncovalent Interactions for Unsolvated Polypeptides. *Theor. Chem. Acc.* **2016**, *135*, 34.

(25) Dračínský, M.; Möller, H. M.; Exner, T. E. Conformational Sampling by Ab Initio Molecular Dynamics Simulations Improves NMR Chemical Shift Predictions. *J. Chem. Theory Comput.* **2013**, *9*, 3806–3815.

(26) Wei, D.; Guo, H.; Salahub, D. R. Conformational Dynamics of an Alanine Dipeptide Analog: An Ab Initio Molecular Dynamics Study. *Phys. Rev. E* **2001**, *64*, 011907.

(27) Zivanovic, S.; Colizzi, F.; Moreno, D.; Hospital, A.; Soliva, R.; Orozco, M. Exploring the Conformational Landscape of Bioactive Small Molecules. *J. Chem. Theory Comput.* **2020**, *16*, 6575–6585.

(28) Duane, S.; Kennedy, A.; Pendleton, B. J.; Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **1987**, *195*, 216–222.

(29) Akhmatskaya, E.; Bou-Rabee, N.; Reich, S. A Comparison of Generalized Hybrid Monte Carlo Methods with and without Momentum Flip. *J. Comput. Phys.* **2009**, *228*, 2256–2265.

(30) Sweet, C. R.; Hampton, S. S.; Skeel, R. D.; Izaguirre, J. A. A Separable Shadow Hamiltonian Hybrid Monte Carlo Method. *J. Chem. Phys.* **2009**, *131*, 174106.

(31) Gelb, L. D. Monte Carlo Simulations Using Sampling from an Approximate Potential. *J. Chem. Phys.* **2003**, *118*, 7747–7750.

(32) Iftimie, R.; Salahub, D.; Wei, D.; Schofield, J. Using a Classical Potential as an Efficient Importance Function for Sampling from an Ab Initio Potential. *J. Chem. Phys.* **2000**, *113*, 4852.

(33) Burger, S. K.; Ayers, P. W.; Schofield, J. Efficient Parameterization of Torsional Terms for Force Fields. *J. Comput. Chem.* **2014**, *35*, 1438–1445.

(34) Sampson, C.; Fox, T.; Tautermann, C. S.; Woods, C.; Skylaris, C.-K. A "Stepping Stone" Approach for Obtaining Quantum Free Energies of Hydration. *J. Phys. Chem. B* **2015**, *119*, 7030–7040.

(35) Cave-Ayland, C.; Skylaris, C.-K.; Essex, J. W. A Monte Carlo Resampling Approach for the Calculation of Hybrid Classical and Quantum Free Energies. *J. Chem. Theory Comput.* **2017**, *13*, 415–424.

(36) Michel, J.; Taylor, R. D.; Essex, J. W. Efficient Generalized Born Models for Monte Carlo Simulations. *J. Chem. Theory Comput.* **2006**, *2*, 732–739.

(37) Leiding, J.; Coe, J. D. Reactive Monte Carlo Sampling with an Ab Initio Potential. *J. Chem. Phys.* **2016**, *144*, 174109.

(38) Mittal, A.; Lyle, N.; Harmon, T. S.; Pappu, R. V. Hamiltonian Switch Metropolis Monte Carlo Simulations for Improved Conformational Sampling of Intrinsically Disordered Regions Tethered to Ordered Domains of Proteins. *J. Chem. Theory Comput.* **2014**, *10*, 3550–3562.

(39) Mittal, A.; Holehouse, A. S.; Cohan, M. C.; Pappu, R. V. Sequence-to-Conformation Relationships of Disordered Regions Tethered to Folded Domains of Proteins. *J. Mol. Biol.* **2018**, *430*, 2403–2421.

(40) Ito, S.; Cui, Q. Multi-Level Free Energy Simulation with a Staged Transformation Approach. *J. Chem. Phys.* **2020**, *153*, 044115.

(41) Hudson, P. S.; Boresch, S.; Rogers, D. M.; Woodcock, H. L. Accelerating QM/MM Free Energy Computations via Intramolecular Force Matching. *J. Chem. Theory Comput.* **2018**, *14*, 6327–6335.

(42) Giese, T. J.; York, D. M. Development of a Robust Indirect Approach for MM $\rightarrow$ QM Free Energy Calculations That Combines Force-Matched Reference Potential and Bennett's Acceptance Ratio Methods. *J. Chem. Theory Comput.* **2019**, *15*, 5543–5562.

(43) Coe, J. D.; Sewell, T. D.; Shaw, M. S. Nested Markov Chain Monte Carlo Sampling of a Density Functional Theory Potential: Equilibrium Thermodynamics of Dense Fluid Nitrogen. *J. Chem. Phys.* **2009**, *131*, 074105.

(44) Coe, J. D.; Sewell, T. D.; Shaw, M. S. Optimal Sampling Efficiency in Monte Carlo Simulation with an Approximate Potential. *J. Chem. Phys.* **2009**, *130*, 164104.

(45) Bandyopadhyay, P. Increasing the Efficiency of Monte Carlo Simulation with Sampling from an Approximate Potential. *Chem. Phys. Lett.* **2013**, *556*, 341–345.

(46) Leiding, J.; Coe, J. D. An Efficient Approach to Ab Initio Monte Carlo Simulation. *J. Chem. Phys.* **2014**, *140*, 034106.

(47) Andricioaei, I.; Straub, J. E. Generalized Simulated Annealing Algorithms Using Tsallis Statistics: Application to Conformational Optimization of a Tetrapeptide. *Phys. Rev. E* **1996**, *53*, R3055–R3058.

(48) Jackson, N. E.; Webb, M. A.; de Pablo, J. J. Layered Nested Markov Chain Monte Carlo. *J. Chem. Phys.* **2018**, *149*, 072326.

(49) Ercolessi, F.; Adams, J. B. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *Europhys. Lett.* **1994**, *26*, 583–588.

(50) Nagai, Y.; Okumura, M.; Kobayashi, K.; Shiga, M. Self-Learning Hybrid Monte Carlo: A First-Principles Approach. *Phys. Rev. B* **2020**, *102*, 041124.

(51) Jadrich, R. B.; Leiding, J. A. Accelerating Ab Initio Simulation via Nested Monte Carlo and Machine Learned Reference Potentials. *J. Phys. Chem. B* **2020**, *124*, 5488–5497.

(52) Bandyopadhyay, P. Accelerating Quantum Mechanical/Molecular Mechanical Sampling Using Pure Molecular Mechanical Potential as an Importance Function: The Case of Effective Fragment Potential. *J. Chem. Phys.* **2005**, *122*, 091102.

(53) Nakayama, A.; Seki, N.; Taketsugu, T. Efficient Sampling for Ab Initio Monte Carlo Simulation of Molecular Clusters Using an Interpolated Potential Energy Surface. *J. Chem. Phys.* **2009**, *130*, 024107.

(54) Bulusu, S.; Fournier, R. Density Functional Theory Guided Monte Carlo Simulations: Application to Melting of Na_{13}. *J. Chem. Phys.* **2012**, *136*, 064112.

(55) Tamanini, E.; Buck, I. M.; Chessari, G.; Chiarparin, E.; Day, J. E. H.; Frederickson, M.; Griffiths-Jones, C. M.; Hearn, K.; Heightman, T. D.; Iqbal, A.; Johnson, C. N.; Lewis, E. J.; Martins, V.; Peakman, T.; Reader, M.; Rich, S. J.; Ward, G. A.;

Williams, P. A.; Wilsher, N. E. Discovery of a Potent Nonpeptidomimetic, Small-Molecule Antagonist of Cellular Inhibitor of Apoptosis Protein 1 (cIAP1) and X-Linked Inhibitor of Apoptosis Protein (XIAP). *J. Med. Chem.* **2017**, *60*, 4611–4625.

(56) Gelb, L. D.; Carnahan, T. N. Isothermal–Isobaric Monte Carlo Simulations of Liquid Lithium Using Density Functional Theory. *Chem. Phys. Lett.* **2006**, *417*, 283–287.

(57) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(58) Tuckerman, M. E. *Statistical Mechanics: Theory and Molecular Simulation*; Oxford University Press, 2010.

(59) Tomar, V. Accelerating the Molecular Time Steps for Nanomechanical Simulations: Hybrid Monte Carlo Method. *J. Appl. Phys. (Melville, NY, U. S.)* **2007**, *101*, 103512.

(60) Marsaglia, G.; Bray, T. A. A Convenient Method for Generating Normal Variables. *SIAM Rev.* **1964**, *6*, 260–264.

(61) Mehlig, B.; Heermann, D. W.; Forrest, B. M. Hybrid Monte Carlo Method for Condensed-Matter Systems. *Phys. Rev. B* **1992**, *45*, 679–685.

(62) 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.

(63) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98–103.

(64) Fang, Y.; Sanz-Serna, J. M.; Skeel, R. D. Compressible Generalized Hybrid Monte Carlo. *J. Chem. Phys.* **2014**, *140*, 174108.

(65) Frenkel, D.; Smit, B. *Understanding Molecular Simulations: From Algorithms to Applications*; Academic Press, 2002.

(66) Izaguirre, J. A.; Hampton, S. S. Shadow Hybrid Monte Carlo: An Efficient Propagator in Phase Space of Macromolecules. *J. Comput. Phys.* **2004**, *200*, 581–604.

(67) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-Consistent-Charge Density-Functional Tight-Binding Method for Simulations of Complex Materials Properties. *Phys. Rev. B* **1998**, *58*, 7260–7268.

(68) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(69) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; SciPy 1.0 Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272.

(70) Wu, D.; Kofke, D. A. Phase-Space Overlap Measures. I. Fail-Safe Bias Detection in Free Energies Calculated by Molecular Simulation. *J. Chem. Phys.* **2005**, *123*, 054103.

(71) Wu, D.; Kofke, D. A. Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations. *J. Chem. Phys.* **2005**, *123*, 084109.

(72) Hourahine, B.; Aradi, B.; Blum, V.; Bonafé, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshaye, M. Y.; Dumitrică, T.; Dominguez, A.; Ehlert, S.; Elstner, M.; van der

Heide, T.; Hermann, J.; Irle, S.; Kranz, J. J.; Köhler, C.; Kowalczyk, T.; Kubař, T.; Lee, I. S.; Lutsker, V.; Maurer, R. J.; Min, S. K.; Mitchell, I.; Negre, C.; Niehaus, T. A.; Niklasson, A. M. N.; Page, A. J.; Pecchia, A.; Penazzi, G.; Persson, M. P.; Řezáč, J.; Sánchez, C. G.; Sternberg, M.; Stöhr, M.; Stuckenberg, F.; Tkatchenko, A.; Yu, V. W.-z.; Frauenheim, T. DFTB+, a Software Package for Efficient Approximate Density Functional Theory Based Atomistic Simulations. *J. Chem. Phys.* **2020**, *152*, 124101.

(73) Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a Sparse Matrix-Based Implementation of the DFTB Method. *J. Phys. Chem. A* **2007**, *111*, 5678–5684.

(74) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(75) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.

(76) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. Hydrogen Bonding and Stacking Interactions of Nucleic Acid Base Pairs: A Density-Functional-Theory Based Treatment. *J. Chem. Phys.* **2001**, *114*, 5149–5155.

(77) Elstner, M.; Jalkanen, K. J.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. Energetics and Structure of Glycine and Alanine Based Model Peptides: Approximate SCC-DFTB, AM1 and PM3 Methods in Comparison with DFT, HF and MP2 Calculations. *Chem. Phys.* **2001**, *263*, 203–219.

(78) Elstner, M. The SCC-DFTB Method and Its Application to Biological Systems. *Theor. Chem. Acc.* **2006**, *116*, 316–325.

(79) Kraft, D. *A Software Package for Sequential Quadratic Programming*; Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt Köln: Forschungsbericht; Wiss. Berichtswesen d. DFVLR, 1988.

(80) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.

(81) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.

(82) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(83) Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121*, 4023–4039.

(84) Lim, V. T.; Hahn, D. F.; Tresadern, G.; Bayly, C. I.; Mobley, D. L. Benchmark Assessment of Molecular Geometries and Energies from Small Molecule Force Fields. *F1000Research* **2020**, *9*, 1390.

(85) Brand, J. C.; Williams, D. R.; Cook, T. J. Vibrational Analysis of the First Ultraviolet Band System of Aniline. *J. Mol. Spectrosc.* **1966**, *20*, 359–380.

(86) Mukherjee, M.; Bandyopadhyay, B.; Biswas, P.; Chakraborty, T. Amine Inversion Effects on the IR Spectra of Aniline in the Gas Phase and Cold Inert Gas Matrixes. *Indian J. Phys.* **2012**, *86*, 201–208.

(87) Bock, C. W.; George, P.; Trachtman, M. A Molecular Orbital Study of Nitrogen Inversion in Aniline with Extensive Geometry Optimization. *Theor. Chim. Acta* **1986**, *69*, 235–245.

(88) Gross, K. C.; Seybold, P. G. Substituent Effects on the Physical Properties and pKa of Aniline. *Int. J. Quantum Chem.* **2000**, *80*, 1107–1115.

(89) Gkeka, P.; Stoltz, G.; Barati Farimani, A.; Belkacemi, Z.; Ceriotti, M.; Chodera, J. D.; Dinner, A. R.; Ferguson, A. L.; Maillet, J.-B.; Minoux, H.; Peter, C.; Pietrucci, F.; Silveira, A.; Tkatchenko, A.; Trstanova, Z.; Wiewiora, R.; Lelièvre, T. Machine Learning Force Fields and Coarse-Grained Variables in Molecular Dynamics: Application to Materials and Biological Systems. *J. Chem. Theory Comput.* **2020**, *16*, 4757–4775.

(90) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential With DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192–3203.

(91) Izaguirre, J. A.; Sweet, C. R.; Pande, V. S. *Biocomputing 2010*; World Scientific, 2009; p 240–251.

(92) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular Dynamics Simulations of Water and Biomolecules With a Monte Carlo Constant Pressure Algorithm. *Chem. Phys. Lett.* **2004**, *384*, 288–294.

(93) Chow, K.-H.; Ferguson, D. M. Isothermal-Isobaric Molecular Dynamics Simulations With Monte Carlo Volume Sampling. *Comput. Phys. Commun.* **1995**, *91*, 283–289.

(94) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An Nlog(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(95) 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.

(96) Bakowies, D.; Thiel, W. Hybrid Models for Combined Quantum Mechanical and Molecular Mechanical Approaches. *J. Phys. Chem.* **1996**, *100*, 10580–10594.

(97) Weiner, S. J.; Singh, U. C.; Kollman, P. A. Simulation of Formamide Hydrolysis by Hydroxide Ion in the Gas Phase and in Aqueous Solution. *J. Am. Chem. Soc.* **1985**, *107*, 2219–2229.

(98) Vreven, T.; Morokuma, K. *Annu. Rep. Comput. Chem.*; Elsevier, 2006; Vol. 2; p 35–51.

(99) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The Resp Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(100) Reynolds, C. A.; Essex, J. W.; Richards, W. G. Atomic Charges for Variable Molecular Conformations. *J. Am. Chem. Soc.* **1992**, *114*, 9075–9079.

(101) Besler, B. H.; Merz, K. M.; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431–439.

(102) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *J. Chem. Theory Comput.* **2011**, *7*, 3143–3161.

(103) 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. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(104) Patel, S.; Brooks, C. L. CHARMM Fluctuating Charge Force Field for Proteins: I Parameterization and Application to Bulk Organic Liquid Simulations. *J. Comput. Chem.* **2004**, *25*, 1–16.

(105) Vanommeslaeghe, K.; MacKerell, A. CHARMM Additive and Polarizable Force Fields for Biophysics and Computer-Aided Drug Design. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 861–871.

(106) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical Fluctuating Charge Force Fields: Application to Liquid Water. *J. Chem. Phys.* **1994**, *101*, 6141–6156.

(107) Rick, S. W.; Berne, B. J. Dynamical Fluctuating Charge Force Fields: The Aqueous Solvation of Amides. *J. Am. Chem. Soc.* **1996**, *118*, 672–679.

(108) Ando, K. A Stable Fluctuating-Charge Polarizable Model for Molecular Dynamics Simulations: Application to Aqueous Electron Transfers. *J. Chem. Phys.* **2001**, *115*, 5228–5237.

(109) Dohn, A. O. Multiscale Electrostatic Embedding Simulations for Modeling Structure and Dynamics of Molecules in Solution: A Tutorial Review. *Int. J. Quantum Chem.* **2020**, *120*, e26343.

(110) Brunk, E.; Rothlisberger, U. Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States. *Chem. Rev. (Washington, DC, U. S.)* **2015**, *115*, 6217–6263.

Graphical TOC Entry