JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

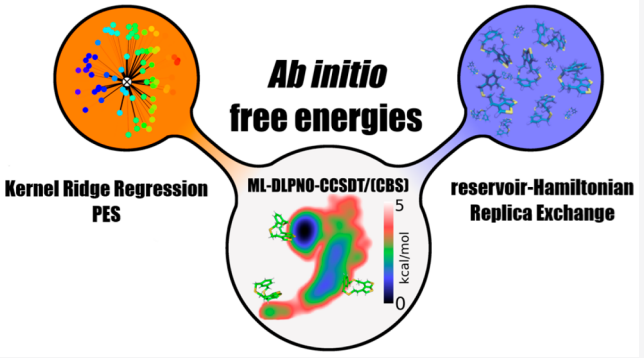
Hamiltonian-Reservoir Replica Exchange and Machine Learning Potentials for Computational Organic Chemistry

Raimon Fabregat, Alberto Fabrizio, Benjamin Meyer, Daniel Hollas, and Clémence Corminboeuf*

Cite This: J. Chem. Theory Comput. 2020, 16, 3084−3094 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: This work combines a machine learning potential energy function with a modular enhanced sampling scheme to obtain statistically converged thermodynamical properties of flexible medium-size organic molecules at high *ab initio* level. We offer a modular environment in the python package MORESIM that allows custom design of replica exchange simulations with any level of theory including ML-based potentials. Our specific combination of Hamiltonian and reservoir replica exchange is shown to be a powerful technique to accelerate enhanced sampling simulations and explore free energy landscapes with a quantum chemical accuracy unattainable otherwise (e.g., DLPNO-CCSD-(T)/CBS quality). This engine is used to demonstrate the relevance of accessing the *ab initio* free energy landscapes of molecules whose stability is determined by a subtle interplay between variations in the underlying potential energy and conformational entropy (i.e., a bridged asymmetrically polarized dithiacyclophane and a widely used organocatalyst) both in the gas phase and in solution (implicit solvent).



1. INTRODUCTION

Machine learning techniques are increasingly used to bypass expensive quantum chemical computations. Typical examples are machine learning-based potentials that are exploited to propagate the dynamical evolution of molecular systems on *ab initio* potentials at a fraction of the cost. The seminal work in the field comes from Behler and Parrinello,[1] who trained a generalized Artificial Neural Network (ANN) capable of predicting density functional theory-based energies and atomic forces and demonstrated its capability on bulk silicon[2] and then on carbon[3,4] and sodium.[5] Behler and Marquetand then applied the same approach to *n*-alkanes[6] and alanine tripeptides.[7] Comparable capabilities were achieved by Csanyi and co-workers, who used kernel-based method (i.e., the Gaussian approximation potential)[8] to propagate the density functional theory (DFT)−molecular dynamics (MD) of bulk crystal,[9] amorphous carbon,[10] and silicon.[11] Kernel ridge methods were also exploited for the "on-the-fly" propagation of the dynamic in the electronic states, circumventing the need for the explicit time dependent DFT or the CASSCF computations.[12] Roitberg et al. modified the Behler−Parrinello symmetry functions and proposed a deep neural network (ANAKIN-ME) to learn the potential of organic molecules approaching the CCSD(T)/CBS accuracy.[13−15] Such a high-accuracy level was also achieved by Tkatchenko and co-workers using gradient-domain machine learning (GDML).[16−18] The SchNet[19−21] deep learning architecture was also exploited to predict the potential energy surface and other quantum chemical properties of molecules and materials. Inspired by SchNet, Meuwly and Unke introduced the PhysNet architecture[22] for predicting energies, forces, or dipole moment of small organic molecules and polyalanines.

Overall, machine learning potentials (neural network or kernel-based) achieving post-Hartree−Fock or DFT accuracy were essentially employed for the molecular dynamics of fairly small and rigid systems (e.g., benzene, ethane and malonadehyde, aspirin, uracil, naphthalene, salicylic acid, and toluene)[16,17,23] or alternatively for larger systems with limited chemical diversity (e.g., peptides made of the same amino-acid type).[7] For these reasons, the associated (free) energy landscapes were explored using standard *ab initio* molecular dynamics without the need of making use of accelerated sampling approaches. Describing more flexible organic molecules (i.e., molecules that possess low-frequency (anharmonic) modes and multiple local minima close in energy) with machine learning potentials sets additional challenges, which influence both the accuracy of the ML potential and the

Received: January 29, 2020
Published: March 26, 2020

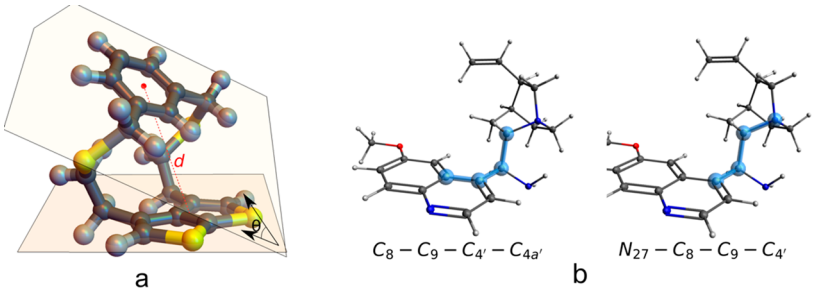


Figure 1. (a) Dithiacyclophane and the collective variables used to characterize its global structure: the distance between the center of masses of each cyclic bulk and the angles between the average planes going through them. (b) Cinchona alkaloid organocatalyst and the two dihedral angles used to characterize its global structure.

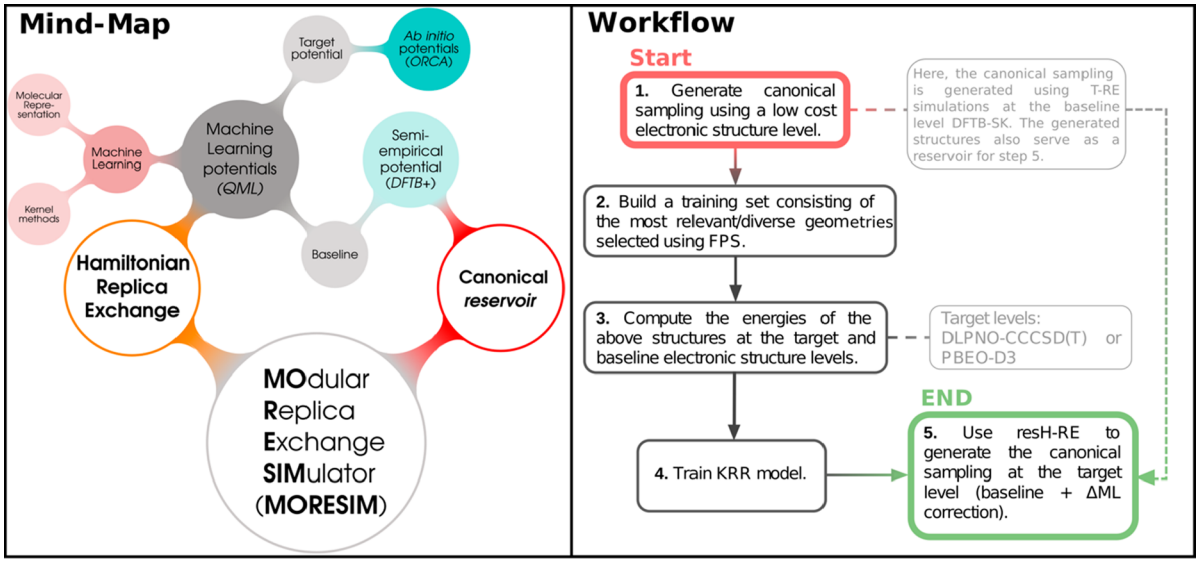


Figure 2. Mind-map and workflow illustrating the proposed methodology.

convergence of the statistical sampling of complex potential energy surfaces.

In 2016, one of us demonstrated[24] the utility of coupling temperature replica exchange[25] (T-RE) molecular dynamic with the most recent variant of density functional tight binding, i.e., DFTB3[26−28] to map the free energy landscapes of fluxional organic molecules in order to address organic chemistry problems that are not solvable solely relying on static electronic structure computations or standard molecular dynamics, the latter being too short to capture the interconversion between different possible states. A replica exchange molecular dynamics (REMD) simulation overcomes problems associated with running insufficiently long simulations by performing a series of energetically independent simulations (named *replicas*) of the same system in different equilibrium conditions and allowing them to occasionally exchange their configuration in a way that still ensures a canonical sampling within each replica. Replica exchange is especially appealing when relevant collective variables essential to a metadynamic[29] simulation are not easily identifiable[24] (see ref 30 for a recent example of metadynamics at the DFTB level). The most common version is T-RE,[25] where the replicas differ by their temperature. The additional insights provided by the coupling of REMD and DFTB3 (REMD@DFTB3)[24] were demonstrated on four examples including reaction energy pathways and conformational free energy differences, characteristic of organocatalysts, and flexible molecular rotors. While REMD@DFTB3 permits thorough exploration of potential energy surfaces at an affordable computational cost, the accuracy of the electronic structure method was sacrificed to ensure statistical convergence. In fact, the incompatibility associated with obtaining both converged statistical sampling and highly accurate energetics has traditionally prevented the ability to improve the quantum chemical description of moderately sized, yet highly flexible molecules that evolve on complex potential energy surfaces,[31−34] sometimes leading to catastrophic results.[35] In the present work, we achieve high-level *ab initio* accuracy by correcting semiempirical potentials with a machine learning model based on kernel ridge regression[36] combined with a more general enhanced sampling scheme connecting Hamiltonian[37] (H-RE) and reservoir[38] (resRE) replica exchange (i.e., resH-RE). With the former, the replicas evolve under a different Hamiltonian instead of a different temperature like in Temperature Replica Exchange, (i.e., T-RE). As for reservoir Replica Exchange, it was originally developed to improve T-RE by replacing the highest-temperature replica with a pool of structures (i.e., reservoir) acting as any other replica but exchanging conformations taken randomly from the pool. The proposed combination of Hamiltonian and reservoir RE (resH-RE) dramatically accelerates the exploration of *ab initio* free energy landscapes of archetypes flexible medium-size organic molecules that are dictated by a subtle energetic interplay originating from both enthalpic contributions and conformational entropy. The

https://dx.doi.org/10.1021/acs.jctc.0c00100
J. Chem. Theory Comput. 2020, 16, 3084−3094

illustrative systems considered herein (Figure 1) are motivated by our previous work[24,39,40] and are (1) the bridged asymmetrically polarized dithiacyclophane, incorporating a thieno[2,3-*b*],[41] and (2) a prototypical cinchona alkaloid organocatalyst.[40,42,43] Specifically, the first molecule is chosen because its relative conformational stability is governed by subtle intramolecular noncovalent interactions that necessitates an accurate *ab initio* treatment, while the large conformational entropy effects can only be accounted for by using accelerated sampling techniques. The free energy landscape of the organocatalyst is a complementary example that depends on individual energy contributions arising from rotational isomerism.

2. METHODS AND COMPUTATIONAL DETAILS

Overview. The proposed protocol is schematically illustrated in the workflow given in Figure 2, whereas all the details on the quantum chemistry, machine learning models, and enhanced sampling approaches are described in the upcoming individual sections. In brief, a low-cost semiempirical approach is used as a quantum chemical baseline, while the targeted free energies are achieved at an accurate *ab initio* target level with machine-learned corrections that learn the difference (Δ) between the baseline and the target (Δ-ML correction).[44] The semiempirical level is first used to generate a canonical sampling using T-RE that serves two purposes: (a) a subset of structures extracted from this T-RE simulation is selected to build a training set of energies/structures to train the Δ-ML model, and (b) the pool of structures associated with the T-RE is used as a reservoir (*vide infra*). The thorough exploration of the free energy landscapes, which is finally performed using a potential corresponding to the semiempirical level + Δ-ML correction, combines two variants of replica exchange that are Hamiltonian and reservoir RE (resH-RE). The resH-RE simulations were performed with a modular in-house python implementation of replica exchange uploaded in git-hub[45] in the form of a Python library under the name Modular Replica Exchange Simulator (MORESIM).

2.1. Quantum Chemical Potentials: Targets and Baseline. DFTB3[26] with the 3OB parameters[27,28] and the Slater−Kirkwood dispersion correction[46] (DFTB-SK) using the DFTB+[47] software is the chosen baseline for the Δ-ML model and for building the reservoir. The target potential is the domain-based local pair natural orbital coupled-cluster with perturbative triples (DLPNO-CCSD(T)/CBS)[48,49] as implemented in ORCA 4.0.[50] Complete Basis Set (CBS) extrapolations are performed following Neese's scheme starting from Dunning basis sets[51,52] (i.e., cc-pVDZ and cc-pVTZ) computations. PBE0[53]-D3[54]/(6-31G) is also used as target. The TeraChem[55,56] implementation serves to provide a comparison with the direct (exact PBE0 as opposed to ML-based) free energy computations at the PBE0-D3/(6-31G) level. For the direct computation, we follow the work from Martinez et al.[56] and utilized an MPI interface between AMBER and TeraChem GPU to accelerate single-point computations performed during the PBE0-D3/(6-31G) and AMBER[57] T-RE simulations. At each step of the dynamics, the converged density from the previous step was passed as the initial point for the SCF computation. These PBE0 simulations enable comparison between the explicit *ab initio* free energy landscapes and the faster ML ansatz, a comparison that is not possible at the DLPNO-CCSD(T) level. Details on the relative computational cost are provided in Figure S1.

Solvent effects were included implicitly using the SMD[58] model (with the dielectric constant of chloroform) at the PBE0-D3/(6-31G) level, also in ORCA 4.0.

2.2. Machine Learning Models. The ML corrections trained to learn the difference between the baseline and target levels are based on Kernel Ridge Regression[36] (KRR) and use the Spectrum of London Axilrod−Teller−Muto (SLATM)[59] molecular representation developed by von Lilienfeld and Huang in the Quantum Machine Learning (QML) package.[60] Among all the tested molecular representations (e.g., Coulomb Matrix,[61] Bag of bonds[62]), SLATM offered the best accuracy for the class of problems investigated herein. The KRR space was generated with a Gaussian kernel. The training set is built on the basis of the most distinct structures extracted from the DFTB-SK T-RE simulations and corresponds to 1500 and 1800 structures and energies for the bridged asymmetrically polarized dithiacyclophane (a), and a prototypical cinchona alkaloid organocatalyst (b), respectively. These sets were divided into a training and a validation set (200 and 300 random structures) and were used for validation for each system, respectively. Among the initial 1500 and 1800 structures/energies, a random subset of 500 was used to optimize the hyper-parameters (i.e., the standard deviation of the Gaussian kernel σ, and the regularization parameter λ) optimized with a Nelder−Mead simplex algorithm.[63] The quality of the trained model is evaluated by the mean absolute errors (MAE; see Figure S2) for the predictions on the test set. Overall, the final Δ-ML models offer an accuracy reaching 1 kcal/mol for both the dithiacyclophane and the cinchona alkaloid organocatalyst in comparison to the electronic energy computed at the exact reference level (see learning curves in Figure S2). Achieving such accuracy was a sufficient reason to favor kernel methods over a neural network.

2.3. Hamiltonian-Reservoir Replica Exchange. Two complementary sampling techniques are used for the exploration of the free energy landscapes computed with the Δ-ML models: Hamiltonian Replica Exchange[37] (H-RE) and reservoir Replica[38] (resRE). Approaches based on replica exchange typically use parallel simulations with modified parameters (temperature, Hamiltonian, atomic masses, ...) to facilitate the crossing of energy barriers and thus accelerate the sampling of canonical probability distributions.[25,64,65] Over the course of the simulation, the original replica, operating at the target conditions, exchange molecular conformations with the modified replicas as a way to introduce significant jumps in the phase-space. To ensure nonvanishing exchange probabilities, a sufficient number of replicas connecting the original conditions with the other extreme is introduced.

In H-RE, the transition between states is accelerated by creating intermediate potentials (i.e., between the baseline and target condition) using a modified Hamiltonian for each of the replicas.[64,66,67] In our implementation, H-RE exploits a reservoir of DFTB-SK structures obtained from previous simulations (*vide infra*). The replicas evolve at the same temperature (300 K) and under a potential $V_\lambda = (1 - \lambda)V_{\text{target}} + \lambda V_{\text{low}}$ that transition from DFTB-SK (low) to DFTB-SK + Δ-ML corresponding to targeted post-Hartree−Fock or DFT accuracy. The replica with $\lambda = 0$ evolves with the pure accurate potential $V_0 = V_{\text{target}}$ = DFTB-SK + Δ-ML, while the replica with $\lambda = 1$ corresponds to the lower-level potential $V_1 = V_{\text{low}}$ = DFTB-SK. In practice, the "highest" replica ($\lambda = 1$) is replaced by an available reservoir, generated with the low-level potential (i.e., DFTB-SK at 300 K), in the spirit of reservoir Replica

https://dx.doi.org/10.1021/acs.jctc.0c00100
J. Chem. Theory Comput. 2020, 16, 3084−3094

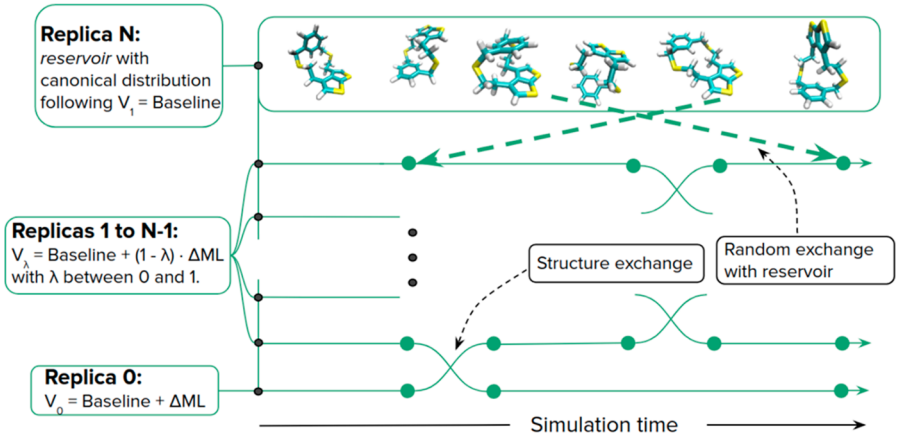


Figure 3. Schematic depiction of Hamiltonian reservoir Replica Exchange.

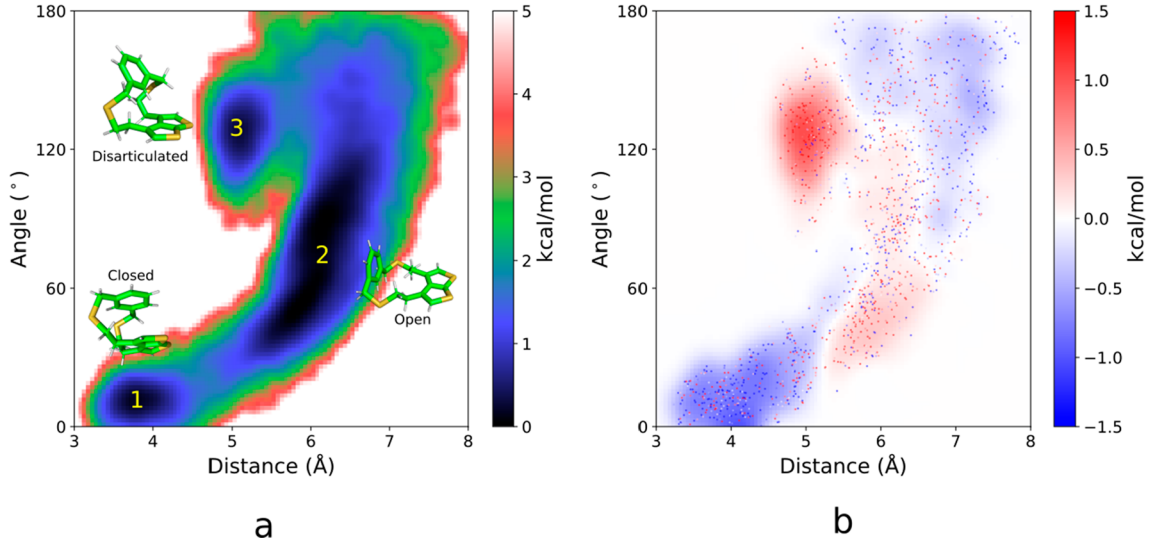


Figure 4. (a) Free energy landscape (DFTB-SK/3OB level) of dithiacyclophane at 300 K (T-RE) projected on the 2D space generated by the collective variables visible in Figure 1a. (b) Projection of the data set made of 1500 dithiacyclophane structures extracted with farthest point sampling from the 300 K canonical ensemble of 40 000 structures and color coded on the basis of the single-point energy difference ΔE = ((DFTB-SK/3OB) − (DLPNO-CCSD(T)/CBS)). The continuous background is plotted using a Gaussian interpolation of the mean energy difference. The smooth histograms were constructed with a Gaussian Kernel Density Estimator (Gaussian KDE) using the SciPy[73] python library.

Exchange (resRE) (see scheme in Figure 3). Here, the canonical DFTB-SK reservoirs (see Figures 4a and 5a) were taken from previous T-RE simulations (300 K) within i-PI.[68,69]

The rationale behind the use of a reservoir is to reduce the trajectory correlation time and thus the number of single point computations needed to achieve convergence.[38] Of course, resRE is only adequate to compute ensemble averages if the reservoir contains samples that follow a known physical distribution (in our case canonical), as otherwise it is not possible to derive a proper exchange probability with the other replicas.

By coupling the reservoir with H-RE rather than T-RE, we are able to explore the accurate *ab initio* level by accelerating the sampling without increasing the temperature, which leads to several advantages. In comparison to T-RE (achievable only at the semiempirical level), for which the jumps between free energy minima are generated by the exchange and acceptance probability with replica evolving at different equilibrium conditions (e.g., temperature), resH-RE exploits an existing reservoir of structures covering the entire conformational space that are accessible to the replica evolving at the target condition. In other words, by taking advantage of an *a priori* canonical sampling performed at an affordable semiempirical level, resH-RE will more efficiently simulate rare events. Additionally, fewer replicas (4/6 (H-RE) vs 16/48 (T-RE) (for the two considered molecules) are necessary to achieve optimal exchange probabilities, suggesting a larger overlap between the potential energy distributions of the replica than in T-RE.

In our context, another key advantage of exchanging with the reservoir is that replicas only serve to induce local diffusion in the phase-space, whereas the swaps between local minima in the free energy landscape (i.e., between basins) and crossings of energy barriers occur through the reservoir. The resH-RE simulation is thus achievable with thermalized Molecular Dynamics, but also with simple Monte Carlo (MC) moves (e.g., random particle moves) that are otherwise largely inefficient for systems with nonlinear potential energy surfaces like those investigated herein.[70]

In fact, many of the existing ML-based potentials have not yet been adapted to run molecular dynamics. With the kernel based approaches, the forces can be obtained from deriving the

https://dx.doi.org/10.1021/acs.jctc.0c00100
J. Chem. Theory Comput. 2020, 16, 3084−3094

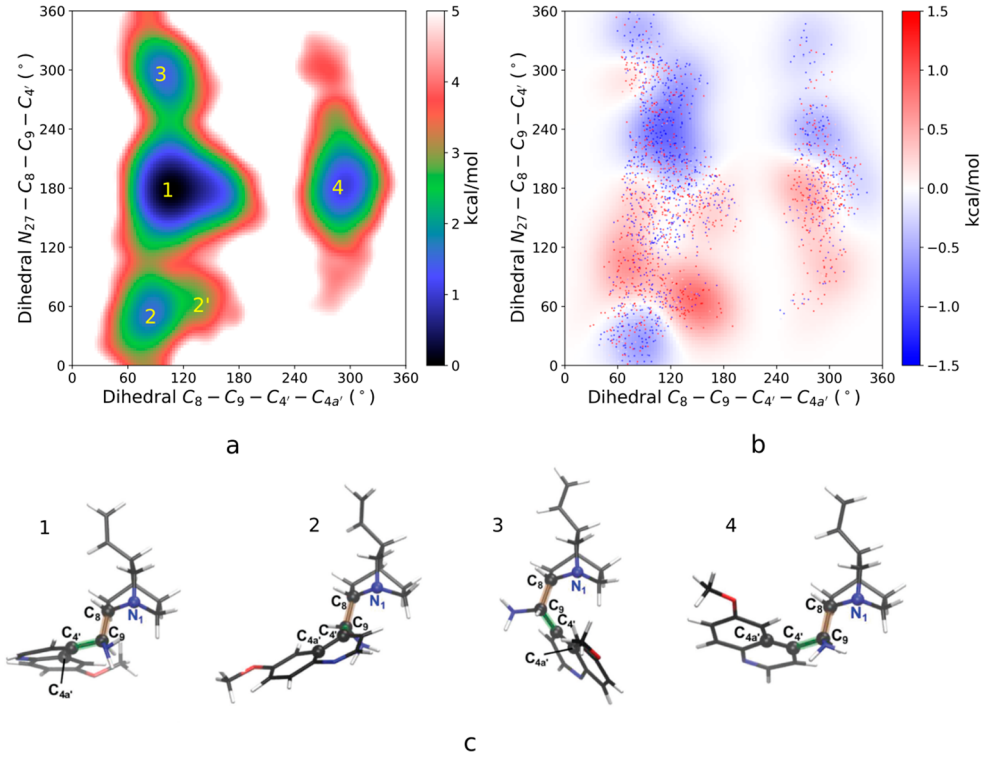


Figure 5. (a) Free energy landscape (DFTB-SK/3OB level) of the cinchona alkaloid organocatalyst at 300 K projected on the 2D space generated by the collective variables visible in Figure 1b. Constructed with canonical structures generated with T-RE simulations with DFTB-SK as potential energy. (b) Projection of the 1800 data set structures obtained with FPS from a canonical ensemble of 32 000 structures at 300 K canonical ensemble and color coded on the basis of the single point energy difference ΔE = ((DFTB-SK/3OB) − (DLPNO-CCSD(T)/CBS)). (c) Structures representing each of the four conformational regions (i.e., basins).

expression of the KRR (and thus of the molecular representation) with respect to the atomic coordinates,[12] but the task is not straightforward for the SLATM representation used here. The alternative, the Gradient Domain ML scheme developed by Müller et al.,[16−18] consists of learning the forces directly is considerably more expensive and only applicable to small molecules.

This work uses resH-RE with MC moves not only because of the unavailability of the forces but also to illustrate that the broad applicability of the sampling scheme with any of the existing ML potentials.

The convergence of free energy computations was evaluated by analyzing the evolution of the estimated relative free energies between basins (see section 4 in the Supporting Information). Given that the crossing between basins represents the slowest dynamical mode, the stabilization of the estimated basin free energies represents a good indicator of convergence. Statistical error boundaries on the estimated free energies were evaluated using a block jackknife with a width of one-tenth of simulation time.[71]

Note that the adopted approaches are applicable to any molecule but is specially designed for situations when free energy perturbation is not sufficient or suffers from convergence problems. resH-RE is indeed not a simple reweighting scheme; the reservoir in resH-RE is used to accelerate jumps over the conformational space, but the data generated by the replica that samples the canonical distribution of the target potential correspond to an unbiased sampling of the adequate probability distribution.

2.4. Technical Details. The 300 K free energy landscape of dithiacyclophane was obtained with a resH-RE simulation using four replicas (λ = 0, 0.33, 0.66, 1) exploiting a reservoir of 40 000 structures taken from a previous DFTB-SK[24] canonical distribution of structures obtained with T-RE at 300 K using the Langevin thermostat. The T-RE DFTB-SK simulation corresponds to 16 replicas with temperatures ranging from 300 to 1500 K in logarithmic intervals, and a time-step of 0.25 fs ran for 2.5×10^6 MD steps. A subset with the 1500 most distinct structures was extracted from the reservoir with Farthest Point Sampling[9] (FPS) (Figure 4b) and used to train the Δ-ML corrections. Six replicas (λ = 0, 0.2, 0.6, 0.8, 1) were required for the resH-RE simulations of the cinchona-based asymmetric organocatalyst, and the 1800 most distinct structures (Figure 5b) extracted from a reservoir of 32 000 structures obtained as discussed above (i.e., T-RE simulations) were used for the training. For both resH-RE simulations, it was ensured that the exchange rate between replicas reaches an optimal ∼30%.[72] Exchange between replicas was attempted every 20 MC steps consisting of a Gaussian random displacement of all atoms (in Cartesian coordinates) with standard deviation σ = 0.003 Å set to 50% acceptance rate. The MC simulations correspond to a total of 10^6 steps for both systems.

The 300 K canonical sampling at the direct PBE0-D3/(6-31G) level was also generated with a T-RE simulation using TeraChem[55,56] in combination with AMBER.[57] The simulation corresponds to 10 replicas with temperatures ranging from 300 to 1500 K in logarithmic intervals, and a time-step of 0.75 fs ran for 1.5×10^6 MD steps.

https://dx.doi.org/10.1021/acs.jctc.0c00100
J. Chem. Theory Comput. 2020, 16, 3084−3094

3. RESULTS

3.1. Dithiacyclophane. The three conformational regions of dithiacyclophane, previously investigated by one of us,[24] and visible in Figure 4a, are controlled by very distinct enthalpic and entropic contributions. The π-stacked "closed" conformer is stabilized by long-range correlation effects and only captured at the DFT level upon the addition of a London dispersion correction.[39,40,74] In sharp contrast, the open conformer is highly flexible and driven by entropy and large anharmonic effects. The third "disarticulated" conformer is rigid but less sensitive to London dispersion forces in comparison to the closed region. Our former T-RE DFTB simulations that captured all the conformational entropy effects, highlighted the limitation of the harmonic approximation for describing the relative stability of the most floppy (i.e., open) conformer (see Figure 4a).

The harmonic approximation fails to account for the full conformational entropy contributions and the anharmonic nature of the open state (**2**) (see ref 75 for relevant examples of approximations for anharmonic free energies) and to a lesser extent of the closed conformational region (**1**). The large entropic contributions characterizing the open region (see Figure 4a) makes it the lowest-energy conformer at the DFTB-SK level at 300 K and temperatures above. Yet, the DFTB relative free energies between the three conformers are very small (within 1 kcal mol^{-1}). Converging the statistical sampling comes with a quantum chemical cost and the affordable semiempirical level is not expected to capture all these subtle energy differences accurately.

The ML correction to DFTB-SK offers access to converged DLPNO-CCSD(T) free energy profiles at a fraction of the cost (*vide supra*). Prior to obtaining the full free energy landscapes with resH-RE, it is interesting to identify the trends emerging from the Δ-ML correction added to the DFTB-SK energy of the structures in the reservoir (Figure 6). The 0.79 regression

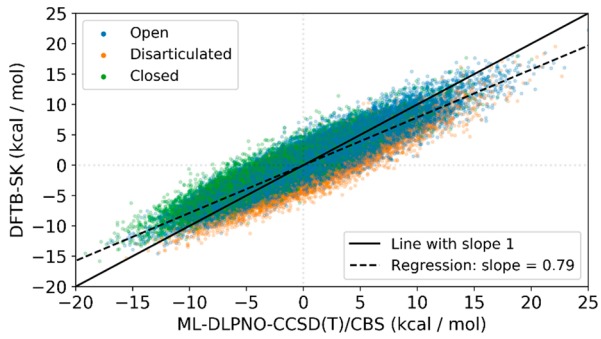


Figure 6. Comparison between the DFTB-SK electronic energy and the ML-DLPNO-CCSD(T)/CBS predictions (i.e., DFTB-SK + ΔML correction) for the 40 000 structures in the reservoir.

slope between DFTB-SK and ML-DLPNO-CCSD(T) is indicative of the much flatter potential energy surface of the former or, in other words, an underestimation of the energy differences and barriers across the energy landscape (with PBE0-D3/(6-31G) the slope is 0.74; see Figure S3).

The consequence of these energy discrepancies is clear when comparing the full free energy landscapes and relative free energies (upon integration within the free energy basins,[24] Figure 7) obtained with resH-RE sampling at different quantum chemical levels. Overall, the shape of the ML-DLPNO-CCSD(T)/CBS and DFTB-SK profiles are very similar but the disarticulated basin is strongly favored by CCSD(T) at the detriment of the close conformer (>2 kcal mol^{-1} higher). In contrast to the flat DFTB-SK free energy profile, the ML-DLPNO-CCSD(T) landscape is clearly uneven highlighting the difficulties of the tight-binding method to reproduce the delicate interaction interplay characterizing the conformational regions of this illustrative system. The PBE0 free energy landscapes are even less flat and favor the disarticulated state even more with the closed structure being around 3 kcal mol^{-1} higher. Yet, the excellent agreement between PBE0-D3/(6-31G) and ML-PBE0-D3/(6-31G) is a proof-of-principle demonstration that this trend is not an artifact from the ML potentials (see Figure 7d,e).

Note that a direct approach using DLPNO-CCSD(T)/CBS is not achievable, given the intrinsic computational cost of the method. For this specific reason, a relatively small basis set was chosen for the DFT profiles. While some of the deviations between DFTB-SK and the higher-level approaches (i.e., much smaller energy differences) are already apparent in the static picture (see the harmonic free energies in Figure 7), the deviations between methods are more pronounced when accounting for thermal fluctuations. With PBE0-D3, the approximated barriers between each conformational regions are over 5 kcal mol^{-1}, which is significantly higher than the DFTB-SK (<2 kcal mol^{-1}).

The Monte Carlo approaches used in this work are not easily compatible with the inclusion of explicit solvent, but the effect of the environment is, of course, essential to decipher the true molecular behavior and its associated PES. As a compromise, solvent effects were incorporated implicitly with the SMD continuum model (with the chloroform dielectric constant[41]) at the PBE0 level. The inclusion of these effects severely affects the relative energetic stability and flatten the entire profile looking much more similar to the gas phase DFTB-SK profile. The limitation associated with the continuum model could be overcome by using an additional potential that models the interaction between the solute and explicit solvent within a dynamic simulation, a possibility that we will explore in future work.

3.2. Cinchona Alkaloid. The same methodology is applied to a common cinchona-based asymmetric organocatalyst for which we also generated the *ab initio* free energy landscape (Figure 8).[24] The 2D conformational map extracted from the 2016 T-RE simulations at the DFTB level revealed four easily accessible conformational regions (**1**−**4**) and one that is much less populated (**2′**). Unexpected from previous static computations was the dihedral angle (open rather than close) characteristic of the conformational state **3**. Other added values from the T-RE simulation were the demonstration of the pronounced entropic nature of **1** at 300 K reversing the relative stability between **1** and **4** in comparison with the static computation and the appearance of region **2′**. Yet, the **1** and **4** conformational regions were within 2 kcal mol^{-1} stressing the importance of improving upon the DFTB level.

With a slope much smaller than 1 (i.e., 0.76), Figure S6 confirms the general flatness of the DFTB-SK potential compared to that of DLPNO-CCSD(T). Note that the DFTB3 underestimation of the rotational barriers and of the relative energy differences, which originates from the limited amount of atomic overlap afforded by the use of a minimal basis set, is reminiscent of other examples in the literature.[27,76]

https://dx.doi.org/10.1021/acs.jctc.0c00100
J. Chem. Theory Comput. 2020, 16, 3084−3094

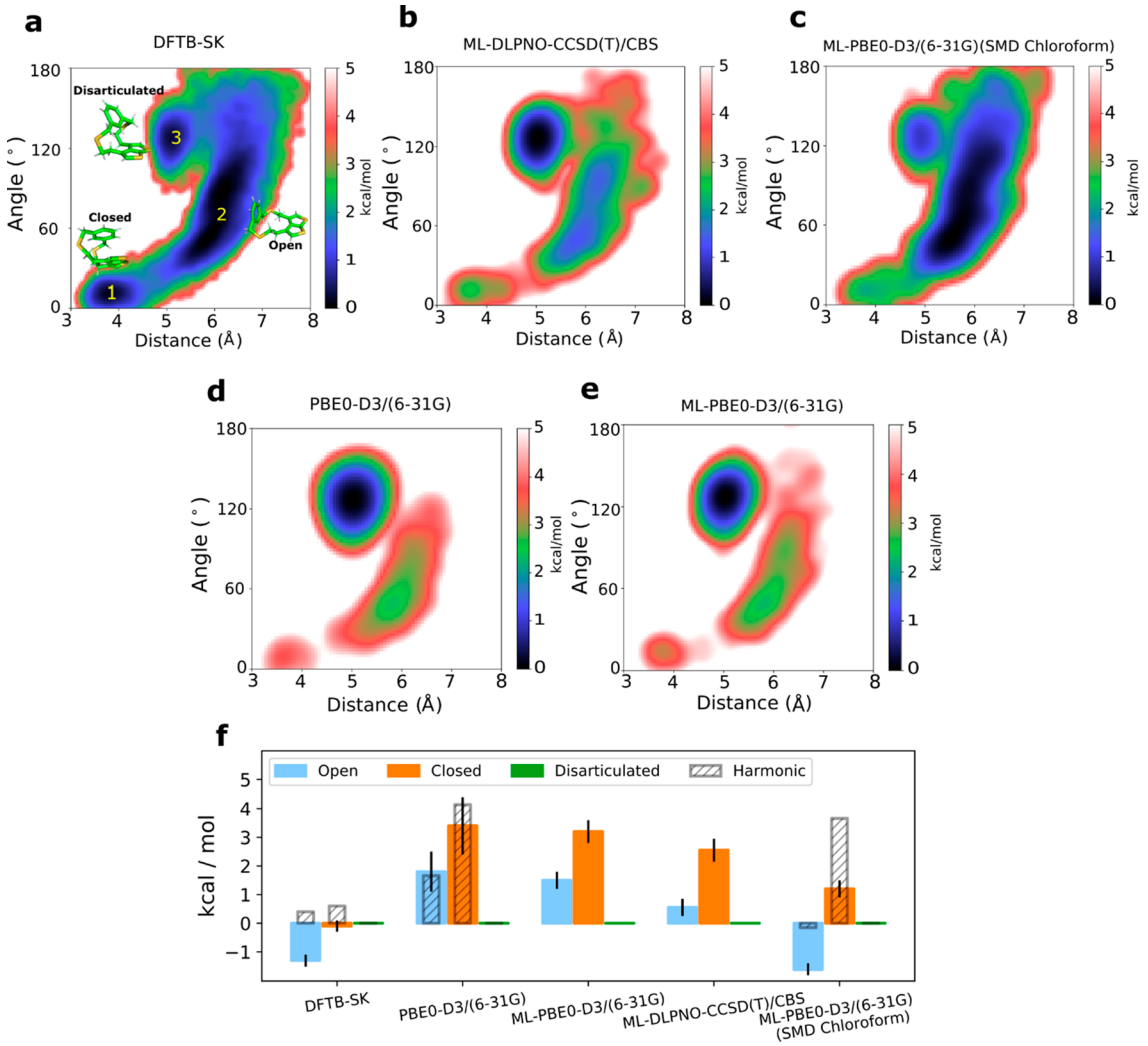


Figure 7. Free energy landscapes at 300 K generated with the potential: (a) DFTB-SK; (b) ML-DLPNO-CCSD(T)/CBS; (c) ML-[PBE0-D3/(6-31G)(SMD Chloroform)]; (d) PBE0-D3/(6-31G); (e) ML-PBE0-D3/(6-31G). (f) Relative free energies by integration within the local minima.[24] The free energies are all given relative to the Disarticulated state except for the solvated system, where the open state is used as a reference. The striped columns correspond to the static relative free energy using the harmonic approximation (for the solvated system the harmonic free energies were computed with the true potential, and not with the machine learning version). All the free energies maps come from resH-RE expect for the direct PBE0, which uses T-RE, as described in the methods section.

Figure 8 compares the full DFTB profile with those obtained with ML-DLPNO-CCSD(T) and ML-PBE0-D3 accounting for the implicit effect of the chloroform environment.[77] The general shape of the ML-DLPNO-CCSD(T)/CBS free energy landscape of the organocatalyst is once again similar to the DFTB-SK one but with significant differences. Specifically, **4** is clearly enthalpically stabilized at the higher level, whereas the flexibility of the conformational region **1** is enhanced (i.e., larger spread over the dihedral angle characteristic of the *syn*/*anti* conformation). The metastable region **2′** is also more populated at this level. Quantitatively, these trends translate into **1** and **4** lying very close in energies (within 0.5 kcal mol^{-1}) with state **3** being nearly 3 kcal mol^{-1} higher and disconnected from region **1** (i.e., a high barrier separating the two regions). Akin to the dithiacyclophane, the PBE0 gas-phase profile (Figure S8) is much closer to ML-DLPNO-CCSD(T)/CBS than the implicit solvated (in chloroform) profile but the flattening of the free energy landscape of the cinchona derivatives upon implicit solvation is less pronounced than for the dithiacyclophane (see Figure 8c and Figure S8). Overall, the effect of the solvent on conformer **3** is negligible but the metastable **2′** specie disappears, while **2**/**4** are more/less populated.

These two complementary examples are associated with different energetic driving forces that are the interplays between pronounced intramolecular vdW interactions and conformational entropy in the first case and the individual contributions arising from rotational isomerism in the second.

4. CONCLUSIONS

In 2016, we highlighted the importance of thorough mappings of the free energy landscapes for solving problems in computational organic chemistry. In this subsequent step, we demonstrate how to exploit a variant of Hamiltonian replica exchange and kernel-based machine learning potentials to achieve a remarkable accuracy/cost ratio and accelerate the accurate predictions of relative free energies, which is one of the most challenging goals in computational quantum chemistry. Overall, our results stress the relevance of improving the entropic and enthalpic description of flexible organic

https://dx.doi.org/10.1021/acs.jctc.0c00100
J. Chem. Theory Comput. 2020, 16, 3084−3094

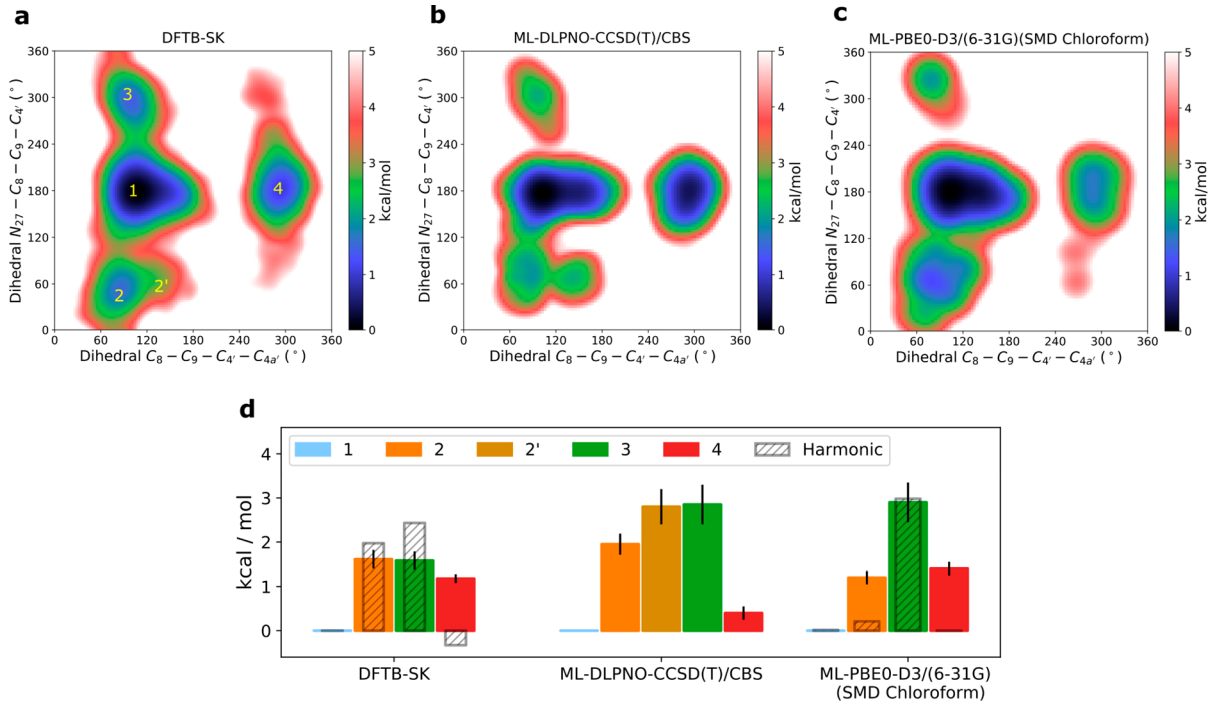


Figure 8. Free energy landscapes at 300 K generated with the potential: (a) DFTB-SK; (b) ML-DLPNO-CCSD(T)/CBS; (c) ML-[PBE0-D3/(6-31G)(SMD Chloroform)]. (d) Free energies upon integration within the free energy basins. The free energies are all given relative to the Disarticulated state. The stripped columns are the free energy predictions of the basins using the static free energies using the harmonic correction.

molecules having complex free energy landscapes dictated by subtle energetic interplays. In particular, based on comparisons between the DFTB-SK baseline and the ML-DLPNO-CCSD-(T) target, one concludes that the semiempirical method generally leads to much flatter free energy landscapes. Similarly, such systems are poorly described by the picture arising from static free energies, which underestimate the conformational entropy of the most flexible conformational regions. For all these reasons, our original combination of Hamiltonian and reservoir replica exchange and its implementation into a modular environment (the python package MORESIM) represents a powerful solution capable of accelerating enhanced sampling simulations involving any machine learning-based or *ab initio* potential energies.

Subsequent objectives will consist of using the same workflow but circumventing the reduced transferability associated with the use of a global molecular machine learning representation by deriving a differentiable kernel approach based on local atomic environment that is also compatible with molecular dynamic simulations.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c00100.

Histogram of the cost of computations, learning curves, energy comparisons, basin potential energies, energies scatter plot, free energy landscapes and convergences, and detailed information on the training procedure (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Clémence Corminboeuf − *Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland;* orcid.org/0000-0001-7993-2879; Email: clemence.corminboeuf@epfl.ch

Authors

Raimon Fabregat − *Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering, École polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*

Alberto Fabrizio − *Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland;* orcid.org/0000-0002-4440-3149

Benjamin Meyer − *Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*

Daniel Hollas − *Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering, École polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland;* orcid.org/0000-0003-4075-6438

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.0c00100

https://dx.doi.org/10.1021/acs.jctc.0c00100
J. Chem. Theory Comput. 2020, 16, 3084−3094

Notes

The authors declare no competing financial interest.

The data that support the findings of this study are openly available in Materials Cloud at https://doi.org/10.24435/materialscloud:2020.0033/v1, while the Modular Replica Exchange Simulator (MORESIM) framework is available on github http://doi.org/10.5281/zenodo.3630553.

■ ACKNOWLEDGMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No 817977). AF. and B.M. acknowledge the National Centre of Competence in Research (NCCR) "Materials' Revolution: Computational Design and Discovery of Novel Materials (MARVEL)" of the Swiss National Science Foundation (SNSF) for financial support.

■ REFERENCES

(1) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98* (14), 146401.

(2) Behler, J.; Marto crnnák, R.; Donadio, D.; Parrinello, M. Metadynamics Simulations of the High-Pressure Phases of Silicon Employing a High-Dimensional Neural Network Potential. *Phys. Rev. Lett.* **2008**, *100* (18), 185501.

(3) Khaliullin, R. Z.; Eshet, H.; Kühne, T. D.; Behler, J.; Parrinello, M. Graphite-Diamond Phase Coexistence Study Employing a Neural-Network Mapping of the Ab Initio Potential Energy Surface. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *81* (10), 100103.

(4) Khaliullin, R. Z.; Eshet, H.; Kühne, T. D.; Behler, J.; Parrinello, M. Nucleation Mechanism for the Direct Graphite-to-Diamond Phase Transition. *Nat. Mater.* **2011**, *10* (9), 693.

(5) Eshet, H.; Khaliullin, R. Z.; Kühne, T. D.; Behler, J.; Parrinello, M. Ab Initio Quality Neural-Network Potential for Sodium. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *81* (18), 184107.

(6) Gastegger, M.; Kauffmann, C.; Behler, J.; Marquetand, P. Comparing the Accuracy of High-Dimensional Neural Network Potentials and the Systematic Molecular Fragmentation Method: A Benchmark Study for All-Trans Alkanes. *J. Chem. Phys.* **2016**, *144* (19), 194110.

(7) Gastegger, M.; Behler, J.; Marquetand, P. Machine Learning Molecular Dynamics for the Simulation of Infrared Spectra. *Chem. Sci.* **2017**, *8* (10), 6924−6935.

(8) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: the Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104* (13), 136403.

(9) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine Learning Unifies the Modeling of Materials and Molecules. *Sci. Adv.* **2017**, *3* (12), No. e1701816.

(10) Caro, M. A.; Deringer, V. L.; Koskinen, J.; Laurila, T.; Csányi, G. Growth Mechanism and Origin of High Sp3 Content in Tetrahedral Amorphous Carbon. *Phys. Rev. Lett.* **2018**, *120*, 166101.

(11) Deringer, V. L.; Bernstein, N.; Bartók, A. P.; Cliffe, M. J.; Kerber, R. N.; Marbella, L. E.; Grey, C. P.; Elliott, S. R.; Csányi, G. Realistic Atomistic Structure of Amorphous Silicon from Machine-Learning-Driven Molecular Dynamics. *J. Phys. Chem. Lett.* **2018**, *9* (11), 2879−2885.

(12) Hu, D.; Xie, Y.; Li, X.; Li, L.; Lan, Z. Inclusion of Machine Learning Kernel Ridge Regression Potential Energy Surfaces in On-the-Fly Nonadiabatic Molecular Dynamics Simulation. *J. Phys. Chem. Lett.* **2018**, *9* (11), 2725−2732.

(13) 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* (4), 3192−3203.

(14) Smith, J. S.; Nebgen, B.; Lubbers, N.; Isayev, O.; Roitberg, A. E. Less Is More: Sampling Chemical Space with Active Learning. *J. Chem. Phys.* **2018**, *148* (24), 241733.

(15) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching Coupled Cluster Accuracy with a General-Purpose Neural Network Potential through Transfer Learning. *Nat. Commun.* **2019**, *10* (1), 2903.

(16) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K. R. Machine Learning of Accurate Energy-Conserving Molecular Force Fields. *Sci. Adv.* **2017**, *3* (5), No. e1603015.

(17) Chmiela, S.; Sauceda, H. E.; Müller, K. R.; Tkatchenko, A. Towards Exact Molecular Dynamics Simulations with Machine-Learned Force Fields. *Nat. Commun.* **2018**, *9* (1), 3887.

(18) Sauceda, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K. R.; Tkatchenko, A. Molecular Force Fields with Gradient-Domain Machine Learning: Construction and Application to Dynamics of Small Molecules with Coupled Cluster Forces. *J. Chem. Phys.* **2019**, *150* (11), 114102.

(19) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, *8* (1), 13890.

(20) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K. R. SchNet-A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **2018**, *148* (24), 241722.

(21) Schütt, K. T.; Kessel, P.; Gastegger, M.; Nicoli, K. A.; Tkatchenko, A.; Müller, K. R. SchNetPack: A Deep Learning Toolbox for Atomistic Systems. *J. Chem. Theory Comput.* **2019**, *15* (1), 448−455.

(22) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15* (6), 3678−3693.

(23) Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K. R. Bypassing the Kohn-Sham Equations with Machine Learning. *Nat. Commun.* **2017**, *8* (1), 872.

(24) Petraglia, R.; Nicolaï, A.; Wodrich, M. D.; Ceriotti, M.; Corminboeuf, C. Beyond Static Structures: Putting Forth REMD as a Tool to Solve Problems in Computational Organic Chemistry. *J. Comput. Chem.* **2016**, *37* (1), 83−92.

(25) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(26) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7* (4), 931−948.

(27) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9* (1), 338−354.

(28) Gaus, M.; Lu, X.; Elstner, M.; Cui, Q. Parameterization of DFTB3/3OB for Sulfur and Phosphorus for Chemical and Biological Applications. *J. Chem. Theory Comput.* **2014**, *10* (4), 1518−1537.

(29) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562−12566.

(30) Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **2019**, *15* (5), 2847−2862.

(31) Kaczor, A.; Reva, I. D.; Proniewicz, L. M.; Fausto, R. Importance of Entropy in the Conformational Equilibrium of Phenylalanine: A Matrix-Isolation Infrared Spectroscopy and Density Functional Theory Study. *J. Phys. Chem. A* **2006**, *110* (7), 2360−2370.

(32) Ess, D. H.; Wheeler, S. E.; Iafe, R. G.; Xu, L.; Çelebi-Ölçüm, N.; Houk, K. N. Bifurcations on Potential Energy Surfaces of Organic Reactions. *Angew. Chem., Int. Ed.* **2008**, *47* (40), 7592−7601.

(33) Rehbein, J.; Carpenter, B. K. Do We Fully Understand What Controls Chemical Selectivity? *Phys. Chem. Chem. Phys.* **2011**, *13* (47), 20906.

(34) Schreiner, P. R.; Reisenauer, H. P.; Ley, D.; Gerbig, D.; Wu, C.-H.; Allen, W. D. Methylhydroxycarbene: Tunneling Control of a Chemical Reaction. *Science* **2011**, *332* (6035), 1300−1303.
(35) Plata, R. E.; Singleton, D. A. A Case Study of the Mechanism of Alcohol-Mediated Morita Baylis-Hillman Reactions. The Importance of Experimental Observations. *J. Am. Chem. Soc.* **2015**, *137* (11), 3811−3826.
(36) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*; Springer Series in Statistics; Springer New York: New York, NY, 2009.
(37) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002**, *116* (20), 9058−9067.
(38) Okur, A.; Roe, D. R.; Cui, G.; Hornak, V.; Simmerling, C. Improving Convergence of Replica Exchange Simulations through Coupling to a High-Temperature Structure Reservoir. *J. Chem. Theory Comput.* **2007**, *3* (2), 557−568.
(39) Brémond, É.; Golubev, N.; Steinmann, S. N.; Corminboeuf, C. How Important Is Self-Consistency for the dDsC Density Dependent Dispersion Correction? *J. Chem. Phys.* **2014**, *140*, 18A516.
(40) Petraglia, R.; Steinmann, S. N.; Corminboeuf, C. A Fast Charge-Dependent Atom-Pairwise Dispersion Correction for DFTB3. *Int. J. Quantum Chem.* **2015**, *115* (18), 1265−1272.
(41) Mashraqui, S. H.; Sangvikar, Y. S.; Meetsma, A. Synthesis and Structures of Thieno[2,3-b]Thiophene Incorporated [3.3]-Dithiacyclophanes. Enhanced First Hyperpolarizability in an Unsymmetrically Polarized Cyclophane. *Tetrahedron Lett.* **2006**, *47* (31), 5599−5602.
(42) Dijkstra, G. D. H.; Kellogg, R. M.; Wynberg, H.; Svendsen, J. S.; Marko, I.; Sharpless, K. B. Conformational Study of Cinchona Alkaloids. A Combined NMR, Molecular Mechanics and x-Ray Approach. *J. Am. Chem. Soc.* **1989**, *111* (21), 8069.
(43) Zhou, J.; Wakchaure, V.; Kraft, P.; List, B. Primary-Amine-Catalyzed Enantioselective Intramolecular Aldolizations. *Angew. Chem., Int. Ed.* **2008**, *47* (40), 7656.
(44) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11* (5), 2087−2096.
(45) Fabregat, R. Modular Replica Exchange Simulatior. http://doi.org/10.5281/zenodo.3630553, **2020**.
(46) 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* (12), 5149−5155.
(47) Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a Sparse Matrix-Based Implementation of the DFTB Method. *J. Phys. Chem. A* **2007**, *111* (26), 5678−5684.
(48) Riplinger, C.; Neese, F. An Efficient and near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method. *J. Chem. Phys.* **2013**, *138* (3), 034106.
(49) Neese, F.; Valeev, E. F. Revisiting the Atomic Natural Orbital Approach for Basis Sets: Robust Systematic Basis Sets for Explicitly Correlated and Conventional Correlated ab initio Methods? *J. Chem. Theory Comput.* **2011**, *7* (1), 33−43.
(50) Neese, F. The ORCA Program System. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2* (1), 73−78.
(51) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90* (2), 1007−1023.
(52) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96* (9), 6796−6806.
(53) Adamo, C.; Barone, V. Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0Model. *J. Chem. Phys.* **1999**, *110* (13), 6158−6170.
(54) 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* (15), 154104.
(55) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5* (10), 2619−2628.
(56) Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Generating Efficient Quantum Chemistry Codes for Novel Architectures. *J. Chem. Theory Comput.* **2013**, *9* (1), 213−221.
(57) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comput. Phys. Commun.* **1995**, *91* (1), 1−41.
(58) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113* (18), 6378−6396.
(59) Huang, B.; von Lilienfeld, O. A. The "DNA" of Chemistry: Scalable Quantum Machine Learning with "Amons". **2017** *arXiv:1707.04146*. https://arxiv.org/abs/1707.04146 (accessed March 23, 2020)
(60) Christensen Faber, H. A Python Toolkit for Quantum Machine Learning. http://doi.org/10.5281/zenodo.817332 **2017**.
(61) Rupp, M.; Tkatchenko, A.; Müller, K. R.; Lilienfeld, V.; Anatole, O. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108* (5), 058301.
(62) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; Von Lilienfeld, O. A.; Müller, K. R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6* (12), 2326−2331.
(63) Nelder, J. A.; Mead, R. A Simplex Method for Function Minimization. *Comp. J.* **1965**, *7* (4), 308−313.
(64) Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002**, *116* (20), 9058−9067.
(65) Swendsen, R. H.; Wang, J.-S. Replica Monte Carlo Simulation of Spin-Glasses. *Phys. Rev. Lett.* **1986**, *57* (21), 2607.
(66) Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. A Novel Hamiltonian Replica Exchange MD Protocol to Enhance Protein Conformational Space Sampling. *J. Chem. Theory Comput.* **2006**, *2* (2), 217−228.
(67) 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* (30), 9431−9438.
(68) Ceriotti, M.; More, J.; Manolopoulos, D. E. I-PI: A Python Interface for Ab Initio Path Integral Molecular Dynamics Simulations. *Comput. Phys. Commun.* **2014**, *185* (3), 1019−1026.
(69) Kapil, V.; Rossi, M.; Marsalek, O.; Petraglia, R.; Litman, Y.; Spura, T.; Cheng, B.; Cuzzocrea, A.; Meißner, R. H.; Wilkins, D. M.; Juda, P.; Bienvenue, S. P.; Fang, W.; Kessler, J.; Poltavsky, I.; Vandenbrande, S.; Wieme, J.; Corminboeuf, C.; Kühne, T. D.; Manolopoulos, D. E.; Markland, T. E.; Richardson, J. O.; Tkatchenko, A.; Tribello, G. A.; Van Speybroeck, V. V.; Ceriotti, M. i-PI 2.0: A universal force engine for advanced molecular simulations. *Comput. Phys. Commun.* **2019**, *236*, 214−223.
(70) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Elsevier: London, UK, 2001.
(71) Kunsch, H. R. The Jackknife and the Bootstrap for General Stationary Observations. *Ann. Stat.* **1989**, *17* (3), 1217−1241.
(72) Rathore, N.; Chopra, M.; de Pablo, J. J. Optimal Allocation of Replicas in Parallel Tempering Simulations. *J. Chem. Phys.* **2005**, *122* (2), 024111.

(73) Jones, E.; Pearson, P.; Travis, O. SciPy: Open Source Scientific Tools for Python. http://www.scipy.org, 2001.
(74) Vannay, L.; Meyer, B.; Petraglia, R.; Sforazzini, G.; Ceriotti, M.; Corminboeuf, C. Analyzing Fluxional Molecules Using DORI. *J. Chem. Theory Comput.* **2018**, *14* (5), 2370−2379.
(75) Kapil, V.; Engel, E.; Rossi, M.; Ceriotti, M. Assessment of Approximate Methods for Anharmonic Free Energies. *J. Chem. Theory Comput.* **2019**, *15* (11), 5845−5857.
(76) Bürgi, T.; Baiker, A. Conformational Behavior of Cinchonidine in Different Solvents: A Combined NMR and Ab Initio Investigation. *J. Am. Chem. Soc.* **1998**, *120* (49), 12920−12926.
(77) Huang, M.; Dissanayake, T.; Kuechler, E.; Radak, B. K.; Lee, T. S.; Giese, T. J.; York, D. M. A Multidimensional B-Spline Correction for Accurate Modeling Sugar Puckering in QM/MM Simulations. *J. Chem. Theory Comput.* **2017**, *13* (9), 3975−3984.

https://dx.doi.org/10.1021/acs.jctc.0c00100
J. Chem. Theory Comput. 2020, 16, 3084−3094