



Cite this: *Phys. Chem. Chem. Phys.*, 2022, 24, 10775

# $\Delta$ -Quantum machine-learning for medicinal chemistry†

Kenneth Atz, <sup>a</sup> Clemens Isert, <sup>a</sup> Markus N. A. Böcker, <sup>a</sup> José Jiménez-Luna <sup>\*ab</sup> and Gisbert Schneider <sup>\*ac</sup>

Many molecular design tasks benefit from fast and accurate calculations of quantum-mechanical (QM) properties. However, the computational cost of QM methods applied to drug-like molecules currently renders large-scale applications of quantum chemistry challenging. Aiming to mitigate this problem, we developed DelFTa, an open-source toolbox for the prediction of electronic properties of drug-like molecules at the density functional (DFT) level of theory, using  $\Delta$ -machine-learning.  $\Delta$ -Learning corrects the prediction error ( $\Delta$ ) of a fast but inaccurate property calculation. DelFTa employs state-of-the-art three-dimensional message-passing neural networks trained on a large dataset of QM properties. It provides access to a wide array of quantum observables on the molecular, atomic and bond levels by predicting approximations to DFT values from a low-cost semiempirical baseline.  $\Delta$ -Learning outperformed its direct-learning counterpart for most of the considered QM endpoints. The results suggest that predictions for non-covalent intra- and intermolecular interactions can be extrapolated to larger biomolecular systems. The software is fully open-sourced and features documented command-line and Python APIs.

Received 19th February 2022,  
 Accepted 5th April 2022

DOI: 10.1039/d2cp00834c

rsc.li/pccp

## Introduction

The electronic structure of drug-like molecules is responsible for various drug-relevant properties, such as molecular recognition in protein–ligand complexes,<sup>1–3</sup> drug-induced phototoxicity,<sup>4,5</sup> reactivity for covalent ligand–protein interaction,<sup>6–8</sup> cell membrane permeability,<sup>9</sup> or three-dimensional (3D) conformation energies.<sup>10</sup> However, despite advances in density functional theory (DFT) approaches,<sup>11,12</sup> which are widely regarded as a compromise between accuracy and computational cost,<sup>13,14</sup> calculating quantum-mechanical (QM) properties at this level of theory for many or for sizable molecules remains a computationally expensive task. Cheaper alternatives such as force fields<sup>15</sup> and semiempirical methods<sup>16,17</sup> have become popular alternatives, albeit with reduced accuracy. To overcome some of these issues, there has been a recent surge of interest in quantum machine-learning (QML), a set of techniques which aim to approximate quantum observables through statistical

modeling approaches.<sup>18–22</sup> Geometric deep learning in particular, a discipline focused on the investigation of neural network architectures that incorporate symmetry information into their design,<sup>23–25</sup> has become an active topic of research. Recent advances in geometric deep learning, such as the development of E(3)-equivariant neural networks, have led to improved prediction accuracy of energies,<sup>26–28</sup> forces for molecular dynamics simulations,<sup>29–31</sup> and wave functions in the form of local bases of atomic orbitals.<sup>32,33</sup>

In parallel to these developments,  $\Delta$ -QML (delta-QML) approaches, which aim to learn corrections between computationally inexpensive QM methods and more accurate, albeit more expensive ones, have been shown to deliver promising results.<sup>35</sup> Machine-learned corrections of this kind have been reported for both coupled cluster theory<sup>36,37</sup> *via* DFT, and for DFT *via* the semiempirical family of methods GFN-xTB,<sup>38,39</sup> as well as for other combinations.<sup>40</sup> However, despite their encouraging performance, to the best of our knowledge there are currently no open-source implementations of  $\Delta$ -QML or readily available trained models, which limits their widespread adoption. Addressing this need, we present DelFTa, an open-source deep-learning toolbox that enables both fast and accurate approximations of molecular electronic properties on the DFT<sup>41,42</sup> level of theory. Models were trained on the QMugs<sup>43</sup> dataset, which consists of  $\sim 2$  M molecular conformers with of a comprehensive array of QM observables both at semiempirical and DFT levels of theory for each structure.

<sup>a</sup> *ETH Zurich, Department of Chemistry and Applied Biosciences, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland. E-mail: joluna@ethz.ch, gisbert@ethz.ch*

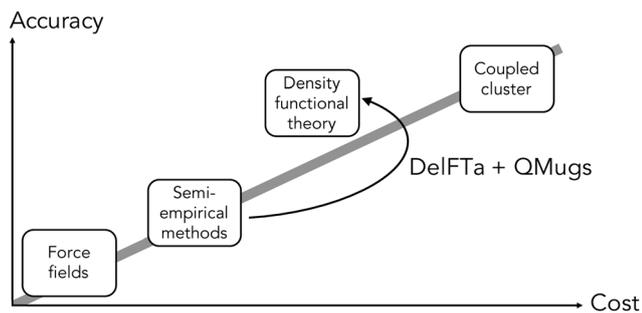
<sup>b</sup> *Department of Medicinal Chemistry, Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Straße 65, 88397 Biberach an der Riss, Germany*

<sup>c</sup> *ETH Singapore SEC Ltd., 1 CREATE Way, #06-01 CREATE Tower, Singapore 138602, Singapore*

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d2cp00834c>

‡ These authors contributed equally to this work.





**Fig. 1** Overall idea behind the DelFTa software package.  $\Delta$ -learning models were trained to predict the correction ( $\Delta$ ) between observables computed with a lower-cost semiempirical baseline method (GFN2-xTB) and the corresponding DFT-level reference method ( $\omega$ B97X-D/def2-SVP). Visualization inspired by ref. 34.

Specifically, QM observables were learned at the  $\omega$ B97X-D/def2-SVP level of theory, either directly or *via*  $\Delta$ -learning through corrections to the semiempirical GFN2-xTB method<sup>17,44–46</sup> (Fig. 1). To this end, E(3)-invariant three-dimensional (3D) message-passing neural networks (MPNNs) were employed, which are able to learn properties at either the global molecular, atom, and bond levels<sup>25</sup> (Fig. 2) and whose predictions are invariant to translations or rotations of the input molecule. The potential utility of the presented DelFTa approach is threefold:

(1) The models provided expand the  $\Delta$ -QML prediction landscape for drug-like molecules by providing access to commonly-used properties such as formation energies, as well as previously-unreported ones, such as energies of the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO, respectively), HOMO/LUMO gaps, dipole moments, Mulliken partial charges, and Wiberg bond orders for covalent and non-covalent bonds.

(2) We investigate several key concepts of QML for applications in drug-relevant chemical space; namely, (i) the advantages and

limitations of  $\Delta$  – compared to direct-learning, (ii) the utility of multi- over single-task learning paradigms for molecular properties, (iii) the dependence of prediction errors on training set size, and (iv) the extrapolation capabilities of the trained machine-learning models to non-covalent intra- and intermolecular interactions in biomolecules.

(3) We provide a fully open-sourced and user-friendly software package *via* command-line and Python APIs that includes all trained models, as well as extensive documentation and tutorials. Interested researchers will be able to use the provided models, train new ones, or build upon them.

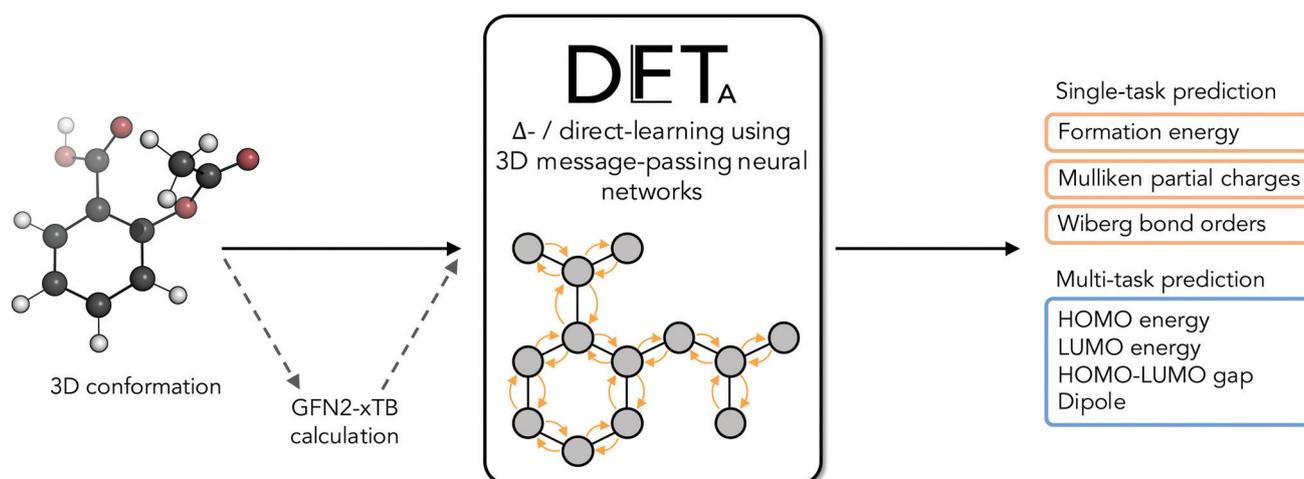
The DelFTa approach enables access to a variety of QM properties at DFT accuracy in a fast and user-friendly manner can be routinely used in numerous relevant applications in molecular modeling and design.

## Methods

### Reference dataset and dataset splits

DelFTa was built upon the QMugs<sup>43</sup> data collection, which comprises  $\sim 2$  M conformers of over 665 k molecules extracted from the ChEMBL database (release 27),<sup>47</sup> to obtain training, validation, and test sets. It includes QM properties at two levels of theory, namely the semiempirical method GFN2-xTB<sup>17,44–46</sup> and DFT ( $\omega$ B97X-D/def2-SVP<sup>41,42</sup>).

Each molecule in the QMugs dataset, associated with a unique ChEMBL identifier, was assigned to either training, validation or test sets, with all conformers of one molecule becoming part of the same set. A validation set composed of  $\sim 29$  k molecules was used for hyperparameter optimization and early stopping. All optimized models were tested on three test sets of  $\sim 29$  k molecules each ( $\sim 88$  k individual conformers). As in the QMugs dataset, some molecules with distinct ChEMBL identifiers are represented with the same SMILES<sup>48</sup> notation (*i.e.*, the same 2D molecular graph), all molecules with the same



**Fig. 2** Schematic of the  $\Delta$ -learning concept and application. A three-dimensional (3D) molecular conformation is used as an input to either a single- or multi-task trained message-passing neural network (MPNN). When a  $\Delta$ -learning endpoint is requested, an additional GFN2-xTB calculation is carried out, and the network is tasked with predicting the correction ( $\Delta$ ) between this baseline value and its  $\omega$ B97X-D/def2-SVP analogue. If a direct-learning prediction is requested, the network outputs an approximation to the  $\omega$ B97X-D/def2-SVP value, without using the GFN2-xTB baseline.



SMILES were assigned to the same test set to avoid information leakage (see ref. 43 for details).

While the production models available in the DelFTa application were trained on the entire QMugs dataset, model performance was also benchmarked for different training set sizes featuring both single molecular conformations per molecule (100, 1 k, 10 k, 100 k and  $\sim 547$  k training samples) as well as multiple ones ( $\sim 1.6$  M individual conformers of  $\sim 547$  k distinct molecules). For the formation energy models, all conformers of the same molecule were grouped within the same train-validation-test splits, correspondingly yielding training set sizes of approximately 300, 3 k, 30 k, 300 k and 1.6 M samples. Fig. S1 (ESI†) shows a schematic of the data splitting approach.

### Neural network architecture and training details

The 3D-MPNNs used in this work are based on the E(3)-Equivariant Graph Neural Network (EGNN) architecture.<sup>27,49</sup>  $\Delta$ -learning models for all endpoints  $y_i$ , at either global, node, or edge levels, associated with the  $i$ -th molecular conformation, were trained to predict the difference  $y_i^\Delta$  between DFT-computed properties ( $y_i^{\text{DFT}} \in \mathbb{R}^k$ ) and GFN2-xTB equivalents ( $y_i^{\text{GFN2-xTB}} \in \mathbb{R}^k$ ), specifically:

$$y_i^\Delta = y_i^{\text{DFT}} - y_i^{\text{GFN2-xTB}}. \quad (1)$$

Direct-learning models were trained on  $y_i^{\text{DFT}}$  values only. Molecular conformations were represented as fully-connected 3D graphs  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ , where  $\mathcal{V}$  corresponds to its set of nodes ( $v_i \in \mathcal{V}$ ),  $\mathcal{E}$  to its set of adjacent edges ( $e_{ij} \in \mathcal{E}$ ) and  $\mathcal{R}$  to associated Cartesian coordinates in 3D space ( $\mathbf{r}_i \in \mathbb{R}^3$ ). Initial node features  $\mathbf{v}_i^0$  were obtained *via* a linear embedding of the respective atom types. Edge features  $\mathbf{r}_{ij}$  were obtained *via* a sinusoidal and cosinusoidal encoding of the pairwise diatomic distances  $\|\mathbf{r}_i - \mathbf{r}_j\|_2^2$  (*i.e.*, a Fourier-like encoding scheme).

An Equivariant Graph-Convolutional Layer (EGCL) was applied over all edges  $e_{ij}$  of the graph. It uses the node embeddings of  $\mathbf{v}_i^l$  at layer  $l$  as well as their respective atomic positions  $\mathbf{r}_i$  to produce updated node representations  $\mathbf{v}_i^{l+1}$ :

$$\mathbf{v}_i^{l+1} = \text{EGCL}(\mathbf{v}_i^l, \mathbf{r}_i, \mathcal{E}), \quad (2)$$

using the following message-passing mechanism:

$$\begin{aligned} \mathbf{m}_{ij} &= \phi_e(\mathbf{v}_i^l, \mathbf{v}_j^l, \mathbf{r}_{ij}), \\ \mathbf{m}_i &= \sum_j \mathbf{m}_{ij}, \\ \mathbf{v}_i^{l+1} &= \phi_h(\mathbf{v}_i^l, \mathbf{m}_i), \end{aligned} \quad (3)$$

where  $\phi_h$ ,  $\phi_e$  are node and edge non-linear transformations, respectively, modeled with multilayer Perceptrons non-linearized with the SiLU activation function,<sup>50</sup>  $\mathbf{m}_{ij}$  the computed edge message features, and  $\mathbf{m}_i$  the aggregated message features per node.

After five message-passing steps, node features  $\mathbf{v}_i$  were summed for extensive properties (*i.e.*, formation energy), and mean-pooled for intensive ones<sup>51</sup> (*i.e.*, dipole, orbital energies, HOMO–LUMO gap) and then mapped to their corresponding

target shapes *via* an additional multi-layer Perceptron. In the specific cases of the node- or edge-based endpoints (*i.e.*, Mulliken partial charges and Wiberg bond orders), the learned node-level  $\mathbf{v}_i$  and message-level features  $\mathbf{m}_{ij}$  were used directly for prediction.

While we employed MSE losses for the intensive properties and Mulliken partial charge models, the ones used for the formation energy and Wiberg bond order endpoints were composed of two terms. Similar to previous work,<sup>38,39,52</sup> in the case of formation energy (eqn (4)) a first loss term minimized the error on the absolute formation energy  $\mathbf{y}_{\text{abs}}$ , while a second term minimized the relative energy differences between the different geometries of the same molecule  $\mathbf{y}_{\text{rel}}$ :

$$\mathcal{L}_{\text{form}} = \|\hat{\mathbf{y}}_{\text{abs}} - \mathbf{y}_{\text{abs}}\|_2^2 + \beta \|\hat{\mathbf{y}}_{\text{rel}} - \mathbf{y}_{\text{rel}}\|_2^2. \quad (4)$$

Specifically training the model to differentiate between conformer energies was motivated by the relevance of this task in identifying the most stable conformers within an ensemble or assessing reaction barriers.<sup>53</sup> In the case of Wiberg bond orders (eqn (5)), the first term minimized the error on covalent bonds  $\mathbf{y}_{\text{cov}}$ , and the second term that on non-covalent interactions  $\mathbf{y}_{\text{non}}$ :

$$\mathcal{L}_{\text{wbo}} = \|\hat{\mathbf{y}}_{\text{cov}} - \mathbf{y}_{\text{cov}}\|_2^2 + \lambda \|\hat{\mathbf{y}}_{\text{non}} - \mathbf{y}_{\text{non}}\|_2^2. \quad (5)$$

Both  $\beta$  and  $\lambda$  values were optimized on the validation set and set to  $\beta = 1$ , and  $\lambda = 5 \times 10^{-2}$  (see ESI† Section 1 for further details). The following network hyperparameters were used in all models considered in this study: (i) node dimension  $\mathbf{v}_i = 128$ , (ii) message passing dimension  $\mathbf{m}_{ij} = 32$  for molecular and atomic models and 64 for the Wiberg bond order models, (iii) number of sinusoidal and cosinusoidal distance encoding features: 32, (iv) number of EGCLs: 5, and (v) number of global multi-layer Perceptrons: 3, each containing 256 hidden units. Because most of the considered endpoints feature different numerical ranges, which could cause optimization instability issues during the training of the multi-task models, a min–max standardization strategy was applied using the 1st and 99th percentiles of each endpoint, thereby also avoiding outlier scaling problems.

Networks for all endpoints, and for both  $\Delta$ - and direct-learning models, were trained using the Adam stochastic gradient descent optimizer<sup>54</sup> with a starting learning rate of  $10^{-3}$  and  $10^{-4}$  for all single-task and multi-task models, respectively. An early-stopping strategy that monitored the monotonic decrease of the chosen loss function on the chosen validation set was adopted (see ESI† Section 1 for further details).

### Processing of biomolecules

The structures were retrieved from the Protein Data Bank (PDB)<sup>55</sup> and preprocessed with the MOE software<sup>56</sup> (version 2019.0102). Since bond orders are intrinsically local properties, and in order to make DFT calculations feasible, atoms which were farther away than one additional residue from the non-covalent interactions of interest were removed and the resulting radicals were padded with hydrogens (see ESI† Section 4 for further details). QM reference values were obtained *via* Psi4<sup>57</sup>



(version 1.3.2) using the  $\omega$ B97X-D functional and the def2-SVP basis set.

## Results

### Predictive performance

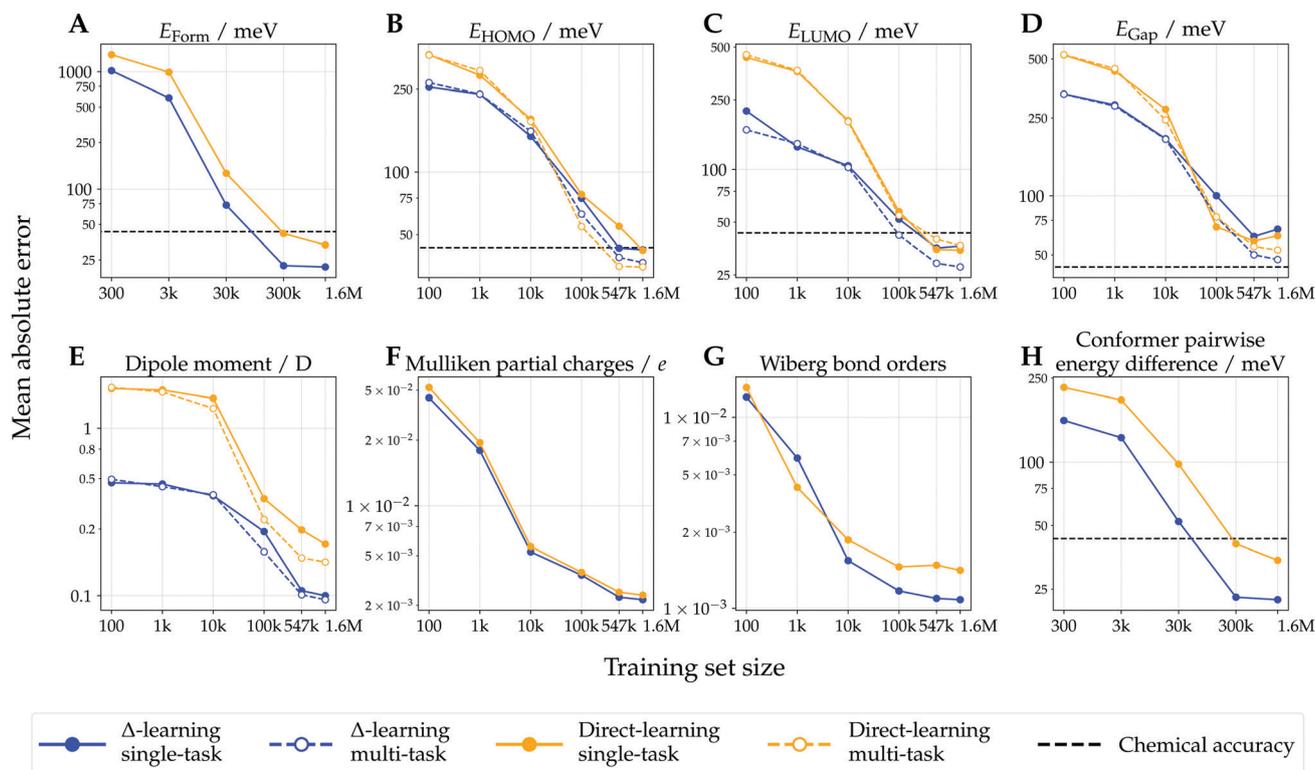
Test-set learning curves of models trained with varying training-set sizes are shown in Fig. 3. Mean absolute errors (MAEs) w.r.t. DFT reference values decreased with increasing training set size, generally resulting in linear correlation recorded on log-log plots due to their inverse power law relationship.<sup>58,59</sup> For most of the considered endpoints, the  $\Delta$ -learning models consistently achieved a better predictive performance (lower MAEs) than their direct-learning counterparts. This performance gap was observed for most training set sizes, highlighting the usefulness of  $\Delta$ -learning in low-data regimes. For the prediction of formation energy and relative conformer energy differences, the  $\Delta$ -learning models achieved performance surpassing chemical accuracy ( $1 \text{ kcal mol}^{-1} \approx 43.4 \text{ meV}$ ) for all training set sizes larger than 300 k conformers, while direct-learning models required 1.6 M training points. For all computationally intensive molecular endpoints (*i.e.* those not depending on the system size, such as orbital energies or dipoles), the performance of multi-task models was superior to that of their single-task counterparts, with the exception of direct-learning on LUMO

energies for which both single- and multi-task models achieved similar performance.

The predictive performance of the models trained on 1.6 M training conformers was analyzed in more detail. Table 1 shows MAEs w.r.t. the  $\omega$ B97X-D/def2-SVP reference values obtained for the test sets, and compares them to those of the semiempirical baseline method GFN2-xTB. For all considered endpoints and models, the DFT reference values were more closely approximated with the proposed machine-learning models than with GFN2-xTB. For most of the considered endpoints, the  $\Delta$ -learning approach yielded lower MAEs than its direct-learning counterpart. However, the direct-learning approach achieved a slightly lower MAE (35.0 meV *vs.* 36.7 meV for  $\Delta$ -learning) for the prediction of HOMO energies, which can be attributed to the fact that HOMO energies calculated with the DFT and GFN2-xTB methods, respectively, correlated to a lesser degree than other considered endpoints. Scatter plots showing  $\Delta$ -predicted properties *versus* their DFT reference values are provided in Fig. 4. Direct-learning approaches also yielded higher accuracies than the semiempirical baseline GFN2-xTB for all considered endpoints (see Fig. S2, ESI†).

### Utility of $\Delta$ -learning

While  $\Delta$ -learning models generally outperformed their direct-learning analogues in our experiments, the observed performance

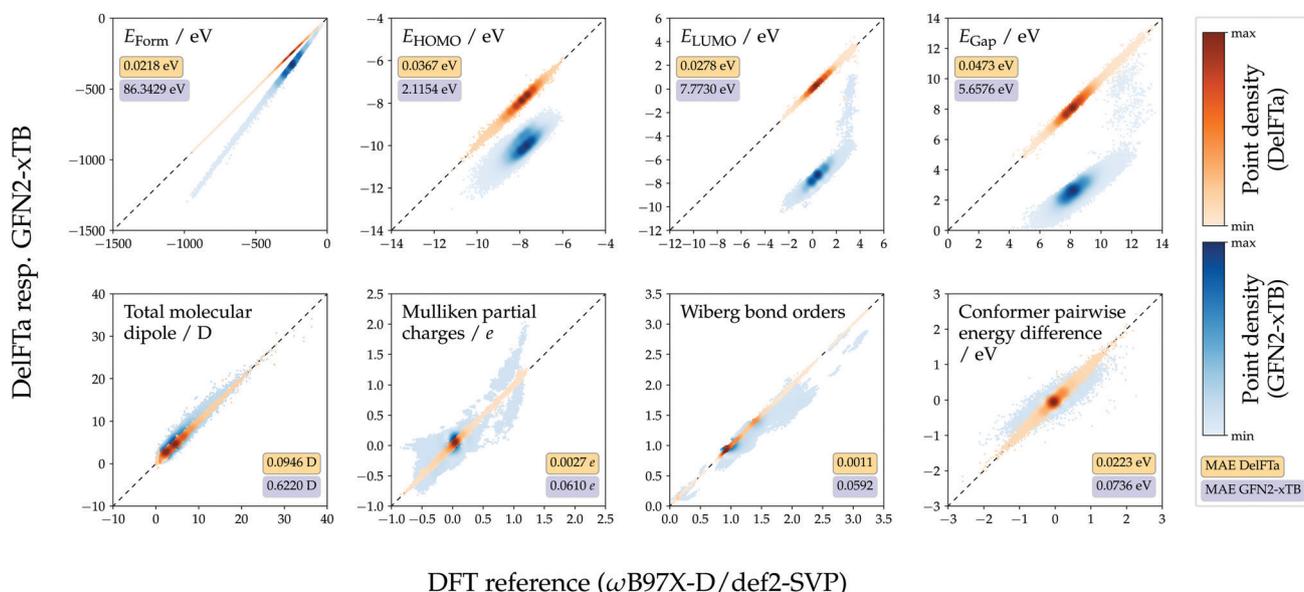


**Fig. 3** Learning curves of  $\Delta$ - and direct-learning model for extensive (panels A and H) and intensive (panels B–E) molecular properties, and for atom- and bond-centered properties (panels F and G). These show the mean absolute error (MAE) as a function of training set size (axes in logarithmic scale) computed on the three test sets comprised of  $\sim 29$  k molecules ( $\sim 88$  k conformers). For the energy models (panels A and H), all conformers of the same molecule are grouped within the same train-validation-test splits (see Methods). For intensive molecular properties (panels B–E), learning curves are shown for both single-task and multi-task learning paradigms. Chemical accuracy thresholds are indicated for energy and orbital energy models ( $1 \text{ kcal mol}^{-1} \approx 43.4 \text{ meV}$ ).<sup>39</sup>



**Table 1** MAEs ( $\pm 1$  standard deviation) for the baseline (GFN2-xTB) as well as the  $\Delta$ - and direct-learning models w.r.t. the DFT reference values ( $\omega$ B97X-D/def2-SVP). Results computed for  $\sim 88$  k molecules ( $\sim 263$  k conformers) from the three test sets. Wiberg bond order results only for bonds where GFN2-xTB values were available. The lowest MAE w.r.t. reference values are highlighted

Property	Unit	GFN2-xTB	DelFTa	
			$\Delta$ -learning	Direct-learning
Formation energy	meV	86343 ( $\pm 39$ )	<b>21.78</b> ( $\pm 0.10$ )	33.50 ( $\pm 0.05$ )
HOMO energy	meV	2115.4 ( $\pm 0.5$ )	36.7 ( $\pm 0.1$ )	<b>35.0</b> ( $\pm 0.1$ )
LUMO energy	meV	7773.0 ( $\pm 0.7$ )	<b>27.8</b> ( $\pm 0.2$ )	36.8 ( $\pm 0.2$ )
HOMO-LUMO gap	meV	5658 ( $\pm 1$ )	<b>47.3</b> ( $\pm 0.2$ )	52.9 ( $\pm 0.1$ )
Total molecular dipole	<i>D</i>	0.622 ( $\pm 0.002$ )	<b>0.0946</b> ( $\pm 0.0006$ )	0.1588 ( $\pm 0.0006$ )
Mulliken partial charges	<i>e</i>	0.0610 ( $\pm 0.0000$ )	<b>0.0027</b> ( $\pm 0.0000$ )	0.0029 ( $\pm 0.0000$ )
Wiberg bond orders	—	0.0592 ( $\pm 0.0001$ )	<b>0.0011</b> ( $\pm 0.0000$ )	0.0017 ( $\pm 0.0000$ )
Conformer pairwise energy difference	meV	73.6 ( $\pm 0.4$ )	<b>22.29</b> ( $\pm 0.06$ )	34.27 ( $\pm 0.08$ )



**Fig. 4** Calculated versus DFT reference values ( $\omega$ B97X-D/def2-SVP). Mean absolute errors (MAEs) computed for  $\Delta$ -learning models and for the GFN2-xTB baseline in  $\sim 88$  k test-set molecules ( $\sim 263$  k conformers).  $\Delta$ -learning predictions obtained using the models trained on the 1.6 M conformer training set, and in the multi-task setting for all intensive endpoints. Wiberg bond order results only for bonds where GFN2-xTB values were available. Colorbars scaled individually for each property.

difference was not uniformly distributed across the endpoints. To investigate under which conditions the  $\Delta$ -learning paradigm is advantageous over direct-learning, we analyzed the relative performance difference between the two approaches as a function of the Pearson correlation coefficient  $r$  between baseline (GFN2-xTB) and reference ( $\omega$ B97X-D/def2-SVP) values. Fig. 5 indicates that the relative performance advantage of  $\Delta$ -learning is positively correlated to the correlation  $r$  between baseline and reference values. This result confirms the intuitive understanding that  $\Delta$ -learning provides a larger performance advantage over direct-learning the more information the baseline method provides.

### Non-covalent interactions in biomolecules

Many tasks in medicinal and bioorganic chemistry encompass the study of non-covalent intra- (*e.g.* secondary-structure-stabilizing) and intermolecular (*e.g.* protein-ligand) interactions.<sup>60,61</sup>

Modelling of non-covalent interactions with semiempirical methods has previously been shown to be challenging.<sup>62</sup> Given the biological relevance of these molecular interactions, it is desirable to develop QML models which can accurately extrapolate from small molecules (*e.g.*, bioactive ligands) to larger biomolecules (*e.g.*, peptides), foregoing the need for the expensive DFT calculations associated with structures of such size.

Towards that end, we preliminarily investigated the generalization capabilities of direct-learning QML models for selected biomolecular systems with crucial non-covalent interactions by comparing predicted Wiberg bond orders to DFT reference values.  $\Delta$ -learning models were not considered in these analyses, as bond order values for many non-covalent interactions of interest are not provided by the GFN2-xTB semiempirical method. The investigated structures (Fig. 6) include a hydrolase  $\beta$ -turn,<sup>63</sup> a  $\beta$ -sheet and an  $\alpha$ -helix of ubiquitin,<sup>64</sup> glutamate in a glutamate dehydrogenase binding pocket,<sup>65</sup> an uracil-adenine base pair in



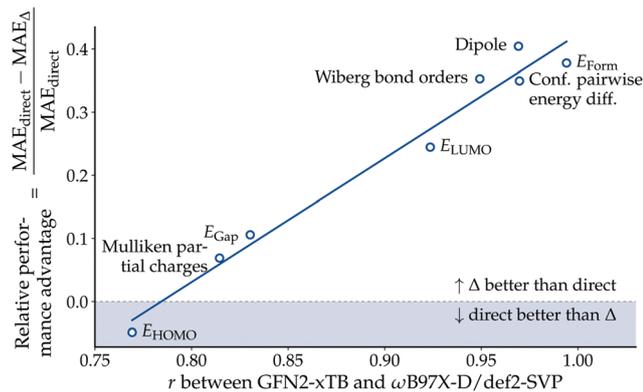


Fig. 5 Relative performance advantage of  $\Delta$ - over direct-learning. The figure shows the model performance against the Pearson correlation coefficient  $r$  between semiempirical (GFN2-xTB) and DFT ( $\omega$ B97X-D/def2-SVP) methods. The solid line shows a least-squares linear fit.

an RNA structure,<sup>66</sup> and a transcription factor binding to a cytosine-guanine base pair of a DNA structure.<sup>67</sup>

By training only on monomers of drug-like molecules and their intramolecular interactions, as included in the QMugs dataset, the models successfully extrapolated to non-covalent intra- and intermolecular interactions in biomacromolecules including monomers and dimers. For instance, weak hydrogen bonds (Wiberg bond order  $< 0.1$ ), such as the ones found in  $\alpha$ -helices (Fig. 6(C)), as well as strong hydrogen bonds (Wiberg bond order  $> 0.1$ ), such as the ones found in the RNA base pairs (Fig. 6(E)),  $\beta$ -turns/sheets (Fig. 6(A) and (B)), and in the glutamate binding pocket (Fig. 6(D)), were accurately predicted. However, we observed reduced predictive capabilities for some interactions such as hydrogen bonds with phosphate groups (Fig. S4, ESI<sup>†</sup>).

## Benchmarking

We compared the implementation of the DelFTa deep-learning architecture used in this work to the one originally reported in ref. 27, for the QM9 dataset,<sup>68</sup> a benchmark used in previous QML studies, which features quantum observables for  $\sim 134$  k small molecules. The same DelFTa model architecture used for the direct-learning of formation energies was retrained on the QM9 training set ( $\sim 100$  k molecules), validated and tested on its respective validation and test sets ( $\sim 15$  k molecules each). The trained models achieved an MAE of  $11.9 \pm 0.7$  meV in three independent model runs, which is comparable to the originally-reported performance (12 meV).<sup>27</sup> The lower overall error of the models trained on QM9 compared to those trained on QMugs (11.9 meV and 33.5 meV, respectively) can be attributed to two key differences between the datasets, namely atom type diversity (10 different atom types in QMugs, 5 in QM9), and molecular size (up to 100 heavy atoms in QMugs, and up to 9 heavy atoms in QM9, respectively).

Since DelFTa models were trained on uncharged molecules but had shown predictive capabilities for charged biomolecules, we quantitatively investigated the models' performance on 176 randomly-sampled charged conformers corresponding to 59 molecules extracted from the test sets (see Table S2 for details, ESI<sup>†</sup>).

Compared to uncharged molecules, the predictive performance decreased moderately for Mulliken partial charges and Wiberg bond orders, and substantially decreased for other endpoints (see Table S3, ESI<sup>†</sup>). Based on these results, we discourage the use of the DelFTa models provided in the accompanying software for out-of-distribution molecules.

**Listing 1** A small snippet highlighting the main predictive capabilities of the DelFTa Python package and its integration with Pybel. Molecules, with or without associated 3D geometry, can be supplied *via* a wide array of file types.

```
from delfta.calculator import DelftaCalculator
from openbabel.pybel import readstring

mol = readstring("smi", "CCO")
calc = DelftaCalculator()
predictions = calc.predict(mol)
print(predictions)
# >> {"E_form": -1.2884707, "E_lumo": 0.18285048, ...
```

Finally, we explored the model performance (trained on  $\omega$ B97X-D/def2-SVP DFT data) on a set of molecules whose DFT reference values were computed with a more comprehensive basis set ( $\omega$ B97X-D/def2-QZVP). Calculations for 2,874 conformations corresponding to 958 distinct molecules with this larger basis set did not indicate superior performance of  $\Delta$ -over direct learning (see Fig. S3 for details, ESI<sup>†</sup>). Furthermore, Mulliken partial charges on the  $\omega$ B97X-D/def2-QZVP level of theory were better approximated using GFN2-xTB than with either of the provided machine-learning models. This was expected as GFN2-xTB better approximates charges computed with the larger basis set than the chosen DFT reference used throughout this study ( $\omega$ B97X-D/def2-SVP).

## Software

DelFTa is fully implemented in the Python programming language<sup>69</sup> and uses the PyTorch<sup>70</sup> (version 1.8.0) and PyTorch Geometric packages<sup>71</sup> (version 1.7.2) to enable model training and inference. A minimalist code example for the usage of the package is provided in Listing 1. Semiempirical calculations at the GFN2-xTB<sup>17,44-46</sup> level of theory are computed *via* open-source xtb binaries. All molecular manipulation routines (including optional generation of initial 3D coordinates and GFN2-xTB geometry optimization) are integrated into DelFTa and handled *via* the Pybel package<sup>72</sup> and OpenBabel<sup>73</sup> Python bindings. The software is fully open-sourced, available on GitHub (<https://github.com/josejimenezluna/delfta>) under a permissive AGPLv3 license, and distributed through the conda package manager.<sup>74</sup> A Docker<sup>75</sup> container is also provided for easier accessibility and to ensure long-term functionality. Furthermore, DelFTa provides extensive documentation for its code and APIs. Tutorials in the form of several didactic Jupyter notebooks<sup>76</sup> are also available.



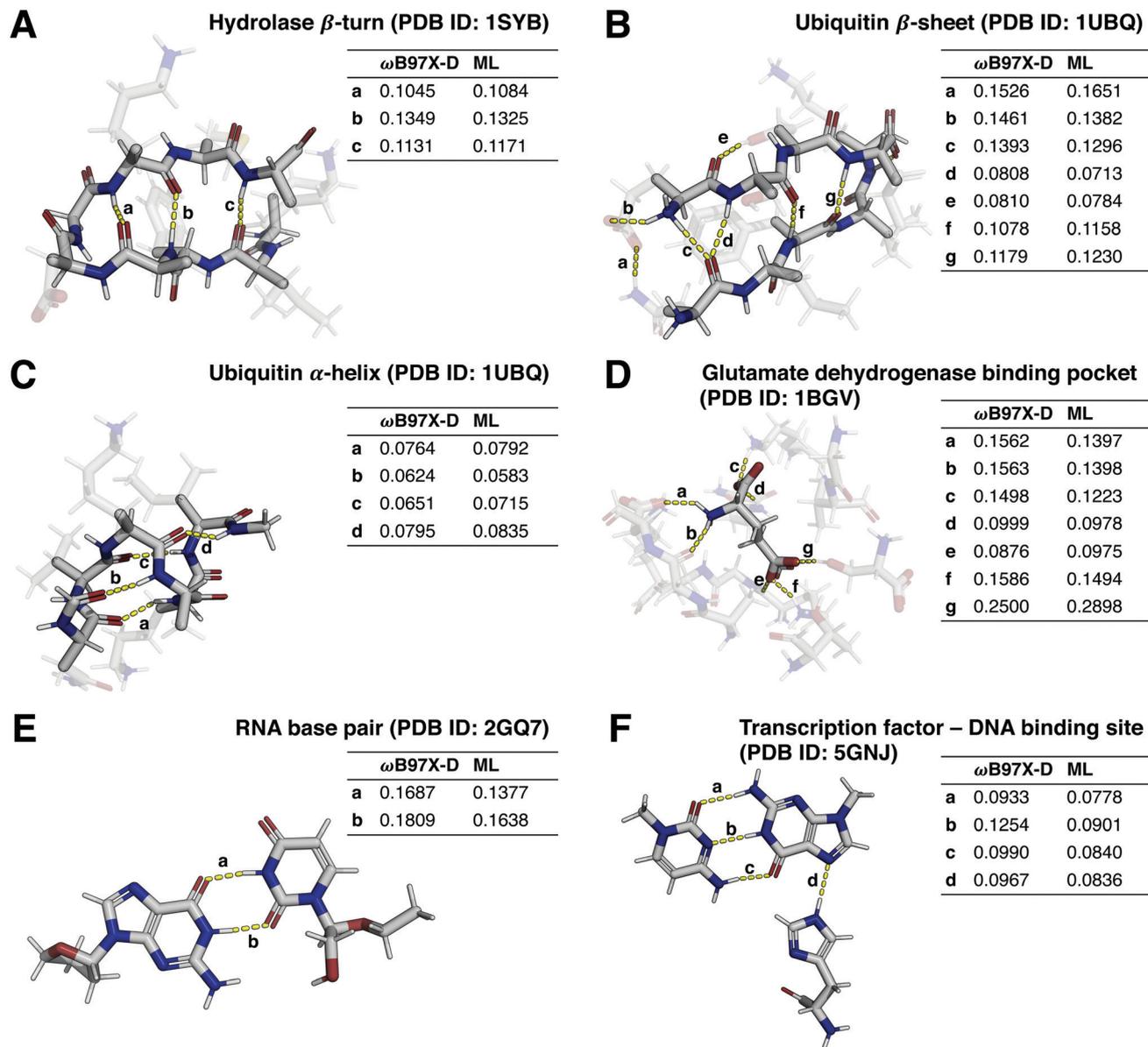


Fig. 6 Non-covalent interactions in selected biomacromolecules. For panels (A–D), non-backbone residues resp. protein pockets shown semi-transparently for visual clarity. Interactions with DFT-calculated Wiberg bond orders between 0.05 and 0.8 shown, and both calculated ( $\omega$ B97X-D/def2-SVP) and predicted (ML, using models trained on 1.6 M datapoints) values tabulated. Interactions with solvent atoms and between atoms spaced less than six covalent bonds apart not shown. White: hydrogen, gray: carbon, red: oxygen, blue: nitrogen.

On a computer with a consumer-grade graphics processing unit, DelFTA predicts all considered endpoints at a speed of approximately 50 and 5 molecules per second for the direct and  $\Delta$ -learning models respectively, with the latter approach mostly bottlenecked by the additionally-required baseline GFN2-xTB calculations.

## Discussion

QML models were trained for a wide variety of endpoints on a large dataset of quantum observables. Models were validated for both  $\Delta$ - and direct-learning, as well as single- and multi-task paradigms. The results suggest that both  $\Delta$ - and direct-learning

models have improved accuracy over the GFN2-xTB baseline in approximating  $\omega$ B97X-D/def2-SVP reference values. For the majority of the considered endpoints,  $\Delta$ -learning models displayed lower MAEs than their direct-learning analogues at roughly the same computational cost as GFN2-xTB. Additionally, the Wiberg bond order models were able to approximate non-covalent interactions in larger biomolecular systems.

We foresee many applications for the hereby provided models in both supervised and generative molecular pipelines. For instance, featurization with quantum-derived properties, such as partial charges and nuclear magnetic resonance shifts, was shown to increase the performance of reactivity prediction with graph neural networks in low-data regimes.<sup>77</sup> Similar



effects may also be anticipated in medicinal chemistry as the electronic structure of drug-like molecules governs many related properties. Potential examples include the influence of (i) HOMO/LUMO energies on phototoxicity,<sup>4,5</sup> (ii) dipole moments on aqueous solubility<sup>78,79</sup> and membrane permeability,<sup>9,80</sup> and (iii) formation energies on 3D-conformer ensembles<sup>9</sup> or site-of-metabolism prediction.<sup>81</sup>

Future prospective applications will reveal the practical applicability and usefulness of these models in drug discovery-related tasks. The current limitations are twofold, concerning the modelling performance and the models' applicability domain. With regard to modelling performance, we noted that while the  $\Delta$ -learning approach affords substantial improvements over its direct-learning analogue w.r.t. target DFT reference values, this does not necessarily hold in the case of HOMO energies, probably owing to the limited correlation of the baseline and reference methods. Furthermore, some of the observations regarding the performance advantages do not necessarily hold for reference values computed with a more comprehensive basis set. Limitations with regard to the applicability domain mostly stem from the underlying QMugs dataset, which was conceived with medicinal chemistry applications in mind. For example, it does not feature organometallic complexes, polymers, crystalline structures, or molecular systems including dimers, radicals, excited electronic states, higher-order spin states, off-equilibrium structures or charged molecules. Specifically for charged molecules we observed substantially decreased predictive performance in this study. Adequate models for these types of molecular structures will require training data that specifically covers the respective chemical space, and therefore remain a subject of future work.

## Conflicts of interest

G. S. is a cofounder of inSili.com LLC, Zurich, and a consultant to the pharmaceutical industry.

## Acknowledgements

We thank N. Weskamp for helpful discussions on this work. This work was financially supported by the ETH RETHINK initiative, the Swiss National Science Foundation (Grant No. 205321\_182176), and Boehringer Ingelheim Pharma GmbH & Co. KG. C.I. acknowledges support from the Scholarship Fund of the Swiss Chemical Industry.

## Notes and references

- M. Khrenova, A. V. Nemukhin, B. L. Grigorenko, A. Krylov and T. Domratcheva, *J. Chem. Theory Comput.*, 2010, **6**, 2293–2302.
- N.-Z. Xie, Q.-S. Du, J.-X. Li and R.-B. Huang, *PLoS One*, 2015, **10**, e0137113.
- A. B. M. Tavares, J. X. L. Neto, U. L. Fulco and E. L. Albuquerque, *Sci. Rep.*, 2018, **8**, 1–13.
- M. Freccero, E. Fasani, M. Mella, I. Manet, S. Monti and A. Albini, *Chem. – Eur. J.*, 2008, **14**, 653–663.
- J. Llano, J. Raber and L. A. Eriksson, *J. Photochem. Photobiol., A*, 2003, **154**, 235–243.
- H. S. Yu, C. Gao, D. Lupyan, Y. Wu, T. Kimura, C. Wu, L. Jacobson, E. Harder, R. Abel and L. Wang, *J. Chem. Inf. Model.*, 2019, **59**, 3955–3967.
- Z. Zhao, Q. Liu, S. Bliven, L. Xie and P. E. Bourne, *J. Med. Chem.*, 2017, **60**, 2879–2889.
- J. Fanfrlik, P. S. Brahmshatriya, J. Rezac, A. Jilkova, M. Horn, M. Mares, P. Hobza and M. Lepsik, *J. Phys. Chem. B*, 2013, **117**, 14973–14982.
- F. Pultar, M. E. Hansen, S. Wolfrum, L. Bösel, R. Fróis-Martins, S. Bloch, A. G. Kravina, D. Pehlivanoglu, C. Schäffer, S. LeibundGut-Landmann, S. Riniker and E. M. Carreira, *J. Am. Chem. Soc.*, 2021, **143**, 10389–10402.
- M. Rupp, M. R. Bauer, R. Wilcken, A. Lange, M. Reutlinger, F. M. Boeckler and G. Schneider, *PLoS Comput. Biol.*, 2014, **10**, e1003400.
- K. Burke, *J. Chem. Phys.*, 2012, **136**, 150901.
- C. D. Sherrill, *J. Chem. Phys.*, 2010, **132**, 110902.
- O. A. Von Lilienfeld, I. Tavernelli, U. Rothlisberger and D. Sebastiani, *Phys. Rev. Lett.*, 2004, **93**, 153004.
- T. Schwabe and S. Grimme, *Acc. Chem. Res.*, 2008, **41**, 569–579.
- S. Wang, J. Witek, G. A. Landrum and S. Riniker, *J. Chem. Inf. Model.*, 2020, **60**, 2044–2058.
- J. J. P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173–1213.
- C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.
- O. A. von Lilienfeld, K.-R. Müller and A. Tkatchenko, *Nat. Rev. Chem.*, 2020, **4**, 347–358.
- B. Huang and O. A. von Lilienfeld, *Nat. Chem.*, 2020, **12**, 945–951.
- O. T. Unke, S. Chmiela, H. E. Saucedo, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K.-R. Müller, *Chem. Rev.*, 2021, **121**(16), 10142–10186.
- O. T. Unke, D. Koner, S. Patra, S. Käser and M. Meuwly, *Mach. Learn.: Sci. Technol.*, 2020, **1**, 013001.
- D. Lemm, G. F. von Rudorff and O. A. von Lilienfeld, *Nat. Commun.*, 2021, **12**, 4468.
- M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam and P. Vandergheynst, *IEEE Signal Process. Mag.*, 2017, **34**, 18–42.
- M. M. Bronstein, J. Bruna, T. Cohen and P. Velicković, 2021, arXiv:2104.13478.
- K. Atz, F. Grisoni and G. Schneider, *Nat. Mach. Intell.*, 2021, **3**, 1023–1032.
- K. T. Schütt, H. E. Saucedo, P.-J. Kindermans, A. Tkatchenko and K.-R. Müller, *J. Chem. Phys.*, 2018, **148**, 241722.
- V. G. Satorras, E. Hoogeboom and M. Welling, 2021, arXiv:2102.09844.
- K. T. Schütt, O. T. Unke and M. Gastegger, 2021, arXiv:2102.03150.
- O. T. Unke and M. Meuwly, *J. Chem. Theory Comput.*, 2019, **15**, 3678–3693.
- S. Batzner, T. E. Smidt, L. Sun, J. P. Mailoa, M. Kornbluth, N. Molinari and B. Kozinsky, 2021, arXiv:2101.03164.



- 31 O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schütt, H. E. Sauceda and K.-R. Müller, 2021, arXiv:2105.00304.
- 32 K. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller and R. J. Maurer, *Nat. Commun.*, 2019, **10**, 5024.
- 33 O. T. Unke, M. Bogojeski, M. Gastegger, M. Geiger, T. Smidt and K.-R. Müller, 2021, arXiv:2106.02347.
- 34 O. A. von Lilienfeld, 31st Conference on Neural Information Processing Systems, 2017.
- 35 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. von Lilienfeld, *J. Chem. Theory Comput.*, 2015, **11**, 2087–2096.
- 36 J. S. Smith, R. Zubatyuk, B. Nebgen, N. Lubbers, K. Barros, A. E. Roitberg, O. Isayev and S. Tretiak, *Sci. Data*, 2020, **7**, 134.
- 37 A. Nandi, C. Qu, P. L. Houston, R. Conte and J. M. Bowman, *J. Chem. Phys.*, 2021, **154**, 051102.
- 38 Z. Qiao, M. Welborn, A. Anandkumar, F. R. Manby and T. F. Miller III, *J. Chem. Phys.*, 2020, **153**, 124111.
- 39 A. S. Christensen, S. K. Sirumalla, Z. Qiao, M. B. O'Connor, D. G. Smith, F. Ding, P. J. Bygrave, A. Anandkumar, M. Welborn and F. R. Manby, *et al.*, *J. Chem. Phys.*, 2021, **155**, 204103.
- 40 P. Zheng, R. Zubatyuk, W. Wu, O. Isayev and P. O. Dral, *Nat. Commun.*, 2021, **12**, 7022.
- 41 J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.
- 42 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
- 43 C. Isert, K. Atz, J. Jiménez-Luna and G. Schneider, 2021, arXiv:2107.00367.
- 44 S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2017, **13**, 1989–2009.
- 45 S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 2847–2862.
- 46 C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2020, e01493.
- 47 A. Gaulton, A. Hersey, M. Nowotka, A. P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis and E. Cibrián-Uhalte, *et al.*, *Nucleic Acids Res.*, 2017, **45**, D945–D954.
- 48 D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.
- 49 EGNN-PyTorch, <https://github.com/lucidrains/egnn-pytorch>, 2021.
- 50 S. Elfwing, E. Uchibe and K. Doya, *Neural Networks*, 2018, **107**, 3–11.
- 51 W. Pronobis, K. T. Schütt, A. Tkatchenko and K.-R. Müller, *Eur. Phys. J. B*, 2018, **91**, 1–6.
- 52 Z. Qiao, A. S. Christensen, F. R. Manby, M. Welborn, A. Anandkumar and T. F. Miller III, 2021, arXiv:2105.14655.
- 53 A. Pung and I. Leito, *J. Phys. Chem. A*, 2017, **121**, 6823–6829.
- 54 D. P. Kingma and J. Ba, 2014, arXiv:1412.6980.
- 55 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.
- 56 C. C. G. ULC, Molecular Operating Environment (MOE), 2019.01, 2020.
- 57 D. G. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio and A. Alenaizan, *et al.*, *J. Chem. Phys.*, 2020, **152**, 184108.
- 58 F. A. Faber, A. S. Christensen, B. Huang and O. A. Von Lilienfeld, *J. Chem. Phys.*, 2018, **148**, 241717.
- 59 K.-R. Müller, M. Finke, N. Murata, K. Schulten and S. Amari, *Neural Comput.*, 1996, **8**, 1085–1106.
- 60 B. Kuhn, P. Mohr and M. Stahl, *J. Med. Chem.*, 2010, **53**, 2601–2611.
- 61 C. Bissantz, B. Kuhn and M. Stahl, *J. Med. Chem.*, 2010, **53**, 5061–5084.
- 62 A. S. Christensen, T. Kubar, Q. Cui and M. Elstner, *Chem. Rev.*, 2016, **116**, 5301–5337.
- 63 T. R. Hynes, R. A. Kautz, M. A. Goodman, J. F. Gill and R. O. Fox, *Nature*, 1989, **339**, 73–76.
- 64 S. Vijay-Kumar, C. E. Bugg and W. J. Cook, *J. Mol. Biol.*, 1987, **194**, 531–544.
- 65 T. Stillman, P. Baker, K. Britton and D. Rice, *J. Mol. Biol.*, 1993, **234**, 1131–1139.
- 66 W. Rypniewski, M. Vallazza, M. Perbandt, S. Klusmann, L. J. DeLucas, C. Betzel and V. A. Erdmann, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2006, **62**, 659–664.
- 67 T.-f. Lian, Y.-p. Xu, L.-f. Li and X.-D. Su, *Cell Rep.*, 2017, **19**, 1334–1342.
- 68 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. Von Lilienfeld, *Sci. Data*, 2014, **1**, 140022.
- 69 The Python Language Reference, <https://docs.python.org/3/reference/>.
- 70 A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein and L. Antiga, *et al.*, *Adv. Neural Inf. Process. Syst.*, 2019, **32**, 8026–8037.
- 71 M. Fey and J. E. Lenssen, ICLR Workshop on Representation Learning on Graphs and Manifolds, 2019.
- 72 N. M. O'Boyle, C. Morley and G. R. Hutchison, *Chem. Cent. J.*, 2008, **2**, 1–7.
- 73 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Cheminformatics*, 2011, **3**, 1–14.
- 74 Conda package manager, <https://conda.io>.
- 75 D. Merkel, *Linux J.*, 2014, **2**.
- 76 T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla and C. Willing, *Jupyter Notebooks – A publishing format for reproducible computational workflows*, IOS Press, 2016, pp. 87–90.
- 77 T. Stuyver and C. W. Coley, 2021, arXiv:2107.10402.
- 78 R. M. Cardoso, P. A. Martins, C. V. Ramos, M. M. Cordeiro, R. J. Leote, K. R. Naqvi, W. L. Vaz and M. J. Moreno, *Biochim. Biophys. Acta, Biomembr.*, 2020, **1862**, 183157.
- 79 S. Darvishmanesh, J. Vanneste, E. Tocci, J. C. Jansen, F. Tasselli, J. Degrève, E. Drioli and B. Van der Bruggen, *J. Phys. Chem. B*, 2011, **115**, 14507–14517.
- 80 A. M. Matuszek and J. Reynisson, *Mol. Inf.*, 2016, **35**, 46–53.
- 81 H. Sun and D. O. Scott, *Chem. Biol. Drug Des.*, 2010, **75**, 3–17.

