Communications in Computational Chemistry
Commun. Comput. Chem., Vol. 1, No. 2, 152–160
doi: 10.4208/cicc.2025.90.02

Regular Article

Benchmarking Universal Machine Learning Force Fields with Hydrogen-Bonding Cooperativity

Xinping Feng[1,2,3], You Xu[1,2,3,*] and Jing Huang[1,2,3,*]

[1]*School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310030, China ;*

[2]*Westlake AI Therapeutics Laboratory, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang 310024, China ;*

[3]*Institute of Biology, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310024, China*

[*] Corresponding authors: huangjing@westlake.edu.cn; xuyou@westlake.edu.cn

Received on 24 April 2025; Accepted on 31 May 2025

Abstract: Machine learning force fields (MLFFs) offer a promising balance between quantum mechanical (QM) accuracy and molecular mechanics efficiency. While MLFFs have shown strong performance in modeling short-range interactions and reproducing potential energy surfaces, their ability to capture long-range cooperative effects remains underexplored. In this study, we assess the ability of three MLFF models — ANI, MACE-OFF, and Orb — to reproduce cooperative interactions arising from environmental induction and dispersion, which are essential for many biomolecular processes. Using a recently proposed framework, we quantify hydrogen bond (H-bond) cooperativity in N-methylacetamide polymers. Our results show that all MLFFs capture cooperativity to some extent, with MACE-OFF yielding the closest agreement with QM data. These findings highlight the importance of evaluating many-body effects in MLFFs and suggest that H-bond cooperativity can serve as a useful benchmark for improving their physical fidelity.

Key words: Machine learning force field, cooperative effects, self-assembly, neural network potential, hydrogen bond.

1. Introduction

Machine learning potentials (MLPs), sometimes also referred to as machine learning force fields (MLFFs), have emerged as a significant advancement in computational chemistry, offering a balance between the high accuracy of quantum mechanics (QM) and the computational efficiency of molecular mechanics (MM) [1,2]. By leveraging machine learning algorithms, MLPs learn the statistical relationship between molecular structures and their potential energies from large datasets. Various methodologies, including kernel-based approaches and neural network (NN) models, have demonstrated notable success in simulating and predicting the properties of complex chemical systems.

NN potentials are particularly promising because deep NNs excel at fitting high dimensional data distributions, enabling them to capture intricate intra- and intermolecular interactions [3–8]. The seminal work of Behler and Parrinello introduced a scheme in which the total potential energy is decomposed into atomic contributions, each predicted by an NN that takes atom-centered environment descriptors as input [9]. The field has progressed rapidly since then, with more advanced network architectures designed to better preserve physical symmetries and improve training efficiency [10–14]. Whereas early MLPs were usually trained bespoke on data generated for the particular system studied, recent efforts have shifted toward pre-training models for broad, off-the-shelf use — much like traditional force fields (FFs) that are carefully parametrized once and then distributed to end-users. Resonating with the wider ML move toward “foundation models”, an increasing number of universal MLFFs covering substantial regions of chemical space are now becoming available.

One of the most popular universal MLFFs is the ANI series developed by Roitberg and co-workers. The ANI-2x model, employs transfer learning: a network first trained on a large DFT dataset is fine-tuned with a smaller, high-level CCSD(T) dataset, achieving accuracy across a broad chemical space that includes C, H, N, O, S, F, and Cl [15], and performing well on a wide range of drug-like

molecules and small peptides. Another notable model, MACE, extends message-passing neural networks (MPNNs) by incorporating higher-order equivariant messages, boosting both efficiency and accuracy [13]. The MACE open force fields (MACE-OFFs) were trained on diverse datasets including organic molecules, water clusters, small peptides and dipeptides, and have demonstrated state-of-the-art performance not only in reproducing potential energy surfaces (PES) and atomic forces, but also in predicting condensed-phase properties such as liquid densities and solvation free energies [16,17]. Meanwhile, the Orb model, designed for large-scale simulations of inorganic and crystalline materials, employs a scalable graph NN architecture that preserves rotational invariance and allows efficient modeling of long-range dispersion interactions via diffusion pre-training and D3 correction [14].

In classical FFs, non-bonded interactions are evaluated throughout the space. In particular, electrostatics are typically handled with the particle meshed Ewald (PME) method, and more recently the LJ-PME method has been adopted to fully account for the van der Waals interactions [18,19]. This contrasts with MLFFs, in which atomic energies are computed from the information contained within a direct cutoff (typically 4.0 Å to 6.0 Å). Interactions beyond this range are presumed to be captured indirectly, for example through multiple passes of message passing, yet the extent to which MLFFs reproduce full long range interaction remains to be systematically benchmarked [20]. Long-range interactions are essential in chemical and biological systems dominated by non-covalent forces, for example proteins. A prime example is the cooperative effects of hydrogen bonds (H-bonds) in stabilizing protein structures. Once a few H-bonds form between residues, subsequent H-bond formation becomes energetically more favorable, facilitating protein folding, assembly, and aggregation [21–23].

H-bond cooperativity arises mainly from electronic induction, reflecting the molecular polarizability in response to the surrounding environment. QM methods such as CCSD(T), MP2, and DFT can accurately capture these effects, as can polarizable MM force fields, which explicitly account for charge redistribution [24–27]. In contrast, additive force fields fail to reproduce such cooperative effects. The quantification of cooperative energy has historically varied due to differences in computational methods, model systems, and definitions of cooperativity. QM calculations of various H-bonded polymers, including N-methylacetamide (NMA) [28], water [29], N-methylformamide (NMF) [30], formamide [31], and alanine peptides [32,33], have yielded cooperative energy estimates ranging from 3 to 26 kcal/mol, depending on the specific formulation used.

Recent methodological advancements have enabled more rigorous quantification of cooperative effects and improved benchmark for classical MM force fields [34]. In this recent study, we introduced a general framework that defines cooperativity as the difference between the interaction energy of an isolated dimer ($V_{A\text{-}B}^{AB}$) and its interaction energy in the presence of a third molecule ($V_{A\text{-}B}^{ABC}$) (Figure 1a). This cooperative energy is computed using the internal energies of optimized geometries (Equation 1).

$$\Delta V_{A\text{-}B} = -(U_{ABC}^{ABC} - U_{BC}^{ABC} - U_{AC}^{ABC} + U_{C}^{ABC} - U_{AB}^{AB} + U_{A}^{AB} + U_{B}^{AB}) \quad (1)$$

As the superscripts indicate, the first four terms (internal energies of *ABC*, *BC*, *AC*, and *C*) are computed using the geometry and basis set of complex *ABC*, while the remaining three terms (internal energies of *AB*, *A*, and *B*) are obtained using the geometry and basis set of complex *AB*. Here, *A*, *B*, and *C* represent any number of molecules. This framework accounts for the cooperative effects arising from both geometric distortion and electrostatic induction within molecules *A* and *B* in response to the addition of *C*, while explicitly excluding any contributions from the internal interactions or polarization of *C* itself. This formulation thus enables a systematic analysis of many-body effects in various molecular environments.

Here, we apply this theoretical framework to evaluate the ability of MLFFs to capture H-bond cooperatives in MLFFs. We used the same model systems and QM references, where NMA chains were optimized at the ωB97XD/cc-pVTZ level and energies computed at the RI-MP2/aug-cc-pVTZ level. Our findings reveal that all MLFFs considered, including the ANI, MACE-OFF, and Orb models, capture cooperativity to some extent but with considerable variance in accuracy. Despite being trained on high-quality QM data covering chemical spaces, these MLFFs differ substantially in how they represent environment-dependent interactions. Our work provides a critical benchmark for MLFFs and sheds light on how these models capture complex cooperative effects.

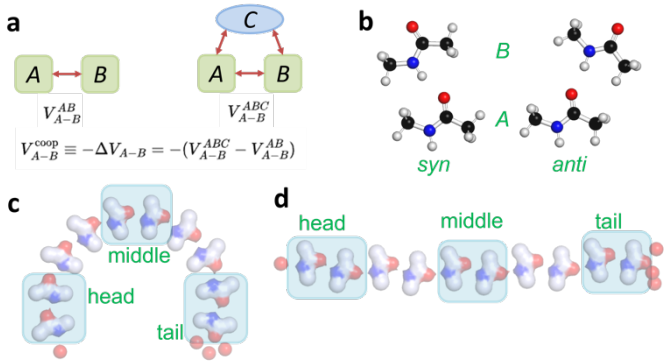

Figure 1. The scheme of cooperativity calculation and models. (a) Cooperativity is the difference of dimer interaction between two-body and three-body systems. (b) Two conformations of NMA dimer optimized at the ωB97XD/cc-pVTZ level of QM. (c) An arc decamer with exclusively syn conformation of dimer blocks. (d) A linear decamer with alternating syn and anti blocks. In (c) and (d), the water capping sites on termini are illustrated; and the cooperative energies calculated for the first, middle, and last H-bonds to evaluate the effect of NMAs extending on B-side, both sides and A-side, respectively, are highlighted with transparent boxes.

2. Method

Homogeneous NMA polymers identical to the previous study were used [34], where NMAs oriented in parallel configuration were arranged to form hydrogen bonds (H-bonds) in a head-to-tail way. As the fundamental H-bonded block, an NMA dimer was used, in which a hydrogen bond was formed between the carbonyl oxygen of molecule A (O_A) and the amide nitrogen of molecule B (N_B). The dimer was constrained to remain planar.

The NMA dimer adopted two distinct conformations. In the syn conformation, atoms N_A and N_B are positioned on the same side of the C_A-O_A axis, corresponding to dihedral ϕ_1 (N_A-C_A-O_A-N_B) of 0°. In contrast, in the anti conformation, N_A and N_B lay on opposite sides of the C_A-O_A axis, with ϕ_1 at 180° (Figure 1b). Two well-ordered NMA polymer patterns were examined: one in which identical dimers formed an arc pattern and another where alternating syn and anti dimers resulted in a linear pattern. All polymer structures were built starting from a syn dimer and extended up to 12