Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Enhancing Molecular Energy Predictions with Physically Constrained Modifications to the Neural Network Potential[†]

Weiqiang Fu,[*a‡] Yujie Mo,[a‡] Yi Xiao,[a] Chang Liu,[a] Feng Zhou,[a] Yang Wang,[a] Jielong Zhou,[*a] and Yingsheng Zhang[*a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Exclusively prioritizing the precision of energy prediction frequently proves inadequate in satisfying multifaceted requirements. A heightened focus is warranted on assessing the rationality of potential energy curves predicted by machine learning-based force fields (MLFF), alongside evaluating the pragmatic utility of these MLFF. This study introduces SWANI, an optimized Neural Network Potential (NNP) stemming from the ANI framework. Through the incorporation of supplementary physical constraints, SWANI aligns more cohesively with chemical expectations, yielding rational potential energy profiles. It also exhibits superior predictive precision compared to the ANI model. Additionally, a comprehensive comparison is conducted between SWANI and a prominent Graph Neural Network (GNN)-based model. The findings indicate that SWANI outperforms the latter, particularly for molecules exceeding the dimensions of the training set. This outcome underscores SWANI's exceptional capacity for generalization and its proficiency in handling larger molecular systems.

1 Introduction

Molecular energy predictions play a pivotal role in a diverse array of scientific and industrial applications, ranging from drug discovery and materials science to environmental chemistry and catalysis.[1,2] The development of accurate and efficient methods for estimating molecular energies has been a longstanding challenge in computational chemistry. Traditionally, computational chemistry techniques can be used to simulate chemical systems through two different approaches. The first approach, which relies on quantum chemical (QC) methods, is computationally demanding and is only suitable for small-sized systems. QC methods are sometimes referred to as first principle or ab initio methods since they approximately solve the molecular Schrödinger equation[3,4]. The second approach, which employs molecular mechanics (MM), can model large systems like proteins but ignores electrons. Classical MM views atoms and bonds as balls and springs, and calculates energy as a function of nuclear positions only[4].

The rapid advancement of computational chemistry has spurred the development of novel techniques for predicting molecular properties, enabling accelerated exploration of chemical space and facilitating insights into complex molecular behaviors.[3–8] Among these approaches, the neural network potential (NNP), a machine learning-based model, has gained significant attention for its capability to predict molecular energies with notable accuracy.[9] The NNP, often referred to as a machine learning force field (MLFF), leverages the power of artificial neural networks to capture intricate relationships between atomic coordinates and corresponding potential energy surfaces.[10–12] This promises a more efficient alternative to traditional quantum mechanical calculations, making it an indispensable tool for molecular simulations, drug discovery, and materials design.[13–16]

The core principle of the NNP lies in its ability to learn the underlying potential energy function directly from a training dataset containing molecular configurations and their corresponding reference energies.[10] By efficiently approximating this complex energy landscape, the NNP bridges the gap between accuracy and computational efficiency, significantly reducing the computational cost compared to ab initio quantum methods. Numerous studies have explored the NNP.[10,13–15,17–19]. Among these studies, kernel-based methods and neural networks (NN)-based methods are the most used to learn the potential enegy of molecules. Kernel-based methods (*e.g.*, GDML[18] and sGDML[20,21]) can get high performence in the case of limited samples, but due to the limitations of Kernel Ridge Regression (KRR), they only gain limited benefits with the dataset grows. Beyond that, they are not designed as unified MLFF models capable of predicting various types of molecules. Pinheiro *et al.* summarized that NN-based methods can be divided into two types based on whether the lo-

---

[a] *Beijing StoneWise Technology Co Ltd., Haidian Street 15, Haidian District, Beijing 100080, China. Tel: +1(508)397-7337; E-mail: xxxx@aaa.bbb.ccc*

[†] Electronic Supplementary Information (ESI) available. See DOI: 00.0000/00000000.

[‡] These authors contributed equally to this work.

cal descriptor is learnable: NN with fixed local descriptor (NN-fLD) using fixed local descriptor such as Behler and Parrinello symmetry functions (BPSFs)[10] and atomic environment vectors (AEVs)[11] as a molecular representation; while NN with learned local descriptor (NN-lLD) assign node vectors to each atom, using message passing to update local chemical environment based on graph (convolutional) neural networks (GNNs)[22].

Despite its remarkable success, the NNP is not without limitations. The inherent black-box nature of neural networks can hinder the model's interpretability and may lead to predictions that lack physical meaning. Moreover, the NNP's performance can degrade when extrapolating beyond the training data range, making it challenging to ensure accurate predictions for molecules in diverse conformational states. Addressing these challenges, recent research has focused on enhancing the NNP's accuracy, transferability, and physical interpretability.[23–26]

In this study, we improved molecular energy predictions by incorporating physically constrained modifications into the Neural Network Potential. By introducing additional physical constraints and insightful adjustments, the modified model, referred to as SWANI, aims to overcome the limitations of the conventional NNP. This study builds upon the success of the ANI[11] approach and leverages physically meaningful features to ensure accurate asymptotic energy properties and overall smoothness. In addition, we revised the training data set ANI-1x.[27] Through extensive evaluation, including comparisons with state-of-the-art models, our experimental results demonstrate the enhanced accuracy and transferability of the proposed model, especially exhibit greater stability during molecular dynamics simulations. We expect our framework to push the prediction of molecular properties to a real-time level and to be statistically meaningful to guide the molecular simulations.

2 Methods

2.1 SWANI model architecture design

SWANI is based on the ANAKIN-ME (ANI) method that is one of the popular transferable neutral network-based molecular potential.[11] Figure 1 provides an overview of the SWANI architecture. To offer a clearer understanding of our model, we will provide a brief overview of the ANI model first. The ANI model consists of a set of atomic number specific NNPs, denoted as $\{F_x\}$. For the i^{th} atom of a molecule with atomic number x, the atomic environment vector (AEV), $\vec{G_i}^x = \{G_1, G_2, G_3, ...G_M\}$, calculated by molecular coordinates, is designed to give a numerical representation of both radial and angular features. The $\vec{G_i}^x$ is then fed into F_x to predict atomic energy contributions, $E_i{}^x$, and then to obtain the total energy of a molecule, E^T.

$$E_T = \sum_{i}^{all\ atoms} E_i = \sum_{i}^{all\ atoms} F^x(\vec{G_i}^x) \qquad (1)$$

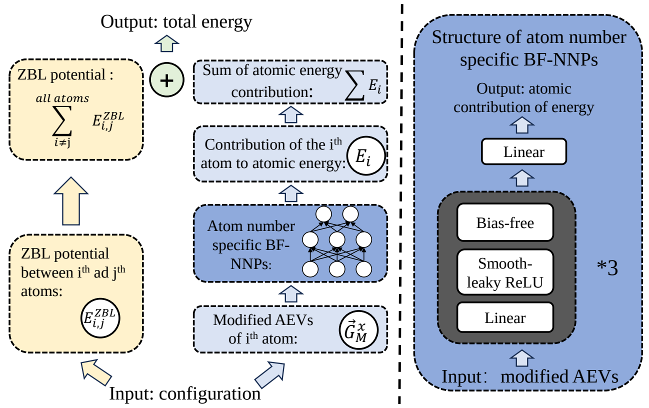

Fig. 1 Illustration of the SWANI architectural overview (left), and the neural network structure (right).

Although ANI demonstrates commendable performance in terms of accuracy and computational efficiency, it tends to generate substantial errors when confronted with configurations falling outside its training range. This limitation is partially attributable to the data set's construction and sampling methodology, which predominantly comprises conformations situated near equilibrium states. It is underrepresented for conformations that are far away from the equilibrium bond length. Including a significant amount of non-equilibrium structures can overcome some of those defects,[28] however, it is not always possible for biomolecular systems. Therefore, the short-range and long-range potential that struggles to be adequately trained via neural networks, should be taken into account at the beginning of the model design. This phenomenon can be readily demonstrated through potential energy curves of binary atomic systems, which exhibits a direct relationship between potential energies and the interatomic separation. As the separation between the atoms increases, the energy approaches zero, indicating the absence of contact between them. Conversely, at extremely close distances, repulsive forces dominate, leading to a steep rise in potential energy. The interplay between attractive and repulsive effects reaches equilibrium at the minimum point of the energy curve. As demonstrated in Figure 2, we conducted dissociation simulations of the H_2 molecule using a range of molecular configurations that deviate from the equilibrium bond length. In the absence of physical constraints, most machine learning force field (MLFF) models struggle to accurately represent potential energy at close distances, as evidenced by the dissociation energy curves for the original ANI model (Figure 2 ANI) and DimeNetPP (Figure 2 DimeNetPP), among others (details in Supporting Information). Additionally, these energy curves exhibit anomalous fluctuations at the equilibrium distance and fail to converge to zero as the interatomic separation increases.

In order to overcome the constraints inherent in the initial model, several modifications were introduced to enhance the original version of the ANI model. The energy of free atoms (atomization energy) should be equal to 0. For the case where two atoms are infinitely separated and there is no interaction between them, the input AEV $\vec{G}$ is 0. To ensure the energy is equal to the sum of the energies of the two free atoms, we removed additive terms (bias) from every stage of the NNPs, resulting in a bias-free NNPs (BF-NNPs). In addition to ensuring the correct prediction of physical properties, bias is considered to be the culprit of overfitting problem, when facing with configuration outside of the train-

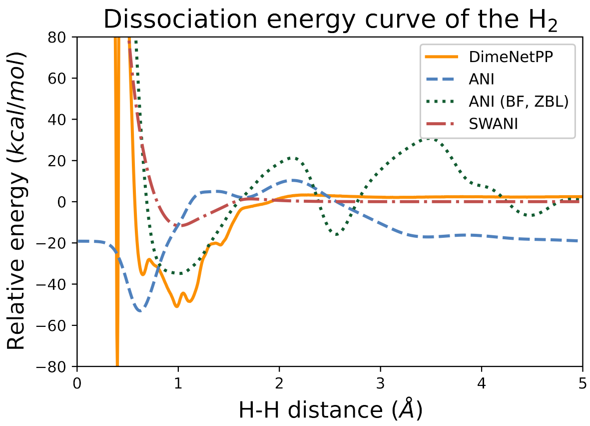

Fig. 2 Dissociation energy curves of the H_2 molecule. The DimeNetPP curve is predicted by the original DimeNetPP model.[17] The ANI curve is predicted by the original ANI model.[11] The ANI(BF, ZBL) curve is calculated by ANI model with bias-free NNs and ZBL potential. The SWANI curve is generated using SWANI model with bias-free, smooth-leakyReLU NNs trained with spectral norm regularization (SN) and addition of ZBL potential. All models were trained on the ANI-1x-re data set.

ing range[29]. Removing all additive constants allows the model to obtain a stronger generalization. To compensate for the reduction in fitting ability resulting from the removal of the bias, we incorporate a wider linear layer in our NNPs.

To better describe the short distance repulsion potential, Ziegler-Biersack-Littmark (ZBL) nuclear repulsion potential[30] was added in the total energy to capture the right physics while atoms near each other. The ZBL potential is calculated using:

$$E_{ij}^{ZBL} = \frac{1}{4\pi\varepsilon_0}\frac{Z_i Z_j e^2}{r_{ij}}\phi(r_{ij}/a) + S(r_{ij}) \tag{2}$$

where e is the electron charge, ε_0 is the electrical permittivity of vacuum, Z_i, Z_j are the nuclear charges of the two atoms, $a = 0.46850/(Z_i^{0.23} + Z_j^{0.23})$, $\phi(x) = 0.18175e^{-3.19980x} + 0.50986e^{-0.94229x} + 0.28022e^{-0.40290x} + 0.02817e^{-0.20162x}$ and $S(r)$ is the switching function. The ZBL terms were added before the model output layer. The total energy of a molecule predicted by SWANI, E_T has the form of a sum over all i “atomic energy contributions” and ZBL potential energy between pairs of atoms:

$$E_T = \sum_{i}^{all\ atoms} F_{BF}^{x}(\vec{G}_i^{\,x}) + \sum_{i,j\neq i}^{all\ atoms} E_{ij}^{ZBL}(Z_i, Z_j, r_{ij}) f_{cutoff}(r_{ij}) \tag{3}$$

Where f_{cutoff} is a cutoff function defined as Equation (6). Figure 2 ANI(BF,ZBL) shows the dissociation energy curve of H_2 computed by the ANI model that has removed the bias and included the ZBL potential, which clearly exhibits the improvements of the predicted potential energy at short and long internuclear distance.

Furthermore, when AEV is 0, the NN should be continuously differentiable, and the forces (first derivative) and the Hessian (second derivative) should be 0. To ensure this property, we proposed a new activation function called smooth-leakyReLU activation function (SL), depicted as follows:

$$\sigma(x) = \begin{cases} x + b_1 & x > x_1 \\ \alpha x^n & x_0 < x \leq x_1 \\ k_{neg} x + b_0 & x \leq x_0 \end{cases} \tag{4}$$

here, α, n, k_{neg} are hyperparameters. α, k_{neg} are positive values. n should be a positive odd value to ensure monotonicity of the activation function. x_0, x_1, b_0, b_1 are values adjusted to α, n, k_{neg} to ensure the continuity and differentiability of the activation function.

2.2 Modified atomic environment vectors

The molecular representation used in this model is the modified version of atomic environment vectors (AEVs) from ANI model.[11] The AEV is composed of radial and angular function to probe the local chemical environment around the atom. To approximate the local atomic environment and the effects attenuating over distance, a piecewise cutoff function is applied:

$$f_{cutoff}(r_{ij}) = \begin{cases} 0.5 \times cos(\frac{r_{ij}\pi}{r_{cutoff}}) + 0.5 & \text{if } r_{ij} \leq r_{cutoff} \\ 0 & \text{if } r_{ij} > r_{cutoff} \end{cases} \tag{5}$$

where r_{ij} is the distance between the atom i and j, and r_{cutoff} is the cutoff radius that is assigned manually.

The original radial and angular environments were examined by two Gaussian distribution functions, which was originally from Behler and Parrinello symmetry function[10]. However, due to the symmetry of Gaussian function, the two sides of the equilibrium distance have the same results, which means the long-range effect is the same as the short-range one within certain limits. Therefore, the sigmoid function was applied in the radial function, instead. The radial function is as shown in Equation (6). The angular function is the same as the ANI method[11] without any modifications as shown in Equation (7). Each function is over a batch of η, r_s, θ_s, and ζ parameters.

$$G_m^R(r_{ij}, r_s) = \Sigma_{j\neq i}^{allatoms} sigmoid(-\eta(r_{ij} - r_s)) \cdot f_C(r_{ij}, r_{cutoff}) \cdot f_C(r_s, 2r_{cutoff})^{0.5} \tag{6}$$

$$G_m^A = 2^{1-\zeta} \Sigma_{j,k\neq i}^{allatoms} (1 + \cos(\theta_{ijk} - \theta_s))^{\zeta} e^{-\eta(\frac{r_{ij}+r_{ik}}{2} - r_s)^2} f_C(r_{ij}) f_C(r_{ik}) \tag{7}$$

2.3 Spectrum norm regularization

In this section, we will elucidate the concept of spectral norm regularization and its role in enhancing the performance of our model. Spectral norm regularization (SN) has been commonly employed to improve the generalizability of deep learning models. Notably, it has been effectively used in generative adversarial networks (GANs) to address issues related to instability during training and the necessity for fine-tuning hyperparameters to ensure convergence.[31] Inspired by this successful application, we anticipate that integrating spectral norm regularization into our SWANI model will result in improved overall performance and enhanced smoothness.

Our primary objective is to achieve a more continuous and robust potential surface. Regarding model performance, We aim to ensure that our model remains robust to small perturbations in the input, as it will not be significantly affected by them. Specifically, when conformations undergo slight changes in position or angle, the energy predictions need to remain continuous and robust.

In our fundamental architecture, we employ a 3-layer bias-free network, which can be described as $x^l = f(W^l x^{l-1})$ where $l = 1,...,L$ denotes the number of layer, x^{l-1} represents the output of the $(l-1)^{th}$ layer, f denotes the activation function, and in mour model, we utilize the smooth-leakyReLU function. The weight matrices for each layer are denoted as $W^l \in R^{n_l \times n_{l-1}}$. The parameter set $\theta = \left\{W^l\right\}_{l=1}^{L}$ characterizes our NN, and it can be described as $f_\theta(x^0) = x^L$, where x^0 is the input and x^L is the output of the final layer L.

Therefore, we consider a small perturbation vector ξ with a small l_2 norm. Employing the smooth-leakyReLU activation function f, which is Lipschitz continuous with a Lipschitz constant of 1, we establish the relationship $\frac{\left\|f(W^l(x+\xi))-f(W^l x)\right\|_2}{\|\xi\|_2} \leq \frac{\left\|W^l \xi\right\|_2}{\|\xi\|_2} \leq$ $\sigma(W^l)$, where $\sigma(W^l)$ represents the spectral norm of W^l. The spectral norm is defined as : $\sigma(W^l) = \max\limits_{\xi \in R^n, \xi \neq 0} \frac{\left\|W^l \xi\right\|_2}{\|\xi\|_2}$, corresponding to the largest singular value of W^l.By controlling W^l to limit its spectral norm, the output of each layer can robustly handle perturbed input.

Subsequently, we delve into the examination of the smoothness property concerning the 3-layer bias-free network implemented in SWANI. As demonstrated by Equation 8, the smoothness of our model can be effectively governed by regulating the spectral norm of each layer. To achieve this, we incorporate the spectral norm regularizer, $\sum_{l=1}^{L} \sigma(W^l)^2$, into the loss function. Through the application of gradient descent, this regularization term actively works to reduce the spectral norms of W^l, thereby fostering enhanced smoothness within the model. Figure 2 SWANI(BF, ZBL, SL, SN) show the dissociation energy curve of H_2 computed by the SWANI with SL, trained with SN. The curve appears to be smoother overall.

$$\frac{\left\|f_\theta(x+\xi) - f_\theta(x)\right\|_2}{\|\xi\|_2^L} \leq \frac{\left\|KW^L KW^{L-1} ... KW^1 \xi\right\|_2}{\|\xi\|_2^L} \leq \prod_{l=1}^{L} \sigma(W^l) \quad (8)$$

2.4 Dataset

Constructing an optimally varied data set tailored for the training of machine learning models is challenging. In this work, we utilized the ANI-1x data set, but recalculated the molecular structures, and implemented further refined adjustments, resulting in the creation of a data set referred to as ANI-1x-revision (ANI-1x-re). The ANI-1x dataset encompasses nearly five million distinct molecular conformations, derived from DFT calculations.[27] This compilation contains a subset of configurations that deviate from equilibrium, including instances of covalent bond rupture scenarios.

The prevalent understanding is that practical utilization of spin-restricted Kohn–Sham density functional theory is limited in its ability to precisely address bond-breaking phenomena.[32] This is primarily due to the single-reference nature of the underlying wavefunction and the absence of functionals that encompass robust correlation effects capable of surpassing this single-reference constraint.[33] Nonetheless, noteworthy advancements have been made recently towards the practical implementation of strong-correlation density functional theory methods.[34,35] However, despite this progress, in scenarios where the introduction of spin-symmetry breaking is acceptable, unrestricted Kohn–Sham (UKS) methods are acknowledged for providing accurate energy outcomes during bond-breaking occurrences along the majority of the potential energy curve, particularly in the case of certain straightforward diatomic molecules.[36] Consequently, within our data set, molecules manifesting bond-breaking attributes were treated utilizing UKS methods.

The calculation of molecular systems characterized by strong static electron correlation (SEC) present considerable computational challenges, necessitating a qualitative comprehension to assess the suitability of approximate quantum chemistry (QC) methods. To address this, Stefan Grimme and his colleagues have introduced a straightforward static electron correlation diagnostic that not only facilitates practical application but also unveils the essence of static electron correlation.[37,38] This diagnostic approach is based on fractional occupation density functional theory (DFT) or finite temperature DFT, termed fractional occupation number weighted electron density (FOD).[37] This diagnostic employs the integration of the FOD across spatial coordinates, yielding a single size-extensive parameter (designated as N_{FOD}) that serves as a global metric for quantifying SEC.[37]

Throughout this study, unless specified otherwise, the optimization of structures was conducted using the r^2SCAN-3c level of theory, a methodology introduced by Stefan Grimme and coworkers in 2021.[39] This approach has demonstrated superior performance compared to prominent hybrid-DFT/QZ approaches, while maintaining computational costs at significantly lower orders of magnitude.[39] The FOD analysis was applied to detect molecular configurations manifesting multi-reference (MR) effects. Notably, a substantial and broadly delocalized FOD, where N_{FOD} exceeds the empirically chosen threshold of 0.05, signifies a genuine MR scenario, which is subsequently treated using UKS-DFT. The FOD analysis was conducted employing the default FOD-theory level (TPSS/def2-TZVP) with a smearing temperature of 5000K. The wavefunction theory computations were conducted employing the ORCA software suite with version 5.0.[40,41]

To further analyze the extensibility and transferability of our model, we employed the COmprehensive Machine-learning Potential (COMP6) as test set.[42] The COMP6 benchmark, proposed by Smith *et al.*, encompasses six rigorous benchmarks that collectively span a wide spectrum of organic and bio-chemical space, for molecules containing C, N, O, and H atoms. These six benchmark subsets are referred to as S66x8[43], GDB7to9, GDB10to13[44–46], Tripeptides, ANI-MD and DrugBank[47].

2.5 Training

We train our SWANI model on a data set comprising over 80% of the 4.95 million data points in the ANI-1x-re data set. The remaining 20% of the data points are used as the validation set. During training, SWANI model was optimized to predict both molecular energies and forces. Notably, the forces predicted by the model correspond to the analytical derivatives of the molecular energies, ensuring the conservation of energy during simulation runs. The loss function consists of the energy term and force term:

$$L = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{{M_i}^{\frac{1}{2}}}(\hat{E}_i - E_i)^2 + \frac{l_0}{M_i}\sum_{j=1}^{M_i}(\hat{f}_{ij} - f_{ij})^2\right] \tag{9}$$

where $\hat{E}_i$ and $\hat{F}_{ij}$ are the energies and forces for a given molecule, and E_i and F_{ij} are the ground truth of energies and forces, respectively. θ is the trainable parameters. l_0 is the weight of force loss, we use $l_0 = 0.1$ to control the component of energy and force during training. N is the number of the molecules, and M_i is the atoms of i-th molecule. The MSE loss of energy and force is appropriately scaled by the square root of the number of atoms and the number of atoms, respectively[48].

In order to train smooth neural NNPs, a spectral norm regularization term was incorporated into the loss function to penalize the upper bounds of the Lipschitz constants. The loss function with spectral norm relularization was:

$$\hat{L} = L + \alpha \sum_{i}^{all\ atoms} e^{-\frac{n^x}{2}} \sum_{l=1}^{L} \sigma(W_x^l)^2 \tag{10}$$

Where $\alpha = 1.0 \times 10^{-6}$ is the base regularization factor, x is the atomic number of i-th atom, $e^{-\frac{n^x}{2}}$ is the atomic number specific regularization factor, where n^x refers to the number of outermost electrons. Considering the principle of size-extensivity, it is expected that the energy will exhibit a positive correlation with the number of electrons. $\sigma(W_x^l)$ is the spectral norm of weight matrix of l-th layer for the atomic number specific NNPs.

The network weights are optimized with SGD using the ADAM optimizer[49]. An initial learning rate of 0.001 is used with the other ADAM parameters set to $\beta_1 = 0.9$, $\beta_2 = 0.99$, and $\varepsilon = 1.0 \times 10^{-8}$, as recommended by the ADAM authors. Training of SWANI model were run for 1000 epochs utilizing the ReduceLRonPlateau schedule with patience of 50 epochs and decay factor of 0.5 with a batch size of 2500.

2.6 Explanation by SHAP based feature analysis

To analyze the different prediction performance of the GNN-based model and AEV-based model, we constructed machine learning model utilizing molecular features to predict the error associated with each model's predictions. We computed a set of 204 molecular descriptors for each molecule using the Descriptors module from the Python RDKit[50], which can be categorized as follows: general descriptors (*e.g.*, molecular weight), topological descriptors (*e.g.*, connectivity index), fragment descriptors (*e.g.*, number of aromatic amines), etc.

To deal with heterogeneous tabular data of descriptors[51], we employed Random Forest (RF) model[52] to predict performance of different model. Mean absolute error (MAE) of the energy of each molecule is used as the regression target of RF model, and we perform log transformation on the MAE to solve the problem that some values of MAE are too large.

Acknowledging the limited interpretability of RF models as black-box models[53]. We employed SHAP[54] (SHapley Additive exPlanations) framework to interpret predictions. Based on cooperative game theory, the concept of Shapley values[55] has been used to develop the SHAP framework, which proposes SHAP values as a comprehensive metric for assessing the feature importance. SHAP values assign a specific contribution to each feature, indicating the impact on predictions when that feature is taken into account. This enables sample-wise explanations, allowing the identification of crucial molecular features that significantly contribute to prediction errors.

3 Results and discussion

To establish the comprehensive precision and adaptability of our model, we subjected it to evaluation across six benchmark datasets: S66X8, GDB-11, GDB-13, Tripeptide, ANI-MD bench and Drugbank. A comparative analysis was performed against the initial ANI model[48] as well as three prominent models grounded in message passing, namely SchNet[56], DimeNetPP[17,57], and SphereNet[58]. Notably, all baseline models were trained employing 80% of the ANI-1x-re dataset's data points, with the remaining 20% reserved for validation. In conformity with established practices, configurations for all baseline methods were either adopted from referenced literature or sourced directly from the original authors.

3.1 Prediction performance of potential energies and forces

To assess the predictive efficacy of potential energy and force, as well as the model's transferability, we subjected both the SWANI model and the baseline models, trained on ANI-1x-re, to evaluation using the COMP6 dataset. Given the GPU memory constraints associated with SphereNet and DimeNetPP, certain molecules with a substantial number of atoms (Max atoms > 110) were omitted from the ANI-MD benchmark. The assessment involved a comparative analysis between the SWANI model and the baseline models, with the root mean squared error (RMSE) serving as the designated metric for energy and force evaluation. The summary of these comparative outcomes in terms of RMSE for both energy (Table 1) and forces (Table 2) is presented.

While SphereNet and DimeNetPP exhibit commendable performance on ANI-1x-re and smaller datasets (s66x8, gdb11, and gdb13), models based on AEV representations, such as ANI and SWANI, surpass them on the ANI-MD bench and Drugbank datasets. Notably, in contrast to Graph Neural Network (GNN)-based models, SWANI demonstrates robustness against variations in atom counts. Furthermore, SWANI demonstrates superior performance relative to ANI in terms of RMSE for energy across all datasets and for forces in the majority of datasets, except for s66x8 and Drugbank. These findings compellingly indicate that our introduced modifications indeed enhance the performance of the original ANI model.

3.2 Molecular feature attribution analysis of model prediction errors

For the purpose of elucidating the factors contributing to model prediction discrepancies, a molecular feature attribution analysis is conducted within the Drugbank dataset. The dataset, comprising 13,379 configurations, is segregated into 761 distinct molecules according to their respective molecular classifications. Subsequently, a collection of 204 two-dimensional (2D) molecular descriptors is employed as features for each molecule, facilitating the in-depth investigation of model prediction errors.

A RF model was employed with a 10-fold cross-validation to estimate the test set error. In each of the 10 cross-validation iterations, SHAP (SHapley Additive exPlanations) values of molecular descriptors were computed for the corresponding test set. Notably, the outcomes of molecular feature attribution analysis obtained from both DimeNetPP and SphereNet, as well as SWANI and ANI, manifest a substantial level of congruence. In our comparative assessment between GNN-based and AEV-based models, our primary focus centers on scrutinizing the distinctions between DimeNetPP and SWANI. This emphasis is reflected in 3a and 3b, which depict correlation plots of prediction errors in DimeNetPP and SWANI against their respective predicted values. Correspondingly, Figure 3c and 3d exhibit SHAP summary plots, offering insight into the feature attribution of prediction errors for both models.

Both SWANI and DimeNetPP demonstrate enhanced accuracy in their predictions for molecules characterized by lower HeavyAtomMolWt, wherein HeavyAtomMolWt signifies the molecular weight excluding hydrogen atoms. Notably, the molecular connectivity indices (Chi3v, Chi2v, Kappa1)[59] exert a pivotal influence on the prediction errors inherent to the DimeNetPP models. DimeNetPP exhibits elevated errors in predicting molecules marked by higher Chi indices, specifically Chi3v. The index order signifies the graph edges within the corresponding subgraph, with the subscript 'v' denoting its valence nature, which is established based on valence delta values. Conversely, SWANI encounters increased errors in predicting molecules associated with higher BertzCT[60] values, a topological index reflecting complexity and escalating with augmented atom numbers, atom types, and bond types. Furthermore, SWANI reveals substantial inaccuracies in predicting molecules characterized by exceedingly low BCUT2D_MRLOW (molar refractivity eigenvalue low)[61] values.

In order to facilitate a more distinct comprehension, our focus narrowed down to two specific molecules: Alectinib and Hydrocortamate. Despite sharing a comparable HeavyAtomMolWt, these molecules exhibit disparate prediction outcomes in DimeNetPP and SWANI. Specifically, Hydrocortamate presents suboptimal predictive results in DimeNetPP, while Alectinib encounters analogous challenges in SWANI predictions. Figure 4a and 4b illustrate the 2D molecular structures of these two molecules, while Figure 4c and 4d exhibit their corresponding SHAP waterfall diagrams.

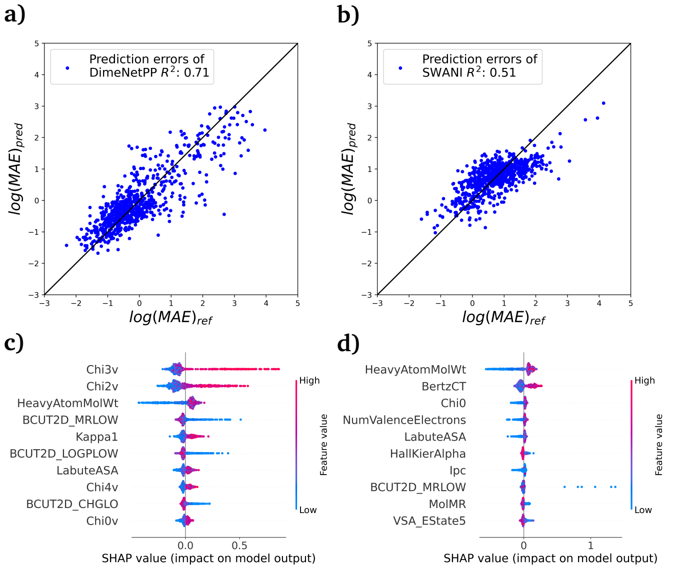

Fig. 3 a) and b) Correlation plots illustrating the relationships between the MAE reference values ($log(MAE)_{ref}$) for DimeNetPP and SWANI, with their corresponding predictions generated by the RF model using molecular features ($log(MAE)_{pred}$). c) and d) The SHAP summary plots showcasing the top-10 important molucule features (sort by gloabl feature importance: $mean(|SHAP|)$) for DimeNetPP and SWANI, respectively. Each dot on the plot represents a molecule, where the color of the dot indicates the size of a feature value associated with the molecule. The x-axis represents the SHAP values (influence of this feature on the prediction result). Larger SHAP values indicate that a molecule's presence of this particular feature poses challenges for the model in making accurate predictions (with higher MAE).

In the context of DimeNetPP, the intricate nature of predicting Hydrocortamate is rooted in its elevated second and third-order chi indices. The elevated chi indices originating from the intricate and extensive interatomic connectivity patterns within the molecule give rise to heightened structural complexity, consequently posing a challenge for precise predictions using the DimeNetPP model. The predictive performance of DimeNetPP is compromised when dealing with larger molecules compared to those present in the training set, primarily due to data limitations, particularly concerning higher Chi indices.

Conversely, Alectinib's molecular configuration encompasses a diverse array of bond types, resulting in a heightened BertzCT value. This heightened complexity poses a challenge for SWANI in precisely forecasting Alectinib's properties, thereby highlighting the inherent difficulty in AEV-based feature engineering when modeling molecules containing an assortment of bond and atom types.

3.3 Molecular Dynamics simulation

In order to assess the efficacy of all models within the context of molecular dynamics (MD) simulations, we adhere to the MD evaluation framework introduced by Fu *et al.*. This evaluation involves the selection of five molecules sourced from the Drugbank dataset, encompassing atom counts ranging from 30 to 71. For each of these molecules, we randomly designated five conformations to serve as the initial inputs. Subsequently, we executed a total of five ab initio molecular dynamics (AIMD) simulations and

Table 1 Comparisons between SWANI model and other models in terms of RMSE of energy in kcal mol^{-1}. The best results are shown with underlines.

RMSE (kcal mol^{-1})	Max atoms	SchNet	SphereNet	DimeNetPP	ANI	SWANI (ours)
ANI-1X (validation)	63	3.64	0.83	1.06	2.14	1.85
s66x8	34	6.37	0.59	0.77	4.73	4.58
gdb11	32	3.04	0.59	0.74	2.28	2.16
gdb13	38	6.28	1.15	1.40	3.95	3.71
tripeptide	69	119.54	4.23	3.83	4.36	3.50
ANI-md bench	103	227.32	9.81	9.88	2.58	2.51
Drugbank	140	76.26	5.75	5.47	4.96	4.70

Table 2 Comparisons between SWANI model and other models in terms of RMSE of forces in kcal mol^{-1} Å^{-1}. The best results are shown with underlines.

RMSE (kcal mol^{-1} Å^{-1})	Max atoms	SchNet	SphereNet	DimeNetPP	ANI	SWANI (ours)
ANI-1X (validation)	63	10.04	2.72	3.22	4.09	3.70
s66x8	34	6.22	5.84	1.00	2.26	2.30
gdb11	32	7.65	1.55	1.98	3.41	3.29
gdb13	38	11.25	2.29	2.79	4.89	4.80
tripeptide	69	29.50	2.53	2.47	4.77	2.93
ANI-md bench	103	21.30	3.50	3.65	2.92	2.71
Drugbank	140	20.87	3.41	4.50	3.34	3.34

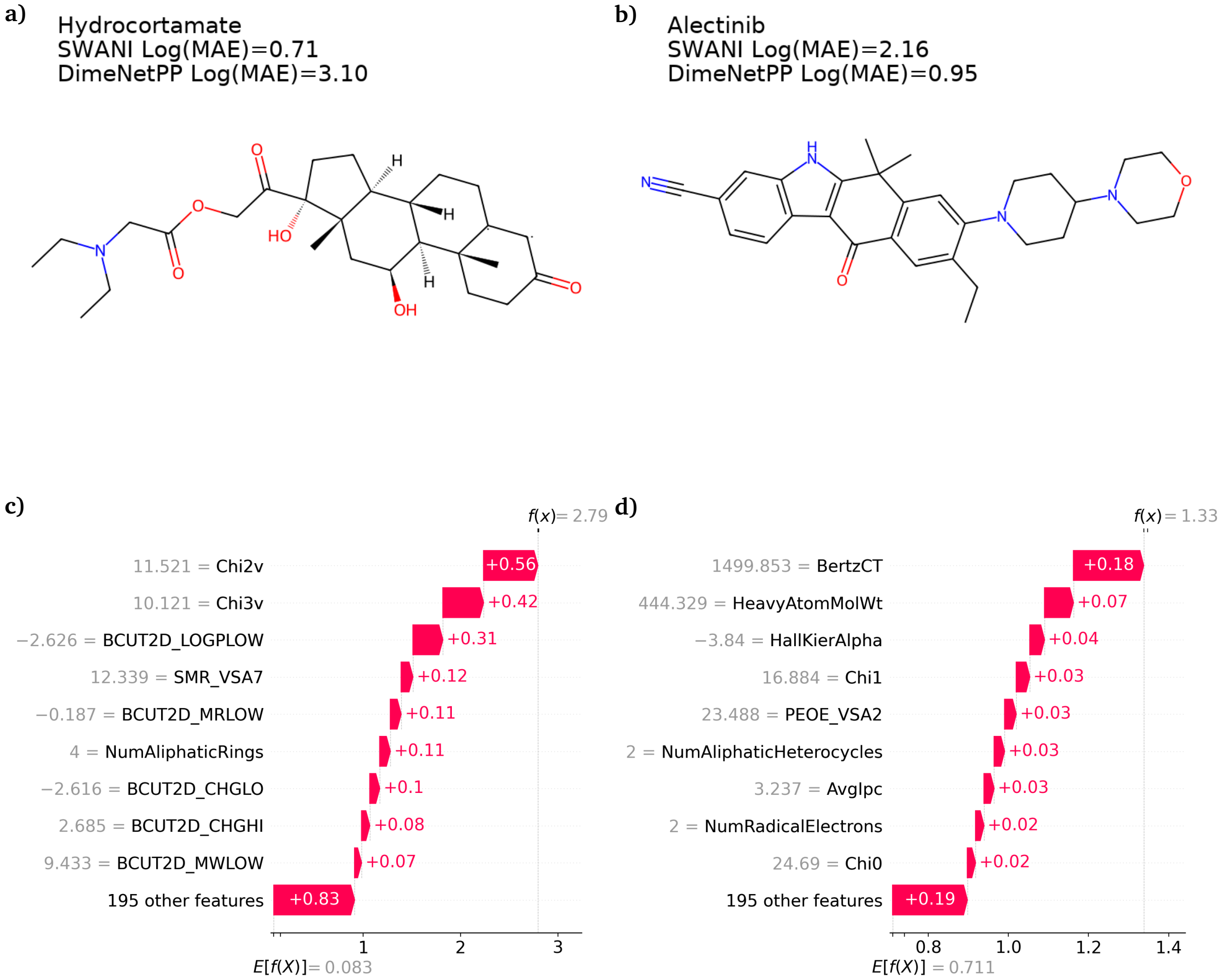

Fig. 4 a) and b) 2D molecular structure diagrams of Hydrocortamate and Alectinib. c) and d) The waterfall plots of Hydrocortamate and Alectinib, illustrating the top-10 important molecule features to predict the model prediction errors of DimeNetPP and SWANI, respectively.

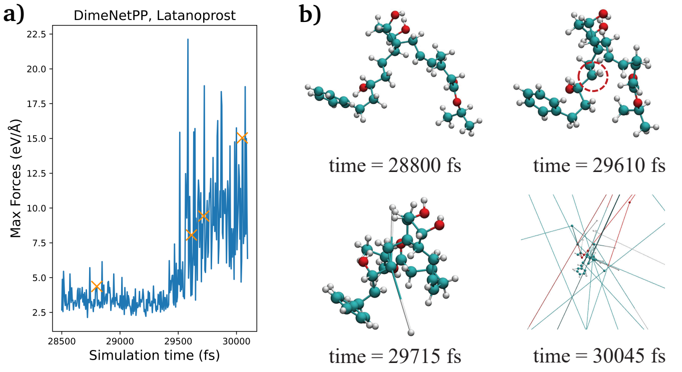

Fig. 5 Examples of simulation collapse when applying DimeNetPP to Latanoprost. a) depicts the time-dependent curve of the maximum force exerted on the atoms within the molecule, showcasing the occurrence of collapse. The orange cross indicates the visualized time steps in b). b) illustrates the process of system collapse in the MD simulation, with the circled portion is the notable nonphysical region.

five MLFF MD simulations for every molecule. To accommodate our computational resources, the AIMD simulations are limited to a duration of 10 picoseconds, whereas the MLFF MD simulations extend to 50 picoseconds. All simulations were conducted under isothermal conditions at a consistent temperature of 300 K, employing a Berendsen thermostat, and utilizing a time step of 0.5 fs.

Furthermore, we note the potential occurrence of simulation collapse within a brief temporal interval subsequent to a phase of stable simulation. This collapse phenomenon is illustrated in Figure 5, which closely resembles a comparable collapse scenario proposed by Fu et al. In specific localized regions (as indicated by the red circle in Figure 5b), the collision of two hydrogen atoms gives rise to an anomalous configuration, subsequently inducing substantial forces causing the hydrogen atoms to repel one another. This initiates a cascading effect, leading to the overall system collapse, thereby causing the disruption of numerous bonds within the molecule.

We employed a stability criterion to further identify instances of simulation collapse.[62] Within the framework of flexible molecules, we assess stability by scrutinizing variations in bond lengths. Specifically, we classify a simulation as entering an "unstable" state at a specific time instant T when the following condition is met:

$$\underset{(i,j)\in B}{max}\left|\left(\left\|x_i(T)-x_j(T)\right\|-b_{i,j}\right)\right|>\Delta \tag{11}$$

In this expression, B denotes the set of all bonds, i and j represent the two terminal atoms of the bond, $b_{i,j}$ signifies the equilibrium bond length, and Δ is a predetermined threshold. For our analysis, we uniformly set Δ to 20 Å for all molecules under consideration.

To assess the credibility of a configuration within the MD trajectory, the distribution of interatomic distances serves as a viable evaluation metric.[62] This metric offers a compact representation of the three-dimensional structure in a reduced-dimensional format. Given a specific configuration denoted as x, the interatomic distance distribution $h(r)$ is computed as follows:

$$h(r)=\frac{1}{N(N-1)}\sum_{i}^{N}\sum_{j\neq i}^{N}\delta\left(r-\left\|x_i-x_j\right\|\right) \tag{12}$$

where r signifies the distance from a reference particle, N denotes the total number of particles, the indices i and j denote the pairs of atoms contributing to the distance statistics, and δ represents the Dirac Delta function utilized for value distribution extraction. The computation and averaging of $h(r)$ are performed exclusively using frames derived from the stable phase within the 10ps total simulation, spanning from beginning to the point of collapse.

The comparison results of MD simulation are summarized in Table 3. When applying the GNN-based model to molecules with a substantial number of atoms, the MD simulation reveals a collapse phenomenon. In contrast, the AEV-based model does not exhibit such behavior, remaining stable throughout the MD simulation. SWANI outperforms the other models in $h(r)$ metrics for three out of five molecules and obtains the highest average score.

The summary of MD simulation comparison results is provided in Table 3. For molecules with a substantial number of atoms, the application of the GNN-based model possibly yields a collapse phenomenon in the MD simulation. In contrast, the AEV-based model displays a conspicuous absence of such behavior, maintaining stability throughout the entirety of the MD simulation. Notably, SWANI surpasses other models in terms of $h(r)$ metrics for three out of the five molecules, achieving the highest average score among them.

4 Conclusions

In this work, we present SWANI, a modified version of NNP based on ANI. By incorporating additional physical constraints, SWANI achieves a better alignment with chemical expectations by obtaining more rational potential energy curves. It also demonstrates significantly higher prediction accuracy when compared to ANI. We also conducted a comparative analysis between SWANI and the currently popular GNN-based model. Our experiment unveiled that SWANI outperforms the latter when dealing with molecules larger than those in the training set. This observation serves to demonstrate SWANI's remarkable generalization performance and its scalability to handle large systems effectively.

However, SWANI exhibits lower accuracy compared to DimeNetPP and SphereNet when dealing with molecules containing a small number of atoms. The result of molecular feature attribution analysis of model prediction errors indicates that BertzCT significantly contributes to limiting the accuracy of SWANI, possibly due to the challenge of AEV in handling multiple types of bonds and atoms. Molecular characterization through AEV is crucial for AEV-based NNPs. Future research should focus on improving AEV's ability to characterize complex molecules and enhancing its discrimination of different bonds and atom representations. These efforts can address SWANI's limitations and enhance its overall performance.

Conflicts of interest

There are no conflicts to declare.

Table 3 Results on DrugBank. The distribution of interatomic distances $h(r)$ MAE is unitless, in 10^{-3}. Standard deviation from 5 simulations is in subscript for applicable metrics. Three of these molecules did not generate enough stable trajectories during the MD simulation to produce h(r) when SchNet was applied to them.

	model	SchNet	SphereNet	DimeNetPP	ANI	SWANI
molecule (Atoms)						
N-Acetylglucosamine	stability	$1.00_{(0)}$	$1.00_{(0)}$	$1.00_{(0)}$	$1.00_{(0)}$	$1.00_{(0)}$
(30)	h(r)	$7.39_{(0.15)}$	$5.76_{(1.96)}$	$4.94_{(1.22)}$	$4.38_{(0.49)}$	$4.82_{(0.57)}$
Spermine	stability	$0.05_{(0.02)}$	$1.00_{(0)}$	$1.00_{(0)}$	$1.00_{(0)}$	$1.00_{(0)}$
(40)	h(r)	-	$7.75_{(3.50)}$	$6.72_{(0.98)}$	$7.84_{(4.25)}$	$6.33_{(1.82)}$
Folic Acid	stability	$0.28_{(0.36)}$	$1.00_{(0)}$	$1.00_{(0)}$	$1.00_{(0)}$	$1.00_{(0)}$
(51)	h(r)	$4.82_{(0)}$	$3.64_{(0.31)}$	$3.41_{(0.38)}$	$5.60_{(3.43)}$	$3.30_{(0.25)}$
Valsartan	stability	$0.01_{(0.00)}$	$0.50_{(0.31)}$	$0.31_{(0.18)}$	$1.00_{(0)}$	$1.00_{(0)}$
(61)	h(r)	-	$4.15_{(1.20)}$	$5.52_{(3.09)}$	$6.55_{(1.36)}$	$3.84_{(0.67)}$
Latanoprost	stability	$0.00_{(0.00)}$	$0.56_{(0.25)}$	$0.53_{(0.09)}$	$1.00_{(0)}$	$1.00_{(0)}$
(71)	h(r)	-	$3.86_{(3.59)}$	$2.51_{(1.93)}$	$4.92_{(3.03)}$	$3.82_{(1.80)}$
Mean	stability	0.27	0.81	0.77	1.00	1.00
(51)	h(r)	6.10	5.03	4.62	5.86	4.42

Acknowledgements

We acknowledge and thank Xiaoman Wang, Jianbo Wang, and Rongfei Xu for their valuable assistance in the management of computational resources and the execution of data set calculations. Their contributions have greatly facilitated our research endeavors.

Notes and references

1 A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi and M. Ceriotti, *Science Advances*, 2017, **3**, e1701816.

2 H. Li, Z. Zhang and Z. Liu, *Catalysts*, 2017, **7**, year.

3 T. Morawietz and N. Artrith, *Journal of Computer-Aided Molecular Design 2020 35:4*, 2020, **35**, 557–586.

4 A. Rimola, S. Ferrero, A. Germain, M. Corno and P. Ugliengo, *Minerals 2021, Vol. 11, Page 26*, 2020, **11**, 26.

5 G. Li, Y. Zhu, Y. Guo, T. Mabuchi, D. Li, S. Huang, S. Wang, H. Sun and T. Tokumasu, *ACS Applied Materials & Interfaces*, 2023, **15**, 5099–5108.

6 Y. Guo, H. Sun, M. An, T. Mabuchi, Y. Zhao and G. Li, *Nanoscale*, 2023, **15**, 12737–12747.

7 J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K. R. Müller and A. Tkatchenko, *Chemical Reviews*, 2021, **121**, 9816–9872.

8 G. B. Goh, N. O. Hodas and A. Vishnu, *Journal of Computational Chemistry*, 2017, **38**, 1291–1307.

9 M. Rupp, *International Journal of Quantum Chemistry*, 2015, **115**, 1058–1073.

10 J. Behler and M. Parrinello, *Physical review letters*, 2007, **98**, 146401.

11 J. S. Smith, O. Isayev and A. E. Roitberg, *Chemical science*, 2017, **8**, 3192–3203.

12 K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller and A. Tkatchenko, *Nature Communications*, 2017, **8**, 13890.

13 L. D. Jacobson, J. M. Stevenson, F. Ramezanghorbani, D. Ghoreishi, K. Leswing, E. D. Harder and R. Abel, *Journal of Chemical Theory and Computation*, 2022, **18**, 2354–2366.

14 H. E. Sauceda, L. E. Gálvez-González, S. Chmiela, L. O. Paz-Borbón, K. R. Müller and A. Tkatchenko, *Nature Communications*, 2022, **13**, year.

15 F. Noé, A. Tkatchenko, K. R. Müller and C. Clementi, *Annual Review of Physical Chemistry*, 2020, **71**, 361–390.

16 M. Rupp, A. Tkatchenko, K. R. Müller and O. A. Von Lilienfeld, *Physical Review Letters*, 2012, **108**, 058301.

17 J. Gasteiger, S. Giri, J. T. Margraf and S. Günnemann, Machine Learning for Molecules Workshop, NeurIPS, 2020.

18 S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt and K.-R. Müller, *Science advances*, 2017, **3**, e1603015.

19 A. P. Bartók, M. C. Payne, R. Kondor and G. Csányi, *Phys. Rev. Lett.*, 2010, **104**, 136403.

20 S. Chmiela, H. E. Sauceda, K.-R. Müller and A. Tkatchenko, *Nature communications*, 2018, **9**, 3887.

21 S. Chmiela, V. Vassilev-Galindo, O. T. Unke, A. Kabylda, H. E. Sauceda, A. Tkatchenko and K.-R. Müller, *Science Advances*, 2022.

22 M. Pinheiro, F. Ge, N. Ferré, P. O. Dral and M. Barbatti, *Chemical Science*, 2021, **12**, 14396–14413.

23 D. Montes de Oca Zapiain, M. A. Wood, N. Lubbers, C. Z. Pereyra, A. P. Thompson and D. Perez, *npj Computational Materials*, 2022, **8**, 189.

24 D. M. Anstine and O. Isayev, *The Journal of Physical Chemistry A*, 2023, **127**, 2417–2431.

25 T. Zubatiuk and O. Isayev, *Accounts of Chemical Research*, 2021, **54**, 1575–1585.

26 S. Käser, L. I. Vazquez-Salazar, M. Meuwly and K. Töpfer, *Digital Discovery*, 2023, **2**, 28–58.

27 S. S. Justin, Z. Roman, B. Nebge and L. Nicholas, 2020, 1–10.

28 J. S. Smith, B. Nebgen, N. Mathew, J. Chen, N. Lubbers, L. Burakovsky, S. Tretiak, H. A. Nam, T. Germann, S. Fensin and K. Barros, *Nature Communications*, 2021, **12**, 1257.

29 S. Mohan, Z. Kadkhodaie, E. P. Simoncelli and C. Fernandez-Granda, *arXiv preprint arXiv:1906.05478*, 2019.

30 J. F. Ziegler and J. P. Biersack, *Treatise on Heavy-Ion Science*,

1985, 93–129.

31 T. Miyato, T. Kataoka, M. Koyama and Y. Yoshida, *arXiv preprint arXiv:1802.05957*, 2018.

32 P. G. Szalay, T. Müller, G. Gidofalvi, H. Lischka and R. Shepard, *Chemical Reviews*, 2012, **112**, 108–181.

33 G. S. Kedziora, S. A. Barr, R. Berry, J. C. Moller and T. D. Breitzman, *Theoretical Chemistry Accounts*, 2016, **135**, year.

34 A. D. Becke, *Journal of Chemical Physics*, 2005, **122**, year.

35 A. D. Becke, *Journal of Chemical Physics*, 2013, **138**, year.

36 M. Fuchs, Y. M. Niquet, X. Gonze and K. Burke, *Journal of Chemical Physics*, 2005, **122**, year.

37 S. Grimme and A. Hansen, *Angewandte Chemie International Edition*, 2015, **54**, 12308–12313.

38 C. A. Bauer, A. Hansen and S. Grimme, *Chemistry – A European Journal*, 2017, **23**, 6150–6164.

39 S. Grimme, A. Hansen, S. Ehlert and J.-M. Mewes, *The Journal of Chemical Physics*, 2021, **154**, 064103.

40 F. Neese, *WIREs Computational Molecular Science*, 2012, **2**, 73–78.

41 F. Neese, *WIREs Computational Molecular Science*, 2022, **12**, e1606.

42 J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev and A. E. Roitberg, *The Journal of chemical physics*, 2018, **148**, 241733.

43 B. Brauer, M. K. Kesharwani, S. Kozuch and J. M. Martin, *Physical Chemistry Chemical Physics*, 2016, **18**, 20905–20925.

44 T. Fink, H. Bruggesser and J.-L. Reymond, *Angewandte Chemie International Edition*, 2005, **44**, 1504–1508.

45 T. Fink and J.-L. Reymond, *Journal of chemical information and modeling*, 2007, **47**, 342–353.

46 L. C. Blum and J.-L. Reymond, *Journal of the American Chemical Society*, 2009, **131**, 8732–8733.

47 V. Law, C. Knox, Y. Djoumbou, T. Jewison, A. C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu *et al.*, *Nucleic acids research*, 2014, **42**, D1091–D1097.

48 X. Gao, F. Ramezanghorbani, O. Isayev, J. S. Smith and A. E. Roitberg, *Journal of chemical information and modeling*, 2020, **60**, 3408–3415.

49 D. P. Kingma and J. Ba, *arXiv preprint arXiv:1412.6980*, 2014.

50 *RDKit: Open-source cheminformatics*, `https://www.rdkit.org`.

51 L. Grinsztajn, E. Oyallon and G. Varoquaux, *Advances in Neural Information Processing Systems*, 2022, **35**, 507–520.

52 L. Breiman, *Machine learning*, 2001, **45**, 5–32.

53 J. Moeyersoms, E. J. de Fortuny, K. Dejaeger, B. Baesens and D. Martens, *Journal of Systems and Software*, 2015, **100**, 80–90.

54 S. M. Lundberg and S.-I. Lee, *Advances in neural information processing systems*, 2017, **30**, year.

55 L. S. Shapley *et al.*, 1953.

56 K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko and K.-R. Müller, *The Journal of Chemical Physics*, 2018, **148**, 241722.

57 J. Gasteiger, J. Groß and S. Günnemann, International Conference on Learning Representations (ICLR), 2020.

58 Y. Liu, L. Wang, M. Liu, Y. Lin, X. Zhang, B. Oztekin and S. Ji, International Conference on Learning Representations (ICLR), 2022.

59 L. H. Hall and L. B. Kier, *Reviews in computational chemistry*, 1991, 367–422.

60 S. H. Bertz, *Journal of the American Chemical Society*, 1981, **103**, 3599–3601.

61 D. M. Bayada, H. Hamersma and V. J. van Geerestein, *Journal of chemical information and computer sciences*, 1999, **39**, 1–10.

62 X. Fu, Z. Wu, W. Wang, T. Xie, S. Keten, R. Gomez-Bombarelli and T. Jaakkola, *arXiv preprint arXiv:2210.07237*, 2022.