Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Digital Discovery*, 2022, **1**, 463

Long-range dispersion-inclusive machine learning potentials for structure search and optimization of hybrid organic–inorganic interfaces†

Julia Westermayr, [a] Shayantan Chaudhuri, [ab] Andreas Jeindl, [c] Oliver T. Hofmann [c] and Reinhard J. Maurer *[a]

The computational prediction of the structure and stability of hybrid organic–inorganic interfaces provides important insights into the measurable properties of electronic thin film devices, coatings, and catalyst surfaces and plays an important role in their rational design. However, the rich diversity of molecular configurations and the important role of long-range interactions in such systems make it difficult to use machine learning (ML) potentials to facilitate structure exploration that otherwise requires computationally expensive electronic structure calculations. We present an ML approach that enables fast, yet accurate, structure optimizations by combining two different types of deep neural networks trained on high-level electronic structure data. The first model is a short-ranged interatomic ML potential trained on local energies and forces, while the second is an ML model of effective atomic volumes derived from atoms-in-molecules partitioning. The latter can be used to connect short-range potentials to well-established density-dependent long-range dispersion correction methods. For two systems, specifically gold nanoclusters on diamond (110) surfaces and organic π-conjugated molecules on silver (111) surfaces, we train models on sparse structure relaxation data from density functional theory and show the ability of the models to deliver highly efficient structure optimizations and semi-quantitative energy predictions of adsorption structures.

Received 25th February 2022
Accepted 3rd June 2022

DOI: 10.1039/d2dd00016d

rsc.li/digitaldiscovery

1. Introduction

Surface nanostructures play a fundamental role in medicine,[1,2] solar cell and fuel cell technologies,[3,4] and photo- or electrocatalysis.[5,6] Several strategies exist to form nanostructures, such as DNA-directed assembly,[7] electrodeposition,[6] or self-assembly at hybrid organic–inorganic interfaces.[8] The molecular composition and molecule–surface interaction strength crucially determine the surface structures that are formed[9–11] and the nucleation and initial growth of nanoclusters (NCs) are crucial steps in controlling a nanostructures' final morphology,[6,12] which itself is important for tuning catalytic selectivity and activity.[13] A better understanding of surface nanostructures can thus advance a wide variety of research fields.[14,15]

Electronic structure theory plays a vital role in the characterization and exploration of organic–inorganic interfaces and materials, but is limited by intrinsic errors such as the lack of long-range dispersion interactions in common density functionals[16–18] and the high computational effort associated with the intrinsic length scale of surface structures. The former issue has been addressed in recent years with the emergence of efficient and accurate long-range dispersion correction methods such as the Grimme[16] and Tkatchenko–Scheffler (TS) families of methods.[19] In case of metal–organic interfaces, the surface-screened van-der-Waals (vdW^{surf})[20] and many-body dispersion (MBD)[21,22] methods, in combination with generalized gradient approximations (GGAs) or range-separated hybrid functionals, have been shown to provide highly accurate predictions of adsorption structures and stabilities.[10,11,18,23–28] Reliable identification and optimization of structures at metal–organic interfaces is a particular challenge due to the structural complexity and the large number of degrees of freedom (molecular orientation, adsorption site, coverage),[15] which creates a particular need for structural exploration methods that are efficient. Examples of simulation methods that can alleviate computational effort compared to DFT include semi-empirical electronic structure methods, such as density functional tight-binding (DFTB),[29] which usually provides a good compromise between accuracy and computational efficiency. Recently, DFTB has been coupled with the vdW and MBD methods[29,30] to incorporate long-range dispersion, but unfortunately few reliable DFTB parametrizations for metal–organic interfaces exist to date.[31]

[a]*Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK. E-mail: r.maurer@warwick.ac.uk*

[b]*Centre for Doctoral Training in Diamond Science and Technology, University of Warwick, Coventry, CV4 7AL, UK*

[c]*Institute of Solid State Physics, Graz University of Technology, 8010 Graz, Austria*

† Electronic supplementary information (ESI) available. See https://doi.org/10.1039/d2dd00016d

Machine learning-based interatomic potentials (MLIPs) offer high computational efficiency whilst retaining the accuracy of the underlying training data based on electronic structure theory. Atomistic MLIP methods include Gaussian approximation potentials[32–34] or neural network (NN) potentials (*e.g.* SchNet,[35–37] PhysNet[38] or Behler–Parinello type NNs[39–41]), which describe atoms in their chemical and structural environment within a cutoff region. MLIPs have the potential to advance structure searches,[42–44] geometry optimizations,[45,46] and molecular dynamics (MD) simulations[40,47–49] of highly complex and large-scale systems comprising many thousands of atoms.[50] However, most established MLIP approaches learn short-range interactions between atoms by introducing a radial cutoff within which the atomic interactions are captured. This can lead to challenges when attempting to capture long-range electrostatic or dispersion interactions.[38]

Recent attempts of accounting for long-range interactions in MLIPs have explicitly treated them as separate additive contributions to the potential,[38,51–54] such as the third and higher generation NN potentials of Behler and co-workers,[55,56] where a charge-equilibration scheme was introduced. Earlier work by Behler and co-workers[51,52] has also shown that the simulation of liquid water can be facilitated with neural networks trained on energies and atomic charges, where the latter was used to correct for electrostatic interactions. This scheme was later complemented with long-range dispersion interactions based on the Grimme D3 correction method.[52] Atomic charges were further used in TensorMol-0.1 (ref. 53) to augment the total energy with Coulomb and vdW corrections. A similar approach was applied by Unke and Meuwly[38] in PhysNet, where the total energy was corrected with additive terms that include electrostatic corrections obtained from partial atomic charges and a D3 dispersion correction term. Recently, this description was extended in SpookyNet, where the total energy is corrected with empirical terms for the nuclear repulsion based on an analytical short-range term, a term for electrostatics and a term for dispersion interactions.[54] The aforementioned approaches have been demonstrated to accurately describe MD or spectroscopic signatures,[53] small clusters on surfaces,[56] water dimers[57] and clusters,[52] crystals,[57] and phase diagrams.[58]

An accurate DFT-based description of hybrid organic–inorganic interfaces is challenging. The Tkatchenko–Scheffler family of methods have proven to provide robust and reliable descriptions of structure and stability. It was recently shown that these electron-density-dependent methods can also be combined with semi-empirical methods such as DFTB[30] and force fields,[59] which suggests that a similar approach could be viable for structure prediction at hybrid organic–inorganic interfaces.

In this work, we present a deep learning approach that combines an NN-based MLIP with an established long-range dispersion method from the TS family of methods to efficiently predict structures and stabilities at metal–organic interfaces for the purpose of high-throughput structural (pre) screening and global energy landscape exploration. As shown in Fig. 1, the short-range description is provided by a local MLIP, whereas the long-range interaction is provided by one of the TS

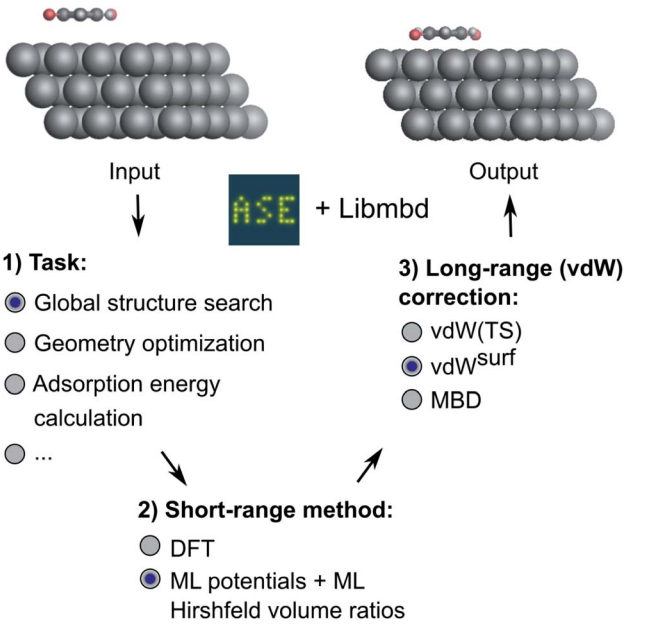

Fig. 1 Overview of the method developed in this work. Different machine learning interatomic potentials (MLIPs) that allow for the computation of Hirshfeld volume ratios can be combined with different flavors of van der Waals (vdW) corrections, *e.g.* screened vdW pairwise interactions[19] and many-body dispersion (MBD).[21] The so-obtained MLIPs are interfaced with the Atomic Simulation Environment (ASE)[60] and can be used for global structure searches, optimizations, energy predictions or other types of simulations implemented within ASE.

methods such as vdW^{surf}. We couple the two approaches by constructing an ML representation of Hirshfeld atoms-in-molecules volumes.[19,61] The atomic volumes are used to rescale atomic polarizabilities that enter the long-range description based on the local chemical environment of the atoms provided by the DFT description of short-range interactions. The method is applicable for the incorporation of additive *a posteriori* dispersion correction schemes, such as vdW(TS), vdW^{surf}, and MBD methods that are implemented in the Libmbd package[62] or the Grimme dispersion methods D3 and D4.[63,64] However, the method in its current form cannot be used for self-consistent vdW corrections such as the vdW-DF family of methods.[65,66] If training on data obtained from vdW-DF methods is sought for, adaptions are needed to accurately model long-range effects, *e.g.*, as it is done in PhysNet[38] or SpookyNet.[54]

We deliver an open-access implementation of our approach by coupling the Atomic Simulation Environment (ASE)[60] with the Libmbd package[62] and the DFT-D3 code.[63] The trained interatomic potentials are independent of the long-range vdW correction. Therefore a trained ML model can be combined with different vdW corrections after model training. To further increase the robustness of our approach, we implement query-by-committee,[39,67,68] which establishes the model variance in energy and force predictions. This allows us to define a dynamic stopping criterion for when the prediction of the MLIP becomes unreliable and structure optimizations have to be continued

with electronic structure theory. This is particularly useful in the context of efficient pre-relaxation of structures to reduce the computational cost associated with structure search. We show the utility of our approach on two systems, namely a global structure search for gold (Au) NCs adsorbed onto a diamond (110) surface and the structural relaxation of large conjugated organic molecules, namely 9,10-anthraquinone (A2O), 1,4-benzoquinone (B2O), and 6,13-pentacenequinone (P2O), summarized as X2O, adsorbed onto a silver (Ag) (111) surface that self-assemble into a variety of surface phases.[9] The model for X2O on Ag(111) is trained on sparse data extracted from open data repositories, which shows the utility of the model to facilitate structure pre-relaxations. We further demonstrate that the ML models trained on these data are transferable to different aromatic organic molecules on the same surface that were not contained in the training data set.

2. Methods

2.1 ML potentials coupled to long-range dispersion corrections

The vdW(TS),[19] vdW^{surf},[20] and MBD[21] methods are commonly used as *a posteriori* corrections to DFT, although they also exist as self-consistent variants.[69] Throughout this section, we refer to vdW(TS) and vdW^{surf} as vdW as they only differ in their parametrization. In the case of the vdW scheme, the dispersion energy contribution is a pairwise potential:[19]

$$E_{\mathrm{vdW}}(\mathrm{R}) = -\sum_{\mathrm{A,B}} f(r_{\mathrm{cut}}, \mathrm{A}, \mathrm{B}) \frac{C_6^{\mathrm{AB}}(\mathrm{R})}{R_{\mathrm{AB}}^6} \quad (1)$$

where R_{AB} is the distance between two atoms, A and B, and f is a damping function that limits the correction to distances beyond a predefined cutoff r_{cut}. The model depends on tabulated free-atom reference parameters such as atomic polarizabilities that are used to calculate C_6^{AB} coefficients and scaled vdW radii that define r_{cut} in the damping function. The C_6^{AB} coefficients explicitly depend on all coordinates of the system R to account for the chemical environment of the atoms. This is achieved by re-scaling the atomic polarizabilities and vdW radii based on the Hirshfeld atoms-in-molecules partitioning scheme.[61] The ratio between effective volume of an atom in a molecule and a free atom is used as re-scaling factor:[19,30]

$$H_{\mathrm{A}} = \frac{V_{\mathrm{A,eff}}}{V_{\mathrm{A,free}}}. \quad (2)$$

The many body dispersion (MBD) scheme is an extension of the vdW method that accounts for long-range correlation interactions beyond the pairwise limit. This is achieved by the construction of a Hamiltonian of dipole-coupled quantum harmonic oscillators located at the positions of the atoms and parametrized to reflect the effective atomic polarizability of the atom in the molecule.

In this work, we couple both the vdW and MBD long-range dispersion schemes to an MLIP by creating an ML model of the Hirshfeld-based scaling ratios (H_{A}) for all atoms A in the system. We note that the range-separation parameter in MBD and damping coefficient used in vdW are the only parameters specific to the employed exchange–correlation functional approximation to which the dispersion correction is coupled. As we train MLIPs to reproduce training data created with a specific exchange–correlation functional, we can retain the same parameters as used for the respective functional for vdW corrections to the generated MLIP.

Throughout this work, we employ the ASE code which offers calculator interfaces to various electronic structure packages.[60] The ML models in this work are based on the continuous-filter convolutional NN SchNet,[35–37] which is a message-passing NN that learns the representation of the atomic environments in addition to its relation to the targeted output. ASE also provides an interface to the deep learning toolbox SchNetPack to employ NN-based MLIPs within ASE.[37] We have implemented an ASE calculator interface for the Libmbd code[62] and further implemented an ASE calculator instance that combines a short-range calculator (*e.g.* electronic structure package or MLIP based on SchNetPack) with a Libmbd calculator instance. This interface calculator passes Hirshfeld scaling ratios predicted by an ML model into the Libmbd calculator to perform vdW- or MBD-corrected SchNet calculations (denoted 'ML+vdW' and 'ML+MBD', respectively). All developed code is freely available on GitHub.[70]

2.2 Training data

2.2.1 Gold nanoclusters on diamond (Au@C). DFT calculations were performed using the all-electron numeric atomic orbital code FHI-aims[71] and the Perdew–Burke–Ernzerhof (PBE)[72] exchange–correlation functional. The numeric atomic orbitals were represented using a 'light' basis set and dispersion effects were accounted for *via* the MBD scheme.[21] The total energy, sum of eigenvalues, charge density, and energy derivatives convergence criteria were set to 1×10^{-6} eV, 1×10^{-2} eV, 1×10^{-5} e/a_0^3, and 1×10^{-4} eV Å^{-1}, respectively. For structure relaxations, the maximum residual force component per atom was set to 1×10^{-2} eV Å^{-1}. Initial structures were constructed using ASE[60] with Au NCs of various sizes adsorbed onto the center of a diamond (110) surface, with all carbon (C) atoms being fully frozen during optimizations. To lower computational costs and memory requirements, we created an aperiodic cluster cut-out of a diamond surface that corresponds to a 7×7 supercell repeat of a 7-layered diamond (110) slab. An example of an Au NC with $n = 50$ (n denotes the number of Au atoms) on a diamond (110) surface can be seen in Fig. 2d.

As starting point for the training data set for Au@C models, we used 62 geometry optimizations of Au NCs on diamond (5, 4, 8, 8, 9, 10, and 18 geometry relaxations were conducted on Au clusters of size n = 15, 20, 30, 35, 40, 45 and 50 atoms, respectively, on the aforementioned diamond (110) surface model). The training data points were collated using every relaxation step of the optimization runs, which therefore included both optimized and not fully-optimized structures. These computations led to an initial training data set comprising 5368 data points, which we used to train four MLIPs (trained on energy and forces). All MLIPs were trained using the

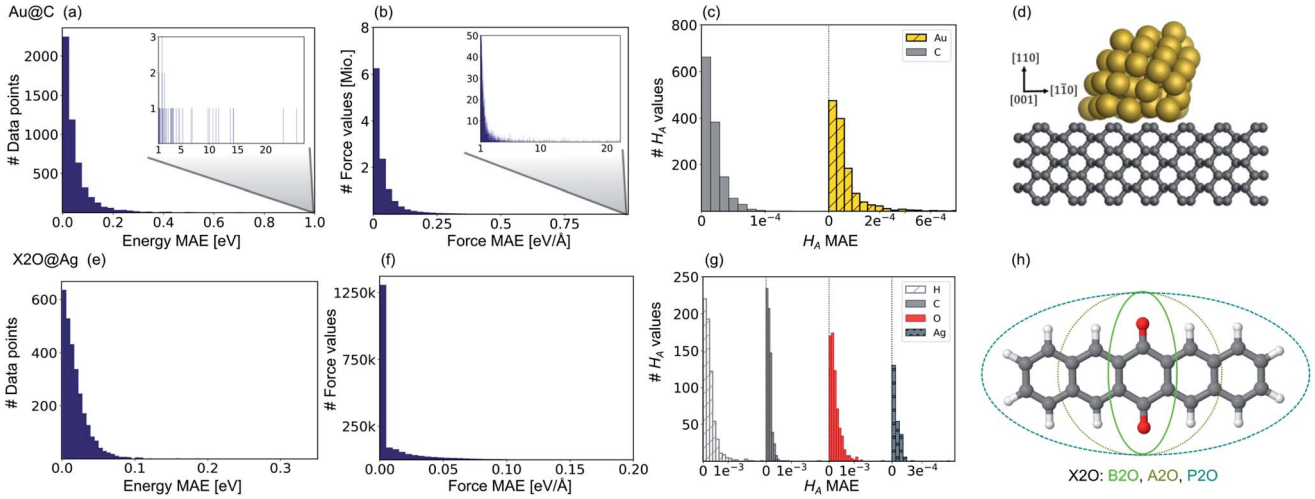

Fig. 2 Prediction errors for gold nanoclusters (NCs) on diamond (110) surfaces (Au@C) on top and for X2O systems on Ag(111) (X2O@Ag) in the bottom. (a and e) Mean absolute errors (MAEs) for energies, (b and f) for forces (middle), and (c and g) Hirshfeld volume ratios, H_A, for Au@C and X2O@Ag, respectively. Bar plots for energies and forces are shown and summarized from four trained machine learning (ML) models. For forces, the error with respect to each force component is shown, *i.e.*, one data point thus contains as many components as thrice the number of atoms (around 2100 values for Au@C and about 200–300 for X2O@Ag systems) for the three orthogonal directions, which are [110], [001] and [1$\bar{1}$0] for Au@C, and [111], [12$\bar{1}$] and [$\bar{1}$01] for X2O@Ag. For Hirshfeld volume ratios, one ML model is used, and the error is split into contributions from the separate atom types. (d) Example of an Au NC with 50 atoms on a diamond (110) surface and (h) X2O systems in the gas phase that are described in this study on Ag(111).

same data set, which was split randomly into training, validation, and test sets. All ML models trained on the initial training data set are denoted as "$ML_{init.}$". MLIPs were used to predict 'local' energies and forces as well as Hirshfeld volume ratios to correct for long-range interactions at the MBD level. For energies and forces, we trained a set of models to use the query-by-committee approach discussed in subsection 2.4, which makes energy predictions more robust by a factor of $\sqrt{q}$, where q is the number of trained ML models. The training process of energies and forces is explained in detail in section S1.1 in the ESI.† The models slightly differed in the weights of energies and forces used in the combined loss function (see eqn (3) and discussion in the next subsection). The model architecture and hyperparameter optimizations for the Hirshfeld model can be found in the ESI in section S1.2.†

To extend the training data set, adaptive sampling[39] was carried out, which was originally developed for molecular dynamics simulations. Importantly, the predictions of the set of ML models are compared at every time step. Whenever the variance of the models exceeded a predefined threshold (with the threshold often being set slightly higher than the root-mean-squared error of the models on a test set[73]), the data point was deemed untrustworthy and recomputed with the reference method. This data point was then added to the training set and the models were retrained. In this work, we applied this concept to a global structure search using the basin-hopping algorithm[74,75] as implemented in ASE[60] rather than MD simulations. After each geometry optimization during the basin-hopping run, the variance of the model predictions was computed and geometries with the largest model variances were selected for further DFT optimizations. These optimizations were then added to the training set. Stopping criteria for ML optimizations are discussed in section 2.4.

In total, three adaptive sampling runs were carried out. The first adaptive sampling run was carried out with the initial ML models, "$ML_{init.}$". After data points were sampled and the data set was extended, ML models were retrained. MLIPs after the first adaptive sampling run (denoted as $ML_{adapt.1}$) were trained on 7700 data points for training and 800 data points for validation. Before adaptive sampling, ML models deviated by several 10s of eV for cluster sizes that were not included in the training set, leading to unphysical structure relaxations. After adding additional data points, the average model variance decreased to around 0.1 eV with maximum errors in the range of 1 eV, when the training regime was left. To further increase the accuracy of the ML models a second adaptive sampling run $ML_{adapt.2}$ was executed with $ML_{adapt.1}$. A total of 9757 data points were collected after the second adaptive sampling run. $ML_{adapt.2}$ models were trained on 8500 data points for training and 800 data points for validation. After the final adaptive sampling run ($ML_{adapt.3}$), there were a total of 15 293 data points. 12 500 data points were used for training and 1500 for validation. More details on the adaptive sampling runs can be found in section S1.1.†

2.2.2 Organic molecules on silver (X2O@Ag). The training data points for X2O@Ag are taken from the NOMAD repository[76–78] and are based on ref. 9. X2O summarizes different functional organic monomers, which are described as monolayers on Ag(111) surfaces (abbreviated as X2O@Ag). As mentioned above, the three different molecules tested were: 9,10-anthraquinone (A2O), 1,4-benzoquinone (B2O), and 6,13-pentacenequinone (P2O) as shown in Fig. 2h. The data set

consists of 8202 data points, where each data point comprises a geometry and the corresponding energies, forces, and Hirshfeld volume ratios. In more detail, the data sets contain 353 data points of the clean substrate in total (about 4% of the data), 1397 data points of P2O molecules, 2249 data points of A2O molecules, and 4156 data points of B2O molecules. The molecules were either in the gas phase, arranged as two-dimensional free-standing overlayers in various unit cells and arrangements (5724 data points; about 70% of the data), or adsorbed onto an 8-layered Ag(111) surface slab (2125 data points; about 26% of the data). Some supercells contained several different molecules adsorbed onto the surface. The reference data points possessed different unit cell sizes and the reference method for the data was vdW^{surf}-corrected DFT (DFT+vdW^{surf}) with the PBE exchange-correlation functional, and a dipole correction also being employed.[79] A 'tight' basis set was used for the top three substrate layers while a 'very light' basis set was used for the five lower lying layers.[9] The data points were taken from 208 geometry relaxations and 6773 single-point calculations. The training data set was generated with FHI-aims in ref. 9, with the total energy, forces, and charge density convergence criteria being set to 1×10^{-5} eV, 1×10^{-3} eV, 1×10^{-2} e/a_0^3, respectively.

For Au@C, four ML models were trained on energies and forces (see section S1.1† for details) and one model on Hirshfeld volume ratios, which was used in all geometry optimizations. As mentioned earlier, adaptive sampling was not carried out for this data set as we wanted to base our models purely on sparse existing data derived from a small set of geometry optimizations to showcase the usability of our model to speed up structure relaxations.

In addition, both DFT and ML structure relaxations of 16 B2O@Ag systems far away from the surface were conducted and served as a test set. These structures are especially challenging to relax as common optimization algorithms often fail for systems that are far away from the optimized structure, even with DFT and long-range interactions. One problem is that vdW forces decrease quickly with the distance of an adsorbate to the surface, and quasi-Newton optimizers with simple Hessian guesses can converge to a geometry that has hardly changed compared to the initial structure. This problem can be overcome by using an improved Hessian approximation for the initialization of the optimization. In this work, we used the Lindh Hessian[71,80] to initialize structure relaxations for DFT+vdW^{surf} and ML+vdW^{surf} calculations. The same optimization criteria were used as in the reference calculations, but we used the ASE calculator with our vdW implementation rather than FHI-aims for consistency.

2.3 Machine Learning Interaction Potentials (MLIPs)

We generate vdW-free SchNet[36,37] MLIPs and a SchNet-based model for the Hirshfeld volume ratios. The local vdW-free potential energy surfaces were obtained by subtracting the vdW corrections from the total energies and forces obtained with FHI-aims. The MLIPs are trained with vdW-free energies (E) and forces (F). The forces are treated as derivatives of the MLIP, $E_{\mathrm{local}}^{\mathrm{ML}}$, with respect to the atomic positions (R) and are trained in addition to the energies using a combined loss function (L_2):

$$L_2 = t\|E_{\mathrm{local}}^{\mathrm{QC}} - E_{\mathrm{local}}^{\mathrm{ML}}\|^2 + (1-t)\left\|F_{\mathrm{local}}^{\mathrm{QC}} - \frac{\partial E_{\mathrm{local}}^{\mathrm{ML}}}{\partial R}\right\|^2, \quad \text{where } E_{\mathrm{local}}^{\mathrm{ML}} = \sum_{\mathrm{A}}^{N} E_{\mathrm{local,A}}^{\mathrm{ML}} \tag{3}$$

The energies are obtained as the sum of atomic contributions with N being the total number of atoms in a system. The trade-off, t, is used to ensure a good balance between energies and forces during training.

In contrast, the Hirshfeld volume ratios were fitted per atom using another SchNet model that was adapted for this purpose. The corresponding loss function, L_2^{H}:

$$L_2^{\mathrm{H}} = \sum_{\mathrm{A}}^{N} \|H_{\mathrm{A}}^{\mathrm{QC}} - H_{\mathrm{A}}^{\mathrm{ML}}\|^2, \tag{4}$$

contains all Hirshfeld volume ratios, allowing for all values to be modeled in one atomistic ML model. The details on the model training and the used parameters for model training can be found in the ESI in section S1.2.†

As mentioned in the previous subsection 2.2.2 the X2O@Ag data was generated using two basis sets for Ag atoms depending on their position. Different basis sets will result in different energies and forces. Therefore, the data set was pre-processed prior to training by representing all the Ag atoms that were described using a 'very light' basis set with a different atom label. This process allowed the MLIPs to be trained on data with mixed basis sets.

2.4 Structure relaxations with MLIPs

For all structure relaxations, local MLIPs and ML Hirshfeld volume ratios were used for additional vdW corrections, and the screened atomic polarizabilities suggested for Ag by Ruiz *et al.*[20] were used to account for the correct dielectric screening of the metal surface. Structure relaxations were carried out using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, as implemented in ASE,[60] which utilizes a maximum atomic force criterion, f_{max}, to decide when the optimization should be stopped. We adopted the decision as to when the optimization should be stopped by further making use of the query-by-committee concept and taking the variance of the ML model predictions for energies into account.

The query-by-committee approach[39,67,68] takes the mean of the predictions of q ML models for a given property, P: $P^{\mathrm{ML}} = \frac{1}{q}\sum_{i=1}^{q} P^{\mathrm{ML}_q}$. In all subsequent calculations, we followed the mean of the potential energy surface and corresponding forces. While the accuracy and robustness of the predictions can be improved by a factor of $\sqrt{q}$,[81] no improvement for the predictive accuracy of other properties such as dipole moments, could be achieved. We also found that the prediction of Hirshfeld volume ratios was not improved by the query-by-

committee approach, so only one ML model was used for learning Hirshfeld volume ratios in the following. The reason for this can be manifold and is likely due to the fact that the accuracy of the Hirshfeld volume ratio models is already very high as compared to the energy models, which is why query-by-committee is unlikely to strongly improve the prediction accuracy of Hirshfeld volume ratios.

A further consequence of having more than one ML model for energies is that this approach allows us to assess the reliability of the ML predictions by computing the model variances,

$$E_{\mathrm{var}}^{\mathrm{ML}} = \sqrt{\frac{1}{1-q}\sum_{i=1}^{q}\left(P^{\mathrm{ML}_q} - P^{\mathrm{ML}}\right)^2}. \quad (5)$$

The assessment of the reliability of predictions is especially important when ML models serve as pre-optimizers and cannot reliably reach a low f_{max} value.

To find optimal stopping criteria of the optimization with ML models, we explored a random grid of 1000 different stopping criterion combinations for structure relaxations of the Au@C test set using $\mathrm{ML}_{\mathrm{init.}}$ and the X2O@Ag test set (see Fig. S2a and b,† respectively). The ability to perform 1000 geometry optimizations as a test further showcases the computational efficiency of the approach. Test runs showed that introducing an additional initial $f_{\mathrm{maxinit.}}$ value as a threshold, after which the ML model variance for energies, $E_{\mathrm{var}}^{\mathrm{ML}}$ (eqn (5)) is monitored, is beneficial with respect to the agreement of the final ML-optimized structure and DFT-optimized structure. The $f_{\mathrm{maxinit.}}$ value was found to be relatively robust and set to 0.15 eV Å^{-1} for the test studies shown in this work, but it can be set to a different value by the user to take into account the requirements of other ML models. We tested different thresholds between 0.1–0.2 eV Å^{-1} for initial models and found that the structures obtained were very similar and differed by less than 0.05 Å root-mean-squared deviation.

As soon as the $f_{\mathrm{maxinit.}}$ value was reached during an optimization, the number of consecutive steps that showed rising energy variances was monitored. The number of consecutive steps that showed rising energy variance was varied in a grid search and we found three consecutive steps of increasing energy variance to be a good criterion to stop the optimization algorithm with final structures closest to the DFT reference minimum (Fig. S1†) for a range of different $f_{\mathrm{maxinit.}}$ values. The energy variance between different ML models will always fluctuate around a small number, even in the case of reliable geometry relaxations. Hence, the energy variance can become larger in consecutive steps without necessarily indicating that the structure relaxation becomes unreliable. Three consecutive steps in which the energy variance was rising was found to be small enough to still ensure that the structure is not already too far away from the last reliable structure. The results obtained when using slightly different parameters (Fig. S1†) for structure optimizations of nanoclusters on surfaces and molecules on surfaces show that parameters are robust and relatively generally applicable to ML models trained on other types of systems. To further ensure that the optimization did not run out of the training regime, we terminated the algorithm after $f_{\mathrm{maxinit.}}$ was reached and after that, whenever the model energy variance reached a high value that we set to 1 eV or when the f_{max} jumped to a value that was larger than 2 eV Å^{-1}. Both events were observed when model predictions ran into regions not supported by training data. These additional parameters are only relevant for models that are trained on a small training set and ensure that the optimization is stopped before the training regime is left. At that point, the remaining optimizations can be carried out with the reference method. For $\mathrm{ML}_{\mathrm{adapt.3}}$ models, an f_{max} value of 0.05 eV Å^{-1} was able to be reached, hence the additional stopping criteria were not required using these refined models.

3. Results

3.1 Model performance

Fig. 2 shows model prediction errors for the vdW-free MLIPs for energies and forces and the Hirshfeld ratio ML models in panels a, b, and c, respectively, for Au@C and panels e, f, and g, respectively, for X2O@Ag models. The mean absolute errors (MAEs) and root-mean-squared errors (RMSEs) on the data points of the hold-out test set shown in Fig. 2 for energies, forces, and Hirshfeld volume ratios can be found in Table S1† in the ESI.

The MAE of the four models ranges from 0.017 to 0.021 eV for energies and 0.021–0.025 eV Å^{-1} for forces for X2O@Ag. ML models trained on Au@C have MAEs of 0.013 to 0.18 eV for energies and 0.014 to 0.26 eV Å^{-1} for forces. As can be seen, there are some outliers in the data set of Au@C with errors on these data points shown in the insets of top panels a and b. These data points are geometries with unfavorable structures and energies far out of the region in which most data points lie. These data points were included to ensure that the model was able to rank structures correctly and predict energetically unfavorable structures with high energies. For training on these data points, the L_2 loss was adapted to a smooth version of the L_1 loss, which is explained and defined in section S1.2.†

Besides data points representing unfavorable Au@C NCs with large vdW-free energies and vdW-free forces that were explicitly introduced into the training set, the ML models predict vdW-free energies, vdW-free forces, and Hirshfeld volume ratios accurately. The MAE for the Hirshfeld volume ratios, a quantity that ranges between about 0.6 and 1.05, is 3.9×10^{-4} and 1.1×10^{-4} for X2O@Ag and Au@C, respectively.

In the following, we will assess the performance of the proposed method by performing structure relaxations of geometries of two additional hold-out test sets for X2O@Ag and Au@C. These hold-out test sets comprise full structure optimizations and none of the geometry optimization steps during the relaxations were included for training.

3.2 Global structure search: gold nanoclusters on diamond (Au@C)

As NCs can exhibit many metastable geometries, we first assess the performance of our model with respect to interatomic distances and then evaluate the applicability of our approach to energetically differentiate between different cluster geometries.

For the first task, we use a test set of Au@C models that contains DFT+MBD optimizations of Au NCs on diamond (110) with cluster sizes of $n = 6, 15, 20, 25, 28, 30, 35, 40, 44, 45, 60$, and 66. On average, 95 optimization steps were required with DFT+MBD for one geometry optimization. All initial starting structures for geometry optimizations of NCs were created with ASE, where the NCs were placed onto the center of a diamond (110) surface. The same starting geometries as used in DFT structure optimizations were taken for structure relaxations with the final model obtained after the third adaptive sampling run, denoted $ML_{adapt.3}$+MBD. The minima found with $ML_{adapt.3}$+MBD were assessed according to the radial atom distributions of the Au NCs in Fig. 3a. Radial atom distributions obtained from structures using the $ML_{adapt.3}$+MBD scheme are similar to those from DFT+MBD. For the Au–Au radial atomic distribution in panel a, distances at values smaller than around 2.6 Å are removed by geometry optimization and the main distance distribution at around 2.8 Å aligns well with DFT+MBD. Slight deviations can be found at 2.5 Å for Au–C in panel b, which can also be seen in the radial atom distributions for the starting structures used for geometry optimizations (denoted as "init."). The peaks of the initial distribution are shifted towards the DFT+MBD peaks upon optimization.

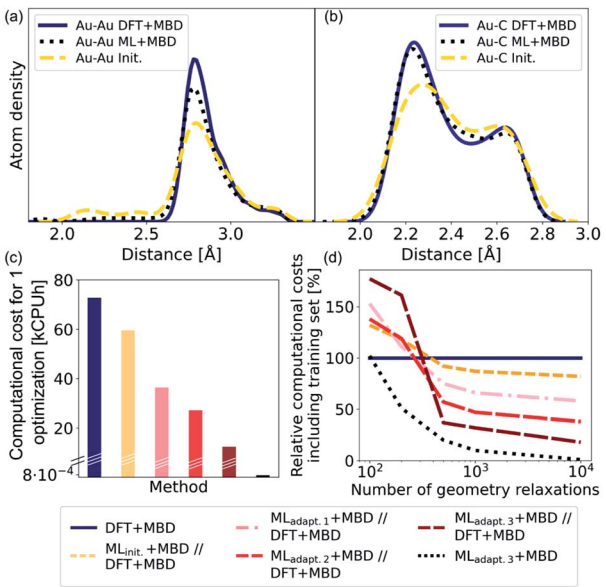

Fig. 3 (a) Kernel density estimate for the radial atom distribution of Au–Au and (b) Au–C bonds of Au@C systems for the optimized structures with DFT that comprise the training set and were computed with DFT+MBD (solid lines, denoted DFT+MBD). The starting structures for geometry optimizations are denoted using "init." and shown using dashed lines and the ML+MBD-optimized ($ML_{adapt.3}$+MBD) structures are shown in dotted lines. (c) Computational costs in kilo central processing unit hours (kCPUh) of a single Au@C structure relaxation performed with DFT+MBD (blue), and prerelaxations with ML+MBD models followed by further optimization with DFT+MBD (denoted ML+MBD//DFT+MBD). (d) Computational cost including model training cost as a function of the number of performed geometry relaxations. Computational costs were assessed by defining an average time per geometry optimization that was based on the initial training data.

The benefit of using ML+MBD instead of DFT+MBD lies in the reduction of computational effort associated with structure relaxations. Fig. 3c and d show the computational costs of structure relaxations with ML+MBD, DFT+MBD and a ML+MBD pre-optimization followed by a DFT+MBD optimization (denoted 'ML+MBD//DFT+MBD'). Panel c shows the cost of a single structure relaxation in kilo-central processing unit hours (kCPUh), recorded on dual AMD EPYC™ Zen2 7742 64-core processors at 2.25 GHz. As can be seen, the computational cost of ML+MBD optimization (black) is about 0.01% of the cost of DFT+MBD. However, it can be argued that the structure relaxations solely conducted with ML+MBD might not be accurate enough for a specific purpose and are not sufficiently close to DFT+MBD. To this aim, we performed DFT+MBD optimizations using the optimized structures obtained from the $ML_{init.}$ (yellow), $ML_{adapt.1}$ (pink), $ML_{adapt.2}$ (red), and $ML_{adapt.3}$ (dark red) models and summed up the computational expenses from respective ML+MBD and additional DFT+MBD calculations. In this approach, ML+MBD acts as a pre-optimization method. As expected, the computational cost increases when combining ML+MBD with DFT+MBD. However, the better the optimized structure resulting from the ML model, the fewer DFT+MBD optimization steps are required. This is why the combination of refined adaptive models with DFT incurs less computational cost for the same task than the initial model in combination with DFT.

Fig. 3d plots the computational cost of performing one to 10 000 structure optimizations of the different models including the cost of generating the training data set for the ML model construction. The costs are extrapolated and are shown relative to DFT+MBD (100%, dark blue). As can be seen from the dotted black lines, using the final ML model, $ML_{adapt.3}$+MBD can greatly reduce the computational costs whilst still achieving good accuracy (see panels a and b). Note that ML+MBD values include the cost of training data generation and model training. In case of large-scale screening studies, where many geometry optimizations are required, it is clearly beneficial to use refined and accurate ML+MBD models. In cases where high accuracy is required, a subsequent re-optimization with DFT+MBD to reach an f_{max} of < 0.01 eV Å^{-1} may be necessary. In this scenario, we find that the ML+MBD//DFT+MBD optimization sequence is only computationally beneficial to standalone DFT+MBD optimization if the number of required structural relaxations is between 100 and 500. In Fig. 3d, $ML_{init.}$ – $ML_{adapt.3}$ refers to models trained on more and more data points. The break-even point in terms of computational cost for ML+MBD//DFT+MBD is similar for all models, but lowest for "adapt.2" (about 100 structure relaxations) and highest for "init." (about 500 structure relaxations). This shows that there is a sweet spot for the construction of MLIPs between the cost of creating an (overly) large training data set and the computational time saving benefit.

To validate the reliability of the structure and adsorption energy prediction of the ML+MBD models for Au@C, three basin-hopping optimization runs that were carried out for the initial adaptive sampling runs for clusters of size $n = 6$, 15 and 40 were selected. The global minimum and two random local

minima were selected from each basin-hopping run for the different cluster sizes. The basin-hopping run for a cluster size of $n = 6$ is shown in Fig. 4a. The three structures used for validation are denoted S1–S3 (yellow in panel b) and were re-optimized with DFT+MBD (blue) and $ML_{adapt.3}$ (red) separately. In panel Fig. 4c, the structures of DFT+MBD are compared to those of $ML_{adapt.3}$+MBD. The structures are very similar to each other with slight deviations visible in geometry S3.

The energies of the three structures are plotted in Fig. 4d relative to the most stable structure. Even though the structures are not exactly the same, the energies are ranked similarly to each other. The ordering of the three structures is also correctly predicted with each method. As expected, the energy ranking of $ML_{adapt.3}$+MBD is closer to the relative energy ordering of DFT+MBD than the initial ML model. Panel d further shows the results of the same procedure carried out for cluster sizes of $n =$ 15 and 40, respectively. The structures for all clusters as predicted by all methods are visualized in Fig. S2† of the ESI. As can be seen, for the Au NC with 15 atoms, the energies are ordered incorrectly according to the initial model. The correct ordering of energies is established with the final model, $ML_{adapt.3}$+MBD, and is similar to DFT. However, the highest energy geometry is predicted to be more stable than in the reference. This result could be an indication that the least favorable structure with a size of 15 is in a region of the potential energy surface that is under-represented in the training set. Indeed, the energy variance according to the query-by-committee approach is 4 times higher for this structure (around 30 meV) than for the other clusters (around 7 meV). For the Au NC with 40 atoms, the initial model suggests three energetically different structures, while the $ML_{adapt.3}$+MBD and DFT+MBD methods suggest that the first two structures are identical in their energy. To conclude, ML combined with a long-range dispersion correction (MBD in this case) can reduce the computational effort of structure relaxations with DFT+MBD substantially. Given the rich diversity of structures and cluster sizes and the relatively few data points required, the model can be utilized as a pre-optimizer that leads to radial atom distributions close to the DFT+MBD optimum and can facilitate fast global structure searches including an approximate energy ranking of structures.

3.3 Adsorption of organic molecules on Ag(111)

Our second application case is based on organic molecules of the X2O family[9] on Ag(111), as shown in Fig. 2h. The existing training data set only includes few data points based on a small set of local geometry optimizations. We have defined a test set that contains randomly selected optimized structures held out from the training set. We removed several full structure optimizations, *i.e.*, the starting geometries, the intermediate steps, and the final optimized structures, from the training set to ensure no structure relevant for the test set is explicitly known by the models. The test set represents a small set of exemplary local minima of X2O molecules on a Ag(111) surface. The structures in the test set are denoted based on the type of organic molecule that is adsorbed on the surface, *i.e.*, B2O, A2O, and P2O. The indices after the molecule abbreviations indicate geometries that differ in their adsorption site, orientation or cell size. One test example shows a unit cell with two B2O molecules. Fig. 5a and c show the adsorption heights and adsorption energies, respectively, of the ML+vdW^{surf}-relaxed structures compared to the DFT+vdW^{surf}-relaxed structures. The adsorption energies were obtained using the ML+vdW^{surf} method and reference adsorption energies were obtained from the DFT+vdW^{surf}-optimized structures. Hence the energies in panel c are not obtained from identical geometries, but from the respective minimum energy structures of the methods. The adsorption energy is defined as $E_{ads+Ag} - E_{ads} - E_{Ag}$, with "ads" referring to the adsorbate and "Ag" to the metal surface. Relaxed geometries of the clean surface and the isolated molecule were used as references in the calculation of the adsorption energy, and a negative adsorption energy value corresponds to an exothermic process. Adsorption heights were computed as

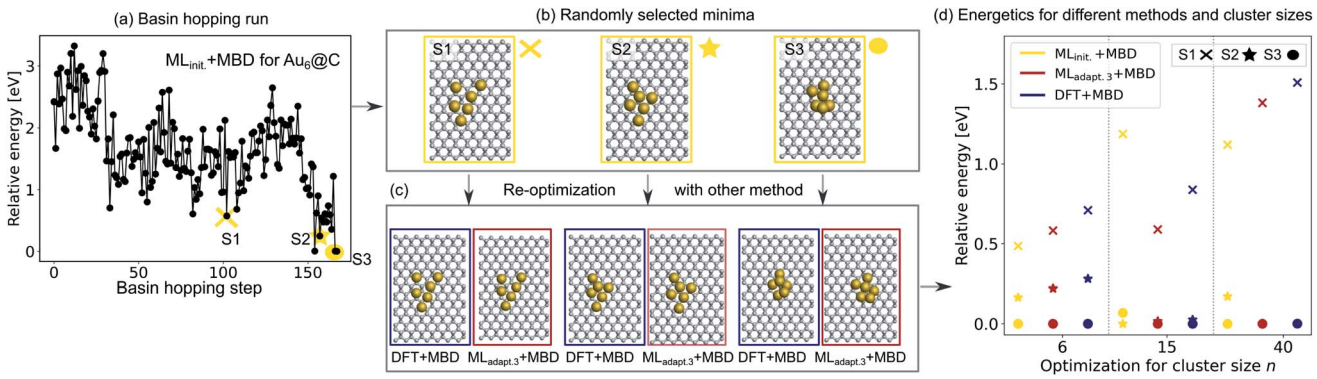

Fig. 4 (a) Basin hopping run with $ML_{init.}$ for Au@C with Au_6 nanoclusters (NCs). Yellow circles indicate (b) 3 selected structures S1–S3 that include the energetically lowest geometry and two randomly selected structures according to $ML_{init.}$ that are (c) reoptimized with DFT+MBD (blue) and $ML_{adapt.3}$+MBD (red). (d) Relative energies reported with respect to the energetically lowest cluster for each method. In addition, energy rankings of the energetically lowest structures and two randomly selected structures from basin hopping runs with NC sizes of 15 and 40 atoms using $ML_{init.}$+MBD (yellow), $ML_{adapt.3}$+MBD (red), and DFT+MBD (blue). Corresponding structures are shown for each method in Fig. S2.†

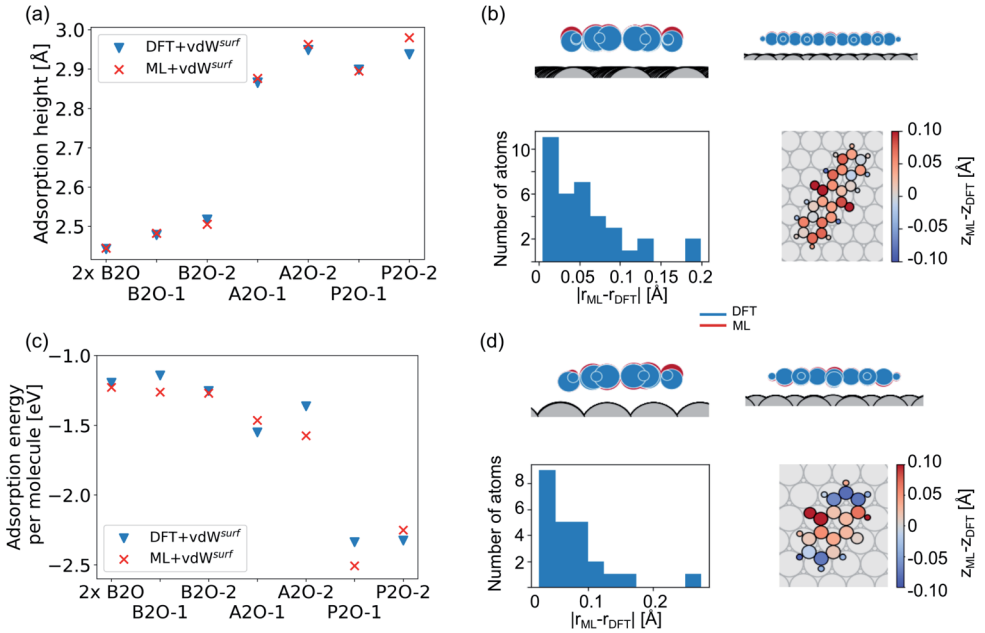

Fig. 5 (a) Adsorption heights (average heights of all atoms in the molecule compared to the average heights of the first Ag layer) and (c) adsorption energies of X2O@Ag of a hold-out test set computed with DFT+vdW^{surf} and ML+vdW^{surf}. The structures are single B2O molecules and two B2O molecules in a unit cell (denoted as "2×B2O"), A2O, and P2O on Ag(111) that differ in adsorption sites and orientations. (b and d) ML+vdW^{surf} structures (P2O-2 and A2O-2) compared to DFT+vdW^{surf} structures of panels (a) and (c).

distances of the average heights of the first Ag layer and the average heights of all atoms in the molecule.

The test to validate the new method is carried out as follows: the same starting geometries were used for ML+vdW^{surf} geometry relaxations as were used in DFT+vdW^{surf} reference optimizations. As can be seen from Fig. 4a, our method reports adsorption heights that are very similar to those obtained with DFT+vdW^{surf}. The structural similarity can be further assessed from panels b (P2O-2) and d (A2O-2), which show the ML+vdW^{surf} compared to DFT+vdW^{surf} structures with the worst agreement in adsorption heights between ML and DFT. The top images show ML+vdW^{surf}-optimized structures in red and DFT+vdW^{surf}-optimized structures in blue. Bottom images show the error of each atom in Å. The ML-predicted minimum energy structures are typically relatively close to DFT-predicted structures with the largest deviations in adsorption height per atom at about 0.2 Å. Most deviations are below 0.05 Å. Noticeably, these are not differences in bond lengths (Fig. S4†) but absolute positions in z direction. Visualizations for the remaining structures presented in 5a and c are shown in Fig. S3† of the ESI.

In addition to the adsorption heights, we sought to assess the adsorption energies for the purpose of relative energy predictions of adsorption phases with respect to each other. As can be seen from panel c, the trend observed in the reference data can mostly be reproduced when comparing different molecules. There is hardly any trend in over- or underestimation of adsorption energies and the mean error on adsorption energies is around 0.10 ± 0.06 eV. While the model can distinguish adsorption energies between different molecules, it fails to distinguish the adsorption energies for energetically beneficial local minima of the same molecule at different symmetry sites. Achieving this would likely require more training data than what was provided in the original NOMAD data repository.

As a more difficult challenge for the model, we generated an additional test set of 16 B2O structures on Ag(111) with DFT+vdW^{surf}, which are far from the surface. These structures required around five to six times more optimization steps than the calculations in the training set and thus provide a test set with initial structures that are much less favorable than those in the training set and the structures tested before. As mentioned briefly in the Methods section, geometry optimization algorithms struggle with geometries far away from the surface and require additional considerations. To counter this problem, a two-fold optimization was conducted with our method. First, all atomic positions of the molecule were fixed apart from motion along the [111] direction, with the Ag(111) substrate fully constrained. After this initial relaxation, the molecule was allowed to relax into all directions and the top three Ag layers were also allowed to relax. The rest of the Ag layers were kept frozen, as in ref. 9. To initialize the optimizations, we used the Lindh–Hessian[71,80] as was done in DFT+vdW^{surf} optimizations. The results are shown in Fig. 6a. Our model gives fair adsorption heights for the systems when compared to the DFT reference and can be used as a computationally efficient pre-relaxation procedure without ever learning from data of systems with large molecule-metal separation, as those were accounted for by the long-range dispersion correction. The mean error for adsorption heights is relatively low and around 0.04 ± 0.02 Å.

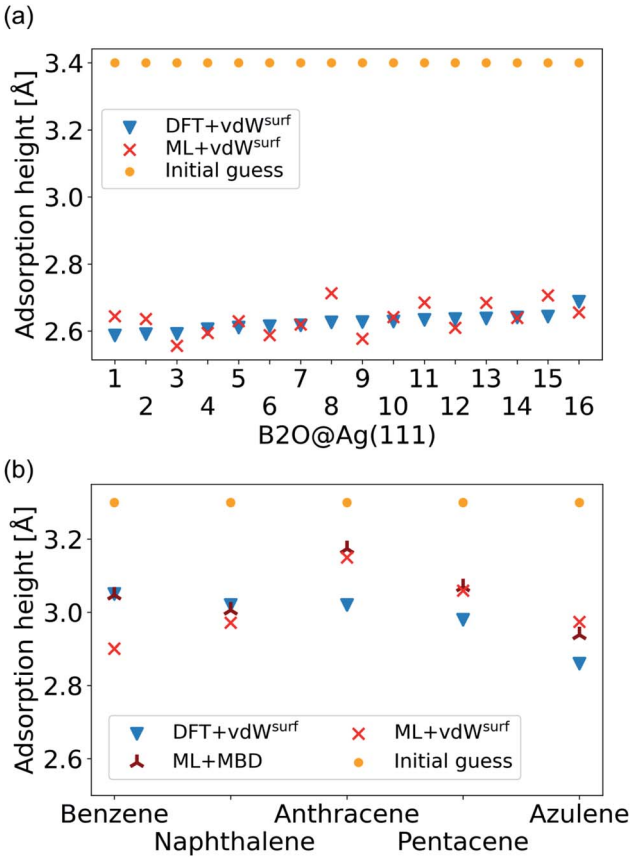

Fig. 6 (a) Adsorption heights of B2O molecules on Ag(111). (b) Adsorption heights of benzene,[82] naphthalene,[83] anthracene,[84] pentacene,[85] and azulene,[83] computed with ML+vdW^{surf} and compared to DFT+vdW^{surf}. The same adsorption sites as mentioned in the cited references (Table 1) are used.

The final challenge was to test our model for transferability to other organic molecules that have not been seen by the model. This would open the possibility to generate a fully transferable MLIP for hybrid metal–organic interfaces to be applied as a general structural pre-relaxation tool. We tested our approach on several different organic molecules adsorbed on Ag(111) that have been experimentally and computationally characterized previously, namely benzene, naphthalene, anthracene, pentacene (all from the acene family), and azulene. According to literature,[24,83,84,86] the most stable symmetry site was selected (indicated in Table 1 in the first column) to compare our results to available DFT data in literature and experimental data. We note that models trained on such sparse data will likely fail to reliably predict energy differences between different adsorption sites. The gas-phase optimized structure of each organic molecule was placed around 3.3 Å away from the surface. A similar two-step optimization procedure was applied as before. As shown in Fig. 6b, the trend in adsorption heights across molecules that is found with DFT+vdW^{surf} (blue triangles) can be reproduced with ML+vdW^{surf} (red crosses). The deviations are in the range of ±0.1 Å vertical adsorption height. Considering that none of the molecules were featured in the training data set, this demonstrates the increased transferability that the model inherits due to the separate treatment of long- and short-range interactions. The molecules that lead to the largest deviations in adsorption heights are azulene and anthracene. Besides low computational costs, a further advantage of the proposed method is that the vdW correction can be changed. To demonstrate the flexibility of our method we further relaxed structures at ML+MBD level and compute the related adsorption heights (dark-red star-like shapes). As can be seen from Fig. 6b, the adsorption heights are very close to ML+vdW^{surf}. Larger deviations are only seen when it comes to benzene. However, the prediction of ML+MBD is in line with the adsorption height of 2.97 Å reported in ref. 82 and 87.

In addition to adsorption heights, we sought to investigate whether the ML+vdW^{surf} method can be used to approximate adsorption energies. Table 1 shows the computed adsorption energies with both, ML+vdW^{surf} and ML+MBD. The trends observed in members of the acene family, *i.e.*, increasing adsorption energy with increasing molecular size, can be reproduced with both methods. However, some energies are overestimated, while others are underestimated with respect to DFT+vdW^{surf}, which correlates with adsorption heights being over- and underestimated, respectively, for all structures except for anthracene. Nevertheless, given the fact that these systems were never seen by the ML models and that the ML models are based on a small amount of data, the results are encouraging for a future development of fully transferable ML models for a wide range of physisorbed structures with only little amount of additional data. This could be applied to large-scale screening studies of organic molecules on surfaces and to perform structural pre-relaxations.

Table 1 Adsorption energies for benzene, naphthalene, anthracene, pentacene, and azulene, on Ag(111) on the most stable symmetry site based on literature, where negative values correspond to an exothermic process. Literature values are based on PBE+vdW^{surf}.[82,84,86,88] Values are compared to those of ML+vdW^{surf} and ML+MBD using the relaxed structures obtained with the respective method

	Adsorption energy [eV]			
Molecule (symmetry)	DFT+vdW^{surf}	ML+vdW^{surf}	DFT+MBD	ML+MBD
Benzene (hcp0)[24,82,87]	−0.75	−0.81	−0.57	−0.77
Naphthalene (top30)[88]	−1.08	−1.19	−0.77	−1.10
Anthracene (hcp0)[84]	−1.38	−1.53	−0.93	−1.12
Pentacene (bridge60)[84,86]	−2.40	−2.12	−1.65	−1.79
Azulene (top30)[88]	−1.37	−1.22	−0.91	−1.07

4. Conclusion

We have developed an approach for the efficient prediction of long-range-corrected potential energy surfaces and forces based on machine learning (ML) potentials and external long-range dispersion corrections based on Hirshfeld atoms-in-molecules partitioning. Different types of long-range van-der-Waals interactions were implemented including the Tkatchenko–Scheffler vdW(TS) and MBD methods to describe nanoclusters on surfaces and organic molecules on metal surfaces. One of the powerful features is, thus, that the type of long-range correction can easily be changed, such that different methods can be employed without the need for retraining.

To apply the method for structure pre-relaxations with ML models trained on little data, we additionally incorporated dynamic stopping criteria that take the variance of machine learning predictions into account and ensure the structure relaxation does not run into unreliable territory. The method was tested for fast (pre-)relaxations of complex hybrid systems. Firstly, we demonstrated our framework on gold nanoclusters on a diamond (110) surface and showed that by adaptively optimizing the ML models, global structure searches can be enabled that would be computationally too expensive without the use of ML.

Secondly, we reused data from ref. 9 of three organic molecules (X2O) on Ag(111) surfaces. The goal of this study was to assess the applicability of ML models based purely on reused data from open data repositories without generating a tailor-made training data set. This reflects the realistic application scenario in which a small set of initial geometry optimizations can be used to construct an ML+vdW model that can computationally expedite structural pre-relaxation. The conducted tests show not only the power of open data for developing new methods, but also demonstrate that the method can be used to semi-quantitatively predict adsorption heights and energies and to pre-relax challenging starting systems. Finally, we tested the transferability of our model to unseen organic molecules on Ag(111).

The approach we present is of general utility for the computational surface science community and has the potential to drastically reduce the computational effort of some of the most common tasks in this field. While our method accounts for long-range dispersion interactions, it does not explicitly treat electrostatic interactions. To account for this, the SchNet+vdW approach could be extended in a similar vein by learning partial atomic charges and using these to predict electrostatic long-range interactions, similar to SpookyNet[54] or Behler's fourth-generation high-dimensional neural networks.[56,89] Our data provides evidence that the construction of a more general and transferable structure relaxation model of hybrid organic–metallic interfaces is feasible and potentially desirable, although small (and rough) system-specific models may be more advantageous in many cases.

Data availability

Input and output files for all Au@C calculations, comprising the training data set and the adaptive run calculations, have been uploaded as a data set to the NOMAD electronic structure data repository and are freely available under **https://doi.org/10.17172/NOMAD/2021.10.28-1**.[90] The molecular geometries and corresponding properties of gold nanoclusters on diamond surfaces are saved in a database format provided by the Atomic Simulation Environment.[60] The data for X2O are obtained from NOMAD.[76–78] In addition, we include a script that shows how to replace the external vdW(TS) correction schemes[19,22] with the D3 method by Grimme *et al.*,[63] which is also interfaced with ASE. Tutorials for training ML models, generating a data set, and making ML-based optimizations with external vdW corrections (vdW(TS), vdW^{surf}, MBD, DFT-D3, and DFT-D4) (Jupyter Notebooks), files to reproduce figures, test data, and additional code to run ML models are available from figshare (**https://doi.org/10.6084/m9.figshare.19134602.v2**).[91]

Code availability

All code developed in this work is made available on figshare (**https://figshare.com/s/78b54de875cfb9cadbdd**) and GitHub including test examples under URL: **https://github.com/maurergroup/SchNet-vdW**. The used version is tagged as v0.1. In addition, the SchNetPack-version we use is available under URL: **https://github.com/juliawestermayr/schnetpack** and has been uploaded to figshare in its current state under **https://doi.org/10.6084/m9.figshare.19134602.v2**. The script to generate the Lindh Hessian for geometry initialization is available *via* FHI-aims.[71] A few other versions of the Lindh Hessian script are available *via* the gensec package[92] on GitHub: **https://github.com/sabia-group/gensec**.

Conflicts of interest

There is no conflict of interest to declare.

Acknowledgements

This work was funded by the Austrian Science Fund (FWF) [J 4522-N and Y 1157-N36], the EPSRC Centre for Doctoral Training in Diamond Science and Technology [EP/L015315/1] and the UKRI Future Leaders Fellowship programme [MR/S016023/1]. We are grateful for use of the computing resources from the Scientific Computing Research Technology Platform of the University of Warwick (including access to Avon, Orac and Tinis), the EPSRC-funded Northern Ireland High Performance Computing service [EP/T022175/1] for access to Kelvin2, and the EPSRC-funded High End Computing Materials Chemistry Consortium [EP/R029431/1] for access to the ARCHER2 UK National Supercomputing Service (**https://www.archer2.ac.uk**).

References

1 Y. Dong, X. Wu, X. Chen, P. Zhou, F. Xu and W. Liang, *Biomed. Pharmacother.*, 2021, **137**, 111236.
2 V. Gewin, *Nature*, 2021, **593**, 470.

3 D. Guo, R. Shibuya, C. Akiba, S. Saji, T. Kondo and J. Nakamura, *Science*, 2016, **351**, 361–365.
4 Z. P. Cano, D. Banham, S. Ye, A. Hintennach, J. Lu, M. Fowler and Z. Chen, *Nat. Energy*, 2018, **3**, 279–289.
5 C. Li, O. J. H. Chai, Q. Yao, Z. Liu, L. Wang, H. Wang and J. Xie, *Mater. Horiz.*, 2021, **8**, 1657–1682.
6 F. Bottari and K. De Wael, *J. Electroanal. Chem.*, 2017, **801**, 521–526.
7 C. H. Lalander, Y. Zheng, S. Dhuey, S. Cabrini and U. Bach, *ACS Nano*, 2010, **4**, 6153–6161.
8 J. A. Lloyd, A. C. Papageorgiou, S. Fischer, S. C. Oh, O. Sağlam, K. Diller, D. A. Duncan, F. Allegretti, F. Klappenberger, M. Stöhr, R. J. Maurer, K. Reuter, J. Reichert and J. V. Barth, *Nano Lett.*, 2016, **16**, 1884–1889.
9 A. Jeindl, J. Domke, L. Hörmann, F. Sojka, R. Forker, T. Fritz and O. T. Hofmann, *ACS Nano*, 2021, **15**, 6723–6734.
10 R. Otero, J. M. Gallego, A. L. V. de Parga, N. Martín and R. Miranda, *Adv. Mater.*, 2011, **23**, 5148–5176.
11 A. Tan and P. Zhang, *J. Phys.: Condens. Matter*, 2019, **31**, 503001.
12 S. J. Cobb, Z. J. Ayres and J. V. Macpherson, *Annu. Rev. Anal. Chem.*, 2018, **11**, 463–484.
13 K. L. Kelly, E. Coronado, L. L. Zhao and G. C. Schatz, *J. Phys. Chem. B*, 2003, **107**, 668–677.
14 J. He, C. He, C. Zheng, Q. Wang and J. Ye, *Nanoscale*, 2019, **11**, 17444–17459.
15 O. T. Hofmann, E. Zojer, L. Hörmann, A. Jeindl and R. J. Maurer, *Phys. Chem. Chem. Phys.*, 2021, **23**, 8132–8180.
16 S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, *Chem. Rev.*, 2016, **116**, 5105–5154.
17 J. Hermann, R. A. DiStasio and A. Tkatchenko, *Chem. Rev.*, 2017, **117**, 4714–4758.
18 R. J. Maurer, V. G. Ruiz, J. Camarillo-Cisneros, W. Liu, N. Ferri, K. Reuter and A. Tkatchenko, *Prog. Surf. Sci.*, 2016, **91**, 72–100.
19 A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, 2009, **102**, 073005.
20 V. G. Ruiz, W. Liu, E. Zojer, M. Scheffler and A. Tkatchenko, *Phys. Rev. Lett.*, 2012, **108**, 146103.
21 A. Tkatchenko, R. A. DiStasio, R. Car and M. Scheffler, *Phys. Rev. Lett.*, 2012, **108**, 236402.
22 A. Ambrosetti, A. M. Reilly, R. A. DiStasio and A. Tkatchenko, *J. Chem. Phys.*, 2014, **140**, 18A508.
23 R. J. Maurer, V. G. Ruiz and A. Tkatchenko, *J. Chem. Phys.*, 2015, **143**, 102808.
24 W. Liu, F. Maaß, M. Willenbockel, C. Bronner, M. Schulze, S. Soubatch, F. S. Tautz, P. Tegeder and A. Tkatchenko, *Phys. Rev. Lett.*, 2015, **115**, 036104.
25 P. J. Blowey, B. Sohail, L. A. Rochford, T. Lafosse, D. A. Duncan, P. T. P. Ryan, D. A. Warr, T.-L. Lee, G. Costantini, R. J. Maurer and D. P. Woodruff, *ACS Nano*, 2020, **14**, 7475–7483.
26 L. Hörmann, A. Jeindl and O. T. Hofmann, *J. Chem. Phys.*, 2020, **153**, 104701.
27 O. A. Vydrov, J. Heyd, A. V. Krukau and G. E. Scuseria, *J. Chem. Phys.*, 2006, **125**, 074106.
28 A. Karolewski, L. Kronik and S. Kümmel, *J. Chem. Phys.*, 2013, **138**, 204115.
29 B. Hourahine, *et al.*, *J. Chem. Phys.*, 2020, **152**, 124101.
30 M. Stöhr, G. S. Michelitsch, J. C. Tully, K. Reuter and R. J. Maurer, *J. Chem. Phys.*, 2016, **144**, 151101.
31 A. Fihey, C. Hettich, J. Touzeau, F. Maurel, A. Perrier, C. Köhler, B. Aradi and T. Frauenheim, *J. Comput. Chem.*, 2015, **36**, 2075–2087.
32 A. P. Bartók, M. C. Payne, R. Kondor and G. Csányi, *Phys. Rev. Lett.*, 2010, **104**, 136403.
33 A. P. Bartók and G. Csányi, *Int. J. Quantum Chem.*, 2015, **115**, 1051–1057.
34 T. Young, T. Johnston-Wood, V. L. Deringer and F. Duarte, *Chem. Sci.*, 2021, **12**, 10944–10955.
35 K. T. Schütt, H. E. Sauceda, P. J. Kindermans, A. Tkatchenko and K. R. Müller, *J. Chem. Phys.*, 2018, **148**, 241722.
36 K. T. Schütt, P. J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko and K. R. Müller, *Advances in Neural Information Processing Systems*, 2017, pp. 992–1002.
37 K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko and K.-R. Müller, *J. Chem. Theory Comput.*, 2019, **15**, 448–455.
38 O. T. Unke and M. Meuwly, *J. Chem. Theory Comput.*, 2019, **15**, 3678–3693.
39 J. Behler, *Int. J. Quantum Chem.*, 2015, **115**, 1032–1050.
40 J. Behler, *Angew. Chem., Int. Ed.*, 2017, **56**, 12828–12840.
41 J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 146401.
42 M. S. Jørgensen, H. L. Mortensen, S. A. Meldgaard, E. L. Kolsbjerg, T. L. Jacobsen, K. H. Sørensen and B. Hammer, *J. Chem. Phys.*, 2019, **151**, 054111.
43 H. L. Mortensen, S. A. Meldgaard, M. K. Bisbo, M.-P. V. Christiansen and B. Hammer, *Phys. Rev. B*, 2020, **102**, 075427.
44 A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu and D. Hassabis, *Nature*, 2020, **577**, 706–710.
45 R. Meyer and A. W. Hauser, *J. Chem. Phys.*, 2020, **152**, 84112.
46 Y. Yang, O. A. Jiménez-Negrón and J. R. Kitchin, *J. Chem. Phys.*, 2021, **154**, 234704.
47 S. Chmiela, H. E. Sauceda, K.-R. Müller and A. Tkatchenko, *Nat. Commun.*, 2018, **9**, 3887.
48 M. Bogojeski, L. Vogt-Maranto, M. Tuckerman, K.-R. Müller and K. Burke, *Nat. Commun.*, 2020, **11**, 5223.
49 V. L. Deringer, N. Bernstein, G. Csányi, C. B. Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold and S. R. Elliott, *Nature*, 2021, **589**, 59–64.
50 K. Heather, *et al.*, *Electron. Struct.*, 2022, DOI: **10.1088/2516-1075/ac572f**.
51 T. Morawietz, V. Sharma and J. Behler, *J. Chem. Phys.*, 2012, **136**, 064103.
52 T. Morawietz and J. Behler, *J. Phys. Chem. A*, 2013, **117**, 7356–7366.
53 K. Yao, J. E. Herr, D. Toth, R. Mckintyre and J. Parkhill, *Chem. Sci.*, 2018, **9**, 2261–2269.

54 O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schütt, H. E. Sauceda and K.-R. Müller, *Nat. Commun.*, 2021, **12**, 7273.
55 T. W. Ko, J. A. Finkler, S. Goedecker and J. Behler, *Acc. Chem. Res.*, 2021, **54**, 808–817.
56 T. W. Ko, J. A. Finkler, S. Goedecker and J. Behler, *Nat. Commun.*, 2021, **12**, 398.
57 L. Zhang, H. Wang, M. C. Muniz, A. Z. Panagiotopoulos, R. Car and W. E., 2021, arXiv:2112.13327.
58 H. Muhli, X. Chen, A. P. Bartók, P. Hernández-León, G. Csányi, T. Ala-Nissila and M. A. Caro, *Phys. Rev. B*, 2021, **104**, 054106.
59 T. Bereau and O. A. von Lilienfeld, *J. Chem. Phys.*, 2014, **141**, 034101.
60 A. H. Larsen, *J. Phys.: Condens. Matter*, 2017, **29**, 273002.
61 F. Hirshfeld, *Theor. Chem. Acc.*, 1977, **44**, 129–138.
62 *Libmbd*, **https://github.com/libmbd/libmbd**.
63 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
64 E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2019, **150**, 154122.
65 M. Dion, H. Rydberg, E. Schröder, D. C. Langreth and B. I. Lundqvist, *Phys. Rev. Lett.*, 2004, **92**, 246401.
66 M. Mura, A. Gulans, T. Thonhauser and L. Kantorovich, *Phys. Chem. Chem. Phys.*, 2010, **12**, 4759–4767.
67 Y. Freund, H. Seung, E. Shamir and N. Tishby, *Mach. Learn.*, 1997, **28**, 133–168.
68 P. Melville and R. J. Mooney, *Proceedings of the 21st International Conference on Machine Learning (ICML-2004)*, 2004, pp. 584–591.
69 N. Ferri, R. A. Distasio, A. Ambrosetti, R. Car and A. Tkatchenko, *Phys. Rev. Lett.*, 2015, **114**, 176802.
70 *SchNet-vdW*, **https://github.com/maurergroup/SchNet-vdW**.
71 V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler, *Comput. Phys. Commun.*, 2009, **180**, 2175–2196.
72 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
73 J. Westermayr, M. Gastegger, M. F. S. J. Menger, S. Mai, L. González and P. Marquetand, *Chem. Sci.*, 2019, **10**, 8100–8107.
74 D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A*, 1997, **101**, 5111–5116.
75 D. J. Wales and H. A. Scheraga, *Science*, 1999, **285**, 1368–1372.
76 A. Jeindl, O. Hofmann and L. Hörmann, *A2O structure search*, 2021, DOI: **10.17172/NOMAD/2021.03.09-2**.
77 A. Jeindl, O. Hofmann and L. Hörmann, *B2O structure search*, 2021, DOI: **10.17172/NOMAD/2021.03.09-3**.
78 A. Jeindl, O. Hofmann and L. Hörmann, *P2O structure search*, 2021, DOI: **10.17172/NOMAD/2021.03.09-1**.
79 J. Neugebauer and M. Scheffler, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1992, **46**, 16067–16080.
80 R. Lindh, A. Bernhardsson, G. Karlström and P.-Å. Malmqvist, *Chem. Phys. Lett.*, 1995, **241**, 423–428.
81 M. Gastegger, J. Behler and P. Marquetand, *Chem. Sci.*, 2017, **8**, 6924–6935.
82 W. Liu, J. Carrasco, B. Santra, A. Michaelides, M. Scheffler and A. Tkatchenko, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **86**, 245405.
83 B. P. Klein, J. M. Morbec, M. Franke, K. K. Greulich, M. Sachs, S. Parhizkar, F. C. Bocquet, M. Schmid, S. J. Hall, R. J. Maurer, B. Meyer, R. Tonner, C. Kumpf, P. Kratzer and J. M. Gottfried, *J. Phys. Chem. C*, 2019, **123**, 29219–29230.
84 J. M. Morbec and P. Kratzer, *J. Chem. Phys.*, 2017, **146**, 034702.
85 S. Duhm, C. Bürker, J. Niederhausen, I. Salzmann, T. Hosokai, J. Duvernay, S. Kera, F. Schreiber, N. Koch, N. Ueno and A. Gerlach, *ACS Appl. Mater. Interfaces*, 2013, **5**, 9377–9381.
86 D. Käfer, PhD thesis, Ruhr-University Bochum, 2008.
87 W. Liu, F. Maaß, M. Willenbockel, C. Bronner, M. Schulze, S. Soubatch, F. S. Tautz, P. Tegeder and A. Tkatchenko, *Phys. Rev. Lett.*, 2015, **115**, 036104.
88 S. R. Kachel, B. P. Klein, J. M. Morbec, M. Schöniger, M. Hutter, M. Schmid, P. Kratzer, B. Meyer, R. Tonner and J. M. Gottfried, *J. Phys. Chem. C*, 2020, **124**, 8257–8268.
89 J. Behler, *Chem. Rev.*, 2022, **121**, 10037–10072.
90 S. Chaudhuri, J. Westermayr and R. J. Maurer, *Au@C for SchNet+vdW*, 2022, DOI: **10.17172/NOMAD/2021.10.28-1**.
91 J. Westermayr, S. Chaudhuri and R. J. Maurer, *SchNet+vdW*, 2022, **https://figshare.com/articles/software/SchNet_vdW/19134602**.
92 D. Maksimov and M. Rossi, *GenSec*, 2021, DOI DOI: **10.5281/zenodo.5651513**.