Open-source Δ-quantum machine learning for medicinal chemistry

Kenneth Atz,[†,§] Clemens Isert,[†,§] Markus N. A. Böcker,[†] José Jiménez-Luna,[*,†,‡] and Gisbert Schneider[*,†,¶]

†Department of Chemistry and Applied Biosciences, RETHINK, ETH Zurich, 8093 Zurich, Switzerland

‡Department of Medicinal Chemistry, Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Straße 65, 88397 Biberach an der Riss, Germany

¶ETH Singapore SEC Ltd., 1 CREATE Way, #06-01 CREATE Tower, Singapore 138602, Singapore

§Contributed equally to this work

E-mail: joluna@ethz.ch; gisbert@ethz.ch

Abstract

Certain 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 compounds currently makes large-scale applications of quantum chemistry challenging. In order to mitigate this problem, we developed DelFTa, an open-source toolbox for predicting small-molecule electronic properties at the density functional (DFT) level of theory, using the Δ-machine learning principle. DelFTa employs state-of-the-art E(3)-equivariant graph neural networks that were trained on the QMugs dataset of QM properties. It provides access to a wide array of quantum observables by predicting approximations to ωB97X-D/def2-SVP values from a GFN2-xTB semiempirical baseline. Δ-learning with DelFTa was shown to outperform direct DFT learning for most of the considered QM endpoints. The software is provided as open-source code with fully-documented command-line and Python APIs.

Introduction

The electronic structure of drug-like molecules is responsible for various drug-relevant properties, such as molecular recognition in protein–ligand complexes,[1–5] drug-induced phototoxicity,[6,7] reactivity for covalent ligand-protein interaction,[8–10] cell membrane permeability,[11,12] or three-dimensional (3D) conformation energies,[13] to name a few.
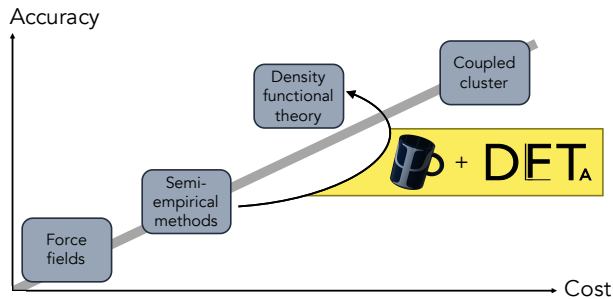

Figure 1: Overall idea behind the DelFTa software package. Equivariant Δ-learning models were trained to predict the correction (Δ) between observables of a lower-cost semiempirical baseline method (GFN2-xTB) and the corresponding DFT-level reference method (ωB97X-D/def2-SVP). Visualization inspired by reference 14.

Despite advances in density functional theory (DFT) approaches,[15,16] which are widely regarded as an attractive compromise between chemical accuracy and computational cost,[17,18] calculating quantum-mechanical (QM) properties at the DFT level of theory for many or for sizable molecules remains a computationally expensive task. Cheaper alternatives such as force fields[19] and semiempirical methods[20,21] 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 try to approximate quantum observables through statistical modeling approaches.[22–26] Geometric deep learning in particular, a discipline focused on the investigation of neural network architectures that incorporate symmetry information into their design,[27–29] 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,[30–37] forces for molecular dynamics simulations,[31,38,39] and wave functions in the form of local bases of atomic orbitals.[40,41]

In parallel to these developments, the so-called Δ-QML (delta-QML) approaches, which aim to learn corrections (Δ values) between computationally inexpensive QM methods and more accurate, albeit more expensive, ones have shown to deliver useful results.[42] Machine-learned corrections of this kind have been reported for both coupled cluster theory[43,44] via DFT, and for DFT via the semiempirical family of methods GFN-xTB.[45,46] However, there are no open-source implementations of Δ-QML or readily available trained models, which severely limits their widespread adoption. Addressing this need, we here present DelFTa, an open-source deep-learning toolbox that enables both fast and accurate approximations of molecular electronic properties on the ωB97X-D/def2-SVP[47,48] level of theory. DelFTa is trained on the QMugs[49] dataset, which enables Δ-learning between a comprehensive array of QM observables both at semiempirical and DFT levels of theory (Figure 1). Specifically, DelFTa learns to predict QM observables on the ωB97X-D/def2-SVP level of theory, either directly or via Δ-learning through corrections to the semiempirical GFN2-xTB method.[21,50–52] DelFTa employs E(3)-equivariant message-passing neural networks (EGNN), which are able to learn properties at the global graph (molecule), node (atom) and edge (bond) levels[29] (Figure 2). The utility of the presented work is threefold:

1. DelFTa expands the Δ-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 Mulliken partial charges, Wiberg bond orders, energies of the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO, respectively), HOMO/LUMO gap, and dipole moments.

2. Results show that Δ-ML achieves higher prediction accuracy compared to direct learning for most of the considered QM endpoints, in both multi- and single-task paradigms.

3. DelFTa provides an open-source package with both command-line and Python APIs, as well as extensive documentation and tutorials, so that interested researchers are able to use the provided models, train new ones, or build upon them.

We believe that the provided application, which enables access to a variety of QM properties at close-to-DFT accuracy, both in a fast and user-friendly manner, can be routinely used in numerous relevant applications in molecular modeling and design.

Methods

Reference dataset

DelFTa builds upon the QMugs[49] data collection, which features ~2M conformers of over
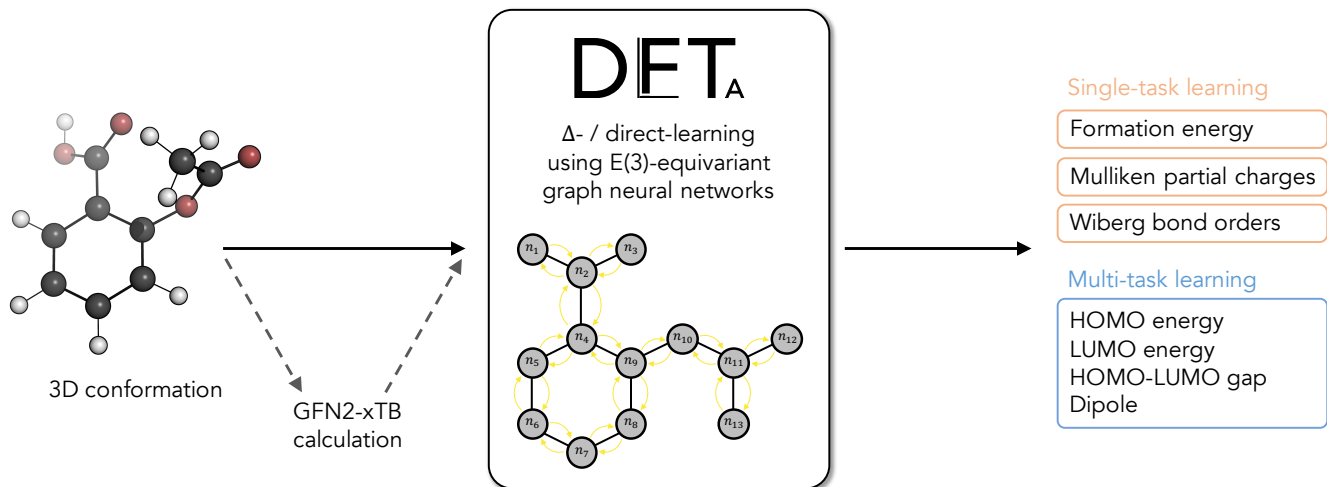
Figure 2: Schematic of the DelFTa concept. A three-dimensional (3D) molecular conformation is used as an input to either single- or multi-task trained E(3)-Equivariant Graph Neural Networks (EGNNs). When a Δ-learning endpoint is requested, an additional GFN2-xTB calculation will be carried out, and the network is tasked with predicting the correction (Δ) between this baseline value and its ωB97X-D/def2-SVP analogue. If a direct-learning prediction is requested, the network will output an approximation to the ωB97X-D/def2-SVP value, without using the GFN2-xTB baseline.

665k molecules extracted from the ChEMBL database (release 27),[53] to extract training, validation, and test sets. It includes QM properties at two levels of theory, namely the semiempirical method GFN2-xTB[21,50–52] and DFT (ωB97X-D/def2-SVP[47,48]).

Neural network architecture and training

The E(3)-Equivariant Graph Neural Network (EGNN) architecture[32,54] was used for learning the desired QM endpoints. EGNNs implement a specific neural message-passing mechanism[55] that has the advantage of transforming learned features equivariantly under the actions of the E(3) Euclidean symmetry group. These include coordinate equivariance under rotations, translations and reflections in 3D Cartesian space, as well as permutations (*i.e.*, node ordering).[29]

Δ-learning models for all endpoints y_i, at either global, node, or edge levels, associated with the i-th molecular conformation, were trained to learn the difference between DFT-computed properties ($y_i^{\mathrm{DFT}} \in \mathbb{R}^k$) and GFN2-xTB equivalents ($y_i^{\mathrm{GFN2-xTB}} \in \mathbb{R}^k$), specifically:

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

Direct-learning models were trained to predict y_i^{DFT}. In the proposed architecture, and for most of the considered endpoints, molecular conformations were represented as fully-connected graphs $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$, alongside their corresponding nodes $v_i \in \mathcal{V}$, edges $e_{ij} \in \mathcal{E}$ and the respective Cartesian coordinates in 3D space $\boldsymbol{r}_i \in \mathbb{R}^3$. Nodes were featurized via a linear embedding of their respective atom types and initial edge features with sinusoidal and cosinusoidal encoding of their pairwise diatomic distances $\|\boldsymbol{r}_i - \boldsymbol{r}_j\|_2^2$ (*i.e.*, a Fourier-like encoding scheme). Contrary to other published Δ-learning approaches, *e.g.*, OrbNet,[45,46] no explicit GFN-xTB edge or node features were used in the DelFTa models, as these did not show substantial performance improvements in preliminary experiments.

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

$$\boldsymbol{v}_i^{l+1} = \mathrm{EGCL}\left(\boldsymbol{v}_i^l, \boldsymbol{r}_i, \mathcal{E}\right), \qquad (2)$$

using the following message-passing mechanism:

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

where $\phi_{h,e}$ are node and edge non-linear transformations modeled with multilayer Perceptrons non-linearized with the SiLU activation function,[56] $\boldsymbol{r}_{ij}$ are the pairwise Euclidean distance features between nodes i and j, $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ their respective atom-level embeddings, $\boldsymbol{m}_{ij}$ the computed edge message features, and $\boldsymbol{m}_i$ the aggregated message features per node.

Networks for all endpoints, and for both direct- and Δ-learning paradigms, were trained using the Adam stochastic gradient descent optimizer[57] with a starting learning rate of 10^{-4} and a mean-squared-error loss. Node features $\boldsymbol{v}_i$ were mean-pooled after five message-passing steps and then mapped to their corresponding target shapes via an additional multi-layer Perceptron. In the specific cases of the node-based endpoints (*i.e.*, Mulliken partial charges), the learned node-level features $\boldsymbol{v}_i$ were used directly for prediction without pooling. In the specific case of Wiberg bond orders, the only considered edge-based endpoint in this study, networks were trained on covalently-connected graphs, using the learned message features $\boldsymbol{m}_{ij}$.

The following network hyper-parameters were used in all models considered in this study: (i) node dimension $\boldsymbol{v}_i$: 128, (ii) message passing dimension $\boldsymbol{m}_{ij}$: 32 for molecular and atomic models and 64 for the Wiberg bond order models, (iii) sinusoidal and cosinusoidal distance encoding features: 32, (iv) number of equivariant graph-convolutional layers: 5, and (v) number of global multi-layer Perceptrons: 3, each containing 256 hidden units. Furthermore, since 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. Additional technical training details are reported in the Supporting Information to this manuscript.

Model validation

While the production models available in the DelFTa application were trained on the entire QMugs dataset, model performance was benchmarked for different training set sizes featuring both single molecular conformations per molecule (100, 1k, 10k, 100k and ~547k training samples) as well as multiple ones (~1.6M individual conformers of ~547k distinct molecules). All models were trained using the Adam stochastic gradient descent optimizer,[57] with an early-stopping strategy that monitored the monotonic decrease of the chosen loss function on a validation set that consisted of ~29k randomly-selected molecules (~88k individual conformers). Each molecule in the QMugs dataset, associated with a unique ChEMBL identifier, was assigned to either the training, validation or test sets, with all conformers of one molecule becoming part of the same set. All optimized models were tested on three similarly-sized and non-overlapping test sets of ~29k molecules each (~88k individual conformers). In the QMugs dataset, some molecules with distinct ChEMBL[53] identifiers are represented with the same SMILES[58] notation (*i.e.*, the same 2D molecular graph). In order to avoid information leakage from the test sets, all molecules that shared their SMILES representation with another one were assigned to the test sets (see reference 49 for details). The remaining test set molecules were chosen at random from the QMugs dataset.

Finally, with the exception of the non-global models considered in this study (*i.e.*, Mulliken charges and Wiberg bond order models), the performance of single-task versus multi-task learning was investigated.
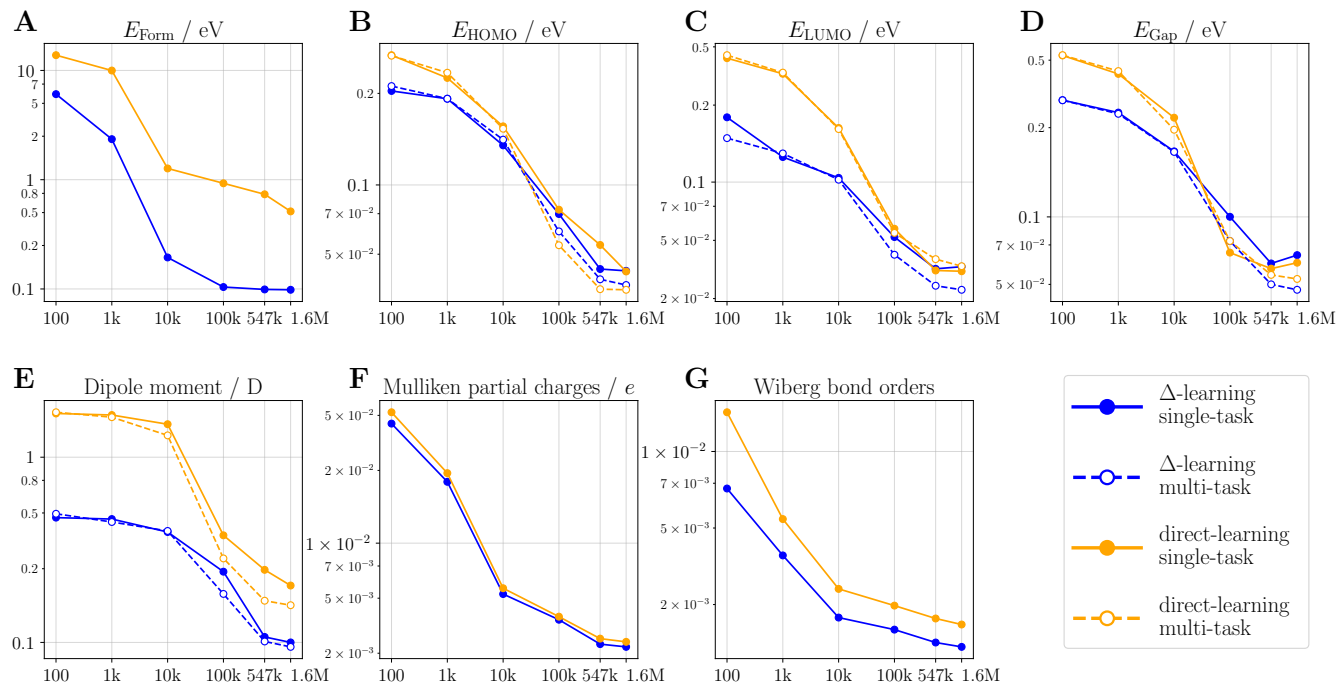
4

Figure 3: Learning curves of both Δ- and direct-learning for different training set sizes. Orbital energies and dipole moment learning curves are shown for both single- and multi-task learning. MAEs computed on the three test sets, each comprised of $\sim$29k molecules ($\sim$88k conformers).

Results

Test-set learning curves evaluated using models trained with varying training-set sizes are shown in Figure 3. For most of the considered endpoints, the Δ-learning models consistently achieved a better predictive performance than the direct-learning models, as measured by lower mean absolute errors (MAEs) w.r.t. DFT reference values. This performance gap was reproduced for most training set sizes, which also highlights the potential usefulness of Δ-learning in low-data regimes. The multi-task models generally outperformed their single-task counterparts, with the notable exception of formation energy (not shown). Due to this discrepancy, the DelFTa production application code implements single-task inference by default for this endpoint, and multi-task inference for the remaining ones, trained explicitly without formation energy data.

The predictive performance of the models trained on 1.6M training conformers was analyzed in more detail. Table 1 shows MAEs w.r.t. the ωB97X-D/def2-SVP reference values obtained for the test sets, and compares them to MAEs of the utilized baseline method GFN2-xTB. For all considered endpoints, the DelFTa application approximates the ωB97X-D/def2-SVP reference values more closely than the semiempirical GFN2-xTB baseline. The direct-learning approach achieved a slightly lower MAE (0.0350 eV) on the prediction of HOMO energies than its Δ-learning counterpart (0.0367 eV). Scatter plots showing Δ-predicted properties versus their DFT reference values are provided in Figure 4.

We compared the implementation of the EGNN models used in this work to the one originally reported.[32] With that goal, we provide results on the QM9 dataset,[59] a well-known benchmark used in previous QML studies which features quantum observables for $\sim$134k small molecules. The same model architecture used in this application for the direct-learning of formation energies was retrained on the QM9 training set (composed of 100k molecules), validated and tested on its respective validation and test

Table 1: Benchmark results showing MAEs ($\pm$ 1 standard deviation) w.r.t. ωB97X-D/def2-SVP reference values for $\sim$88k molecules ($\sim$263k conformers) from the three test sets. Relative energy difference computed as the average pairwise energy difference between the individual conformers of a molecule. Wiberg bond order results for covalent bonds only. Bold numbers highlight the method with the lowest MAE w.r.t. reference values.

Property	Unit	GFN2-xTB	DelFTa	
			Δ-learning	direct-learning
Formation energy	eV	86.34 ($\pm$ 0.04)	**0.0983** ($\pm$ 0.0004)	0.513 ($\pm$ 0.001)
HOMO energy	eV	2.1154 $\pm$ (0.0005)	0.0367 ($\pm$ 0.0001)	**0.0350** ($\pm$ 0.0001)
LUMO energy	eV	7.7730 ($\pm$ 0.0007)	**0.0278** ($\pm$ 0.0002)	0.0368 ($\pm$ 0.0002)
HOMO-LUMO gap	eV	5.658 ($\pm$ 0.001)	**0.0473** ($\pm$ 0.0002)	0.0529 ($\pm$ 0.0001)
Total molecular dipole	D	0.622 ($\pm$ 0.002)	**0.0946** ($\pm$ 0.0006)	0.1588 ($\pm$ 0.0006)
Mulliken partial charges	e	0.0610 ($\pm$ 0.0000)	**0.0027** ($\pm$ 0.0000)	0.0029 ($\pm$ 0.0000)
Wiberg bond orders	-	0.0634 ($\pm$ 0.0001)	**0.0017** ($\pm$ 0.0000)	0.0021 ($\pm$ 0.0000)
Conformer pairwise energy difference	eV	0.0736 ($\pm$ 0.0004)	**0.0612** ($\pm$ 0.0004)	0.197 ($\pm$ 0.001)

sets (each comprised of $\sim$15k molecules). The trained models achieved an MAE of 15.1 ($\pm$0.4) meV on four independent model runs, which is roughly in the same ballpark as the originally-reported performance (12 meV), and suggests that the chosen architecture was successfully reimplemented. The lower overall error of the models trained on QM9 compared to those trained on QMugs can be attributed to two key differences between the sets, namely atom type diversity (10 different atom types in QMugs, 5 in QM9), and molecular size (up to 100 and 9 heavy atoms for QMugs and QM9, respectively).

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.

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

mol = readstring("smi", "CCO")
calc = DelftaCalculator()
predictions = calc.predict(mol)
print(predictions)
# >> {"E_form": -1.2982624, "E_lumo
    ": 0.18278737, ...
```

Software

DelFTa is fully implemented in the Python programming language[60] and uses the PyTorch package[61] (version 1.8.0) and PyTorch Geometric package[62] (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[21,50–52] level of theory are computed via open-source `xtb` binaries. All molecular manipulation routines are integrated into DelFTa and handled via the Pybel package[63] and OpenBabel[64] Python bindings. The DelFTa package is fully open-sourced, available on GitHub (https://github.com/josejimenezluna/delfta) under a permissive AGPLv3 license, and distributed through the conda package manager.[65] A Docker[66] 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[67] are also available.

On a computer with a single Nvidia GTX 1080Ti graphics processing unit, DelFTA predicts all considered endpoints at a speed of approximately 50 and 5 molecules per second for the direct and Δ-learning models respectively, with the latter approach mostly bot-
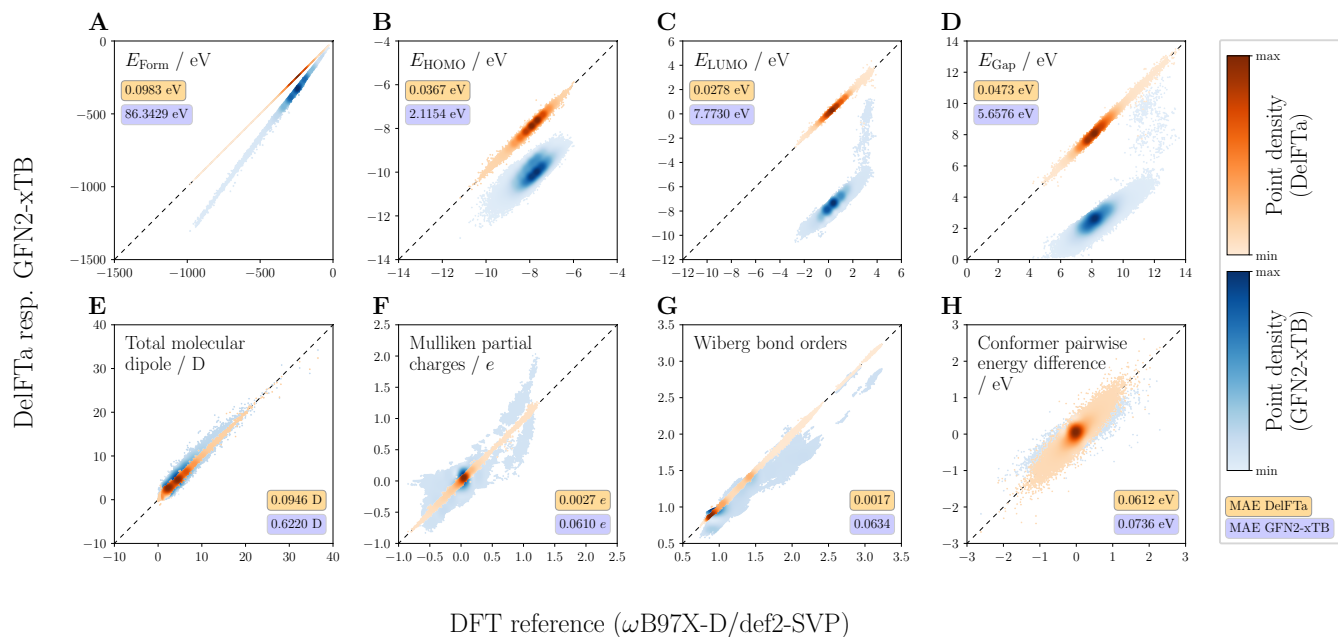
Figure 4: Scatter plots illustrating the accuracy of the predictions provided by the trained Δ-learning models and the GFN2-xTB baseline, w.r.t. DFT reference properties (ωB97X-D/def2-SVP) for $\sim$88k test set molecules ($\sim$263k conformers). Δ-learning predictions obtained using the models trained on the 1.6M conformer training set, and with single-/multi-task settings as described in the Methods section. MAEs w.r.t. DFT reference. Wiberg bond orders for covalent bonds only. Colorbars scaled individually for each property.

tlenecked by the additionally-required baseline GFN2-xTB calculations.

Discussion

The models provided under the DelFTa application pave the way for large-scale applications of quantum chemistry in molecular modeling for small-molecule drug discovery by providing fast and accurate property predictions at close-to-DFT accuracy. The trained machine-learning models were validated for both Δ- and direct-, as well as single- and multi-task learning paradigms. The results shown in this study suggest that the DelFTA application shows improved accuracy over its GFN2-xTB baseline in approximating ωB97X-D/def2-SVP reference values for many of the considered endpoints on previously unseen bioactive molecules, at roughly the same computational cost.

These models enable fast and accurate access to drug-relevant QM features for many molecules, thereby facilitating their integration into small-molecule computational pipelines for both predictive and generative tasks. 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.[68] As the electronic structure of drug-like molecules governs many related properties, many of such effects can be anticipated in medicinal chemistry. Potential examples include the influence of (i) HOMO/LUMO energies on phototoxicity,[6,7] (ii) dipole moments on aqueous solubility[69,70] as well as on membrane permeability,[11,12] (iii) formation energies on 3D-conformer ensembles[11] or site-of-metabolism prediction,[71] and (iv) partial charges and bond orders on non-covalent interactions.[72]

Future prospective application will reveal the practical applicability and usefulness of DelFTa for drug discovery-related tasks. The current limitations are twofold, concerning the modelling performance and applicability domain.

With regards to modelling performance, we note that while the Δ-learning approach affords substantial improvements over the direct-learning approach w.r.t. ωB97X-D/def2-SVP, this does not hold in the case of HOMO energies. Additionally, while the conformer pairwise energy Δ-learning model clearly outperforms its direct-learning analogue, the performance of the former is arguably close to that one of GFN2-xTB. This issue could be mitigated in future work by the use of a loss function that explicitly takes into account conformer energy differences at the training stage.[45,46,73] Finally, the previously-mentioned conclusions do not necessarily hold for reference values computed with a more comprehensive basis set, namely those obtained via ωB97X-D/def2-QZVP. A preliminary test computed for $2,874$ conformations corresponding to 958 distinct molecules did not indicate superior performance of Δ-learning over direct learning (see Supporting Information). Furthermore, Mulliken partial charges and conformer pairwise energy differences on the ωB97X-D/def2-QZVP level of theory were better approximated using GFN2-xTB than with either of the provided DelFTa models.

Limitations in regards to applicability domain issues mostly stem from the underlying QMugs training dataset, as it was conceived with medicinal chemistry applications in mind. For example, it does not feature organometallic complexes, polymers, crystalline structures, or molecular systems which include dimers, radicals, excited electronic states, higher order spin states, charged molecules, or off-equilibrium structures. Predictive models for these types of molecular structures would require training data that specifically covers the respective chemical space, and therefore remain a subject of future work.

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

References

(1) Khrenova, M.; Nemukhin, A. V.; Grigorenko, B. L.; Krylov, A.; Domratcheva, T. Quantum chemistry calculations provide support to the mechanism of the light-induced structural changes in the flavin-binding photoreceptor proteins. *J. Chem. Theory Comput* **2010**, *6*, 2293–2302.

(2) Xie, N.-Z.; Du, Q.-S.; Li, J.-X.; Huang, R.-B. Exploring strong interactions in proteins with quantum chemistry and examples of their applications in drug design. *PLOS ONE* **2015**, *10*, e0137113.

(3) Schaftenaar, G.; Vlieg, E.; Vriend, G. Molden 2.0: quantum chemistry meets proteins. *J. Comput.-Aided Mol. Des* **2017**, *31*, 789–800.

(4) Tavares, A. B. M.; Neto, J. X. L.; Fulco, U. L.; Albuquerque, E. L. Inhibition of the checkpoint protein PD-1 by the therapeutic antibody pembrolizumab outlined by quantum chemistry. *Sci. Rep* **2018**, *8*, 1–13.

(5) Gundelach, L.; Fox, T.; Tautermann, C. S.; Skylaris, C.-K. Protein–ligand free energies of binding from full-protein DFT calculations: Convergence and choice of exchange–correlation functional. *Phys. Chem. Chem. Phys.* **2021**, *23*, 9381–9393.

(6) Freccero, M.; Fasani, E.; Mella, M.; Manet, I.; Monti, S.; Albini, A. Modeling the photochemistry of the reference phototoxic drug lomefloxacin by steady-state and time-resolved experiments, and DFT and post-HF calculations. *Chem. Eur. J.* **2008**, *14*, 653–663.

(7) Llano, J.; Raber, J.; Eriksson, L. A. Theoretical study of phototoxic reactions of psoralens. *J. Photochem. Photobiol. A* **2003**, *154*, 235–243.

(8) Yu, H. S.; Gao, C.; Lupyan, D.; Wu, Y.; Kimura, T.; Wu, C.; Jacobson, L.; Harder, E.; Abel, R.; Wang, L. Toward atomistic modeling of irreversible covalent inhibitor binding kinetics. *J. Chem. Inf. Model.* **2019**, *59*, 3955–3967.

(9) Zhao, Z.; Liu, Q.; Bliven, S.; Xie, L.; Bourne, P. E. Determining cysteines available for covalent inhibition across the human kinome. *J. Med. Chem.* **2017**, *60*, 2879–2889.

(10) Fanfrlik, J.; Brahmkshatriya, P. S.; Rezac, J.; Jilkova, A.; Horn, M.; Mares, M.; Hobza, P.; Lepsik, M. Quantum mechanics-based scoring rationalizes the irreversible inactivation of parasitic Schistosoma mansoni cysteine peptidase by vinyl sulfone inhibitors. *J. Phys. Chem. B.* **2013**, *117*, 14973–14982.

(11) Pultar, F.; Hansen, M. E.; Wolfrum, S.; Böselt, L.; Fróis-Martins, R.; Bloch, S.; Kravina, A. G.; Pehlivanoglu, D.; Schäffer, C.; LeibundGut-Landmann, S.; Riniker, S.; Carreira, E. M. Mutanobactin D from the human microbiome: Total synthesis, configurational assignment, and biological evaluation. *J. Am. Chem. Soc.* **2021**, *143*, 10389–10402.

(12) Matuszek, A. M.; Reynisson, J. Defining known drug space using DFT. *Mol. Inform.* **2016**, *35*, 46–53.

(13) Rupp, M.; Bauer, M. R.; Wilcken, R.; Lange, A.; Reutlinger, M.; Boeckler, F. M.; Schneider, G. Machine learning estimates of natural product conformational energies. *PLoS Comput. Biol.* **2014**, *10*, e1003400.

(14) von Lilienfeld, O. A. Quantum Machine Learning. *31st Conference on Neural Information Processing Systems* **2017**,

(15) Burke, K. Perspective on density functional theory. *J. Chem. Phys.* **2012**, *136*, 150901.

(16) Sherrill, C. D. Frontiers in electronic structure theory. *J. Chem. Phys.* **2010**, *132*, 110902.

(17) Von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. Optimization of effective atom centered potentials for London dispersion forces in density functional theory. *Phys. Rev. Lett.* **2004**, *93*, 153004.

(18) Schwabe, T.; Grimme, S. Theoretical thermodynamics for large molecules: Walking the thin line between accuracy and computational cost. *Acc. Chem. Res.* **2008**, *41*, 569–579.

(19) Wang, S.; Witek, J.; Landrum, G. A.; Riniker, S. Improving conformer generation for small rings and macrocycles based on distance geometry and experimental torsional-angle preferences. *J. Chem. Inf. Model.* **2020**, *60*, 2044–2058.

(20) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173–1213.

(21) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.

(22) von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nat. Rev. Chem.* **2020**, 1–12.

(23) Huang, B.; von Lilienfeld, O. A. Quantum machine learning using atom-in-molecule-based fragments selected on the fly. *Nat. Chem.* **2020**, *12*, 945–951.

(24) Christensen, A. S.; Bratholm, L. A.; Faber, F. A.; Anatole von Lilienfeld, O. FCHL revisited: Faster and more accurate quantum machine learning. *J. Chem. Phys.* **2020**, *152*, 044107.

(25) Heinen, S.; von Rudorff, G. F.; von Lilienfeld, O. A. Toward the design of chemical reactions: Machine learning barriers of competing mechanisms in reactant space. *J. Chem. Phys.* **2021**, *155*, 064105.

(26) Lemm, D.; von Rudorff, G. F.; von Lilienfeld, O. A. Machine learning based energy-free structure predictions of molecules, transition states, and solids. *Nat. Commun* **2021**, *12*, 1–10.

(27) Bronstein, M. M.; Bruna, J.; LeCun, Y.; Szlam, A.; Vandergheynst, P. Geometric deep learning: going beyond euclidean data. *IEEE Signal Process. Mag.* **2017**, *34*, 18–42.

(28) Bronstein, M. M.; Bruna, J.; Cohen, T.; Veličković, P. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv:2104.13478* **2021**,

(29) Atz, K.; Grisoni, F.; Schneider, G. Geometric deep learning on molecular representations. *arXiv:2107.12375* **2021**,

(30) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. Schnet–a deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148*, 241722.

(31) Schütt, K. T.; Unke, O. T.; Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. *arXiv:2102.03150* **2021**,

(32) Satorras, V. G.; Hoogeboom, E.; Welling, M. E-(n) equivariant graph neural networks. *arXiv:2102.09844* **2021**,

(33) Anderson, B.; Hy, T.-S.; Kondor, R. Cormorant: Covariant molecular neural networks. *arXiv:1906.04015* **2019**,

(34) Thomas, N.; Smidt, T.; Kearnes, S.; Yang, L.; Li, L.; Kohlhoff, K.; Riley, P. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv:1802.08219* **2018**,

(35) Klicpera, J.; Groß, J.; Günnemann, S. Directional message passing for molecular graphs. *arXiv:2003.03123* **2020**,

(36) Klicpera, J.; Giri, S.; Margraf, J. T.; Günnemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv:2011.14115* **2020**,

(37) Miller, B. K.; Geiger, M.; Smidt, T. E.; Noé, F. Relevance of rotationally equivariant convolutions for predicting molecular properties. *arXiv:2008.08461* **2020**,

(38) Unke, O. T.; Meuwly, M. PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693.

(39) Klicpera, J.; Becker, F.; Günnemann, S. GemNet: Universal directional graph meural metworks for molecules. *arXiv:2106.08903* **2021**,

(40) Unke, O. T.; Bogojeski, M.; Gastegger, M.; Geiger, M.; Smidt, T.; Müller, K.-R. SE (3)-equivariant prediction of molecular wavefunctions and electronic densities. *arXiv:2106.02347* **2021**,

(41) Schütt, K.; Gastegger, M.; Tkatchenko, A.; Müller, K.-R.; Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural

network for molecular wavefunctions. *Nat. Commun.* **2019**, *10*, 1–10.

(42) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big data meets quantum chemistry approximations: The Δ-machine learning approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.

(43) Smith, J. S.; Zubatyuk, R.; Nebgen, B.; Lubbers, N.; Barros, K.; Roitberg, A. E.; Isayev, O.; Tretiak, S. The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules. *Sci. Data* **2020**, *7*, 1–10.

(44) Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Bowman, J. M. Δ-machine learning for potential energy surfaces: A PIP approach to bring a DFT-based PES to CCSD (T) level of theory. *J. Chem. Phys.* **2021**, *154*, 051102.

(45) Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F. R.; Miller III, T. F. OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *J. Chem. Phys.* **2020**, *153*, 124111.

(46) Christensen, A. S.; Sirumalla, S. K.; Qiao, Z.; O'Connor, M. B.; Smith, D. G.; Ding, F.; Bygrave, P. J.; Anandkumar, A.; Welborn, M.; Manby, F. R., et al. Orb-Net Denali: A machine learning potential for biological and organic chemistry with semi-empirical cost and DFT accuracy. *arXiv:2107.00299* **2021**,

(47) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(48) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

(49) Isert, C.; Atz, K.; Jiménez-Luna, J.; Schneider, G. QMugs: Quantum Mechanical Properties of Drug-like Molecules. *arXiv:2107.00367* **2021**,

(50) Grimme, S.; Bannwarth, C.; Shushkov, P. A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (Z= 1–86). *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.

(51) Grimme, S. Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.

(52) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, e01493.

(53) Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cibrián-Uhalte, E., et al. The ChEMBL database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954.

(54) EGNN-PyTorch. https://github.com/lucidrains/egnn-pytorch, 2021.

(55) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural message passing for quantum chemistry. International Conference on Machine Learning. 2017; pp 1263–1272.

(56) Elfwing, S.; Uchibe, E.; Doya, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Netw.* **2018**, *107*, 3–11.

(57) Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization. *arXiv:1412.6980* **2014**,

(58) Weininger, D. SMILES, a Chemical Language and Information System: 1: Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.

(59) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **2014**, *1*, 1–7.

(60) The Python Language Reference. https://docs.python.org/3/reference/.

(61) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L., et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* **2019**, *32*, 8026–8037.

(62) Fey, M.; Lenssen, J. E. Fast Graph Representation Learning with PyTorch Geometric. ICLR Workshop on Representation Learning on Graphs and Manifolds. 2019.

(63) O'Boyle, N. M.; Morley, C.; Hutchison, G. R. Pybel: A Python wrapper for the OpenBabel cheminformatics toolkit. *Chem. Cent. J.* **2008**, *2*, 1–7.

(64) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminformatics* **2011**, *3*, 1–14.

(65) Conda package manager. https://conda.io.

(66) Merkel, D. Docker: Lightweight Linux containers for consistent development and deployment. *Linux J.* **2014**, 2.

(67) Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; Ivanov, P.; Avila, D.; Abdalla, S.; Willing, C. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*; Loizides, F., Schmidt, B., Eds.; 2016; pp 87 – 90.

(68) Stuyver, T.; Coley, C. W. Quantum chemistry-augmented neural networks for reactivity prediction: Performance, generalizability and interpretability. *arXiv:2107.10402* **2021**,

(69) Cardoso, R. M.; Martins, P. A.; Ramos, C. V.; Cordeiro, M. M.; Leote, R. J.; Naqvi, K. R.; Vaz, W. L.; Moreno, M. J. Effect of dipole moment on amphiphile solubility and partition into liquid ordered and liquid disordered phases in lipid bilayers. *Biochim. Biophys. Acta Biomembr.* **2020**, *1862*, 183157.

(70) Darvishmanesh, S.; Vanneste, J.; Tocci, E.; Jansen, J. C.; Tasselli, F.; Degrève, J.; Drioli, E.; Van der Bruggen, B. Physicochemical characterization of solute retention in solvent resistant nanofiltration: the effect of solute size, polarity, dipole moment, and solubility parameter. *J. Phys. Chem. B* **2011**, *115*, 14507–14517.

(71) Sun, H.; Scott, D. O. Structure-based drug metabolism predictions for drug design. *Chem. Biol. Drug Des.* **2010**, *75*, 3–17.

(72) Sahai, M. A.; Biggin, P. C. Quantifying water-mediated protein–ligand interactions in a glutamate receptor: A DFT study. *J. Phys. Chem. B* **2011**, *115*, 7085–7096.

(73) Qiao, Z.; Christensen, A. S.; Manby, F. R.; Welborn, M.; Anandkumar, A.; Miller III, T. F. UNiTE: Unitary N-body tensor equivariant network with applications to quantum chemistry. *arXiv:2105.14655* **2021**,