



Cite this: *Phys. Chem. Chem. Phys.*,
2026, **28**, 3949

Advancing density functional tight-binding method for large organic molecules through equivariant neural networks

Leonardo Medrano Sandonas,^{ib}*^a Mirela Puleva,^{id}^{bc} Zekiye Erarşlan,^{id}^a
 Ricardo Parra Payano,^d Martin Stöhr,^{bef} Gianauelio Cuniberti^{id}^{agh} and
 Alexandre Tkatchenko^{ib}*^{bc}

Semi-empirical quantum-mechanical (QM) methods have become valuable tools for studying complex (bio)molecular systems due to their balance between computational efficiency and accuracy. A key aspect of these methods is their parameterization, which not only governs the reliability of the results but also provides an opportunity to enhance their overall performance. In our previous work [*J. Phys. Chem. Lett.*, 2021, **11**, 16], we advanced the third-order semi-empirical density functional tight-binding (DFTB3) method for computing multiple properties of small molecules by developing the machine learning (ML) potential NN_{rep} to bridge the gap between DFTB3 electronic components and those of the hybrid DFT-PBE0 functional. To overcome the limitations of NN_{rep} , we introduce the EquiDTB framework, which leverages physics-inspired equivariant neural networks (NN) to parameterize scalable and transferable many-body Δ_{TB} potentials, replacing the standard pairwise DFTB repulsive potential. This advancement extends the applicability of our ML-corrected DFTB approach to larger molecules and non-covalent systems (including only C, N, O, and H atoms), going beyond the chemical space represented in the training QM datasets. The enhanced performance of EquiDTB over the standard TB methods is demonstrated by the accurate computation of the atomic forces of S66x8 molecular dimers, as well as their interaction energies. Moreover, EquiDTB can be effectively employed to explore the potential energy surfaces of large and flexible drug-like molecules—for example, to determine the minimum energy path between isomers, analyze structural transitions during dynamical simulations, compute vibrational modes, and investigate energetic rankings. The performance for single molecules slightly decreases when the DFTB electronic energy is reduced to first-order but remains superior to standard TB methods. Our work thus demonstrates that an optimal integration of an equivariant NN with QM datasets can advance the DFTB method while maintaining high efficiency, paving the way for reliable (bio)molecular simulations.

Received 6th January 2026,
Accepted 16th January 2026

DOI: 10.1039/d6cp00038j

rsc.li/pccp

1 Introduction

The high computational costs associated with using first-principles electronic structure methods to accurately investigate the properties of (bio)molecular systems (*e.g.*, drugs, proteins, nucleic acids, molecular crystals) have driven the development of machine learning (ML)-based methods in computational (bio)chemistry.^{1–4} Indeed, great efforts have been put into improving molecular representations and neural network (NN) architectures to construct robust ML potentials with quantum-mechanical (QM) accuracy, enabling reliable simulations of large and complex systems that were previously inaccessible.^{5,6} Among them, a major breakthrough was the development of SE(3)-equivariant NNs,^{7–13} which provide a better mapping between molecular structures and their corresponding atomic forces (and total energies), enhancing

^a *Institute for Materials Science and Max Bergmann Center of Biomaterials, TUD Dresden University of Technology, 01062 Dresden, Germany.*

E-mail: leonardo.medrano@tu-dresden.de

^b *Department of Physics and Materials Science, University of Luxembourg, L-1511 Luxembourg City, Luxembourg. E-mail: alexandre.tkatchenko@uni.lu*

^c *Institute for Advanced Studies, University of Luxembourg, Campus Belval, L-4365 Esch-sur-Alzette, Luxembourg*

^d *Universidad Nacional de Ingeniería, Av. Túpac Amaru 210, Rimac, Lima 15333, Peru*

^e *Department of Chemistry and The PULSE Institute, Stanford University, Stanford, CA-94305, USA*

^f *SLAC National Accelerator Laboratory, Menlo Park CA-94025, USA*

^g *Cluster of Excellence CARE, TU Dresden and RWTH Aachen, Germany*

^h *Cluster of Excellence CeTI, TU Dresden, Germany*



the efficiency, accuracy, scalability, and transferability (EAST) requirements¹ of ML potentials. Although equivariant ML potentials trained on absolute energies and forces have already facilitated the high-fidelity analysis of energetic and structural properties in systems like poly-alanine peptide and crambin in aqueous solution,^{14–16} their transferability remains limited when applied to (bio)molecules containing chemical moieties unseen during training. Instead of incorporating additional QM data to mitigate this issue, a more physics-based solution has recently gained considerable attention: the development of ML-augmented semi-empirical (SE) methods.^{17,18} By integrating ML techniques with SE methods, researchers have aimed to improve the predictive capabilities of these models while maintaining computational efficiency.^{18–32}

Building upon this idea, the performance of a widely used SE method such as the density functional tight-binding (DFTB) method³³ has been recently improved using diverse ML techniques.^{18,23–29,34–37} DFTB is a flexible cost-effective QM method derived from density functional theory (DFT) that can be applied to investigate multiple local and global properties of organic molecules and materials.^{33,38,39} However, the inherent approximations of DFTB introduce certain limitations that may affect its ability to accurately model complex (bio)molecular systems and their interactions. For instance, the minimal basis of localized atomic orbitals (LCAO) can lead to inaccuracies in describing electron delocalization and charge transfer processes.⁴⁰ Moreover, the truncation of the Taylor expansion of the density functional to the second or third order around a reference density can lead to errors in describing systems with significant charge transfer or highly reactive conditions.^{41,42} While the parameterization of Slater–Koster integrals and pairwise repulsive potentials (pw_{rep}) enables efficient calculations, it also introduces empirical biases that can affect transferability across different systems. Indeed, parameters optimized for small molecules may however not perform well for large (bio)molecular systems without further adjustments to adequately capture missing electronic effects.^{41,43} Accordingly, ML techniques have been employed to accurately parameterize electronic DFTB Hamiltonian,^{23,34–37} to develop more transferable and reliable repulsive potentials,^{18,25,27} and to implement delta-learning frameworks to achieve a higher QM level.^{24,26,28,29}

In particular, Gaussian process regression repulsive potentials (GPrep-DFTB) have been constructed to improve the accuracy in computing the properties of diverse small molecules,²⁷ Ruthenium oxide,⁴⁴ and Lithium-Intercalated graphite.⁴⁵ Deep tensor neural networks were also used in our previous work to train many-body repulsive potentials resulting in the DFTB- NN_{rep} model that yielded accurate predictions of atomization and isomerization energies, equilibrium geometries, vibrational frequencies, and dihedral rotational profiles for a large variety of organic molecules up to nine heavy atoms.¹⁸ On the other hand, DFTB has also been combined with graph-convolutional NN in a delta-learning framework to improve the description of long-range interactions, which are crucial for condensed-phase systems but challenging for traditional ML models to capture accurately.²⁸ In the same vein, a hierarchically interacting particle

neural network potential has also been developed to correct the DFTB method and accurately calculate the energetic properties of thorium–oxygen nanoclusters²⁹—a nontrivial system for DFTB due to complex electronic effects. Despite advancements, challenges remain in parameterizing a general-purpose ML potential to enhance the DFTB method for accurately investigating more flexible and non-covalent molecular systems.

To address this challenge, and building upon our previous work,¹⁸ we introduce here the EquiDTB framework (short for “Equivariant networks for Delta Tight-Binding”) to investigate how equivariant NNs influence the parameterization of many-body Δ_{TB} potentials, which are intended to replace pw_{rep} in the third-order DFTB method (DFTB3). To this end, we consider QM property data of both (equilibrium and non-equilibrium) small single molecules and molecular dimers contained in QM7-X⁴⁶ and DES15K⁴⁷ datasets, respectively, for training the Δ_{TB} potentials (including only C, N, O, and H). To determine the most reliable Δ_{TB} potential, we have also explored modern SE(3)-equivariant NNs such as SpookyNet,⁸ Allegro,¹¹ and MACE,⁹ each employing distinct strategies to model interactions within local chemical environments. The Δ_{TB} potentials are trained to reproduce the accuracy of the hybrid DFT-PBE0 functional. However, given the importance of long-range interactions in our target systems, we augment all calculations with a many-body dispersion (MBD)^{48,49} treatment. The inclusion of molecular dimers enables us to assess how non-covalent systems influence the parameterization of Δ_{TB} potentials, which so far has been developed only considering covalently-bonded systems.^{18,27} This is an important step toward identifying the key conditions for constructing Δ_{TB} potentials that can accurately capture both short- and long-range physical interactions, which the electronic DFTB Hamiltonian does not fully describe.

The comprehensive investigation of equivariant Δ_{TB} potentials conducted in this work has uncovered several critical factors that must be considered to enhance the performance of DFTB3 method. The analyses span a range of tasks, including the exploration of the potential energy surface of flexible molecules beyond those in QM7-X and DES15K datasets (*e.g.*, tyrosine, zaprinast, ligand 2Q5k⁵⁰); the prediction of energetic ranking for drug-like compounds from the Aquamarine dataset;⁵¹ the calculation of vibrational modes of α -amino acids; and the evaluation of atomic forces and interaction energies for molecular dimers from S66x8 dataset.⁵² This series of rigorous experiments was primarily designed to assess the scalability and transferability of the EquiDTB models—limitations previously identified in our earlier model, NN_{rep} .¹⁸ Our results demonstrate that equivariant NNs can considerably reduce errors in parameterizing the Δ_{TB} potential. However, we also find that achieving the most accurate Δ_{TB} potential does not necessarily lead to optimal performance across all of the diverse evaluated tasks. Notably, the best-performing EquiDTB model employs the MACE architecture and is trained on molecular conformations from QM7-X dataset. This model not only outperforms standard TB methods (DFTB3 and GFN2-xTB) but also surpasses an ML potential trained with MACE on absolute energies and atomic forces of both small single molecules



and molecular dimers. Our work thus suggests that a hybrid QM/ML approach—combining an equivariant NN with an SE method, as implemented in the EquiDTB framework—enables more reliable (bio)molecular simulations than baseline SE methods or pure ML potentials.

2 Computational methods

2.1 EquiDTB: a hybrid ML/DFTB framework

To improve the accuracy and transferability of the density functional tight-binding (DFTB) method, the standard pairwise repulsive potential (p_{wrep}) will be replaced by a ML-based many-body potential (named as Δ_{TB} potentials throughout the text) that will predict the target energies ΔE_{TB} and atomic forces $\Delta \mathbf{F}_{\text{TB}}$ such that:

$$\Delta E_{\text{TB}} = E_{\text{DFT}}^{(\text{at})} - E_{\text{DFTB}}^{(\text{el,at})} \quad \text{and} \quad \Delta \mathbf{F}_{\text{TB}} = \mathbf{F}_{\text{DFT}} - \mathbf{F}_{\text{DFTB}}^{(\text{el})}, \quad (1)$$

with $E_{\text{DFT}}^{(\text{at})}$ and \mathbf{F}_{DFT} as the DFT atomization energy and the DFT atomic forces of a given molecular system. $E_{\text{DFTB}}^{(\text{el,at})}$ and $\mathbf{F}_{\text{DFTB}}^{(\text{el})}$ are the DFTB electronic atomization energy and atomic forces, respectively. The electronic components are obtained with the third-order self-consistent charge DFTB (*i.e.*, DFTB3)^{53–55} method using 3ob parameters.^{56,57} The parameterization of Δ_{TB} potentials with ML methods represents a complex multidimensional fitting problem to reproduce DFT reference results, which renders it the most challenging step in the development of our methodology. Accordingly, building upon our previous work,¹⁸ SE(3)-equivariant neural networks (ENN) are used here to improve the performance of Δ_{TB} potentials to meet the EAST requirements. These networks ensure that quantities like energies and forces transform correctly under 3D rotations, translations, and permutations, thereby preserving physical consistency. Among all modern ENNs, the EquiDTB framework currently involves SpookyNet⁸ (SP), Allegro (AG),¹¹ and MACE (MC),⁹ but it can be extended to consider others. The three models are built on an equivariant message passing NN (MPNN); however, they differ in their approaches to represent the chemical environment and to model interactions. For instance, SpookyNet incorporates electronic degrees of freedom and nonlocal interactions through an attention mechanism within a transformer architecture. Allegro, in contrast, focuses on strictly local interactions by utilizing geometric-tensor features within a cutoff radius, effectively capturing local atomic interactions while preserving a global perspective of the system. MACE⁹ adopts a different approach by leveraging the mathematical structure of atomic cluster expansion (ACE) method, using multiplicative interactions between geometric and atomic features. This diversity in their design is crucial to determine the most reliable EquiDTB model for accurately predicting ΔE_{TB} and $\Delta \mathbf{F}_{\text{TB}}$. After constructing the Δ_{TB} potentials, the predicted ΔE_{TB} and $\Delta \mathbf{F}_{\text{TB}}$ values are added to the DFTB electronic components computed with the DFTB+ code,³³ yielding the final ML-corrected energies and forces (see Fig. 1(a)). Both contributions were integrated into a single calculator instance *via* a locally modified QM/ML calculator within the atomic simulation environment (ASE) package,⁵⁸ which was employed to perform all the computational tasks in this work.

2.2 Parameterization of equivariant Δ_{TB} potentials

One question we seek to address in this work is how the inclusion of non-covalent systems influences the development of the Δ_{TB} potential. In doing so, we have considered equilibrium and non-equilibrium conformations of both small single molecules and molecular dimers (only including C, N, O, and H atoms), which were extracted from QM7-X⁴⁶ and DES15K⁴⁷ datasets, respectively. QM7-X dataset contains property data of approximately 4 M small drug-like molecules of up to seven heavy atoms computed through PBE0 hybrid functional^{59,60} in conjunction with the tightly converged numeric atom-centered basis sets.⁶¹ Since the property data of the 7565 small molecular dimers from the DES15K dataset was calculated using a different QM method, we carried out single-point calculations at the PBE0 level to homogenize the reference data. Initially, we developed Δ_{TB} potentials using only QM7-X molecules, for which we add the label '1' after the name of the ENN, *e.g.*, SP1 for the SpookyNet model. For Δ_{TB} potentials trained on both datasets, we will add the label '2', *e.g.*, for the SpookyNet model, it would be SP2.

The comparison in performance of different equivariant NN architectures for the parameterization of Δ_{TB} potentials is not a trivial task. To conduct a fair examination, we have used the training set size of 500k conformations as it was used in our previous work where we developed the many-body repulsive potential, NN_{rep} , using SchNet architecture¹⁸ and QM7-X molecules. The validation set comprised 50k molecular structures, while the remaining structures were used as the test set. Training samples were randomly selected. Moreover, the default hyperparameters (or alternatives suggested by developers) for each ENN were considered. We only set the number of features to 128 and the cut-off radius to 5 Å. More details about the hyperparameters employed in the training procedure can be found in Section 1 of the SI.

To thoroughly evaluate the equivariant Δ_{TB} potentials, we designed a series of rigorous experiments to assess their scalability and transferability in investigating the properties of larger, more flexible systems, as well as non-covalent complexes (*vide infra*). These experiments require the inclusion of a many-body treatment of vdW/dispersion interactions (MBD) due to its relevance in describing long-range effects that are not properly captured by the PBE0 hybrid functional.^{48,49,62} To this end, the energy and forces resulting from the MBD formalism have been added to the ML-corrected DFTB energies and forces using libMBD package, which is already implemented in the DFTB+ code.³³ The results obtained using the Δ_{TB} potentials will be compared to those produced by two well-established semi-empirical methods: the DFTB3+MBD method with pairwise repulsive potentials and the GFN2-xTB, which already includes the D4⁶³ correction in its design. Additionally, to further highlight the importance of developing Δ_{TB} potentials, we compare their performance with that of a reference ML potential (referred to as rMLP potential throughout the text), which was trained using MACE architecture on absolute PBE0+MBD energies and forces for approximately 500k conformations extracted from the QM7-X and DES15K datasets.



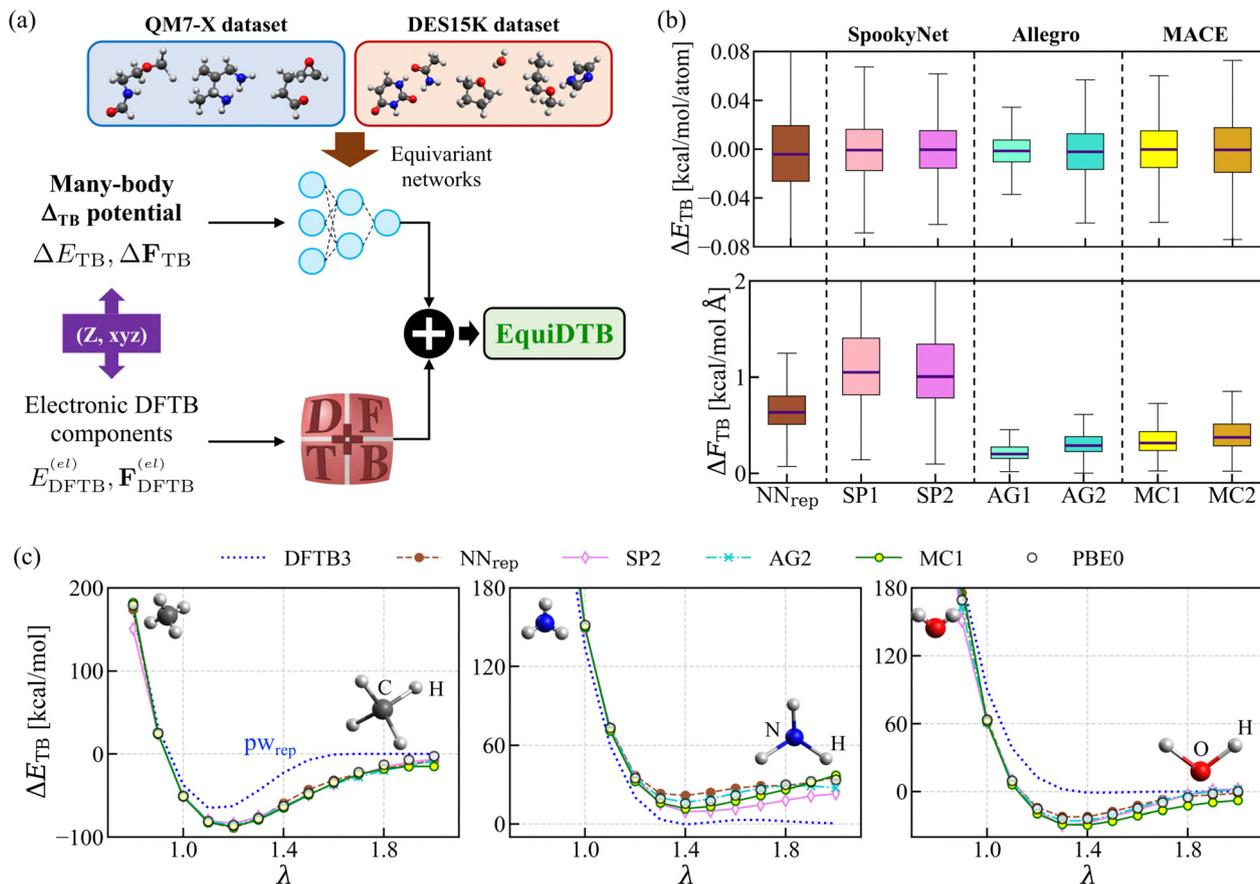


Fig. 1 (a) Schematic representation of the EquiDTB framework (which stands for “Equivariant networks for Delta Tight-Binding”) used in this work. Many-body Δ_{TB} potentials (ΔE_{TB} , $\Delta \mathbf{F}_{\text{TB}}$) are developed using SE(3)-equivariant neural networks (ENN) to replace the standard pairwise DFTB3 repulsive potentials (pW_{rep}). Quantum–mechanical datasets of small single molecules (QM7-X) and molecular dimers (DES15K) were used to train these potentials. The reference level of theory for these datasets is PBE0 hybrid functional. EquiDTB uses atomic numbers (Z) and atomic positions (xyz) as input for molecular simulations. (b) Boxplot of the error in predicting the correction for energies (ΔE_{TB}) and atomic forces ($\Delta \mathbf{F}_{\text{TB}}$) in QM7-X molecules, computed using Δ_{TB} potentials trained with the SpookyNet (SP), Allegro (AG), and MACE (MC) architectures. Results are shown for models trained on QM7-X (label ‘1’) and both datasets (label ‘2’). (c) Variation of ΔE_{TB} as a function of the relative bond length factor λ in methane, ammonia, and water molecules. The curves were computed using the PBE0 hybrid functional and selected Δ_{TB} potentials. Additionally, the results obtained with the DFTB3 pW_{rep} method are shown. In graphs (b) and (c), we also present the results obtained with our previously developed model NN_{rep}, trained using SchNet architecture and QM7-X molecules.

3 Benchmarking datasets

3.1 S66x8 molecular dimers

S66x8 dataset⁵² contains QM energetic and structural data of 66 small organic molecular dimers at eight different dimer separation distances $q \times d_{\text{eq}}$ (with d_{eq} as the equilibrium dimer distance and $q =$ values of 0.90, 0.95, 1.00, 1.05, 1.10, 1.25, 1.50, and 2.00), generating a total of 528 conformations. Since our target level of theory is PBE0, E_{int} and atomic force F_{at} values for the molecular dimers have been recalculated using PBE0 hybrid functional supplemented with a many-body dispersion (MBD) treatment, which serves as our reference data. E_{int} is here calculated using the supramolecular approach,

$$E_{\text{int}} = E_{\text{dim}} - (E_{\text{mono1}} + E_{\text{mono2}}), \quad (2)$$

where E_{dim} and E_{mono1} (E_{mono2}) are the total energies of the dimer configuration and the monomer 1 (monomer 2), respectively. To better understand the results for molecular dimers,

error values were computed for the entire dataset and the four dimer groups, categorized based on their dominant non-covalent interaction, *i.e.*, H-bond (184 confs.), π - π (80 confs.), London (104 confs.), and Mixed (160 confs.).

3.2 Non-equilibrium flexible molecules

We have considered the rotational energy profile of paracetamol ($\text{C}_8\text{H}_9\text{NO}_2$) and tyrosine ($\text{C}_9\text{H}_9\text{NO}_3$) to evaluate the performance of Δ_{TB} potentials in predicting properties of molecules beyond the scope of the training set (see Fig. 4(a)). These systems are good candidates for analyzing transferability in chemical space and scalability with system size, as they exhibit greater flexibility and complexity compared to QM7-X molecules. Paracetamol and tyrosine contain 11 and 13 heavy atoms, respectively, and their molecular motifs are absent in the training set. To build the reference data for rotational profiles, Nudged Elastic Band (NEB) calculations at the PBE0+MBD level



were carried out to determine the minimum energy path for the rotation of the dihedral connecting the aromatic ring and the linear-type structure on the molecules, *i.e.*, C–C–N–C and C–C–C–C dihedrals for paracetamol and tyrosine, respectively. These calculations were performed using FHI-aims calculator implemented in the ASE package. The relative energy of the rotated structures with respect to the initial structure is given by $\Delta E_{\theta_i} = E_{\theta_i} - E_{\theta_0}$, with E_{θ_i} as the energy of the structure at the i th NEB interpolation step corresponding to a dihedral angle θ .

The capabilities of the Δ_{TB} potentials are also examined by exploring the potential energy surface (PES) of larger molecules with biological relevance. In doing so, we performed molecular dynamics (MD) simulations at constant temperature for three molecules of increasing size and flexibility, namely alanine dipeptide (22 atoms), zaprinast (33 atoms), and ligand 2Q5k (94 atoms) (see Fig. 4(b)). The MD simulations were conducted at 300 K for 100 ps using the Langevin thermostat, as implemented in the ASE package. The simulation timestep was set to 0.5 fs, with a friction coefficient of 2×10^{-3} . All simulations using the DFTB method were supplemented with an MBD treatment. The MD simulations with GFN2-xTB were also performed at 300 K but using a Berendsen thermostat with a timestep of 0.5 fs, which is available in the current version of xTB code. Before starting the MD run, all geometries were initially optimized with the corresponding approach. The optimized geometries obtained with EquiDTB model are shown in Fig. 4(b). To better understand the efficiency of each approach, we analyzed the atomic forces F_{at} and structures of 200 randomly selected conformations extracted from the last 40 ps of each MD trajectory. Accordingly, single-point calculations at the PBE0+MBD level were carried out using FHI-aims code for each conformation to obtain the reference atomic force data.

We further assessed the performance of our ML-corrected DFTB method in predicting vibrational modes of biomolecular building blocks, specifically α -amino acids. A total of 18 neutral amino acids composed of C, N, O, and H atoms were considered (see Table S5 in the SI for the complete list). All molecular geometries were first optimized using the respective method, followed by vibrational mode calculations using the Vibrations module from the ASE package. Notice that, since we are combining the DFTB calculator with the corresponding ML calculator, the Hessians are obtained numerically. For GFN2-xTB, we employed the corresponding routine provided in the current version of the xTB code. As reference data, vibrational frequencies and geometries were obtained at the PBE0+MBD level using the FHI-aims calculator within ASE.

3.3 Energetic ranking of drug-like molecules

It is well known that certain drug-like molecules can alter their biological properties by undergoing conformational changes. To address this task, we have considered a subset of large drug-like molecules from the Aquamarine (AQM) dataset,⁵¹ containing only C, N, O, and H atom types. Energies and atomic forces of each molecule in AQM dataset were computed at the PBE0+MBD level. This subset includes 63 distinct molecules and their respective conformers, resulting in a total of 2486 equilibrium structures, with the number of atoms N ranging from 50 to 90.

Additionally, to verify the efficiency of predicting energies and atomic forces of non-equilibrium conformations of large and diverse molecules, we explored the PES of each equilibrium conformation by performing MD simulations at different temperatures using DFTB3+MBD level (see schematic in Fig. 5(a)). This level of theory was also used to generate the structures in AQM dataset. Then, the temperature of the Nose–Hoover thermostat was increased from 100 K to 1500 K in increments of 100 K every 250 fs using the DFTB+ calculator in the ASE package. The timestep was set to 0.25 fs. From each simulation, we considered 30 non-equilibrium conformations selected at different temperatures. Subsequently, single-point calculations at the PBE0+MBD level were performed for each of these conformations, and the QM property data for the successfully converged cases were stored. This step generated the AQM-X dataset, which includes a total of 71 783 equilibrium and non-equilibrium conformations across 63 distinct large drug-like molecules.

4 Results and discussion

4.1 Examination of the Δ_{TB} potentials

We first assess the performance of the Δ_{TB} potentials in predicting the corrections in energy ΔE_{TB} and atomic force ΔF_{TB} of the equilibrium and non-equilibrium conformations of both QM7-X single molecules and DES15K molecular dimers. Fig. 1(b) depicts the boxplots of the errors for ΔE_{TB} and ΔF_{TB} computed using the Δ_{TB} potentials trained with SpookyNet (SP), MACE (MC), and Allegro (AG) and considering only QM7-X (label '1') and both datasets (label '2'). For comparison, the results obtained with our previously developed NN_{rep} model are also presented. One can see that the inclusion of equivariant networks for predicting properties of QM7-X molecules-independent of the training datasets-results in a median of the error distribution of ΔE_{TB} that is closer to zero, along with a reduced data spread. This finding is verified by calculating the mean absolute errors (MAE), see Table 1. For the prediction of

Table 1 Performance of Δ_{TB} potentials in predicting the corrections in energies ΔE_{TB} and in atomic forces ΔF_{TB} for molecules from QM7-X and DES15K datasets. We report the mean absolute errors for our previously developed model NN_{rep},¹⁸ trained using SchNet (SN), as well as for the new EquiDTB models trained using equivariant neural networks (NNs) such as SpookyNet (SP), Allegro (AG), and MACE (MC). The number following the name of each NN architecture indicates whether the model was trained on QM7-X only (label '1') or on both datasets (label '2'). Errors for energies and atomic forces are given in kcal per mol per atom and kcal mol⁻¹ Å⁻¹, respectively

Model	Δ_{TB} potential	QM7-X		DES15K	
		ΔE_{TB}	ΔF_{TB}	ΔE_{TB}	ΔF_{TB}
NN _{rep}	SN1	0.031	0.70	0.874	5.76
EquiDTB	SP1	0.025	1.24	0.863	6.78
	AG1	0.013	0.24	0.863	5.31
	MC1	0.023	0.38	0.852	5.12
	SP2	0.022	1.17	0.017	6.12
	AG2	0.020	0.34	0.094	0.71
	MC2	0.028	0.44	0.198	2.18



ΔF_{TB} , the Allegro and MACE models displayed the best performance on QM7-X molecules, with AG1 and AG2 yielding the lowest MAE values. Interestingly, the models trained with these equivariant NNs on both reference datasets exhibit higher MAE values than those trained solely on QM7-X. Moreover, although training on both datasets reduces the errors in predicting ΔE_{TB} and ΔF_{TB} for small molecular dimers (see Table 1), the models still present only moderate performance, as evidenced by large MAE values and considerable data spread. It is worth noting that the Allegro models are the most accurate on both datasets, a result that can be attributed to the strict locality feature in their design. These results highlight the challenge that equivariant NNs pose for simultaneously learning ΔE_{TB} and ΔF_{TB} of covalently- and non-covalently bonded systems. Accordingly, we have selected SP2, AG2, and MC1 models as representative equivariant Δ_{TB} potentials for further investigation based on an exhaustive analysis described in Sections S1 and S2 of the SI.

To evaluate the efficiency of these Δ_{TB} potentials in capturing physical effects beyond the standard pairwise repulsive potential of the DFTB method (pw_{rep}), we investigated the variation of ΔE_{TB} as a function of the bond length for key small molecules, such as CH_4 , NH_3 , and H_2O (see Fig. 1(c)). In doing so, we have compressed and elongated the covalent bond in these systems by multiplying the equilibrium bond length by the factor λ . When contrasting the target energies, ΔE_{TB} , required to reach the DFT-PBE0 level of theory (black circles) with the DFTB3 repulsive energies (blue dotted line), it becomes evident that the standard pw_{rep} fails to adequately

capture the electronic effects present in these simple molecular systems. On the contrary, the Δ_{TB} potentials, which account for many-body interactions by design, yield results closer to the target energies, regardless of the NN architecture used for training the potential. AG2 model demonstrated the best overall performance in estimating the bond length dependence of ΔE_{TB} for the three molecules, with a mean absolute relative error (MARE) of 8.3%, followed by NN_{rep} model with 10.9%. This analysis provides additional compelling evidence that many-body potentials are better suited to characterize electronic quantum effects that are not fully captured by the DFTB electronic Hamiltonian, with the equivariant network serving as a key component to further enhance their performance. This, in turn, may have crucial implications for their transferability and scalability (*vide infra*).

4.2 Properties of non-covalent systems

Next, we examine the generalizability of the Δ_{TB} potentials by calculating the interaction energy E_{int} of equilibrium and non-equilibrium small molecular dimers from the S66x8 dataset,⁵² which includes C, H, N, and O atoms. Fig. 2(a) shows the error in the prediction of E_{int} for the different ML approaches investigated in this work. For comparison, the results obtained with widely used tight-binding (TB) methods (DFTB3 and GFN2-xTB) are also presented. We have split the analysis into compressed ($q \leq 1$) and elongated ($q > 1$) dimers to analyze their performance in capturing a diverse range of non-covalent interactions. Overall, the equivariant Δ_{TB} potentials perform

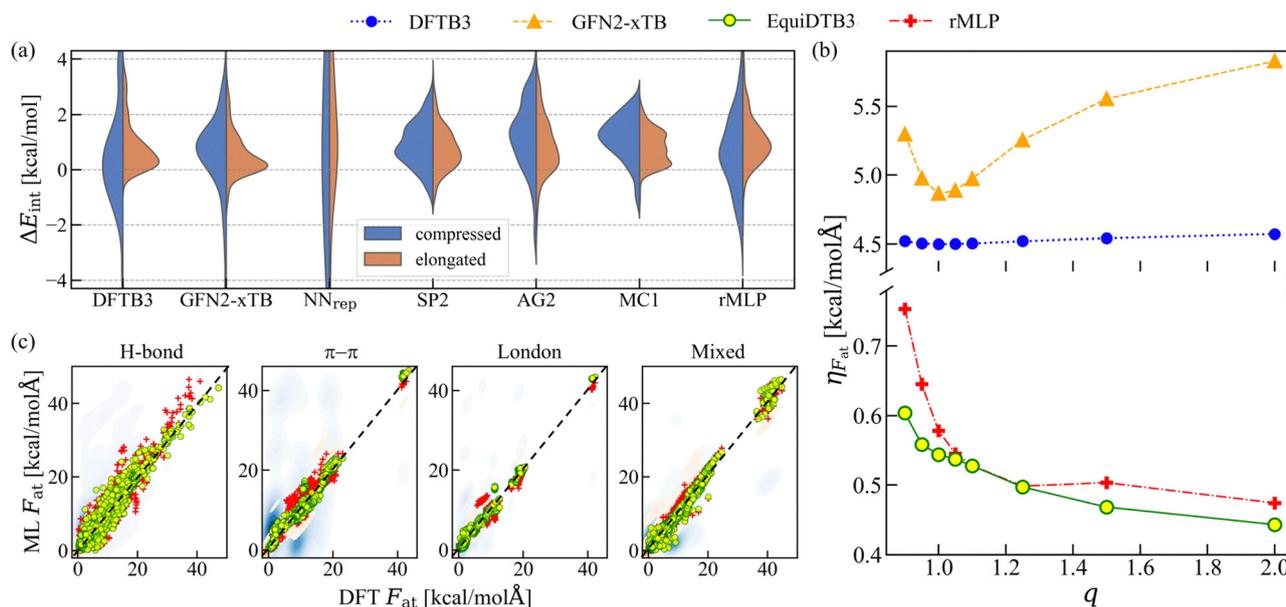


Fig. 2 Benchmarking Δ_{TB} potentials for calculating interaction energies (E_{int}) and atomic forces (F_{at}) in small molecular dimers from the S66x8 dataset. (a) Distributions of the error in computing E_{int} values for compressed ($q \leq 1.0$) and elongated ($q > 1.0$) molecular dimers. We show the distributions for widely used TB methods (DFTB3 and GFN2-xTB), NN_{rep} model, selected equivariant Δ_{TB} potentials (SP2, AG2, MC1), and the reference ML potential (rMLP). (b) Variation of $\eta_{F_{\text{at}}}$ (see eqn (3)) as a function of the relative distance between monomers (q). (c) Correlation plots between DFT and ML F_{at} values for each dimer group. Graphs (b) and (c) primarily show the results for the best-performing EquiDTB3 model (*i.e.*, MC1 model) and the rMLP potential. For comparison, we also include the corresponding values obtained with DFTB3 and GFN2-xTB. Reference values for E_{int} and F_{at} were calculated using PBE0+MBD. All calculations include a many-body dispersion treatment, except for xTB, which considers D4 correction.



better than the NN_{rep} model for both compressed and elongated dimers. In particular, MC1 model produces narrower error distributions than those obtained by standard TB methods, SP2 and AG2 models, and the reference ML potential (rMLP). Fig. 3 compares MAE values in the prediction of E_{int} and atomic forces, F_{at} , for the S66x8 molecular dimers. The symbol size encodes the MAE in the prediction of F_{at} for the QM7-X dataset. The resulting Pareto front highlights the trade-offs among the three error metrics and shows that the MC1 model achieves the most balanced overall performance. Accordingly, the MC1 model is selected for further validation and will hereafter be referred to as the EquiDTB3 model (where '3' denotes the use of DFTB3 electronic parameters). When analyzing the mean error values across the entire dataset, listed in Table 2, one can observe a slightly superior performance of GFN2-xTB (MAE = 0.86 kcal mol⁻¹) compared to EquiDTB3 model (MAE = 0.97 kcal mol⁻¹). A consequence of the more accurate description of the global structural and energetic features of London and mixed dimers when using GFN2-xTB, see Fig. S1 of the SI. These results also show that EquiDTB3 model performs better than rMLP potential in predicting non-covalent properties, demonstrating that combining the electronic DFTB Hamiltonian with an equivariant Δ_{TB} potential offers better generalizability than a ML potential trained on absolute total energies and atomic forces of both single molecules and molecular dimers.

To understand the applicability of the developed Δ_{TB} potentials in (bio)molecular simulations, we have analyzed the

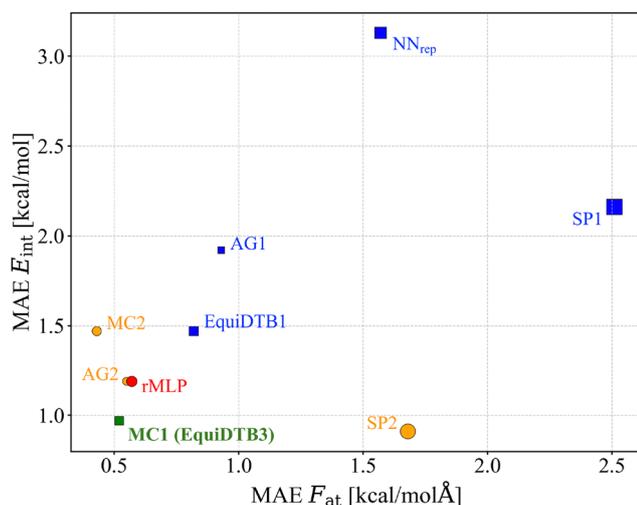


Fig. 3 Evaluation of the performance of the Δ_{TB} potentials within the EquiDTB framework across the QM7-X and S66x8 datasets. Here, we present the mean absolute errors (MAEs) for predicting interaction energies (E_{int}) and atomic forces (F_{at}) of S66x8 molecular dimers. The symbol size reflects the MAE values for the prediction of atomic forces of QM7-X molecules. Models trained using only the QM7-X dataset are represented by squares, while models trained on both the QM7-X and DES15K datasets are represented by circles. For comparison, results from the EquiDTB1 model and the reference ML potential (rMLP) are also shown. The model with the best trade-off among these three metrics is MC1 (green square), which will be referred to as the EquiDTB3 model throughout the remainder of this work.

accuracy in computing atomic forces F_{at} of molecular dimers at different relative distances between monomers, q , by defining the parameter η as,

$$\eta_{F_{\text{at}}} = \frac{1}{66} \sum_{i=1}^{66} \left(\frac{1}{N_i} \sum_{j=1}^{N_i} |F_j^{\text{X}} - F_j^{\text{DFT}}| \right), \quad (3)$$

with N_i as the number of atoms in a given dimer i . F_j^{DFT} and F_j^{X} are the atomic forces computed using the DFT reference method and the X = DFTB3, GFN2-xTB, EquiDTB3, and rMLP models. Interestingly, the atomic forces obtained using DFTB3 and GFN2-xTB deviate significantly from the reference values obtained at the PBE0+MBD level, regardless of the q value (see Fig. 2(b)). On the contrary, EquiDTB3 and rMLP show close agreement with the reference data, with MAE values for force prediction across the entire dataset of 0.52 and 0.57 kcal mol⁻¹ Å⁻¹, respectively. The error in the forces for these models also decreases as a function of q , highlighting their higher accuracy in investigating the properties of single molecules compared to non-covalent systems. By performing the analysis per dimer group (see Fig. 2(c) and Table 2), we found that EquiDTB3 outperforms all other models for all dimer groups with the exception of the H-bond dimers, where rMLP exhibits slightly better performance with an MAE of 0.65 kcal mol⁻¹ Å⁻¹. Although the AG2 model exhibited the lowest errors in predicting F_{at} values for both training datasets (*i.e.*, QM7-X and DES15K), the selected EquiDTB3 model outperforms it when computing F_{at} values for equilibrium and non-equilibrium S66x8 dimers (see Fig. S2 and Table S4 of the SI). Among the different cases in which EquiDTB3 has improved on the DFTB accuracy, molecular dimers involving ethyne C₂H₂ (which contains C atoms with $F_{\text{at}} > 35$ kcal mol⁻¹ Å⁻¹ in π - π , London, and mixed dimers) are the most representative. This is because standard TB methods cannot properly describe their triple-bond electronic configuration (see ethyne-based systems in Fig. S3 of the SI). This analysis confirms the robustness of the EquiDTB3 models in computing local features that influence the overall structure of non-covalent systems unseen by the model.

Table 2 Performance of Δ_{TB} potentials in predicting the interaction energies E_{int} and atomic forces F_{at} for equilibrium and non-equilibrium small molecular dimers from S66x8 dataset. We show the mean absolute errors (MAE) mean absolute relative error (MARE) for the previously developed NN_{rep} ¹⁸ model and the best-performing EquiDTB3 model. For comparison, the error values for EquiDTB1 model, widely used TB methods (DFTB3 and GFN2-xTB), and the reference ML potential (rMLP) are also presented. All calculations include a many-body dispersion treatment, except for xTB, which considers D4 correction. The error values for energies and forces are given in kcal mol⁻¹ and kcal mol⁻¹ Å⁻¹, respectively

Model	E_{int}		F_{at}				
	MAE	MARE	H-Bond	π - π	London	Mixed	Total
DFTB3	1.04	38.0	5.06	6.32	3.08	3.93	4.52
GFN2-xTB	0.86	36.0	5.46	6.22	2.37	6.25	5.20
NN_{rep}	3.13	118.4	2.00	1.88	1.14	1.21	1.57
EquiDTB3	0.97	33.9	0.70	0.59	0.26	0.46	0.52
EquiDTB1	1.47	43.4	0.95	1.43	0.52	0.55	0.82
rMLP	1.19	48.0	0.65	0.75	0.30	0.55	0.57



4.3 Physical complexity of Δ_{TB} potentials

In the previous sections, we discussed EquiDTB models trained using DFTB3 electronic components (*i.e.*, EquiDTB3). This corresponds to a Taylor expansion of the density functional around the reference density ρ_0 truncated after the third-order term.⁶⁴ Accordingly, the DFTB total energy can be expressed as:

$$E_{\text{DFTB3}} = \sum_{iab} \sum_{\mu \in a} \sum_{\nu \in b} n_i c_{\mu i} c_{\nu i} H_{\mu\nu}^0 - \frac{1}{2} \sum_{ab} \Delta q_a \Delta q_b \gamma_{ab} + \frac{1}{3} \sum_{ab} \Delta q_a^2 \Delta q_b \Gamma_{ab} + E_{\text{rep}}^3 \quad (4)$$

where $\Delta q_x = q_x - q_x^0$ is the net charge fluctuation on atom $x = a$ or b (often Mulliken charges), with q_x^0 being the charge of the valence electrons in the neutral atom and q_x the charge of that atom within the molecule. γ represents a function accounting for electron–electron interactions, while Γ describes the variation of the Hubbard parameters with respect to charge fluctuations.⁶⁴ E_{rep}^3 denotes the repulsive energy associated with the DFTB3 electronic components. The DFTB3 approach is a more accurate method developed to overcome the limitations of lower-order DFTB formulations, particularly when applied to anionic and cationic systems. In this context, the EquiDTB framework can also be employed to improve lower-order DFTB electronic components, such as the first-order DFTB (known as DFTB1). DFTB1 is a non-self-consistent (non-SCC) approximation that neglects the second- and higher-order terms in the density fluctuations. Consequently, the generalized eigenvalue problem needs to be diagonalized only once, and the DFTB total energy can be expressed as:

$$E_{\text{DFTB1}} = \sum_{iab} \sum_{\mu \in a} \sum_{\nu \in b} n_i c_{\mu i} c_{\nu i} H_{\mu\nu}^0 + E_{\text{rep}}^1 \quad (5)$$

The first term represents the energy contribution from the tight-binding Hamiltonian and depends only on the reference density. E_{rep}^1 denotes the repulsive energy associated with the DFTB1 electronic components, which by definition differs from E_{rep}^3 .

A new DFTB dataset was generated by calculating electronic energies and forces of the QM7-X molecules using the DFTB1 method. The corresponding ΔE_{TB1} and $\Delta \mathbf{F}_{\text{TB1}}$ values were then obtained according to eqn (1), and, as expected, ΔE_{TB1} differ from the energy correction derived from DFTB3 components (see Fig. S4 of the SI). However, the similarity between the distributions of $\Delta \mathbf{F}_{\text{TB1}}$ and $\Delta \mathbf{F}_{\text{TB3}}$ already indicates that the Δ_{TB} potential corresponding to DFTB1 data will exhibit a different training complexity. The reference atomic energies were also computed with DFTB1 to ensure internal consistency. Using these data, we then trained the EquiDTB1 model (where ‘1’ denotes the use of DFTB1 electronic parameters) with MACE, considering only QM7-X molecules as discussed in Sections 4.1 and 4.2. The MAEs for E_{TB1} and $\Delta \mathbf{F}_{\text{TB1}}$ on small single molecules are 0.026 kcal per mol per atom and 0.44 kcal mol⁻¹ Å⁻¹, respectively. These errors are only slightly higher than those obtained with the EquiDTB3 model (see Table 1), confirming the good accuracy of EquiDTB1. In contrast, significantly larger errors are observed for molecular dimers in the DES15K dataset

(29.98 kcal per mol per atom for energies and 10.21 kcal mol⁻¹ Å⁻¹ for forces), and similar trends appear for the S66x8 benchmark (see Table 2). Note that, despite the increased errors, EquiDTB1 still outperforms standard TB methods and our previously developed NN_{rep} model in force predictions. These results indicate that while EquiDTB1 performs comparably to EquiDTB3 for single molecules, its accuracy deteriorates when investigating equilibrium and non-equilibrium non-covalent systems where electronic effects can be more complex and challenging to capture. In brief, this demonstrates that including second- and third-order DFTB electronic terms helps capture essential physical and chemical effects, thereby simplifying the learning of (ΔE_{TB} , $\Delta \mathbf{F}_{\text{TB}}$) and improving overall model accuracy. Additional validation benchmarks presented in the following sections further strengthen this outcome.

4.4 Exploring potential energy surfaces of flexible molecules

We now evaluate the efficiency of the EquiDTB approach in predicting the rotational energy profiles ΔE_θ of two molecules beyond the scope of the training set: paracetamol (C₈H₉NO₂) and tyrosine (C₉H₁₁NO₃). The atomic configurations of selected cases are embedded in the graphs in Fig. 4(a). In the case of paracetamol, the rotational energy profile is qualitatively well described by all approaches, *i.e.*, the largest ΔE_θ is located around $\theta = 90^\circ$. However, some discrepancies exist in defining this barrier height, with $\Delta E_{90^\circ} = 7.08$ kcal mol⁻¹ from the DFTB3 method and $\Delta E_{90^\circ} = 3.77$ kcal mol⁻¹ from the EquiDTB3 model being the farthest and the closest to the PBE0+MBD reference value (3.4 kcal mol⁻¹), respectively. This is also reflected in the evaluation of the Pearson correlation coefficient (ρ) and the MAE for ΔE_θ , see Table 3. Overall, ρ values are close to 1.0, indicating an almost perfect linear correlation between predicted and reference ΔE_θ values. The key difference lies in the MAE values, where the EquiDTB3 model outperforms the other approaches with an error of 0.38 kcal mol⁻¹, followed by EquiDTB1 model with an error of 0.42 kcal mol⁻¹. Similarly, ρ values are close to 1.0 for predicting the rotational energy profile of a more flexible molecule like tyrosine (compared to paracetamol, see additional carboxyl group in its linear molecular building block), with the best performance achieved by the EquiDTB3 model and the rMLP potential, yielding MAE values of 0.23 kcal mol⁻¹ and 0.24 kcal mol⁻¹, respectively. Notice that for both molecules, GFN2-xTB exhibits higher accuracy than DFTB3, which uses pairwise repulsive potentials. It is worth mentioning that NN_{rep} model has a moderate prediction for ΔE_θ values of paracetamol with an MAE of 0.83 kcal mol⁻¹, however, it predicts poorly the rotational energy profile for tyrosine with $\rho = 0.64$ and an MAE of 1.26 kcal mol⁻¹ (see Fig. S5 of the SI).

The capabilities of the EquiDTB models are then examined by exploring the PES of larger molecules with biological relevance, namely alanine dipeptide (22 atoms), zaprinast (33 atoms), and ligand 2Q5k (94 atoms). The analysis of the time evolution of the total energy of the studied molecules shows that the MD trajectories at 300 K are stable for all approaches, with no abrupt energy change due to bond breaking or formation (see Fig. S6 of SI). Based on the correlation plots between true and predicted F_{at}



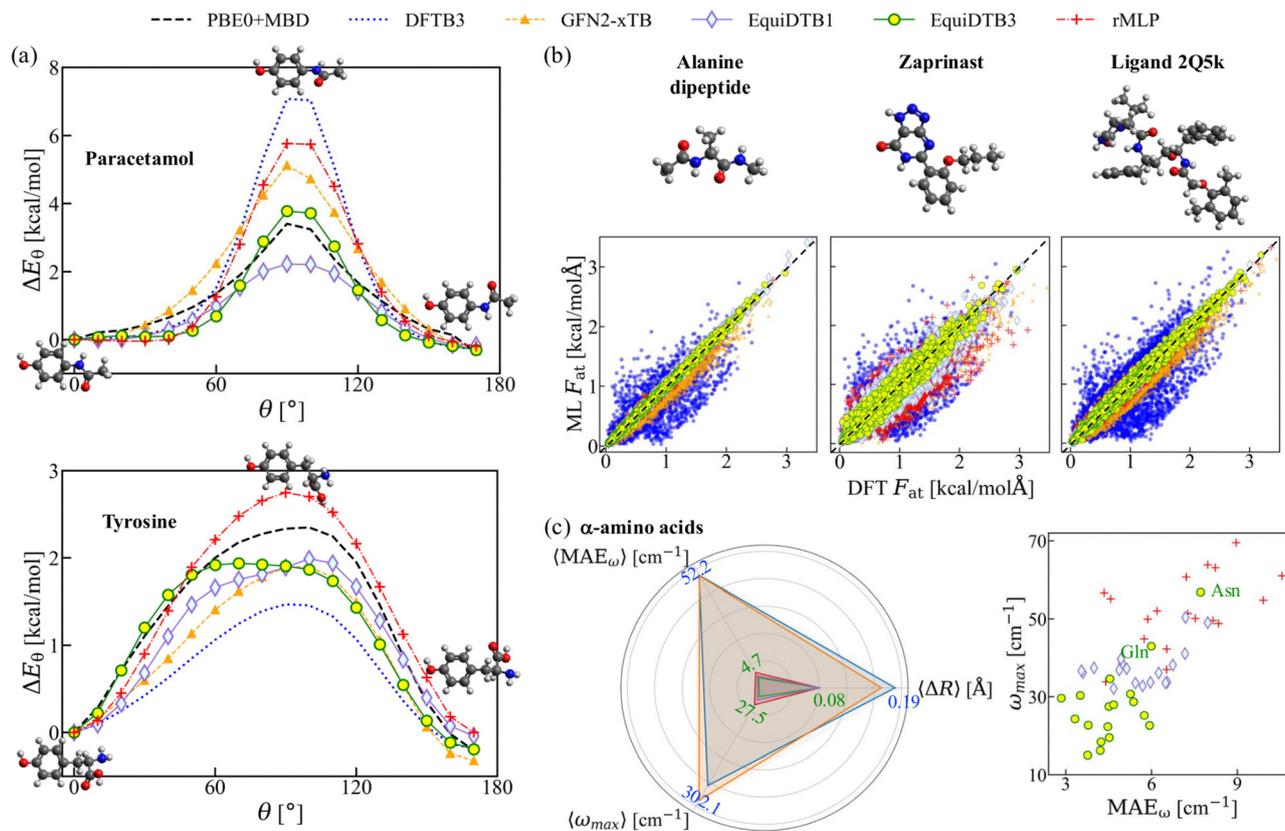


Fig. 4 Assessing predictions of non-equilibrium properties of flexible molecules. (a) Minimum energy path for the rotation of the dihedral connecting the aromatic ring and the linear-type structure in paracetamol and tyrosine. The rotational profiles were computed performing Nudged Elastic Band (NEB) calculations. (b) Analysis of atomic forces F_{at} during an MD trajectory of 100 ps at 300 K for alanine dipeptide, zaprinast, and ligand 2Q5k. We present correlation plots between DFT and ML values of F_{at} for 200 randomly selected conformations sampled during the last 40 ps of the trajectory. (c) Polar plot of the average mean absolute error of the vibrational frequencies ($\langle MAE_{\omega} \rangle$), the average maximum deviation in frequency values ($\langle \omega_{max} \rangle$), and the average root-mean-squared deviation of the optimized structures ($\langle \Delta R \rangle$). These averages were calculated over 18 α -amino acids. We also show the correlation plot between ω_{max} and MAE_{ω} values for these amino acids. For EquiDTB3 model, we highlight the two cases with the largest corresponding values, *i.e.*, Asparagine (Asn) and Glutamine (Gln). In all graphs, we present the results obtained by the widely used TB methods (DFTB3 and GFN2-xTB), EquiDTB1 model, EquiDTB3 model, and rMLP potential. Reference values for energies, forces, and vibrational frequencies were calculated using PBE0+MBD. All calculations include a many-body dispersion treatment, except for xTB, which considers D4 correction.

Table 3 Performance of EquiDTB models in computing the relative energies (ΔE_{θ}) along the minimum energy path between isomers, the atomic forces (F_{at}) of conformations extracted from MD trajectories, and the vibrational modes (ω) of unseen flexible molecules. We show the Pearson correlation factor (ρ) and mean absolute error (MAE) for ΔE_{θ} and F_{at} obtained by the widely used TB methods (DFTB3 and GFN2-xTB), the EquiDTB1 and EquiDTB3 models, and the rMLP potential. For the vibrational analysis, we report the average values of MAE for ω predictions, the maximum deviation ω_{max} , and the root-mean-square deviation (ΔR) between optimized geometries, all quantities computed per α -amino acid. Reference values for relative energies, atomic forces, and vibrational frequencies were calculated using PBE0+MBD. All calculations include a many-body dispersion treatment, except for xTB, which considers D4 correction. The error values for energies and forces are given in kcal mol⁻¹ and kcal Å⁻¹ mol⁻¹, respectively. Frequency errors are expressed in cm⁻¹, while ΔR is given in Å

	System	Metric	DFTB3	GFN2-xTB	EquiDTB3	EquiDTB1	rMLP
Rotational profile (ΔE_{θ})	Paracetamol	ρ	0.98	0.99	0.97	0.98	0.97
		MAE	1.05	0.67	0.38	0.42	0.84
	Tyrosine	ρ	0.99	0.99	0.97	0.99	0.98
		MAE	0.64	0.41	0.23	0.28	0.24
MD trajectory (F_{at})	Alanine dipeptide	ρ	0.81	0.98	1.00	1.00	1.00
		MAE	4.00	1.98	0.31	0.35	0.43
	Zaprinast	ρ	0.76	0.94	0.98	0.97	0.93
		MAE	4.71	3.15	1.06	1.53	2.16
	Ligand 2Q5k	ρ	0.87	0.98	1.00	1.00	0.99
		MAE	3.24	2.00	0.43	0.41	0.63
Vibrational modes (ω)	α -Amino acids	$\langle MAE_{\omega} \rangle$	52.2	51.9	4.7	5.6	7.1
		$\langle \omega_{max} \rangle$	302.1	348.2	27.5	37.6	52.5
		$\langle \Delta R \rangle$	0.19	0.17	0.08	0.08	0.08



for the 200 selected conformations (see Fig. 4(b)), it can be concluded that EquiDTB3 model provides a more accurate description of the PES for these molecules. EquiDTB1 model also performs well, showing a level of accuracy comparable to that of EquiDTB3. As expected, the rMLP potential demonstrates moderate performance, which is still superior to that of standard TB-based methods. Similar to the analysis performed for ΔE_0 , these observations are quantified by computing ρ and MAE for F_{at} predictions (values listed in Table 3). Specifically, we obtained an average ρ of ≈ 0.99 over the three molecules for EquiDTB1 and EquiDTB3 models, followed by ≈ 0.97 for rMLP potential and GFN2-xTB. However, analysis of the MAE values reveals a clear performance difference between the EquiDTB models and rMLP, with EquiDTB3 exhibiting the lowest MAE values for alanine dipeptide and zaprinast, while EquiDTB1 performs slightly better for ligand 2Q5k. These findings highlight the superior scalability and transferability of the ML-corrected DFTB method compared to an ML potential trained on absolute energies and forces of molecules from the QM7-X and DES15K datasets. Among the studied molecules, zaprinast is the most challenging system for the DFTB method, primarily due to the presence of the aromatic heterocycle triazole. DFTB is known to have certain shortcomings in describing aromaticity and delocalized π -systems, especially in molecules containing N#N interactions. Accordingly, we have demonstrated that replacing the pairwise repulsive potential with an equivariant many-body Δ_{TB} potential can considerably improve the description of these electronic configurations, as reflected in the reduction of MAE values from $4.71 \text{ kcal } \text{\AA}^{-1} \text{ mol}^{-1}$ with DFTB3 to $1.06 \text{ kcal } \text{\AA}^{-1} \text{ mol}^{-1}$ with the EquiDTB3 model. Additionally, the triazole ring in zaprinast was found to break apart after 80 ps of MD simulation at 600 K using DFTB3 (see Fig. S7 of the SI), whereas Δ_{TB} potentials help preserve the structure for longer simulation time—another compelling example of the robustness of our methodology.

We now evaluate the structural evolution of these molecules during the MD trajectory. To this end, the radius of gyration R_g of a molecular structure was compared to its root-mean-squared deviation ΔR every 10 fs (see Fig. S8 of the SI). ΔR values were computed using the initial optimized structure as a reference for each approach. For alanine dipeptide, we found that EquiDTB3 and rMLP potential drive the molecule to visit two well-defined structural states: one where the atomic structure is more extended and another where it is compressed. On the other hand, for more complex molecules, the difference between the conformational states gets more pronounced. In particular, for the branched ligand 2Q5k, the MD simulations for both models start from very similar conformations, with an initial $\Delta R = 1.0 \text{ \AA}$ between the optimized structures. Nevertheless, their structural evolution differs, reaching a $\Delta R \approx 3.4 \text{ \AA}$ between the corresponding structures at 100 ps. Moreover, the conformations for the EquiDTB3 model exhibit smaller fluctuations in both R_g and ΔR .

Besides the improved performance in relative energies and atomic forces, we demonstrate that the EquiDTB models can accurately predict the vibrational modes of relevant biomolecular building blocks. In Fig. 4(c), a polar plot shows the average values of the MAE and the maximum deviation for the vibrational frequency (ω) prediction of 18 α -amino acids. The average ΔR values associated with the optimized structures are also displayed. Our results show that both EquiDTB models and rMLP potential outperform the TB methods, and yield highly similar optimized structures of the amino acids, with $\langle \Delta R \rangle = 0.08 \text{ \AA}$. However, EquiDTB3 achieves the lowest errors in computing ω values (see Table 3), with $\langle \text{MAE}_\omega \rangle = 4.7 \text{ cm}^{-1}$ and $\langle \omega_{\text{max}} \rangle = 27.5 \text{ cm}^{-1}$, followed by EquiDTB1 with $\langle \text{MAE}_\omega \rangle = 5.6 \text{ cm}^{-1}$ and $\langle \omega_{\text{max}} \rangle = 37.6 \text{ cm}^{-1}$. Moreover, by evaluating individual cases, we found that 16 out of 18 α -amino acids exhibit $\text{MAE}_\omega < 6.0 \text{ cm}^{-1}$ and $\omega_{\text{max}} < 35.0 \text{ cm}^{-1}$;

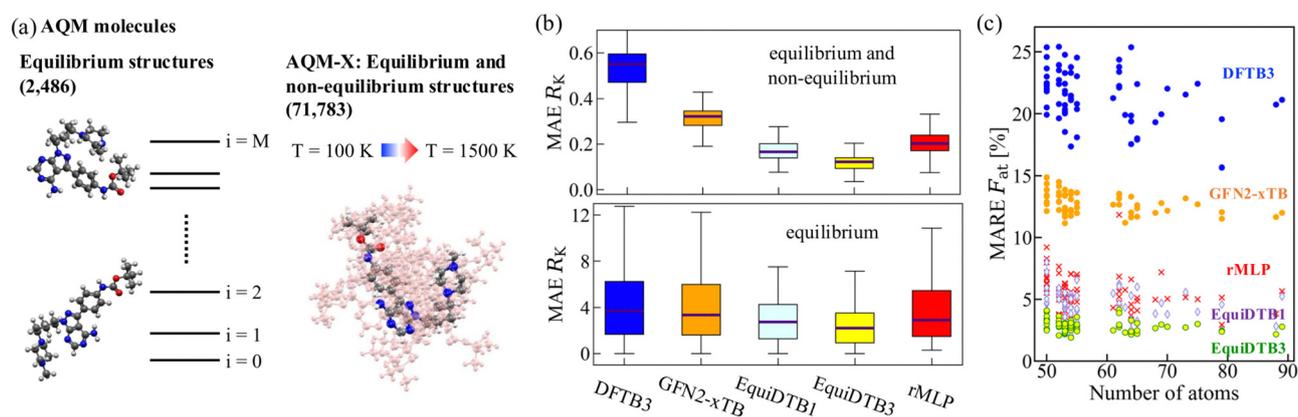


Fig. 5 (a) Schematic representation of the energetic ranking for molecular conformations of equilibrium drug-like molecules and the non-equilibrium conformations generated using molecular dynamics simulations. For each equilibrium conformation, the temperature was increased from 0 to 1500 K and, subsequently, 30 non-equilibrium conformations were selected for the ranking analysis. (b) Boxplot of the mean absolute errors (MAE) of the energetic ranking R_K for subsets considering only equilibrium (lower panel) and both equilibrium and non-equilibrium conformations (upper panel) of large drug-like molecules. (c) Mean relative error in atomic forces ΔF_{at} for equilibrium and non-equilibrium conformations per unique molecule as a function of the number of constituent atoms. For graphs (b) and (c), we show the results obtained using widely used TB models (DFTB3 and GFN2-xTB), the EquiDTB1 and EquiDTB3 models, and the rMLP potential. Reference values for energetic ranking and atomic forces were calculated using PBE0+MBD. All calculations include a many-body dispersion treatment, except for xTB, which considers D4 correction.



only Glutamine (Gln) and Asparagine (Asn) fall outside this range (see Fig. 4(c)). These results highlight the relevance of computing non-equilibrium properties of unseen molecules to better assess the capabilities of an ML model, rather than relying solely on standard error metrics such as MAE of energies and forces.

4.5 Energetic ranking of drug-like molecular conformers

Another important application of our ML-corrected DFTB method, besides molecular simulations, is the calculation of energetic rankings R_K for conformations of large drug-like molecules. Fig. 5(b) shows the boxplots of the MAE values for each distinct drug-like molecule for the change in energetic ranking ΔR_K : in the upper panel considering both equilibrium and non-equilibrium conformations and in the lower panel only the equilibrium ones. The analysis of ΔR_K for equilibrium conformations revealed that the EquiDTB3 model resolves energy values of the discrete spectrum of molecules more accurately, producing a smaller spread of data points and an average MAE of 2.7, followed by EquiDTB1 and rMLP potential with an average MAE of 3.3 and 4.0, respectively (see Table S6 of the SI). Moreover, the average of the maximum deviation in the energetic ranking, $(\Delta R_K)_{\max}$, for the EquiDTB3 model is 10.4, which is considerably smaller than the corresponding values for DFTB3 and GFN2-xTB. By incorporating non-equilibrium conformations into the evaluation of ΔR_K , we obtain a more continuous spectrum that clearly demonstrates the improved accuracy of EquiDTB models. These models achieve average MAE of 0.12 for EquiDTB3 and 0.18 for EquiDTB1—substantially lower than the 0.53 observed for standard DFTB3. To complement the analysis of equilibrium and non-equilibrium conformations in a chemically diverse set of drug-like molecules, we also evaluated the prediction of their atomic forces, F_{at} —a key quantity for investigating their thermal stability. As expected from the preceding discussion, the EquiDTB3 model achieves the lowest MARE values for each distinct molecule in the calculation of F_{at} , as illustrated in Fig. 5(c). Indeed, a comparison of the average MARE reveals a reduction by an order of magnitude, from 21.5% with DFTB3 to 5.1% and 2.8% with EquiDTB1 and EquiDTB3 models. Notice that the accuracy of EquiDTB3 persists across all compositions and molecular sizes, demonstrating its established transferability and scalability.

5 Conclusions

In this work, we present EquiDTB, a framework that parameterizes many-body Δ_{TB} potentials using physics-inspired equivariant NNs, aiming to improve the performance of the semi-empirical (SE) DFTB3 method by replacing the standard pairwise repulsive potential. To this end, we conducted a systematic study that integrates modern equivariant NNs with QM datasets of small molecules (QM7-X) and molecular dimers (DES15K) to identify the most reliable and generalizable Δ_{TB} potential for achieving hybrid DFT-PBE0 level accuracy (supplemented with

a many-body dispersion treatment *a posteriori*). Our results demonstrated that equivariant Δ_{TB} potentials better capture QM interactions that are not adequately described by the DFTB3 electronic Hamiltonian, outperforming our previous model NN_{rep} based on deep tensor NNs. While all models performed well on the training datasets, some Δ_{TB} potentials exhibited limited scalability and transferability when applied to unseen or more flexible molecules. Through a series of analyses on neural network architectures and both datasets, the EquiDTB model offering the best trade-off between accuracy and transferability was identified as the one based on the MACE architecture and trained on small molecules with up to seven heavy atoms (referred to as EquiDTB3; see Fig. 3). Although it is marginally outperformed by some other models in certain metrics, EquiDTB3 consistently enabled accurate investigation of multi-property properties in larger molecules and molecular dimers.

Furthermore, to assess how the physical terms in the DFTB total energy influence the performance of Δ_{TB} potentials, an additional EquiDTB model was developed based on DFTB1 electronic components (referred to as EquiDTB1). The capabilities of both EquiDTB models have been successfully validated through different and rigorous computational tasks involving molecules beyond those in QM7-X and DES15K datasets. Indeed, our ML-corrected DFTB approach can now effectively explore the potential energy surface of more flexible molecules (*e.g.*, tyrosine, zaprinast, ligand 2Q5k); predict the energetic ranking for large drug-like molecules (*e.g.*, AQM molecules); calculate the vibrational modes of α -amino acids; and determine atomic forces and interaction energies for small molecular dimers (*e.g.*, S66x8 dataset). Notably, the EquiDTB3 model outperforms the widely used TB methods (DFTB3 and GFN2-xTB) and the reference ML potential (rMLP) across all evaluated tasks. EquiDTB1 exhibits performance similar to that of EquiDTB3 for single-molecule systems; however, it fails to adequately correct electronic interactions in non-covalent systems. This highlights the crucial role of the second- and third-order DFTB terms in constructing more robust Δ_{TB} potentials. Our findings thus indicate that with an optimal combination of an equivariant NN and QM datasets, the EquiDTB approach can advance the DFTB method to achieve PBE0+MBD level accuracy with high computational efficiency on diverse structural, energetic, dynamic, and vibrational properties of molecular systems.

Although EquiDTB3 calculations require, on average, only 1.1 times the time of DFTB3 calculations (based on the slopes of linear fits in Fig. 6), we expect the integration of more efficient large-scale ENN models—such as SO3krates^{12,13} which also includes the JAX library⁶⁵—can help mitigate this overhead and facilitate the development of the next generation of NN-enhanced TB methods. Moreover, building on the performance analysis of the EquiDTB1 and EquiDTB3 models, a similar strategy could be extended to other SE methods in which the total energy and atomic forces are decomposed into multiple components (*e.g.*, Neglect of Differential-Diatomic Overlap (NDDO)^{66,67} Modified Neglect of Diatomic Overlap (MNDO)^{68,69}). While this study focuses exclusively on gas-phase molecular systems, the EquiDTB framework can be adapted to investigate condensed-phase



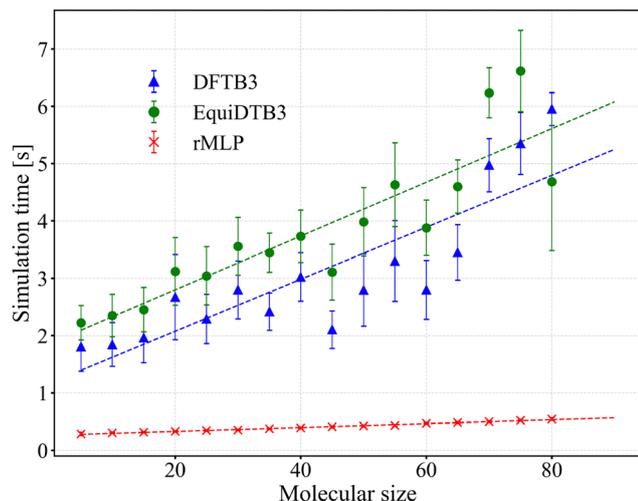


Fig. 6 Comparison of the computational scaling of the standard DFTB3 method, our EquiDFTB3 method, and the reference ML potential (rMLP). As expected, simulations using the rMLP potential are faster than those with DFTB3 or EquiDFTB3. However, EquiDFTB3 exhibits better scalability and accuracy than rMLP across all benchmark datasets, see Tables 2 and 3. To construct the scaling curves, we considered approximately 123 000 equilibrium and non-equilibrium molecular structures extracted from QM7-X and AQM-X datasets, containing 3 to 80 atoms (C, N, O, H). DFTB3 calculations were performed using a single SCC-DFTB iteration with a convergence criterion of 1×10^{-3} , and all calculations included the MBD method to correct van der Waals interactions. Error bars represent the standard deviation of simulation times within a window of ± 2.5 atoms for each data point in the graph. Dashed lines indicate linear fits of the computational scaling curves.

systems³ and organic materials.^{70,71} This can be achieved either by fine-tuning the best-performing EquiDFTB3 model or by parameterizing a new model that incorporates relevant QM datasets, although further validation may be required. Hence, our work provides valuable insights and establishes a concrete framework for integrating and advancing SE and ML methods toward the development of generalizable and reliable data-driven electronic structure approaches for (bio)molecular simulations.

Author contributions

The work was initially conceived by LMS and MP, and designed with contributions from MS and AT. LMS developed the GitHub repository, performed the model training (with SpookyNet and MACE), and analyzed the model performance in diverse applications. RP trained the models with Allegro and analyzed their performance. MP performed preliminary evaluations of the models. ZE generated the DFTB1 dataset and subsequently examined the EquiDFTB1 model. AT and GC supervised and revised all stages of the work. LMS and AT drafted the original manuscript. All authors discussed the results and contributed to the final manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

The dataset files containing the target energies ΔE_{TB} and atomic forces $\Delta \mathbf{F}_{\text{TB}}$ for small single molecules⁴⁶ and molecular dimers⁴⁷ are available in the EquiDFTB GitHub repository (<https://github.com/lmedranos/EquiDFTB>). The models, codes, and additional benchmarking datasets used in this work can also be found in this repository.

Supplementary information (SI) is available. See DOI: <https://doi.org/10.1039/d6cp00038j>.

Acknowledgements

MP acknowledges with gratitude financial support from the Institute for Advanced Studies (IAS) Luxembourg for the PhD “Young Academics” program. LMS thanks Prof. Thomas Heine at TU Dresden for fruitful discussions that helped improve this work. The results presented in this publication have been partially obtained using the HPC facilities of the University of Luxembourg and MeluXina supercomputer (PoC project). This research also used computational resources provided by the Center for Information Services and High-Performance Computing (ZIH) at TU Dresden. An award for computer time was provided by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program. This research used resources from the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. This work was partially supported by the German Research Foundation (DFG) under the Cluster of Excellence CeTI: Centre for Tactile Internet with Human-in-the-Loop (EXC 2050), the Cluster of Excellence REC²: Responsible Electronics in the Climate Change Era (EXC 3035), and the Cluster of Excellence CARE: Climate-Neutral And Resource-Efficient Construction (EXC 3115), project number 533767731.

Notes and references

- B. Huang, G. F. von Rudorff and O. A. von Lilienfeld, *Science*, 2023, **381**, 170–175.
- H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu, P. Chandak, S. Liu, P. Van Katwyk, A. Deac, A. Anandkumar, K. Bergen, C. P. Gomes, S. Ho, P. Kohli, J. Lasenby, J. Leskovec, T.-Y. Liu, A. Manrai, D. Marks, B. Ramsundar, L. Song, J. Sun, J. Tang, P. Veličković, M. Welling, L. Zhang, C. W. Coley, Y. Bengio and M. Zitnik, *Nature*, 2023, **620**, 47–60.
- A. B. Poma, A. Hinostroza Caldas, L. F. Cofas-Vargas, M. S. Jones, A. L. Ferguson and L. Medrano Sandonas, *Biophys. J.*, 2025, **124**, 1–17.
- M. Kulichenko, B. Nebgen, N. Lubbers, J. S. Smith, K. Barros, A. E. A. Allen, A. Habib, E. Shinkle, N. Fedik, Y. W. Li, R. A. Messerly and S. Tretiak, *Chem. Rev.*, 2024, **124**, 13681–13714.
- M. Pinheiro, F. Ge, N. Ferré, P. O. Dral and M. Barbatti, *Chem. Sci.*, 2021, **12**, 14396–14413.
- O. T. Unke, S. Chmiela, H. E. Saucedo, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K.-R. Müller, *Chem. Rev.*, 2021, **121**, 10142–10186.



- 7 I. Batatia, S. Batzner, D. P. Kovács, A. Musaelian, G. N. C. Simm, R. Drautz, C. Ortner, B. Kozinsky and G. Csányi, *Nat. Mach. Intell.*, 2025, **7**, 56–67.
- 8 O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schütt, H. E. Sauceda and K.-R. Müller, *Nat. Commun.*, 2021, **12**, 7273.
- 9 I. Batatia, D. P. Kovacs, G. Simm, C. Ortner and G. Csányi, *Adv. Neural Inf. Process. Syst.*, 2022, **35**, 11423–11436.
- 10 S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt and B. Kozinsky, *Nat. Commun.*, 2022, **13**, 2453.
- 11 A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth and B. Kozinsky, *Nat. Commun.*, 2023, **14**, 579.
- 12 T. Frank, O. Unke and K.-R. Müller, *Adv. Neural Inf. Process. Syst.*, 2022, **35**, 29400–29413.
- 13 J. T. Frank, O. T. Unke, K.-R. Müller and S. Chmiela, *Nat. Commun.*, 2024, **15**, 6539.
- 14 O. T. Unke, M. Stöhr, S. Ganscha, T. Unterthiner, H. Maennel, S. Kashubin, D. Ahlin, M. Gastegger, L. Medrano Sandonas, J. T. Berryman, A. Tkatchenko and K.-R. Müller, *Sci. Adv.*, 2024, **10**, eadn4397.
- 15 D. P. Kovács, J. H. Moore, N. J. Browning, I. Batatia, J. T. Horton, Y. Pu, V. Kapil, W. C. Witt, I.-B. Magdau, D. J. Cole and G. Csányi, *J. Am. Chem. Soc.*, 2025, **147**, 17598–17611.
- 16 A. Kabylda, J. T. Frank, S. Suárez-Dou, A. Khabibrakhmanov, L. Medrano Sandonas, O. T. Unke, S. Chmiela, K.-R. Müller and A. Tkatchenko, *J. Am. Chem. Soc.*, 2025, **147**, 33723–33734.
- 17 N. Fedik, B. Nebgen, N. Lubbers, K. Barros, M. Kulichenko, Y. W. Li, R. Zubatyuk, R. Messerly, O. Isayev and S. Tretiak, *J. Chem. Phys.*, 2023, **159**, 110901.
- 18 M. Stöhr, L. Medrano Sandonas and A. Tkatchenko, *J. Phys. Chem. Lett.*, 2020, **11**, 6835–6843.
- 19 P. Zheng, R. Zubatyuk, W. Wu, O. Isayev and P. O. Dral, *Nat. Commun.*, 2021, **12**, 7022.
- 20 E. H. E. Farrar and M. N. Grayson, *Chem. Sci.*, 2022, **13**, 7594–7603.
- 21 M. Nováček and J. Rezáč, *J. Chem. Theory Comput.*, 2025, **21**, 678–690.
- 22 Y. Chen, W. Yan, Z. Wang, J. Wu and X. Xu, *J. Chem. Theory Comput.*, 2024, **20**, 9500–9511.
- 23 W. Sun, G. Fan, T. van der Heide, A. McSloy, T. Frauenheim and B. Aradi, *J. Chem. Theory Comput.*, 2023, **19**, 3877–3888.
- 24 J. Zhu, V. Q. Vuong, B. G. Sumpter and S. Irlé, *MRS Commun.*, 2019, **9**, 867–873.
- 25 J. J. Kranz, M. Kubillus, R. Ramakrishnan, O. A. von Lilienfeld and M. Elstner, *J. Chem. Theory Comput.*, 2018, **14**, 2341–2352.
- 26 A. McSloy, G. Fan, W. Sun, C. Hölzer, M. Friede, S. Ehlert, N.-E. Schütte, S. Grimme, T. Frauenheim and B. Aradi, *J. Chem. Phys.*, 2023, **158**, 034801.
- 27 C. Panosetti, A. Engelmann, L. Nemeč, K. Reuter and J. T. Margraf, *J. Chem. Theory Comput.*, 2020, **16**, 2181–2191.
- 28 A. Hofstetter, L. Bösel and S. Riniker, *Phys. Chem. Chem. Phys.*, 2022, **24**, 22497–22512.
- 29 D. J. Burrill, C. Liu, M. G. Taylor, M. J. Cawkwell, D. Perez, E. R. Batista, N. Lubbers and P. Yang, *J. Chem. Theory Comput.*, 2025, **21**, 1089–1097.
- 30 A. S. Christensen, S. K. Sirumalla, Z. Qiao, M. B. O'Connor, D. G. A. Smith, F. Ding, P. J. Bygrave, A. Anandkumar, M. Welborn, F. R. Manby, I. Miller and F. Thomas, *J. Chem. Phys.*, 2021, **155**, 204103.
- 31 G. Zhou, N. Lubbers, K. Barros, S. Tretiak and B. Nebgen, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2120333119.
- 32 Z. Qiao, A. S. Christensen, M. Welborn, F. R. Manby, A. Anandkumar and T. F. Miller, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2205221119.
- 33 B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Deshayé, T. Dumitrica and A. Dominguez, *et al.*, *J. Chem. Phys.*, 2020, **152**, 124101.
- 34 G. Fan, A. McSloy, B. Aradi, C.-Y. Yam and T. Frauenheim, *J. Phys. Chem. Lett.*, 2022, **13**, 10132–10139.
- 35 F. Hu, F. He and D. J. Yaron, *J. Chem. Theory Comp.*, 2023, **19**, 6185–6196.
- 36 Q. Gu, Z. Zhouyin, S. K. Pandey, P. Zhang, L. Zhang and E. Weinan, *Nat. Commun.*, 2024, **15**, 6772.
- 37 P. Stishenko, A. McSloy, B. Onat, B. Hourahine, R. J. Maurer, J. R. Kermode and A. Logsdail, *J. Chem. Phys.*, 2024, **161**, 012502.
- 38 A. S. Christensen, T. Kubar, Q. Cui and M. Elstner, *Chem. Rev.*, 2016, **116**, 5301–5337.
- 39 L. Medrano Sandonas, R. Gutierrez, A. Pecchia, A. Croy and G. Cuniberti, *Entropy*, 2019, **21**, 735.
- 40 G. Seifert, *J. Phys. Chem. A*, 2007, **111**, 5609–5613.
- 41 A. S. Krishnapriyan, A. S. Krishnapriyan, P. Yang, A. M. N. Niklasson and M. J. Cawkwell, *J. Chem. Theory Comput.*, 2017, **13**, 6191–6200.
- 42 Q. Wang, M. Gu, C. Michel, N. Goldman, T. Niehaus and S. N. Steinmann, *J. Chem. Theory Comput.*, 2025, **21**, 5267–5278.
- 43 N. Goldman, L. E. Fried, R. Lindsey, C. H. Pham and R. Dettori, *J. Chem. Phys.*, 2023, **158**, 144112.
- 44 C. Panosetti, Y. Lee, A. Samtsevych and C. Scheurer, *J. Chem. Phys.*, 2023, **158**, 224115.
- 45 C. Panosetti, S. B. Anniés, C. Grosu, S. Seidlmayer and C. Scheurer, *J. Phys. Chem. A*, 2021, **125**, 691–699.
- 46 J. Hoja, L. Medrano Sandonas, B. G. Ernst, A. Vazquez-Mayagoitia, R. A. DiStasio Jr. and A. Tkatchenko, *Sci. Data*, 2021, **8**, 43.
- 47 A. G. Donchev, A. G. Taube, E. Decolvenaere, C. Hargus, R. T. McGibbon, K.-H. Law, B. A. Gregersen, J.-L. Li, K. Palmo and K. Siva, *et al.*, *Sci. Data*, 2021, **8**, 55.
- 48 A. Tkatchenko, R. A. DiStasio Jr, R. Car and M. Scheffler, *Phys. Rev. Lett.*, 2012, **108**, 236402.
- 49 A. Ambrosetti, A. M. Reilly, R. A. DiStasio Jr. and A. Tkatchenko, *J. Chem. Phys.*, 2014, **140**, 18A508.
- 50 A. Pecina, J. Fanfrlk, M. Lepšák and J. Rezáč, *Nat. Commun.*, 2024, **15**, 1127.
- 51 L. Medrano Sandonas, D. Van Rompaey, A. Fallani, M. Hilfiker, D. Hahn, L. Perez-Benito, J. Verhoeven, G. Tresadern, J. Kurt Wegner and H. Ceulemans, *et al.*, *Sci. Data*, 2024, **11**, 742.
- 52 J. Rezáč, K. E. Riley and P. Hobza, *J. Chem. Theory Comp.*, 2011, **7**, 3466–3470.
- 53 G. Seifert, D. Porezag and T. Frauenheim, *Int. J. Quantum Chem.*, 1996, **58**, 185–192.



- 54 M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1998, **58**, 7260.
- 55 M. Gaus, Q. Cui and M. Elstner, *J. Chem. Theory Comput.*, 2011, **7**, 931–948.
- 56 M. Gaus, A. Goetz and M. Elstner, *J. Chem. Theory Comput.*, 2013, **9**, 338–354.
- 57 M. Gaus, X. Lu, M. Elstner and Q. Cui, *J. Chem. Theory Comput.*, 2014, **10**, 1518–1537.
- 58 S. R. Bahn and K. W. Jacobsen, *Comput. Sci. Eng.*, 2002, **4**, 56–66.
- 59 J. P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982–9985.
- 60 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.
- 61 V. Havu, V. Blum, P. Havu and M. Scheffler, *J. Comput. Phys.*, 2009, **228**, 8367–8379.
- 62 M. Puleva, L. Medrano Sandonas, B. D. Lörincz, J. Charry, D. M. Rogers, P. R. Nagy and A. Tkatchenko, *Nat. Commun.*, 2025, **16**, 8583.
- 63 E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2019, **150**, 154122.
- 64 P. Koskinen and V. Mäkinen, *Comput. Mater. Sci.*, 2009, **47**, 237–253.
- 65 J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne and Q. Zhang, *JAX: composable transformations of Python + NumPy programs*, 2018, <https://github.com/google/jax>.
- 66 J. A. Pople, D. P. Santry and G. A. Segal, *J. Chem. Phys.*, 1965, **43**, S129–S135.
- 67 T. Husch and M. Reiher, *J. Chem. Theory Comput.*, 2018, **14**, 5169–5179.
- 68 M. J. S. Dewar and W. Thiel, *J. Am. Chem. Soc.*, 1977, **99**, 4899–4907.
- 69 M. J. S. Dewar and W. Thiel, *Theor. Chim. Acta*, 1977, **46**, 89–104.
- 70 M. Mortazavi, J. G. Brandenburg, R. J. Maurer and A. Tkatchenko, *J. Phys. Chem. Lett.*, 2018, **9**, 399–405.
- 71 L. Chen, L. Medrano Sandonas, P. Traber, A. Dianat, N. Tverdokhleba, M. Hurevich, S. Yitzchaik, R. Gutierrez, A. Croy and G. Cuniberti, *Sci. Data*, 2025, **12**, 324.

