Molecular Simulations with a Pretrained Neural Network and Universal Pairwise Force Fields

Adil Kabylda*,[1] J. Thorben Frank*,[2,3] Sergio Suarez Dou,[1] Almaz Khabibrakhmanov,[1] Leonardo Medrano Sandonas,[4] Oliver T. Unke,[5] Stefan Chmiela,[2,3] Klaus-Robert Müller,[2,3,5,6,7,*] and Alexandre Tkatchenko[1,†]

[1]*Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg City, Luxembourg*
[2]*Machine Learning Group, Technische Universität Berlin, 10587 Berlin, Germany*
[3]*Berlin Institute for the Foundations of Learning and Data – BIFOLD, Germany*
[4]*Institute for Materials Science and Max Bergmann Center of Biomaterials, TU Dresden, 01062 Dresden, Germany*
[5]*Google DeepMind, Berlin, Germany*
[6]*Max Planck Institute for Informatics, Stuhlsatzenhausweg, 66123 Saarbrücken, Germany*
[7]*Department of Artificial Intelligence, Korea University, Anam-dong, Seongbuk-gu, Seoul 02841, Korea*

Machine Learning Force Fields (MLFFs) promise to enable general molecular simulations that can simultaneously achieve efficiency, accuracy, transferability, and scalability for diverse molecules, materials, and hybrid interfaces. A key step toward this goal has been made with the GEMS approach to biomolecular dynamics [Sci. Adv. 10, eadn4397 (2024)]. This work introduces the SO3LR method that integrates the fast and stable SO3krates neural network for semi-local interactions with universal pairwise force fields designed for short-range repulsion, long-range electrostatics, and dispersion interactions. SO3LR is trained on a diverse set of 4 million neutral and charged molecular complexes computed at the PBE0+MBD level of quantum mechanics, ensuring a comprehensive coverage of covalent and non-covalent interactions. Our approach is characterized by computational and data efficiency, scalability to 200 thousand atoms on a single GPU, and reasonable to high accuracy across the chemical space of organic (bio)molecules. SO3LR is applied to study units of four major biomolecule types, polypeptide folding, and nanosecond dynamics of larger systems such as a protein, a glycoprotein, and a lipid bilayer, all in explicit solvent. Finally, we discuss the future challenges toward truly general molecular simulations by combining MLFFs with traditional atomistic models.

INTRODUCTION

The desire to perform quantitative molecular dynamics simulations based solely on nuclear charges and electron numbers has been expressed by many researchers, including Schrödinger [1], Dirac [2], and Feynman [3]. Despite a century filled with groundbreaking advances, this vision has yet to be fully realized in the realm of molecular simulations. Existing approaches often make significant trade-offs concerning **E**fficiency, **A**ccuracy, **S**calability, or **T**ransferability (EAST) [4]. In this manuscript, we argue that several methodological advances in the field of atomistic modeling have coalesced to bring us closer to achieving fully quantitative, quantum-accurate molecular simulations. While the journey toward this ultimate goal may be lengthy and complex, it is a pursuit that is undeniably worthwhile and requires a collaborative community-based effort.

A key challenge in molecular simulations is the construction of an atomistic force field (FF) model that satisfies the EAST requirements mentioned above [5–12]. Traditionally, FFs are obtained either from approximate but fast mechanistic expressions, or accurate but computationally prohibitive *ab initio* electronic-structure calculations. Both approaches compromise either accuracy or efficiency, restricting the scope of problems that can be addressed. Recently, machine-learned force fields (MLFFs) have started to bridge this gap by exploiting statistical models with high flexibility [10–14]. Unlike classical FFs, MLFFs exhibit unprecedented transferability across chemical space; however, scalability with system size remains an issue.

Many challenges remain to be addressed to enable EAST-compliant and MLFF-driven general molecular simulations. Among these we mention the development of data and computationally efficient semilocal interatomic interaction models [15–24], explicit treatment of (many-body) long-range interactions [12, 25], building datasets with comprehensive coverage of chemical space [26–34], and development of modern GPU-enabled molecular simulation frameworks [35–37]. Within this work, we take decisive steps towards solving the aforementioned challenges for organic (bio)molecules. Our solution combines recent advances from chemical and computational physics, machine learning (ML), and established techniques from the FF community. Semi-local interactions are described by the SO3krates ML model [38] using a many-body anharmonic treatment. The physical pairwise terms include short-range Ziegler-Biersack-Littmark repulsion [39], long-range electrostatic interactions, and a recently derived universal interatomic van der Waals dispersion potential [40]. Complementarity between the different terms is achieved through careful parametrization on a curated and comprehensive dataset of 4M molecular structures, leading to the SO3LR model (we suggest pronunciation "solar").

We demonstrate the applicability of SO3LR in nanosecond-long simulations of small biomolecular units,

polyalanine systems, bulk water, crambin protein, N-linked glycoprotein, and a lipid bilayer. SO3LR can be scaled to simulations involving up to $\sim$200k atoms with a latency of $\sim$3 μs/atom/step on a single H100 GPU, thus approaching sizes and timescales relevant for realistic biomolecules.

RESULTS

SO3LR Components

Generally applicable molecular simulations can be directly related to an accurate description of interactions across systems and length-scales. To achieve these objectives, SO3LR decomposes the potential energy into four contributions (Fig. 1A):

$$E_{\text{Pot}} = \underbrace{E_{\text{ZBL}}}_{\text{short-range}} + \underbrace{E_{\text{SO3k}}}_{\text{semi-local}} + \underbrace{E_{\text{Elec}} + E_{\text{Disp}}}_{\text{long-range}}, \quad (1)$$

where E_{ZBL} is a short-ranged term inspired by Ziegler-Biersack-Littmark (ZBL) repulsion between nuclei (see "Materials and Methods"), E_{SO3k} is the semi-local many-body potential learned by the SO3krates model, and E_{Elec} and E_{Disp} are the long-ranged electrostatic and dispersion energies, respectively. All potential terms influence each other, and a careful optimization procedure based on a diverse dataset of $\sim$4 million points ensures broad applicability. The proposed combination of model design, dataset curation and joint optimization, resolves the trade-offs in the EAST requirement and is described in the following paragraphs.

EA – SO3krates. The cornerstone of our approach, which enables high computational efficiency and accuracy (**EA**ST), is the SO3krates model [22, 38] – an MLFF based on an equivariant graph neural network. Given atomic positions R, atomic numbers Z, total charge Q, and total spin S, it predicts atomic quantities

$$E_i, q_i, h_i = \text{SO3krates}(R, Z, Q, S), \quad (2)$$

where E_i are atomic energies, q_i are partial charges and h_i are Hirshfeld ratios (ratio of effective and free-atom volume, $V_{\text{eff}}/V_{\text{free}}$) [41]. The semi-local energy contribution is then calculated as the sum over the atomic energies

$$E_{\text{SO3k}} = \sum_{i=1}^{N} E_i. \quad (3)$$

The predicted atomic energies contain information about atoms in the direct local neighborhood *and* beyond via mean field updates, which is why we refer to the energy prediction as semi-local. The mean-field nature of these updates cannot account for all types of interactions and is limited by an effective cutoff, which is *upper bounded*

by the local cutoff times the number of update steps (the effective cutoff in SO3LR is 13.5 Å).

S – Long-range dispersion and electrostatics. To improve the description of long-range effects and extend the description beyond semi-local environments, we explicitly incorporate electrostatics and universal pairwise interatomic vdW potentials. They are made trainable by using the partial charges and Hirshfeld ratios predicted by SO3krates as an input. At the same time, both contributions are based on analytically derived, physical equations such that they follow the correct asymptotic (at this moment pairwise) decay. This is an inevitable requirement for the scalability (EA**S**T) to length-scales exceeding the ones covered by the training data.

Dispersion interactions are calculated using universal pairwise interatomic vdW potentials derived from quantum Drude oscillators (QDO) [40]:

$$E_{\text{Disp}} = -\sum_{i<j}\sum_{n=3}^{5} \frac{C_{2n}^{ij}}{r_{ij}^{2n} + R_{d,ij}^{2n}}, \quad (4)$$

where C_{2n}^{ij} are long-range interatomic dispersion coefficients, and $R_{d,ij}^{2n}$ are the vdW radii of the Becke-Johnson (BJ) damping function [42]. The radii are defined based on atomic polarizabilities as [43]

$$R_{d,ij} = \gamma R_{\text{vdW}}^{ij} = 2\gamma \left(\frac{a_0^4 \alpha_{\text{fsc}}^{-4/3}}{4\pi\varepsilon_0} \frac{\alpha_i + \alpha_j}{2} \right)^{1/7}, \quad (5)$$

where a_0 is the Bohr radius, $\alpha_{\text{fsc}} = e^2/4\pi\varepsilon_0\hbar c$ is the fine-structure constant, with γ being a single tunable parameter in the dispersion module that controls the damping strength. Atomic polarizabilities α_i and dipole-dipole dispersion coefficients C_6^{ij} are obtained using the Tkatchenko-Scheffler method [44] with the ML-predicted Hirshfeld ratios h_i, whereas the scaling relations from the QDO model [40, 45] are applied to generate higher-order dispersion coefficients C_8^{ij} and C_{10}^{ij}.

Electrostatic interactions are modeled using a damped Coulomb potential

$$E_{\text{Elec}} = \sum_{i<j} q_i q_j \frac{\text{erf}\left(r_{ij}/\sigma\right)}{r_{ij}}, \quad (6)$$

where q_i are the ML-predicted partial charges, and σ is a hyperparameter that controls the damping strength. We remark that the semi-local SO3k module has the capacity to accurately describe multipolar interactions, hence we limit our model to leading-order electrostatics.

Coupling between long-range and semi-local energy contributions arises from the structure of the potential energy prediction (Eq. 1). The long-range modules have a non-zero energy contribution for all atomic pairs, including those within the local cutoff of the MLFF. As such, the functional forms of the long-range potentials
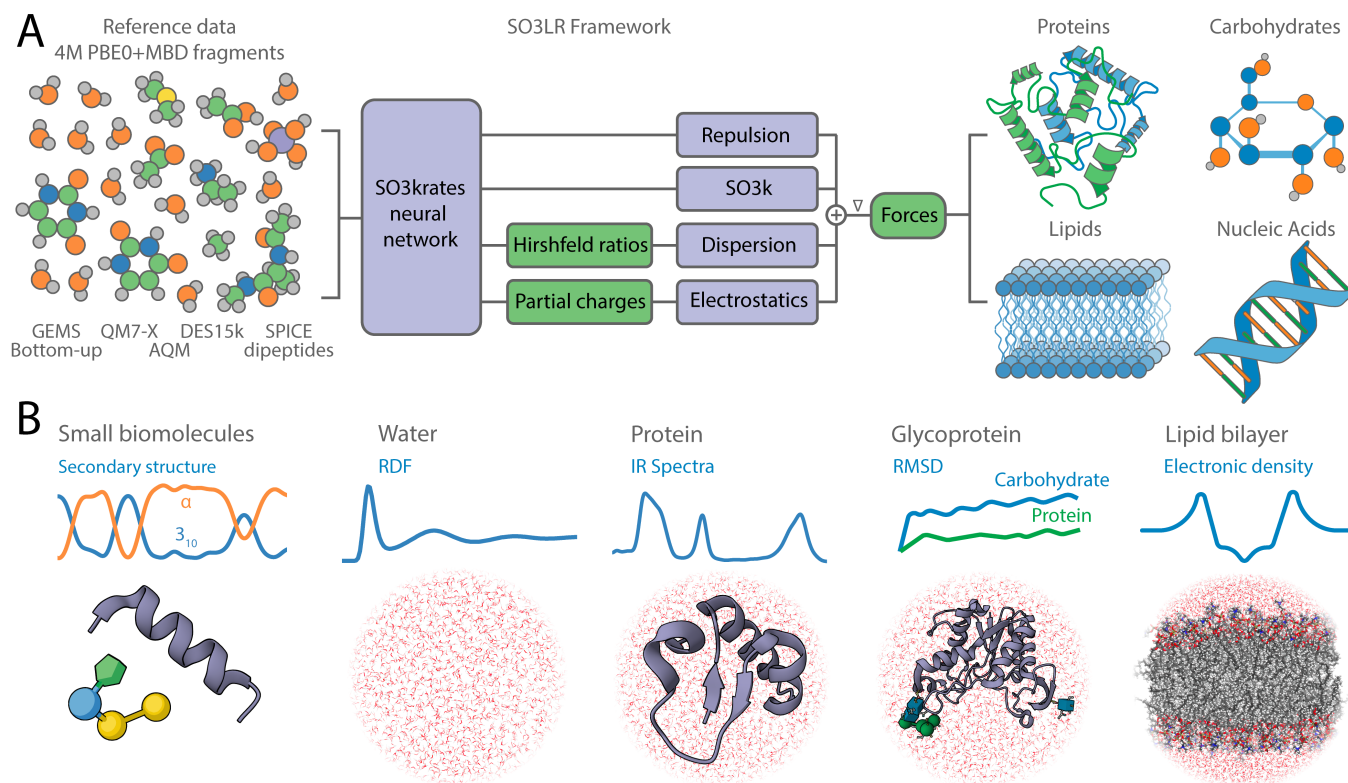
Fig. 1. Overview of the SO3LR model and simulation results. A) SO3LR combines the SO3krates neural network with physically inspired interactions, including ZBL repulsion, electrostatics, and universal pairwise van der Waals potential for dispersion which interact directly with the neural network model. All building blocks are jointly trained on a carefully curated data set which covers a broad range of chemical space and interaction classes. SO3LR enables simulations of small biomolecular units of all four major types of biomolecules, and large-scale simulations of three types. **B)** This includes large-scale simulations of liquid water, protein, glycoprotein, and a lipid bilayer.

alter the potential which is learned by the SO3krates model. The choice of damping hyperparameters γ and σ controls the fine balance between semi-local, electrostatic and dispersion interactions. In principle SO3krates can learn to correct for arbitrary choices of σ and γ up to the local cutoff, but yet their choice affects the overall model performance. This can be attributed to the fact that semi-local and long-range interactions are coupled non-linearly through parameter and hyperparameter optimization in both modules. Hence, the damping hyperparameters were fine-tuned on the S66x8 benchmark dataset [46].

T – Optimization on diverse training data. All SO3LR modules are jointly optimized on a diverse dataset that spans a broad chemical space and various interaction classes. This enables transferability (EAS**T**) between all four major types of biomolecules.

The comprehensive dataset has been a key factor in the development of our MLFF. It is a collection of extensive quantum mechanical data from both small and large molecules, as well as non-covalent systems with and without explicit solvation. To this end, we combined five datasets: 2.7 M bottom-up GEMS fragments [47], 1 M QM7-X molecules [31], 60 k AQM gas-phase molecular drugs [34], 33 k SPICE dipeptides [33], and 15 k DES molecular dimers [30] (see Fig. S1 and Tab. S1 for more details). The first three datasets were computed using the non-empirical hybrid PBE0 density functional with a many-body treatment of van der Waals interactions (PBE0+MBD). To maintain consistent references, we recomputed the last two datasets at the same level of theory. All reference calculations were performed using tightly converged numeric atom-centered orbitals as implemented in the `FHI-aims` code [48, 49].

The datasets are complementary in terms of conformational space and chemical diversity, covering 8 elements predominantly present in biosystems (H, C, N, O, F, P, S, and Cl). Specifically, the QM7-X dataset encompasses the chemical space of small organic molecules, while the AQM dataset includes medium-sized drug-like molecules. DES molecular dimers were incorporated to improve the description of non-covalent interactions.

SPICE dipeptide structures were added to enhance the accuracy for the protein-containing systems. Lastly, the GEMS bottom-up dataset contains gas-phase and explicitly micro-solvated protein fragments, as well as structures with gas-phase water clusters.

Optimization of the model parameters is done by minimizing a combined loss

$$\mathcal{L} = \frac{\lambda_F}{B} \sum_{b=1}^{B} \frac{1}{N_b} \sum_{i=1}^{N_b} \left|\left| \vec{F}_{i,\text{true}} - \vec{F}_{i,\text{pred}} \right|\right|_2^2$$
$$+ \frac{\lambda_\mu}{B} \sum_{b=1}^{B} \left|\left| \vec{\mu}_{b,\text{true}} - \vec{\mu}_{b,\text{pred}} \right|\right|_2^2 \qquad (7)$$
$$+ \frac{\lambda_h}{B} \sum_{b=1}^{B} \frac{1}{N_b} \sum_{i=1}^{N_b} \left(h_{i,\text{true}} - h_{i,\text{pred}} \right)^2,$$

where $\vec{F}_i$ are atomic forces, $\vec{\mu}$ are molecular dipoles, and h_i are Hirshfeld ratios, with λ as trade-off parameters between the individual loss terms. The predicted forces are obtained as the gradient *w.r.t.* the atomic positions of the potential energy (Eq. 1) and Hirshfeld ratios and partial charges are the ones predicted by SO3krates (Eq. 2). The partial charges are indirectly trained based on dipole moments, instead of direct fitting to reference partial charges. This approach reduces the model's sensitivity to the choice of charge-equilibration scheme and enhances transferability [50]. It should be noted that the model is trained on forces, rather than on energies and forces, which ensures accuracy of relative energy predictions only. Further training details can be found in the "Materials and Methods" section.

SO3LR Evaluation

A force field that is truly EAST-compliant should be able to accurately simulate systems of varying nature and size. To demonstrate SO3LR's capabilities and limitations, we first evaluate its performance on test and benchmark sets to assess its accuracy in predicting forces, binding energies, dipole moments, and Hirshfeld ratios. This is followed by an analysis of the dynamics of small biomolecular units from the MD22 benchmark dataset [51]. We then investigate the folding and stability of polyalanine systems *in vacuo*, which depend on a delicate interplay of various interactions. Before transitioning to simulations of larger biosystems, we conduct a detailed analysis of water dynamics. Finally, we extend the evaluation to large-scale molecular dynamics simulations of more complex systems, including a protein, a glycoprotein, and a lipid bilayer, all in explicit water (see Fig. 1B).

Test set and benchmark errors. We begin the evaluation of the model by analysing its accuracy *w.r.t.* quantum mechanical reference data (Tab. I). The test set com-

TABLE I. Root Mean Square Error of the model on various test sets. Force (eV/Å), dipole moment vector (e × Å), and Hirshfeld ratios. Dash indicates no data.

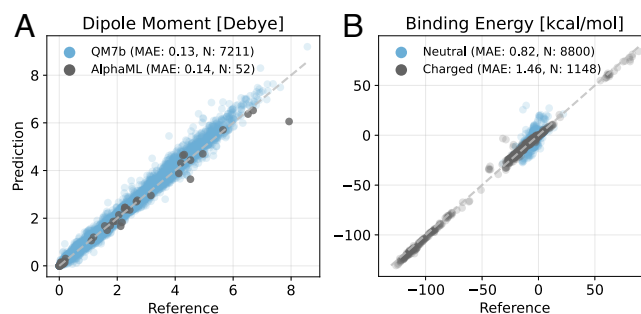| Dataset	Size	# atoms	Force	Dipole	Hirsh. rat.
QM7-X	10000	6-23	0.069	0.031	0.012
GEMS bottom-up	10000	2-120	0.094	0.048	-
AcAla$_3$NMe	100	42	0.052	0.051	0.012
DHA	100	56	0.053	0.072	0.012
AT-AT	100	60	0.170	0.238	0.025
Stachyose	100	87	0.105	0.119	0.016
Buckyball Catcher	100	148	0.387	4.032	0.030
AcAla$_{15}$NMe	312	162	0.055	-	-
Crambin top-down	5624	230-321	0.057	-	-

Fig. 2. Evaluation of the SO3LR long-range modules' performance. A) Evaluation of the model on dipole moment prediction for 7k QM7b molecules and AlphaML showcase database [52]. **B)** Performance of the model evaluated on the unseen SAPT10k dataset [53], separated into neutral and charged subsets.

prises 10k randomly sampled structures from each of the QM7-X and GEMS bottom-up fragments (all other training sets were fully utilized during training). Additionally, we recalculated 100 random structures from five MD22 benchmark molecules at the PBE0+MBD/tight level of theory. Furthermore, we assessed the model using ∼300 AcAla$_{15}$NMe structures and ∼5600 crambin top-down fragments that were employed in the training of system-specific models in Ref. 47.

The model demonstrates good performance in predicting forces, dipole moments, and Hirshfeld ratios. A closer examination of the table reveals two key observations. First, fragments from curved carbon-based systems, such as the buckyball catcher, are absent from the training set, which is reflected in the increased errors. This suggests that further expansion of the dataset would be necessary to achieve full transferability across the chemical space (also see Fig. S2). Nevertheless, the molecular dynamics simulations of all five MD22 molecules remain stable, as
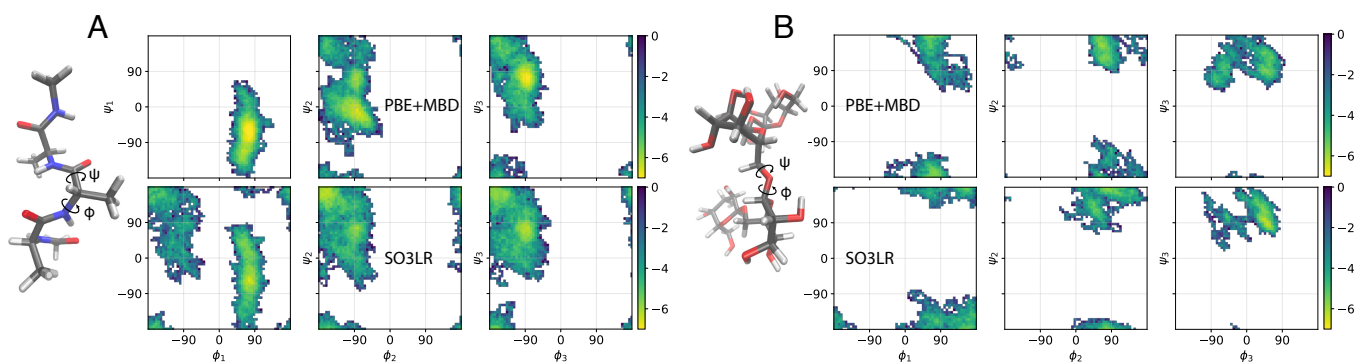
Fig. 3. Simulations of small biomolecular fragments. Ramachandran plots (ϕ/ψ dihedrals) for **(A)** AcAla$_3$NMe and **(B)** stachyose from the MD22 dataset [51]. PBE+MBD and SO3LR simulations at 500 K with 85 ps for AcAla$_3$NMe and 27 ps dynamics for stachyose. SO3LR simulations of 500 ps are shown in Fig. S3. Trajectory is sampled every 1 fs. The Boltzmann-inverted scale is shown in kcal/mol. The comparison between PBE+MBD from MD22 and SO3LR (trained on PBE0+MBD forces) is only shown as a guide to the eye.

discussed in the following subsection. It is important to note that the typically reported MD22 errors [38, 54] correspond to system-specific models. Second, by comparing the errors of AcAla$_3$NMe, AcAla$_{15}$NMe, and crambin top-down fragments, which are identical, we conclude that the SO3LR model is scalable to large solvated protein fragments and that long-range modules effectively describe intermolecular interactions, despite being trained only on small fragments.

To assess the quality of electrostatic interactions, we benchmarked partial charge prediction using the QM7b and AlphaML benchmark datasets, which were computed at the LR-CCSD/d-aug-cc-pVDZ level of theory [52]. SO3LR accurately predicts dipole moments with MAEs of 0.13 D in magnitude and 5.1° in angles (Fig. 2A). This performance is comparable to the best semi-empirical methods [55]. Our training set contains molecules from the QM7x dataset, which includes perturbed structures from QM7b. The AlphaML benchmark, on the other hand, contains a wider set of compounds, including DNA/RNA nucleobases, amino acids, carbohydrates, drugs, and hydrocarbons. We achieve a MAE of 0.14 D on this dataset (Fig. 2A), showcasing transferable and accurate dipole moment prediction crucial for calculating reliable electrostatic interactions.

Next, we evaluate noncovalent interaction energies on a comprehensive SAPT10k benchmark computed at the SAPT2+(3)(CCD)/aug-cc-pVTZ level of theory [53]. It consists of 70 subsets, featuring challenging binding motifs dominated by electrostatics and/or dispersion interactions and offering substantial diversity across chemical space. We exclude 34 out of 9982 complexes because they contain atom types beyond the 8 elements our model was trained on (hence, predictions on those structures are not meaningful). Overall, the model performs well, achieving sub-chemical accuracy with a MAE of 0.89 kcal/mol

(Fig. 2B). Rare outliers include ClO$_4$$^-$-$\pi$, NO$_3$$^-$-$\pi$ and SO$_2$-π complexes. This is an impressive performance overall, particularly given that part of the error comes from the PBE0+MBD reference data.

Simulations of small biomolecular units. Molecular dynamics simulations are the ultimate test for evaluating force fields. We simulated five molecular systems from the MD22 benchmark, encompassing four major biomolecule types and one supramolecular complex: the AcAla$_3$NMe tetrapeptide, stachyose tetrasaccharide, AT-AT DNA base pairs, docosahexaenoic fatty acid (DHA), and the Buckyball Catcher. The first two systems underwent 500 ps of simulation at 500 K to compare with the PBE+MBD references computed at 500 K, other systems were simulated at 300 K. The model demonstrated robust conformational exploration across all molecules. In particular, the free energy surface exploration of tetraalanine and stachyose closely aligns with MD22 *ab initio* results, computed at the PBE+MBD level of theory, as shown in Ramachandran plots (Fig. 3AB). Note that in this figure we report only short molecular dynamics simulations, and the comparison between PBE+MBD and SO3LR dynamics is only provided as a guide to the eye. Full 500 ps trajectories are shown in Fig. S3 [56]. The tetrapeptide explores all 'allowed' (ϕ/ψ) regions found in experimental protein structures [56]. The buckyball catcher complex remained stable and retained the ball (Fig. S4), despite larger errors on the test set. These results suggest that our model can reliably explore conformational landscapes of small molecules even in the absence of the system-specific training data.

Simulations of polyalanine systems. We further investigate polyalanines, focusing on the folding of extended AcAla$_{15}$NMe and the stability of the folded AcAla$_{15}$LysH$^+$ at elevated temperatures. These systems present significant challenges due to the delicate
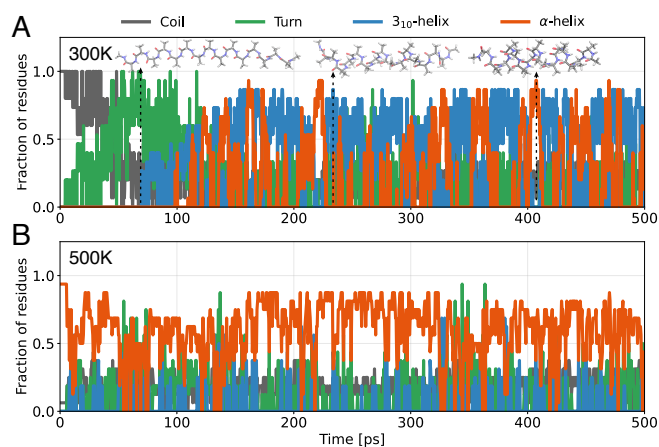
Fig. 4. Simulations of polyalanines. A) Secondary structural motifs observed along a typical folding trajectory of AcAla$_{15}$NMe at 300 K in the gas phase. **B)** Secondary structural motifs observed along a trajectory of AcAla$_{15}$LysH$^+$ at 500 K in the gas phase, starting from the folded α-helix conformation.

interplay of hydrogen bonding, polarization, and dispersion interactions. Previous attempts to simulate them without incorporating top-down fragments either failed to correctly fold AcAla$_{15}$NMe or overstabilized the α-helix, and in some cases, predicted diminished stability for AcAla$_{15}$LysH$^+$ [47, 54].

For each system, we performed four runs of 500 ps. The extended AcAla$_{15}$NMe structure folded in all cases (Fig. 4A and Fig. S5A). The timescales and folding mechanisms were similar to those observed in Ref. 47: initially, the peptide primarily consists of turns, then passes through a "wavy" intermediate, and finally folds into a helical form with dynamic transitions between α- and 3_{10}-helices. The latter is particularly noteworthy, as empirical force fields tend to overestimate the stability of α-helices [61, 62].

For the folded AcAla$_{15}$LysH$^+$, we observe that the α-helical motifs are preserved up to 500–600 K (Fig. 4B and Fig. S5B). These findings agree with experimental measurements, which observe scattering cross sections for AcAla$_{15}$LysH$^+$ consistent with an α-helical structure up to $\approx$725 K when subject to interactions with the helium buffer gas [63]. Direct comparison with gas-phase experiments would have to explicitly include the helium environment and quantum nuclear effects. Overall, the two polyalanine systems provide a good evaluation of scalability to medium-sized systems in dynamics, complementing the observed scalability in terms of test errors.

Simulations of liquid water. Liquid water plays a crucial role in biosystems, making it an essential subject for SO3LR's evaluation. We performed a simulation of a water box containing 4096 water molecules in the NPT ensemble. Observables were averaged over 300 ps follow-

ing an initial 200 ps equilibration phase. Our analysis focused on three aspects: radial distribution function, density convergence, and self-diffusion coefficient.

The radial distribution function shows the expected shell structure (Fig. 5A and Fig. S6), which indicates, however, that the liquid phase is slightly overstructured. Increasing the temperature to 330 K allows for an approximate treatment of missing nuclear quantum effects and improves agreement with the experimental data [57, 64].

We investigated the convergence of the density as a function of the cutoff for long-range interactions (Fig. 5B, see methods for details). The water density varies between 1.04 and 0.97 g/cm^3 for long-range cutoffs of 10–20 Å. We adopted a cutoff of 12 Å for all subsequent biosimulations in explicit water, balancing accuracy and computational efficiency.

To provide further insights into water dynamics, we calculated the self-diffusion coefficient of water (Fig. 5A, labels). The diffusion coefficient is 0.079 Å^2/ps at 300 K and 0.224 Å^2/ps at 330 K with the 12 Å long-range cutoff. For comparison, the experimental diffusion coefficient at room temperature is 0.23 Å^2/ps [58].

While the model does not achieve exact agreement with experimental values, the computed density and self-diffusion coefficients underscore its transferability and scalability. This is notable considering that the training set contains only gas-phase clusters of water with at most 40 molecules ($\sim$10k clusters or $\sim$0.26% of the combined dataset). It is also well known that even explicit *ab initio* MD simulations with the PBE0+vdW functional struggle to fully capture all experimental properties of water [64]. For biomolecules, accurate modeling of water close to their surface may hold greater practical relevance than exact replication of bulk water behavior.

Simulations of large biomolecules. Finally, we showcase the potential of the SO3LR by simulating large biomolecules in explicit water. The selected systems encompass various biomolecular classes, each characterized by distinct structural and functional properties that can be validated against existing simulations or experimental data. The systems include the crambin protein, glycoprotein (PDB: 1K7C), and the POPC lipid bilayer.

For crambin (25k atoms including water), we compute the power spectrum from 125 ps of dynamics at a temporal resolution of 2.5 fs, after 1 ns equilibration period. The experimental water vibrations at 1640 cm^{-1} and 3200–3600 cm^{-1} are reproduced in SO3LR with better agreement than GEMS, AMOEBA and AmberFF (Fig. 5C). We further examined the root mean square deviation RMSD(t, t+Δt) averaged over three 3 ns simulations, which indicates that SO3LR shows slightly increased protein mobility on longer timescales, consistent with the GEMS model (Fig. 5D). Nevertheless, we find that the overall structure stays close to the starting folded state, without any indication of unfolding (Fig. S7). To
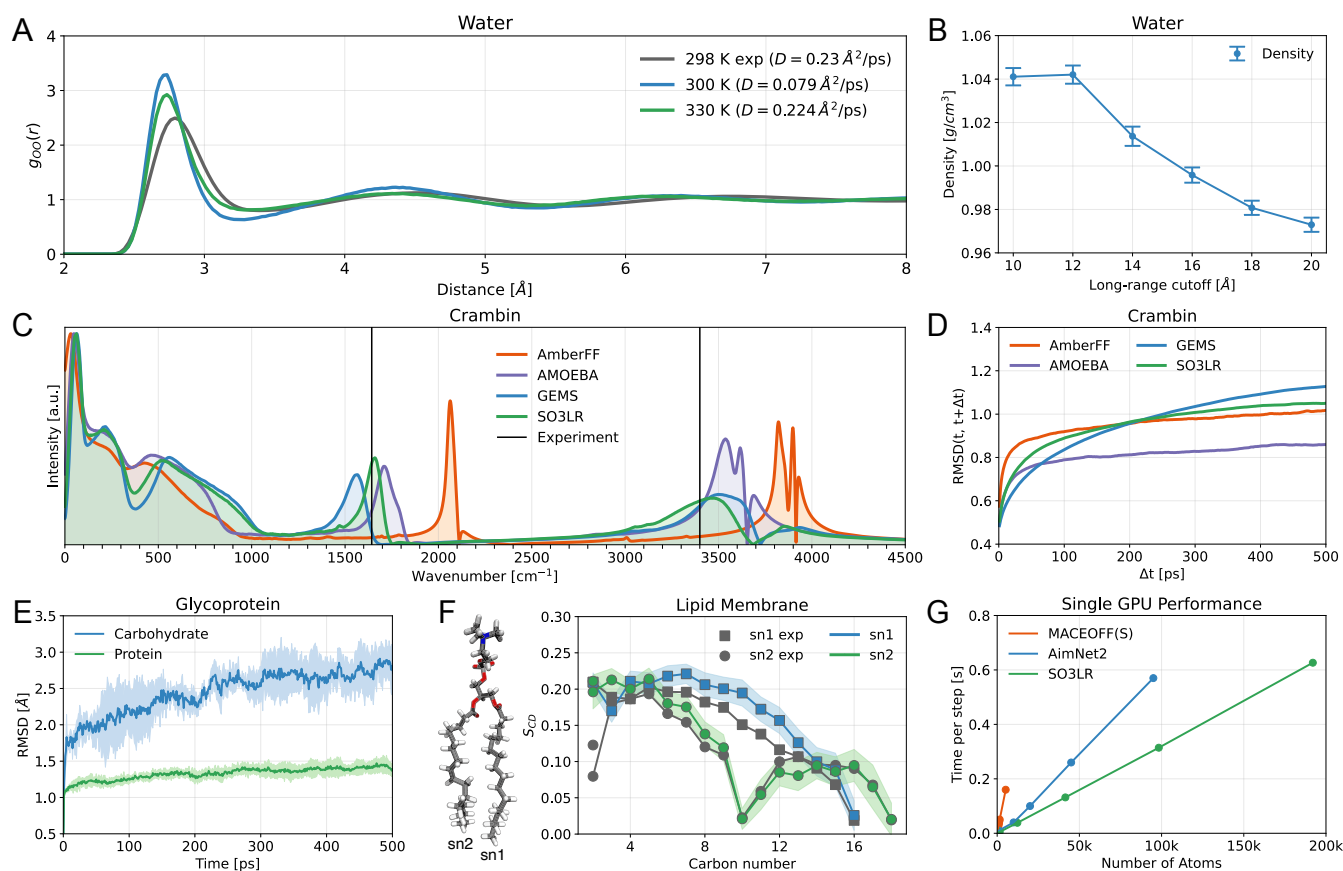
Fig. 5. Simulations of bulk water and explicitly solvated biomolecules. A) Oxygen-oxygen radial distribution function for bulk water (4096 water molecules). The diffusion coefficients of water at 300 and 330 K, obtained using the model with a 12 Å long-range cutoff, are specified in the legend. The experimental values were taken from Refs. 57, 58. **B)** Dependence of density of water on long-range cutoff at 300 K. **C)** Power spectrum of crambin in water obtained from 125 ps of dynamics. AmberFF and GEMS results are taken from Ref. 47. **D)** Root mean square deviation (RMSD) of Crambin, excluding hydrogen atoms, between conformations sampled at times t and $t + \Delta t$ averaged over three 3 ns runs. **E)** RMSD of protein and carbohydrate segment of glycoprotein averaged over three 500 ps runs. **F)** Tail group NMR order parameters from SO3LR simulation of 128 POPC Lipid Bilayer and from experiment [59]. The standard deviation is shown with background color. **G)** Single GPU performance. SO3LR latencies were measured based on liquid water molecular dynamics using JAX-MD [36] in the NVT ensemble on the H100 80 GB GPU. The slope is 3.25×10^{-6} s/atom/step. MACE-OFF(S) and AimNet2 latencies, measured on A100 and H100, respectively, were taken from Refs. 54, 60.

visualize the conformational space sampling, we applied two-dimensional Uniform Manifold Approximation and Projection (UMAP) [65]. The projection of the paths reveals that SO3LR and GEMS sample conformational space more extensively than AmberFF and AMOEBA (Fig. S8), which aligns well with high conformational variability derived from NMR measurements [66].

For glycoprotein (48k atoms including water) we conducted a 500 ps simulation at 300 K. This system, comprising both protein and carbohydrate segments, presented a challenge for SO3LR due to the absence of carbohydrates in its training data. Despite this, the model successfully inferred increased carbohydrate flexibility, as evidenced by the greater RMSD observed for

the carbohydrate segment compared to the protein segment (Fig. 5E). These findings align with results from the specifically parameterized CHARMM force field [67]. However, the simulation revealed limitations in sampling conformations of the N-linkage. Specifically, the Cγ-Cβ-Cα-N dihedral, located at the protein-carbohydrate junction, can adopt three conformations: $g+$ (60°), $anti$ (180°), and $g-$ (300°). Our simulation only sampled the anti-conformation out of these three possible states. Longer simulations would be required to determine whether the model can explore other conformations without carbohydrate-protein linkages in the training set.

Lastly, we modeled a homogeneous POPC lipid bilayer (33k atom system consisting of 128 lipids and 5120 water

TABLE II. POPC Lipid Bilayer Structural Properties: area per lipid (Å^2) and bilayer thicknesses D_{HH} (Å).

	S/lipid	D_{HH}
Experiment [68]	64.3±1.3	36.5
Lipid21 [69]	63.92±0.09	38.50±0.20
CHARMM36/LJ-PME [70]	65.4±0.5	37.3±0.30
SO3LR	58.0±0.1	37.0

molecules). We performed a 500 ps simulation at 303 K and examined the structural properties: area per lipid, bilayer thickness, and lipid tail order parameters. These properties are critical measures of the accuracy of lipid simulations and are highly sensitive to factors such as hydrophilic attraction between head groups, hydrophobic repulsion between lipid tails, and interactions with surrounding water molecules. We found that SO3LR is in good agreement with experimental data and with empirical force fields specifically fine-tuned to lipid simulations (Tab. II). The 10% underestimation of the area-per-lipid likely stems from the isotropic NPT ensemble currently used in SO3LR simulations, compared to the semi-isotropic NPT used for empirical force fields. NMR lipid tail order parameters are another important quantitative measure that describe the degree of order within the acyl chains of lipids in a bilayer. The order parameters averaged over the last 250 ps suggest that the bilayer structure is in suitable agreement with the NMR experiments (Fig. 5F) [59].

DISCUSSION

A long-held vision in the atomistic simulation community is the development of force fields (FFs) with a unified functional form that can be applied across diverse chemical spaces – such as solvents, proteins, DNA, RNA, sugars, and lipids. These FFs should closely approximate quantum-mechanical behavior while remaining efficient and scalable enough to model realistic biomolecular complexes under various conditions (e.g., pressure, temperature, and external environments). In this work, we presented significant advancements towards fulfilling these criteria through the SO3LR model, which is embedded within an openly accessible and fully transparent framework. This framework integrates reliable and diverse quantum-mechanical datasets [30, 31, 33, 34, 47], a fast and stable SO3krates machine-learning architecture [38], universal long-range interaction modules [40], a JAX-MD simulation engine [36], and robust analysis tools. Together, these components facilitate quantum-accurate molecular simulations across an extended biomolecular chemical space.

Our developments aim towards enabling general molecular simulations and similar goals have been pursued by the seminal efforts in the empirical force field community over many decades [71–79]. SO3LR yields a 6-fold improved accuracy compared to the AMBER FF for polyalanine [47] when benchmarked on the PBE0+MBD atomic forces. At the same time, SO3LR is about 40 times slower on a single GPU than the GROMOS FF [47]. Our extended assessment on energies, forces, dipoles, polarizabilities, as well as our analysis of nanosecond-long MD trajectories demonstrates that SO3LR is highly transferable throughout biochemical space and scalable to hundreds of thousands of atoms. Such transferability and scalability are achieved without the need to specify atom types, impose harmonic constraints, or introduce bespoke functional forms for interatomic interactions in different biomolecular entities. The bottom-up training on quantum mechanical data ensures that our simulations are transferable to a wider range of conditions than previously possible. This is confirmed by polyalanine simulations from 300 to 800 K, accurate structural and spectroscopic observables for high and low vibrational frequencies obtained for solvated crambin, as well as the local and global structural properties for the 1K7C glycoprotein and the POPC lipid bilayer. The toolset developed in this work complements the existing and quickly growing machinery of successful biomolecular modeling tools. Our presented advancements would not have been possible without building on a wealth of existing landmark methods, many of which were developed by the empirical force field community.

One noteworthy component of our proposed SO3LR model is a successful combination of explicit physical knowledge, such as short- and long-range force modules, coupled with a semi-local many-body potential. Importantly, all of these contributions are carefully balanced by SO3LR via learning from data. Thus, the known physical interactions do not need to be learned from data, but SO3LR can – under the correct hard-coded inductive biases (repulsion, electrostatics, and dispersion energies) – focus its nonlinear expressive power mainly on learning both: the complex many-body contributions and the appropriate balance of the diverse energy terms from Eq. 1.

Despite recent progress in establishing "foundation models" for atomistic systems [54, 60, 80–82], many challenges remain in achieving truly general molecular simulations. Key areas for enhancing the SO3LR model include: (i) expanding the DFT+MBD training sets to encompass a broader spectrum of (bio)chemical entities, such as ions, sugars, lipids, DNA, supramolecules, and a variety of solvents, (ii) generating higher-level coupled cluster [83] or quantum Monte Carlo [84] reference data for small fragments, (iii) refining long-range interaction modules to effectively account for anisotropic many-body interactions [85], (iv) optimizing SO3LR for multi-GPU

architectures [86], (v) extending simulations to treat nuclear quantum effects [87, 88] beyond classical Newtonian molecular dynamics. This is a non-exhaustive list of research directions, all of which are subject of ongoing efforts in the community.

As atomistic simulations are highly sensitive to the intricacies of the underlying force fields and simulation parameters, it is imperative to establish a standardized set of benchmarks for quantum-accurate machine-learning force fields. Such benchmarks will ensure reproducibility of results and enable robust modeling of experimentally relevant phenomena across realistic time and length scales.

MATERIALS AND METHODS

Ziegler-Biersack-Littmark repulsion. The short-range repulsion between nuclei is modeled via a term inspired by the ZBL repulsion [21, 39]:

$$
\begin{aligned}
E_{\text{ZBL}} = k_e \sum_i \sum_{j \in \mathcal{N}_i} \frac{Z_i Z_j}{r_{ij}} f_{\text{cut}}(r_{ij}) \\
\cdot \sum_{m=1}^4 c_m e^{-a_m r_{ij}(Z_i^p + Z_j^p)/d},
\end{aligned}
\tag{8}
$$

where k_e is the Coulomb constant, Z_i are the atomic numbers, and a_m, c_m, p, and d are free parameters. The term $\mathcal{N}_i$ denotes the neighborhood of the i-th atom, and f_{cut} is a cutoff function that smoothly transitions between one and zero when atoms leave (or enter) the neighborhood. The ZBL term ensures a correct description of nuclear repulsion, which improves the stability of the potential for short bond-distances.

SO3krates. Given initial atomic representations $\boldsymbol{f}_i^{[0]} \in \mathbb{R}^F$ they are iteratively refined via UpdateLayers as

$$
\boldsymbol{f}_i^{[t+1]} = \text{UpdateLayer}\Big[\boldsymbol{f}_i^{[t]}, \boldsymbol{f}_j^{[t]}, \vec{r}_{ij} \Big]_{\forall j \in \mathcal{N}_i},
\tag{9}
$$

where $\mathcal{N}_i$ denotes the neighborhood of atom i containing all atoms within a cutoff sphere of radius r_{cut}. The final representations $\boldsymbol{f}_i^{[T]} \in \mathbb{R}^F$ are used to predict per-atom energies $E_i \in \mathbb{R}$ which are summed to yield per-structure energies $E_{\text{SO3k}} = \sum_{i=1}^N E_i$.

The initial atomic representations encode information about the atomic types Z, the total charge Q and the multiplicity S of the system, such that

$$
\boldsymbol{f}_i^{[0]} = e_{i,Z} + e_{i,Q} + e_{i,S}.
\tag{10}
$$

Each summand is an F-dimensional embedding for each atom i in the system. Atomic numbers are encoded in a per-atom fashion (see Ref. [38]) whereas total charge and multiplicity are encoded globally following the strategy described in Ref. 21.

Partial Charges and Dipoles. Following Ref. 21, partial charges are obtained as

$$
q_i = q_{Z_i} + \tilde{q}_i - \left(Q - \frac{1}{N} \sum_{j=1}^N (q_{Z_j} + \tilde{q}_j) \right),
\tag{11}
$$

where $\tilde{q}_i \in \mathbb{R}$ are predicted from the final atomic representations $\boldsymbol{f}_i^{[T]} \in \mathbb{R}^F$ via a two-layered multi layer perceptron network with silu non-linearity and $q_{Z_i} \in \mathbb{R}$ is an element dependent bias. The charge correction with the total charge Q ensures charge conservation. The partial charges can be used to predict molecular dipole moments (used in the loss function, see Eq. 7).

$$
\vec{\mu} = \sum_{i=1}^N q_i \vec{r}_i,
\tag{12}
$$

where $\vec{r}_i \in \mathbb{R}^3$ are the atomic positions.

Training details. SO3krates models (v1.0) were trained on a combined loss of forces, dipole moments, and Hirshfeld ratios with a weighting factor of 10:1:1, respectively. We used the AMSGrad optimizer [89] with an initial learning rate of 10^{-3} and an exponential learning rate decay every 500k steps by a factor of 0.85. The global norm of the gradient updates is clipped at 10.

The model uses a 4.5 Å cutoff, feature dimension of $F = 128$, and a maximal degree of $L_{\max} = 4$ for the Euclidean variables and $T = 3$ message passing layers, electrostatics damping coefficient of $\sigma = 4$, and dispersion damping coefficient of $\gamma = 1.2$. After each attention update, a two-layered multi-layer perceptron (MLP) with silu non-linearity refines the invariant features. This increases the number of trainable parameters and thus model expressiveness, which is important in the large data regime. To stabilize training and improve gradient flow, layer normalization [90] is applied to the invariant features after the attention and the interaction block. The model was trained for 86 h (corresponding to 5.125M gradient steps) with a batch size of $B = 200$.

Long-Range Cutoff. For large structures with tens to hundreds of thousands of atoms, considering all pairs of atoms becomes computationally infeasible and necessitates the introduction of a long-range cutoff. Additionally, if simulations are performed in a box (e.g. with water) the largest meaningful long-range cutoff is directly connected to the box size. As such, the system under investigation and the simulation parameters, determine different values for the long-range cutoff. To account for this, we carefully designed a switching function for the long-range potentials, which allows to choose between different cutoff values up to no long-range cutoff at the time of simulation. The choice does not affect the first two terms in Eq. 1 or intermediate properties, partial charges and Hirshfeld ratios, which are used as inputs to calculate the last two terms.

Both the dispersion and the electrostatic potential have infinite range and take on a non-zero value at the

long-range cutoff (Fig. S9). This results in a discontinuity in the forces at the cutoff value, leading to energy drift during MD simulations [91]. To ensure smoothness of the PES at the long-range cutoff we modify the pairwise electrostatic potential as

$$\tilde{u}(r) = \frac{q_i q_j k_e}{2} \cdot f_{\text{sw}}(r) \cdot u_{\text{ES}}(r) + \big(1 - f_{\text{sw}}(r)\big) \cdot u_{\text{FS}}(r), \tag{13}$$

where $u_{\text{ES}}(r)$ is the energy-shifted potential, $u_{\text{FS}}(r)$ is the force-shifted potential (see below) and $f_{\text{sw}}(r)$ is a switching function that smoothly interpolates between 1 and 0 on a given interval from r_{on} to r_{off}. By switching between the energy- and the force-shifted term the potential at short-range is the same as the one the model was trained with (no long-range cutoff for training), and enables to maintain the learned balance between different terms (Fig. S9). The shifted potentials are given as [91, 92]

$$u_{\text{ES}}(r) = \begin{cases} u(r) - u(R_c), & r < R_c \\ 0, & r > R_c \end{cases} \tag{14}$$

and

$$u_{\text{FS}}(r) = \begin{cases} u(r) - u(R_c) - u(r - R_c) \cdot u'(R_c) & r < R_c \\ 0, & r > R_c \end{cases} \tag{15}$$

where $u(r)$ is the unmodified pairwise electrostatic potential (Eq. 6) and R_c is the long-range cutoff. Dispersion interactions are smoothly switched to zero as

$$\tilde{v}(r) = f_{\text{sw}}(r) \cdot v(r), \tag{16}$$

where $v(r)$ is the pairwise potential in Eq. 4. The switching function parameters $(r_{\text{on}}, r_{\text{off}})$ were set to $(R_c \times 0.45, R_c)$ for electrostatic interactions and $(R_c - 2, R_c)$ for dispersion interactions. The parameters were chosen to prevent clumping artifacts at the 10 Å long-range cutoff.
Binding energy calculation. Binding energy was calculated as the difference between the bound dimer and the non-interacting monomers (separated by a distance larger than the long-range cutoff) with charges assigned for each monomer in isolation.
Simulation details. All simulations were conducted using the NVT ensemble for gas-phase systems and the NPT ensemble for periodic systems, with a timestep of 0.5 fs. Nosé–Hoover Chains (3 chains) were used for thermostat and barostat coupling, as implemented in JAX-MD [36], with default parameters: 1000 timesteps for the barostat and 100 timesteps for the thermostat [93–95]. Prior to simulation, all structures were pre-optimized using the FIRE algorithm [96].
MD22 molecules and polyalanines. Gas-phase simulations were performed for 500 ps at 500 K for stachyose

and AcAla₃NMe, as well as at 300 K for other MD22 molecules. Simulations of AcAla₁₅NMe were carried out at 300 K, whereas AcAla₁₅LysH⁺ was simulated at 500–800 K with a step of 100 K, each for 500 ps. Secondary structure assignment of polyalanines was conducted using the STRIDE algorithm [97].

Water simulations. Simulations were run for 500 ps, with observables averaged over the final 300 ps. The diffusion coefficient was determined from the positions of oxygen atoms using Einstein diffusion equation [98, 99]. Double-precision was employed to enhance numerical stability.

Crambin. The initial structure was obtained from PDB ID: 2FD7 [100], with mutated residues reverted to the wild-type sequence. The system was solvated with 8205 explicit water molecules. Simulations were performed for 1 ns at 300 K, excluding the first 500 ps for equilibration. The root mean square deviation of Crambin, RMSD(t, t+Δt), was obtained excluding hydrogen atoms from three 3 ns runs. Power spectra were computed from atomic velocities sampled over a 125 ps trajectory with a time resolution of 2.5 fs using schnetpack package [101, 102].

Glycoprotein. The starting structure was taken from the PDB ID: 1K7C [103]. 15008 water molecules were used for solvating the system and the pH was set to 3.7 to guarantee charge neutralization. The RMSD was calculated based on three runs of 500 ps.

POPC Lipid bilayer. The starting structure, consisting of 128 lipids and 5120 water molecules, was obtained from Ref. 69. The system was equilibrated over 250 ps using a combination of geometry relaxations and NVT simulations. Observables were then averaged over an additional 250 ps at 303 K in an isotropic NPT ensemble implemented in JAX-MD. The initial box dimensions were adjusted manually to mimic semi-isotropic NPT ensemble. The area per lipid was calculated from the simulation box dimensions. Bilayer thickness (D_{HH}), derived from electron density profiles, and NMR order parameters were both calculated using CPPTRAJ [104]. Double-precision was employed to enhance numerical stability.

DATA, MODEL AND CODE AVAILABILITY

The model, datasets, and codes used in this work are available at https://github.com/general-molecular-simulations/so3lr. The repository contains notebooks with tutorial simulations.

ACKNOWLEDGEMENTS

CONTRIBUTIONS

A.K. and J.T.F. contributed equally. A.K. and A.T. conceived the project. A.K. implemented and trained the model, designed, performed and analyzed the simulation results. J.T.F. contributed to and guided the software implementation. S.S.D. contributed to software implementation, simulations, and their analysis. A.Kh. contributed to the implementation and analysis of the long-range modules. L.M.S. performed the reference calculations. O.T.U. contributed to the simulation analysis. A.K., J.T.F., and A.T. drafted the manuscript. A.K. created the figures with the help from J.T.F. All authors discussed the results and contributed to editing the manuscript. K.R.M. and A.T. supervised the project at all stages.

---

[*] klaus-robert.mueller@tu-berlin.de
[†] alexandre.tkatchenko@uni.lu

[1] E. Schrödinger, *What is Life? The Physical Aspect of the Living Cell* (Cambridge University Press, 1974).

[2] P. A. M. Dirac, Quantum mechanics of many-electron systems, Proc. R. Soc. Lond. A **123**, 714 (1929).

[3] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman lectures on physics, Vol. I: The new millennium edition: mainly mechanics, radiation, and heat*, Vol. 1 (Basic books, 2015).

[4] B. Huang, G. F. von Rudorff, and O. A. von Lilienfeld, The central role of density functional theory in the ai age, Science **381**, 170 (2023).

[5] P. Dauber-Osguthorpe and A. T. Hagler, Biomolecular force fields: where have we been, where are we now, where do we need to go and how do we get there?, J. Comput. Aided Mol. Des. **33**, 133 (2019).

[6] A. T. Hagler, Force field development phase ii: Relaxation of physics-based criteria. . . or inclusion of more rigorous physics into the representation of molecular energetics, J. Comput. Aided Mol. Des. **33**, 205 (2019).

[7] A. D. MacKerell Jr, Empirical force fields for biological macromolecules: overview and issues, J. Comput. Chem. **25**, 1584 (2004).

[8] W. F. Van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, *et al.*, Biomolecular modeling: goals, problems, perspectives, Angew. Chem. Int. Ed. **45**, 4064 (2006).

[9] T. Schlick and S. Portillo-Ledesma, Biomolecular modeling thrives in the age of technology, Nat. Comput. Sci. **1**, 321 (2021).

[10] F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi, Machine learning for molecular simulation, Annu. Rev. Phys. Chem. **71**, 361 (2020).

[11] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, Machine learning force fields, Chem. Rev. **121**, 10142 (2021).

[12] I. Poltavsky and A. Tkatchenko, Machine learning force fields: Recent advances and remaining challenges, J. Phys. Chem. Lett. **12**, 6551 (2021).

[13] K. T. Schütt, S. Chmiela, O. A. Von Lilienfeld, A. Tkatchenko, K. Tsuda, and K.-R. Müller, Machine learning meets quantum physics, Lecture Notes in Physics 10.1007/978-3-030-40245-7 (2020).

[14] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Muller, and A. Tkatchenko, Combining machine learning and computational chemistry for predictive insights into chemical systems, Chem. Rev. **121**, 9816 (2021).

[15] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys. Rev. Lett. **98**, 146401 (2007).

[16] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian Approximation Potentials: the accuracy of quantum mechanics, without the electrons, Phys. Rev. Lett. **104**, 136403 (2010).

[17] J. Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, J. Chem. Phys. **134**, 074106 (2011).

[18] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, SchNet – a deep learning architecture for molecules and materials, J. Chem. Phys. **148**, 241722 (2018).

[19] O. T. Unke and M. Meuwly, PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges, J. Chem. Theory Comput. **15**, 3678 (2019).

[20] A. S. Christensen, L. A. Bratholm, F. A. Faber, and O. Anatole von Lilienfeld, FCHL revisited: faster and more accurate quantum machine learning, J. Chem. Phys. **152**, 044107 (2020).

[21] O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schütt, H. E. Sauceda, and K.-R. Müller, SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects, Nat. Commun. **12**, 7273 (2021).

[22] T. Frank, O. Unke, and K.-R. Müller, So3krates: Equivariant attention for interactions on arbitrary length-scales in molecular systems, Adv. Neural Inf. Process. Syst. **35**, 29400 (2022).

[23] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, E (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials, Nat. Commun. **13**, 2453 (2022).

[24] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, and G. Csányi, Mace: Higher order equivariant message passing neural networks for fast and accurate force fields, Adv. Neural Inf. Process. Syst. **35**, 11423 (2022).

[25] D. M. Anstine and O. Isayev, Machine learning interatomic potentials and long-range physics, J. Phys. Chem. A **127**, 2417 (2023).

[26] L. C. Blum and J.-L. Reymond, 970 million druglike small molecules for virtual screening in the chemical universe database gdb-13, J. Am. Chem. Soc. **131**, 8732 (2009).

[27] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. Von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules, Sci. Data **1**, 1 (2014).

[28] J. S. Smith, O. Isayev, and A. E. Roitberg, Ani-1, a data set of 20 million calculated off-equilibrium conformations for organic molecules, Sci. Data **4**, 1 (2017).

[29] J. S. Smith, R. Zubatyuk, B. Nebgen, N. Lubbers, K. Barros, A. E. Roitberg, O. Isayev, and S. Tretiak, The ani-1ccx and ani-1x data sets, coupled-cluster and density functional theory properties for molecules, Sci. Data **7**, 134 (2020).

[30] A. G. Donchev, A. G. Taube, E. Decolvenaere, C. Hargus, R. T. McGibbon, K.-H. Law, B. A. Gregersen, J.-L. Li, K. Palmo, K. Siva, *et al.*, Quantum chemical benchmark databases of gold-standard dimer interaction energies, Sci. Data **8**, 55 (2021).

[31] J. Hoja, L. Medrano Sandonas, B. G. Ernst, A. Vazquez-Mayagoitia, R. A. DiStasio Jr, and A. Tkatchenko, Qm7-x, a comprehensive dataset of quantum-mechanical properties spanning the chemical space of small organic molecules, Sci. Data **8**, 43 (2021).

[32] C. Isert, K. Atz, J. Jiménez-Luna, and G. Schneider, Qmugs, quantum mechanical properties of drug-like molecules, Sci. Data **9**, 273 (2022).

[33] P. Eastman, P. K. Behara, D. L. Dotson, R. Galvelis, J. E. Herr, J. T. Horton, Y. Mao, J. D. Chodera, B. P. Pritchard, Y. Wang, *et al.*, Spice, a dataset of drug-like molecules and peptides for training machine learning potentials, Sci. Data **10**, 11 (2023).

[34] L. Medrano Sandonas, D. Van Rompaey, A. Fallani, M. Hilfiker, D. Hahn, L. Perez-Benito, J. Verhoeven, G. Tresadern, J. Kurt Wegner, H. Ceulemans, *et al.*, Dataset for quantum-mechanical exploration of conformers and solvent effects in large drug-like molecules, Sci. Data **11**, 742 (2024).

[35] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, *et al.*, Openmm 7: Rapid development of high performance algorithms for molecular dynamics, PLOS Comput. Biol. **13**, e1005659 (2017).

[36] S. Schoenholz and E. D. Cubuk, Jax md: a framework for differentiable physics, Adv. Neural Inf. Process. Syst. **33**, 11428 (2020).

[37] R. P. Pelaez, G. Simeon, R. Galvelis, A. Mirarchi, P. Eastman, S. Doerr, P. Thölke, T. E. Markland, and G. De Fabritiis, Torchmd-net 2.0: Fast neural network potentials for molecular simulations, J. Chem. Theory Comput. **20**, 4076 (2024).

[38] J. T. Frank, O. T. Unke, K.-R. Müller, and S. Chmiela, A euclidean transformer for fast and stable machine learned force fields, Nat. Commun. **15**, 6539 (2024).

[39] J. F. Ziegler, J. P. Biersack, and U. Littmark, The stopping and range of ions in solids (Pergamon Press, New York, 1985).

[40] A. Khabibrakhmanov, D. V. Fedorov, and A. Tkatchenko, Universal pairwise interatomic van der waals potentials based on quantum drude oscillators, J. Chem. Theory Comput. **19**, 7895 (2023).

[41] F. L. Hirshfeld, Bonded-atom fragments for describing molecular charge densities, Theor. Chim. Acta **44**, 129 (1977).

[42] E. R. Johnson and A. D. Becke, A post-hartree-fock model of intermolecular interactions: Inclusion of higher-order corrections, J. Chem. Phys. **124**, 174104 (2006).

[43] D. V. Fedorov, M. Sadhukhan, M. Stöhr, and A. Tkatchenko, Quantum-mechanical relation between atomic dipole polarizability and the van der waals radius, Phys. Rev. Lett. **121**, 183401 (2018).

[44] A. Tkatchenko and M. Scheffler, Accurate molecular van der waals interactions from ground-state electron density and free-atom reference data, Phys. Rev. Lett. **102**, 073005 (2009).

[45] A. P. Jones, J. Crain, V. P. Sokhan, T. W. Whitfield, and G. J. Martyna, Quantum drude oscillator model of atoms and molecules: Many-body polarization and dispersion interactions for atomistic simulation, Phys. Rev. B **87**, 144103 (2013).

[46] J. Rezác, K. E. Riley, and P. Hobza, S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures, J. Chem. Theory Comput. **7**, 2427 (2011).

[47] O. T. Unke, M. Stöhr, S. Ganscha, T. Unterthiner, H. Maennel, S. Kashubin, D. Ahlin, M. Gastegger, L. Medrano Sandonas, J. T. Berryman, *et al.*, Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments, Sci. Adv. **10**, eadn4397 (2024).

[48] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, Ab initio molecular simulations with numeric atom-centered orbitals, Comput. Phys. Commun. **180**, 2175 (2009).

[49] X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter, and M. Scheffler, Resolution-of-identity approach to hartree–fock, hybrid density functionals, rpa, mp2 and gw with numeric atom-centered orbital basis functions, New J. Phys. **14**, 053020 (2012).

[50] C. G. Staacke, S. Wengert, C. Kunkel, G. Csányi, K. Reuter, and J. T. Margraf, Kernel charge equilibration: efficient and accurate prediction of molecular dipole moments with a machine-learning enhanced electron density model, Mach. Learn.: Sci. Technol. **3**, 015032 (2022).

[51] S. Chmiela, V. Vassilev-Galindo, O. T. Unke, A. Kabylda, H. E. Sauceda, A. Tkatchenko, and K.-R. Müller, Accurate global machine learning force fields for

molecules with hundreds of atoms, Sci. Adv. **9**, eadf0873 (2023).

[52] Y. Yang, K. U. Lao, D. M. Wilkins, A. Grisafi, M. Ceriotti, and R. A. DiStasio Jr, Quantum mechanical static dipole polarizabilities in the qm7b and alphaml showcase databases, Sci. Data **6**, 152 (2019).

[53] C. Villot and K. U. Lao, Ab initio dispersion potentials based on physics-based functional forms with machine learning, J. Chem. Phys. **160**, 184103 (2024).

[54] D. P. Kovács, J. H. Moore, N. J. Browning, I. Batatia, J. T. Horton, V. Kapil, I.-B. Magdău, D. J. Cole, and G. Csányi, Mace-off23: Transferable machine learning force fields for organic molecules, arXiv preprint arXiv:2312.15211 10.48550/arXiv.2312.15211 (2023).

[55] A. Soyemi and T. Szilvási, Benchmarking semiempirical qm methods for calculating the dipole moment of organic molecules, J. Phys. Chem. A **126**, 1905 (2022).

[56] S. C. Lovell, I. W. Davis, W. B. Arendall III, P. I. De Bakker, J. M. Word, M. G. Prisant, J. S. Richardson, and D. C. Richardson, Structure validation by cα geometry: ϕ, ψ and cβ deviation, Proteins Struct. Funct. Bioinform. **50**, 437 (2003).

[57] A. K. Soper, The radial distribution functions of water as derived from radiation total scattering experiments: is there anything we can say for sure?, Int. Sch. Res. Notices **2013**, 279463 (2013).

[58] R. Mills, Self-diffusion in normal and heavy water in the range 1-45. deg., J. Phys. Chem. **77**, 685 (1973).

[59] T. M. Ferreira, F. Coreta-Gomes, O. S. Ollila, M. J. Moreno, W. L. Vaz, and D. Topgaard, Cholesterol and popc segmental order parameters in lipid membranes: solid state 1 h–13 c nmr and md simulation studies, Phys. Chem. Chem. Phys. **15**, 1976 (2013).

[60] D. Anstine, R. Zubatyuk, and O. Isayev, Aimnet2: a neural network potential to meet your neutral, charged, organic, and elemental-organic needs, ChemRxiv preprint chemrxiv-2023-296ch-v2 10.26434/chemrxiv-2023-296ch-v2 (2024).

[61] G. L. Millhauser, C. J. Stenland, P. Hanson, K. A. Bolin, and F. J. van de Ven, Estimating the relative populations of 310-helix and α-helix in ala-rich peptides: a hydrogen exchange and high field nmr study, J. Mol. Biol. **267**, 963 (1997).

[62] K. A. Bolin and G. L. Millhauser, α and 310: the split personality of polypeptide helices, Acc. Chem. Res. **32**, 1027 (1999).

[63] M. Kohtani, T. C. Jones, J. E. Schneider, and M. F. Jarrold, Extreme stability of an unsolvated α-helix, J. Am. Chem. Soc. **126**, 7420 (2004).

[64] R. A. DiStasio, B. Santra, Z. Li, X. Wu, and R. Car, The individual and collective effects of exact exchange and dispersion interactions on the ab initio structure of liquid water, J. Chem. Phys. **141**, 084502 (2014).

[65] L. McInnes, J. Healy, and J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction, arXiv preprint arXiv:1802.03426 10.48550/arXiv.1802.03426 (2018).

[66] H.-C. Ahn, N. Juranić, S. Macura, and J. L. Markley, Three-dimensional structure of the water-insoluble protein crambin in dodecylphosphocholine micelles and its minimal solvent-exposed surface, J. Am. Chem. Soc. **128**, 4398 (2006).

[67] O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison, and A. D. MacKerell Jr, Charmm additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling, J. Chem. Theory Comput. **7**, 3162 (2011).

[68] N. Kučerka, M.-P. Nieh, and J. Katsaras, Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature, Biochim. Biophys. Acta **1808**, 2761 (2011).

[69] C. J. Dickson, R. C. Walker, and I. R. Gould, Lipid21: complex lipid membrane simulations with amber, J. Chem. Theory Comput. **18**, 1726 (2022).

[70] Y. Yu, A. Kramer, R. M. Venable, B. R. Brooks, J. B. Klauda, and R. W. Pastor, Charmm36 lipid force field with explicit treatment of long-range dispersion: parametrization and validation for phosphatidylethanolamine, phosphatidylglycerol, and ether lipids, J. Chem. Theory Comput. **17**, 1581 (2021).

[71] D. A. Case, T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, The amber biomolecular simulation programs, J. Comput. Chem. **26**, 1668 (2005).

[72] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, Development and testing of a general amber force field, J. Comput. Chem. **25**, 1157 (2004).

[73] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al., Charmm: the biomolecular simulation program, J. Comput. Chem. **30**, 1545 (2009).

[74] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, et al., Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields, J. Comput. Chem. **31**, 671 (2010).

[75] X. Zhu, P. E. Lopes, and A. D. MacKerell Jr, Recent developments and applications of the charmm force fields, Wiley Interdiscip. Rev. Comput. Mol. Sci. **2**, 167 (2012).

[76] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids, J. Am. Chem. Soc. **118**, 11225 (1996).

[77] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6, J. Comput. Chem. **25**, 1656 (2004).

[78] N. Schmid, A. P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A. E. Mark, and W. F. Van Gunsteren, Definition and testing of the gromos force-field versions 54a7 and 54b7, Eur. Biophys. J. **40**, 843 (2011).

[79] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio Jr, et al., Current status of the amoeba polarizable force field, J. Phys. Chem. B **114**, 2549 (2010).

[80] R. Zubatyuk, J. S. Smith, J. Leszczynski, and O. Isayev, Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network, Sci. Adv. **5**, eaav6490 (2019).

[81] I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, W. J. Baldwin, N. Bernstein, et al., A foundation

model for atomistic materials chemistry, arXiv preprint arXiv:2401.00096 10.48550/arXiv.2401.00096 (2023).

[82] T. Plé, L. Lagardère, and J.-P. Piquemal, Force-field-enhanced neural network interactions: from local equivariant embedding to atom-in-molecule properties and long-range effects, Chem. Sci. **14**, 12554 (2023).

[83] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, Towards exact molecular dynamics simulations with machine-learned force fields, Nat. Commun. **9**, 3887 (2018).

[84] E. Slootman, I. Poltavsky, R. Shinde, J. Cocomello, S. Moroni, A. Tkatchenko, and C. Filippi, Accurate quantum monte carlo forces for machine-learned force fields: Ethanol as a benchmark, J. Chem. Theory Comput. **20**, 6020 (2024).

[85] M. Stöhr and A. Tkatchenko, Quantum mechanics of proteins in explicit water: The role of plasmon-like solute-solvent interactions, Sci. Adv. **5**, eaax0024 (2019).

[86] Y. Park, J. Kim, S. Hwang, and S. Han, Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations, J. Chem. Theory Comput. 10.1021/acs.jctc.4c00190 (2024).

[87] H. E. Sauceda, V. Vassilev-Galindo, S. Chmiela, K.-R. Müller, and A. Tkatchenko, Dynamical strengthening of covalent and non-covalent molecular interactions by nuclear quantum effects at finite temperature, Nat. Commun. **12**, 442 (2021).

[88] F. Musil, I. Zaporozhets, F. Noé, C. Clementi, and V. Kapil, Quantum dynamics using path integral coarse-graining, J. Chem. Phys. **157**, 181102 (2022).

[89] S. J. Reddi, S. Kale, and S. Kumar, On the convergence of adam and beyond, arXiv preprint arXiv:1904.09237 10.48550/arXiv.1904.09237 (2019).

[90] J. L. Ba, J. R. Kiros, and G. E. Hinton, Layer normalization, arXiv preprint arXiv:1607.06450 10.48550/arXiv.1607.06450 (2016).

[91] C. J. Fennell and J. D. Gezelter, Is the ewald summation still necessary? pairwise alternatives to the accepted standard for long-range electrostatics, J. Chem. Phys. **124**, 234104 (2006).

[92] D. Wolf, P. Keblinski, S. Phillpot, and J. Eggebrecht, Exact method for the simulation of coulombic systems by spherically truncated, pairwise r- 1 summation, J. Chem. Phys. **110**, 8254 (1999).

[93] S. Nosé, A unified formulation of the constant temperature molecular dynamics methods, J. Chem. Phys **81**, 511 (1984).

[94] W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, Phys. Rev. A **31**, 1695 (1985).

[95] G. J. Martyna, M. L. Klein, and M. Tuckerman, Nosé–hoover chains: The canonical ensemble via continuous dynamics, J. Chem. Phys **97**, 2635 (1992).

[96] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, Structural relaxation made simple, Phys. Rev. Lett. **97**, 170201 (2006).

[97] D. Frishman and P. Argos, Knowledge-based protein secondary structure assignment, Proteins Struct. Funct. Bioinform. **23**, 566 (1995).

[98] G. Pranami and M. H. Lamm, Estimating error in diffusion coefficients derived from molecular dynamics simulations, J. Chem. Theory Comput. **11**, 4586 (2015).

[99] E. J. Maginn, R. A. Messerly, D. J. Carlson, D. R. Roe, and J. R. Elliot, Best practices for computing transport properties 1. self-diffusivity and viscosity from equilibrium molecular dynamics [article v1. 0], Living J. Comp. Mol. Sci. **1**, 6324 (2019).

[100] D. Bang, V. Tereshko, A. A. Kossiakoff, and S. B. H. Kent, Role of a salt bridge in the model protein crambin explored by chemical protein synthesis: X-ray structure of a unique protein analogue, [v15a]crambin-α-carboxamide, Mol. BioSyst. **5**, 750 (2009).

[101] K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller, SchNetPack: A Deep Learning Toolbox For Atomistic Systems, J. Chem. Theory Comput. **15**, 448 (2019).

[102] K. T. Schütt, S. S. P. Hessmann, N. W. A. Gebauer, J. Lederer, and M. Gastegger, SchNetPack 2.0: A neural network toolbox for atomistic machine learning, J. Chem. Phys. **158**, 144801 (2023).

[103] A. Mølgaard and S. Larsen, A branched N-linked glycan at atomic resolution in the 1.12Å structure of rhamnogalacturonan acetylesterase, Acta Crystallogr. D **58**, 111 (2002).

[104] D. R. Roe and T. E. Cheatham III, Ptraj and cpptraj: software for processing and analysis of molecular dynamics trajectory data, J. Chem. Theory Comput. **9**, 3084 (2013).

[105] B. K. Rai, V. Sresht, Q. Yang, R. Unwalla, M. Tu, A. M. Mathiowetz, and G. A. Bakken, Torsionnet: A deep neural network to rapidly predict small-molecule torsional energy profiles with the accuracy of quantum mechanics, J. Chem. Inf. Model. **62**, 785 (2022).
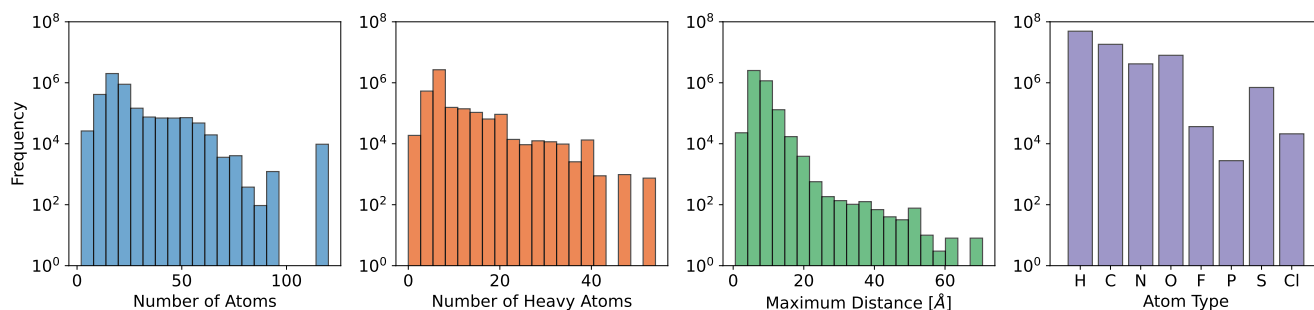
Fig. S1. Statistics on a combined dataset of 3.9 million molecular fragments. Histograms of the number of atoms, number of heavy atoms, maximum distance in each fragment, and atom types.

TABLE S1. Properties present in the combined datasets.

Dataset	Size	Forces	Dipoles	Hirsh. rat.
GEMS bottom-up	2.7m	✓	✓	✗
QM7-X	1m	✓	✓	✓
AQM	60k	✓	✓	✓
SPICE Dipeptides	33k	✓	✓	✓
DES15k	15k	✓	✓	✓
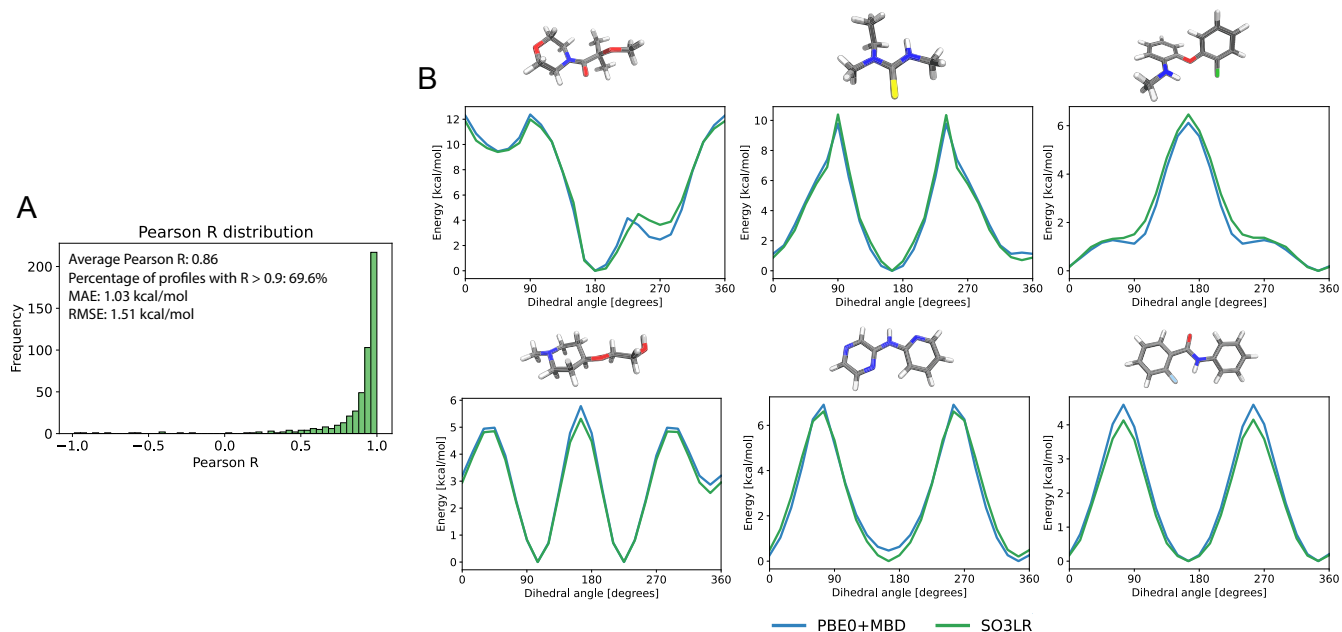
Fig. S2. TorsionNet500 Benchmark. Comparison of energies predicted by SO3LR with the TorsionNet500 benchmark [105], recomputed at the PBE0+MBD level of theory. **A** Histogram of Pearson R coefficients, with additional metrics shown in the inset. **B** Torsional profiles for six molecules. The absence of certain functional groups (e.g., triazole and trifluoromethylthio groups) in the training set leads to higher average errors. In contrast, torsional profiles commonly encountered in biosimulations are predicted accurately.
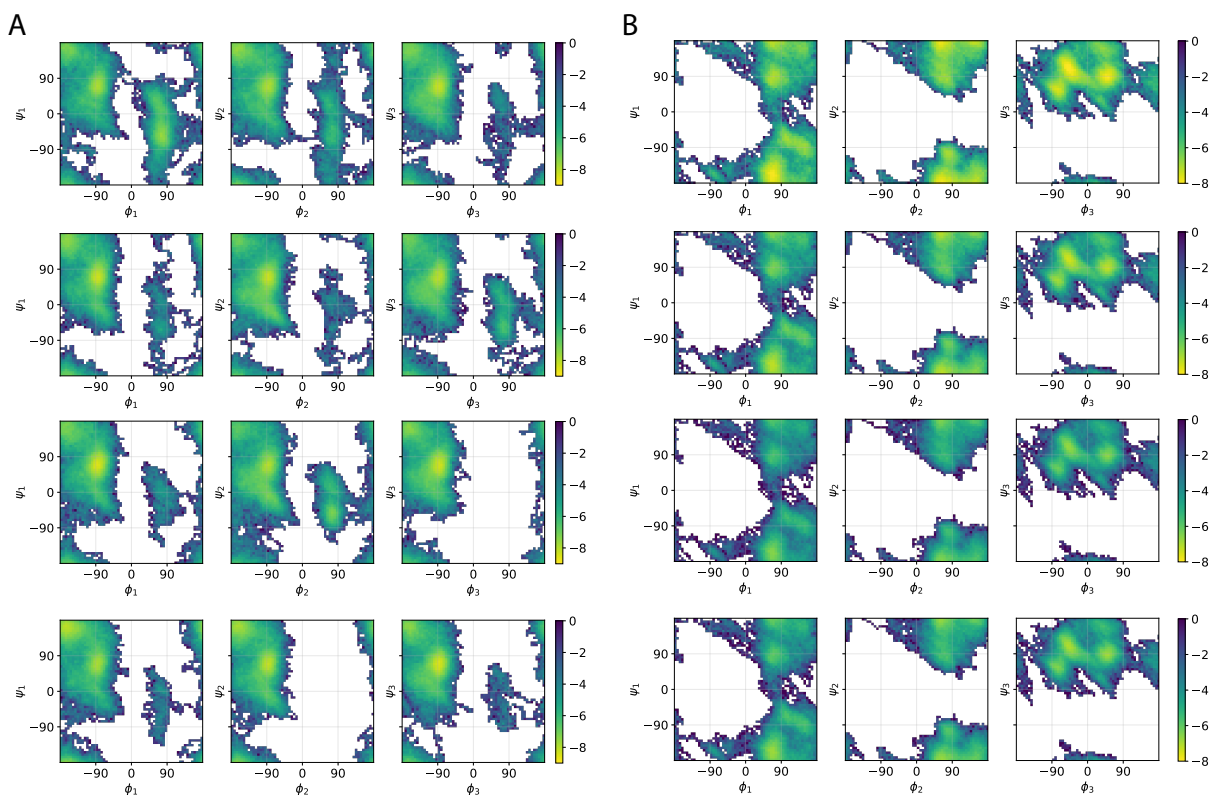
Fig. S3. Simulations of small biomolecular fragments. Ramachandran plots (ϕ/ψ dihedrals) for **(A)** AcAla$_3$NMe and **(B)** stachyose molecules from MD22. SO3LR simulations at $500\,\mathrm{K}$ for $500\,\mathrm{ps}$. Trajectory is sampled every $1\,\mathrm{fs}$. The Boltzmann-inverted scale is shown in kcal/mol.
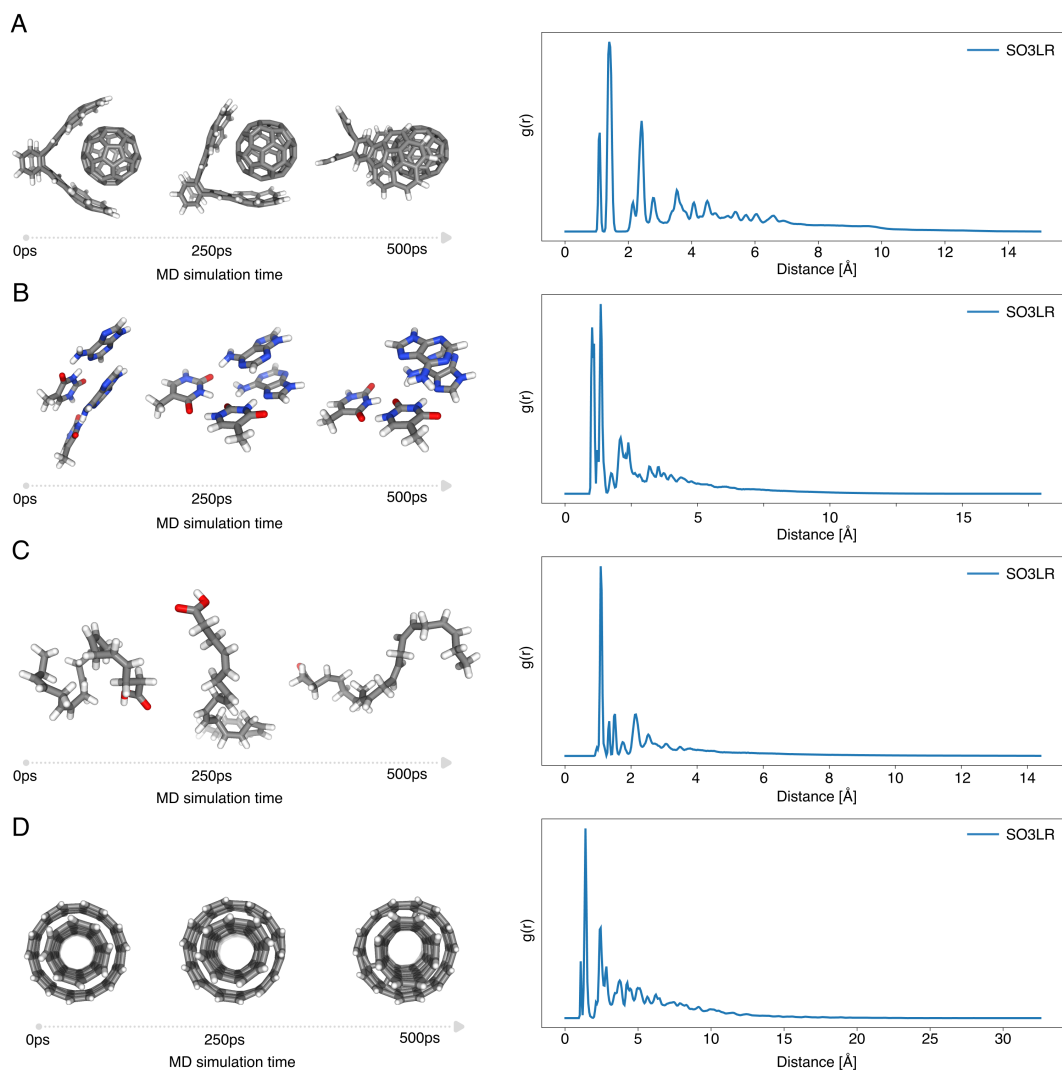
Fig. S4. Simulation of structures from the MD22 benchmark. Snapshots of the simulation at 0 ps, 250 ps and 500 ps (left) and the corresponding radial distribution function $g(r)$ computed over frames sampled every 1 ps (right) for the **(A)** buckyball catcher, **(B)** AT-AT, **(C)** DHA, and **(D)** nanotube. Simulations were performed for 500 ps at 300 K.
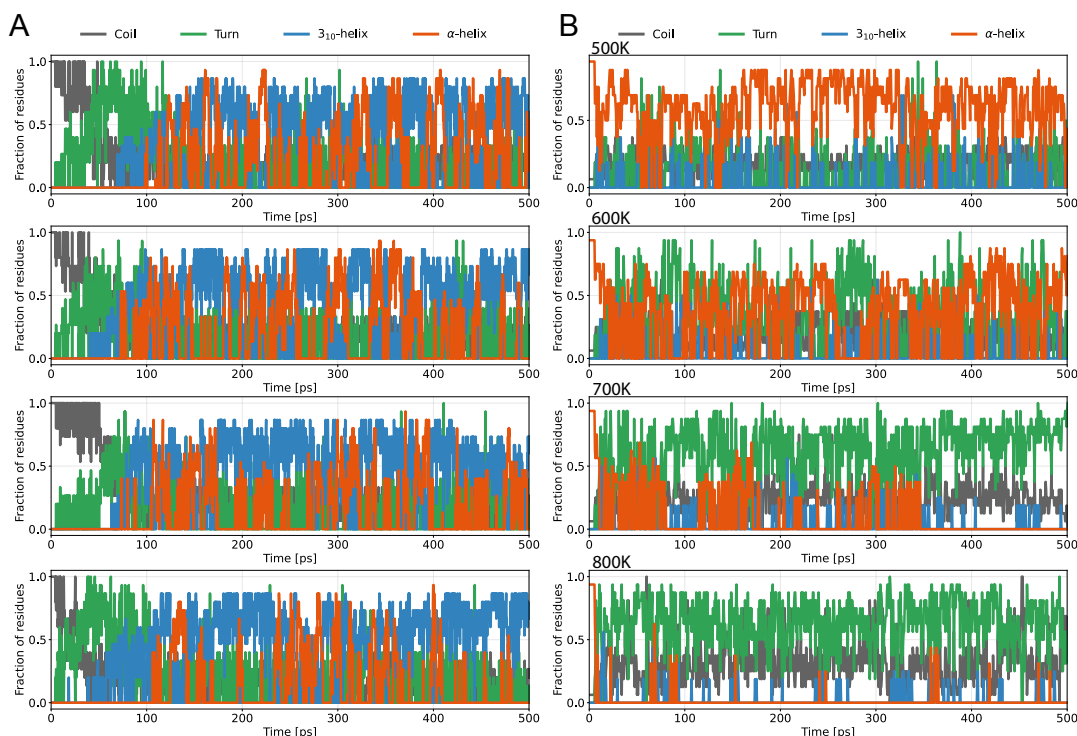
Fig. S5. Polyalanine simulation. Secondary structural motifs of **(A)** four folding trajectories of extended AcAla$_{15}$NMe at 300K in the gas phase and **(B)** four trajectories starting from the folded AcAla$_{15}$LysH$^+$ at 500, 600, 700, and 800K. STRIDE was used for secondary structure assignment [97].
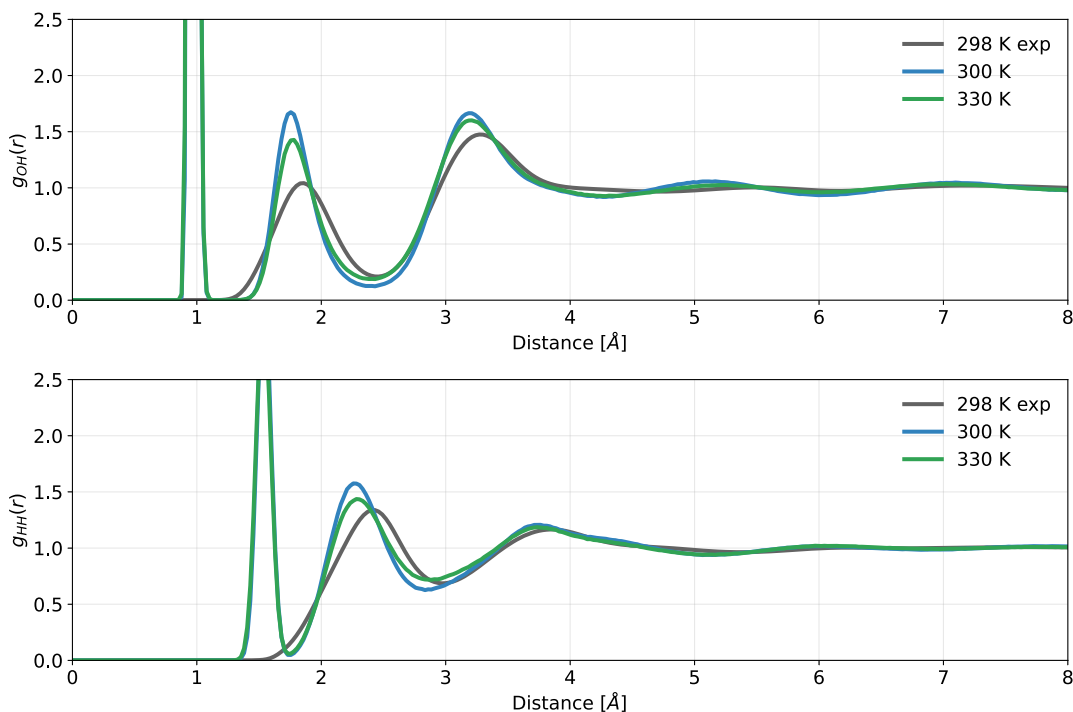

Fig. S6. Water radial distribution functions. Oxygen-hydrogen and hydrogen–hydrogen radial distribution functions for bulk water (4096 water molecules), simulated at 300 and 330 K and compared with experimental results [57]. Intramolecular contributions in the experimental data are not included.
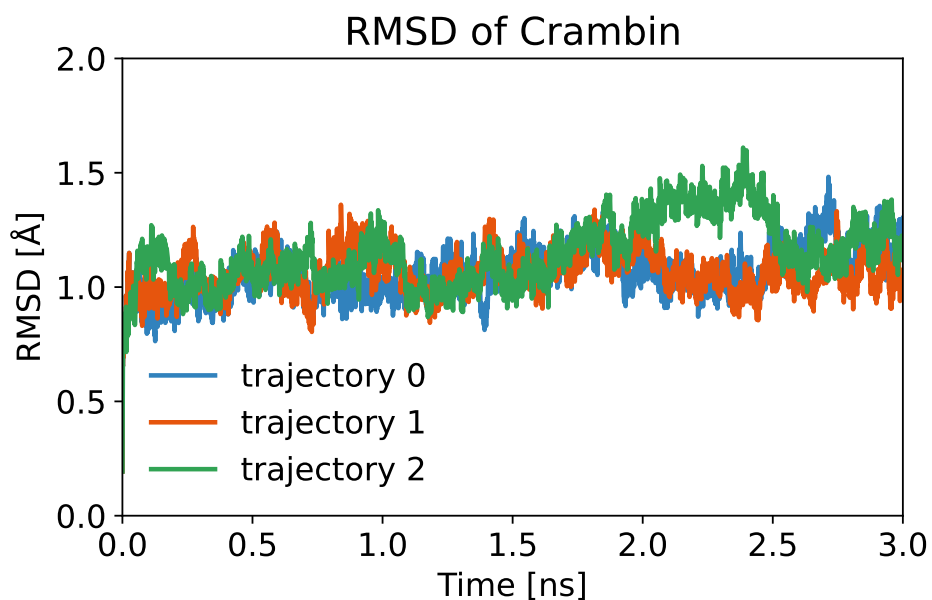
Fig. S7. Crambin RMSD. Root mean square deviations of three crambin trajectories simulated with SO3LR with respect to the initial frame.
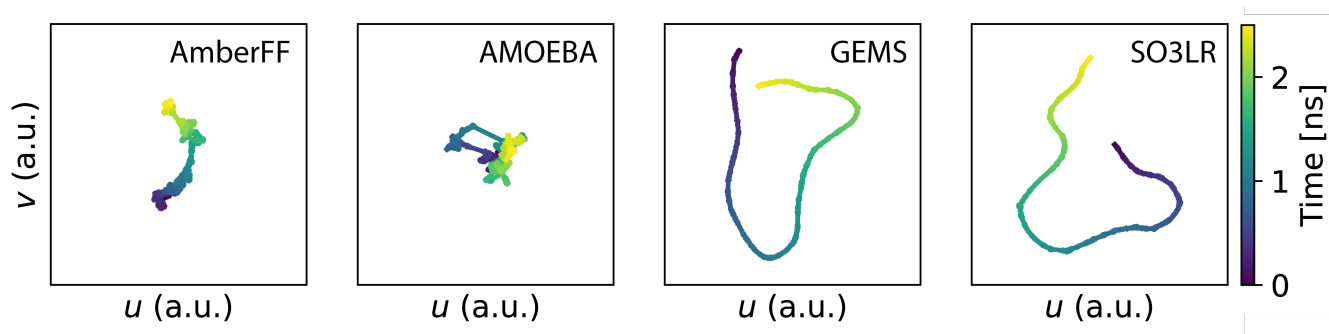

Fig. S8. 2D UMAP. Calculated using UMAP package [65] from a 3 ns trajectory of crambin in aqueous solution. The initial 500 ps were discarded for system equilibration. The projection was calculated using non-hydrogen protein atoms with UMAP hyperparameters set to 30 neighbors and a minimum distance of 0.1. All subplots share identical latent space coordinates.
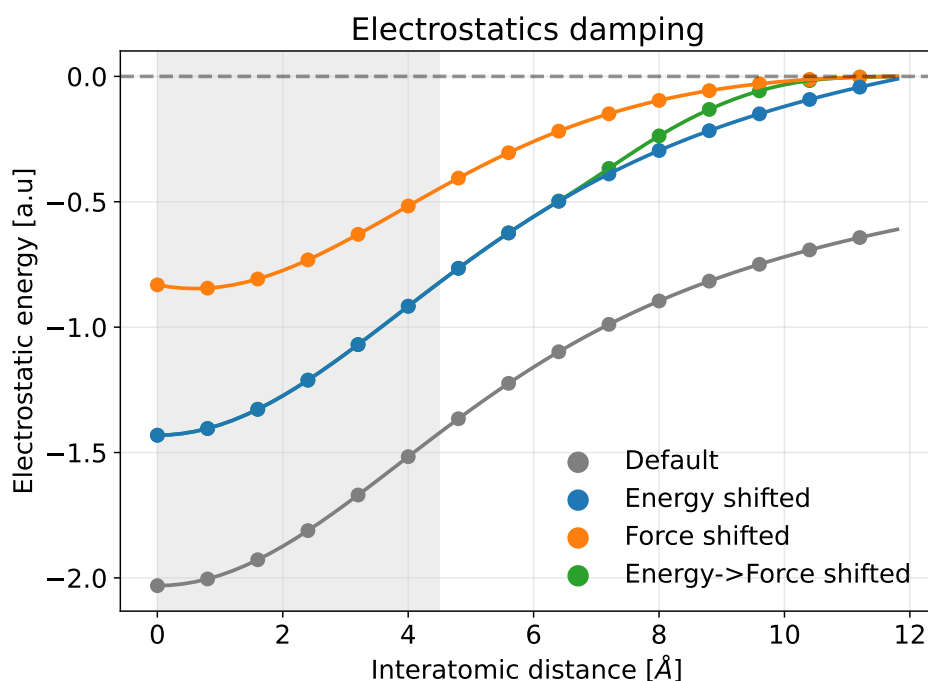
Fig. S9. Electrostatics switched interactions. The model was trained with a damped $\mathrm{erf}(r_{ij}/4)/r_{ij}$ electrostatic potential on gas-phase data, with a $100\,\text{Å}$ long-range cutoff that recovers all neighbours and a $4.5\,\text{Å}$ short-range cutoff. In simulations with periodic boundary conditions, we employ a long-range cutoff of $12\,\text{Å}$ to balance accuracy and computational efficiency. The potential at short-range should be the same as the one the model was trained with to maintain the learned balance between different terms. Simultaneously, the potential should smoothly transition to zero at the long-range cutoff to ensure that the potential is the exact integral of the force and to avoid introducing discontinuities in the forces. Therefore, we smoothly switch between the energy-shifted (blue curve) and force-shifted (orange curve) potentials to obtain the final potential (green curve). Dispersion interactions are smoothly energy-shifted starting $2\,\text{Å}$ before long-range cutoff.