


 Cite this: *RSC Adv.*, 2024, 14, 4492

# Exploring protein–ligand binding affinity prediction with electron density-based geometric deep learning†

 Clemens Isert,  Kenneth Atz,  Sereina Riniker  and Gisbert Schneider \*

Rational structure-based drug design relies on accurate predictions of protein–ligand binding affinity from structural molecular information. Although deep learning-based methods for predicting binding affinity have shown promise in computational drug design, certain approaches have faced criticism for their potential to inadequately capture the fundamental physical interactions between ligands and their macromolecular targets or for being susceptible to dataset biases. Herein, we propose to include bond-critical points based on the electron density of a protein–ligand complex as a fundamental physical representation of protein–ligand interactions. Employing a geometric deep learning model, we explore the usefulness of these bond-critical points to predict absolute binding affinities of protein–ligand complexes, benchmark model performance against existing methods, and provide a critical analysis of this new approach. The models achieved root-mean-squared errors of 1.4–1.8 log units on the PDBbind dataset, and 1.0–1.7 log units on the PDE10A dataset, not indicating significant advantages over benchmark methods, and thus rendering the utility of electron density for deep learning models context-dependent. The relationship between intermolecular electron density and corresponding binding affinity was analyzed, and Pearson correlation coefficients  $r > 0.7$  were obtained for several macromolecular targets.

 Received 18th December 2023  
Accepted 19th January 2024

DOI: 10.1039/d3ra08650j

[rsc.li/rsc-advances](https://rsc.li/rsc-advances)

## Introduction

A key requirement for most drug candidates is a high binding affinity of the investigated ligand to its desired biological target.<sup>1</sup> However, the experimental determination of binding affinity can be time- and resource-intensive, often requiring synthesis of the ligand and a suitable experimental assay. Aiming to limit the number of laboratory experiments and enable more rapid ligand design, several computational approaches have been developed for *in silico* binding affinity prediction.<sup>2</sup> Structure-based deep learning has received particular attention for this specific task,<sup>3–5</sup> as well as for binding site identification,<sup>6,7</sup> molecular docking,<sup>8,9</sup> and *de novo* molecular design.<sup>10</sup>

Although different deep learning-based techniques for predicting binding affinity have achieved success in computational drug design studies,<sup>11,12</sup> certain deep learning models have been criticized for specific challenges regarding their generalization abilities and their potential to adequately capture the fundamental physical principles governing intermolecular interactions.<sup>2,13–15</sup> Much simpler methods such as nearest-

neighbor analysis or prediction of random values have been shown to achieve a similar predictive performance.<sup>16</sup> Furthermore, existing datasets such as PDBbind<sup>17,18</sup> commonly used for model development have been criticized as biased because similar prediction accuracies were achieved irrespective of whether the whole protein–ligand complex, only the protein, or only the ligand was considered.<sup>2,19,20</sup> This potential bias in dataset composition makes it challenging to meaningfully assess and compare model performance.<sup>2,16,19,20</sup> As a potential remedy to the challenge of learning underlying dataset biases rather than meaningfully capturing physical interactions, Rognan and coworkers suggested to consider “only noncovalent interactions while omitting their protein and ligand atomic environments”.<sup>2</sup>

In addition to geometric deep learning approaches,<sup>5,23–26</sup> also simulation-based approaches (*e.g.*, free energy perturbation,<sup>27,28</sup> MM/PBSA<sup>29–31</sup>) are frequently used for binding affinity prediction. While simulation-based methods are, by design, rooted in physical principles, deep-learning methods often incur lower computational cost for predictions and can leverage pre-existing data. Herein, we combine a commonly used simulation technique, the Quantum Theory of Atoms in Molecules (QTAIM),<sup>32</sup> with geometric deep learning neural networks.

The QTAIM analyzes the topology of the electron density surrounding a molecular geometry.<sup>32</sup> The electron density is highest at the nuclei, decreases as one moves away from

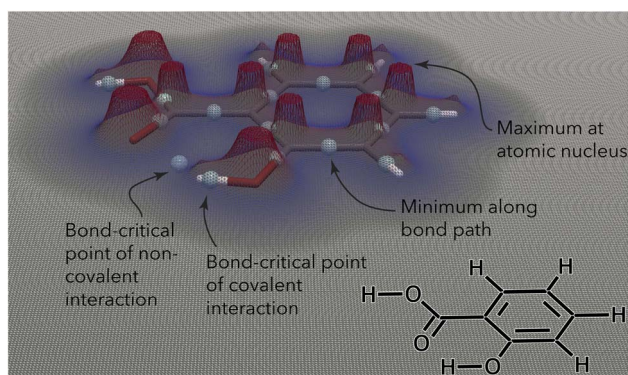
ETH Zurich, Department of Chemistry and Applied Biosciences, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland. E-mail: [gisbert@ethz.ch](mailto:gisbert@ethz.ch); Tel: +41 44 633 73 27

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d3ra08650j>



a nucleus, and eventually increases again as another nucleus is approached (Fig. 1). The QTAIM partitions the electron density in three-dimensional (3D) space into atomic basins, the boundaries between which are surfaces of zero flux of the gradient vector of the electron density. A bond path is formed between two interacting nuclei, following the line along which the electron density is maximal relative to its local surroundings. The point along this line with the lowest electron density (*i.e.*, the point at which the bond path crosses the zero-flux interatomic surface) is termed a “bond-critical point” (BCP).<sup>32,34</sup> A range of quantum-mechanical (QM) properties (Table 1) can be evaluated at BCPs and have been connected to physically observable quantities.<sup>32,34,41–43</sup>

Herein, we investigate the hypothesis that an electron density-based, interaction-centric view of protein–ligand complexes may be exploited with geometric deep learning to predict binding affinities. Our objective is to evaluate whether this approach offers advantages in comparison to established methods for predicting binding affinity. Choosing a molecular representation directly rooted in the electron density is motivated by the following three considerations: First, the electron density surrounding a molecule is uniquely mapped to all observable properties of that molecule.<sup>21</sup> Second, the electron density is a fundamental description of physical and chemical phenomena.<sup>32</sup> By providing deep learning models with a molecular description that is more closely connected to the observed phenomenon itself than commonly used model concepts (such as atomic nuclei as points in space), we hypothesized that the experimentally measured binding affinity might potentially be more accurately predicted. Third, the electron density and other derived properties at intermolecular BCPs were successfully employed in previous quantitative structure–activity relationship (QSAR) studies.<sup>32,34,41–48</sup> While the idea of combining QM with machine learning (ML) is not new (*e.g.*, ML to predict QM-calculated properties,<sup>49–54</sup> or QM-calculated features being used as ML inputs,<sup>55–59</sup>) this work represents, to the best of our knowledge, the first combination of BCPs with 3D-aware



**Fig. 1** Contour plot of the electron density (at  $z = 0$ , in plane with the phenyl ring) surrounding the salicylic acid molecule. The electron density reveals local maxima at the atomic nuclei and decreases as one moves away from the nuclei. The point with the lowest electron density along a bond path (either covalent or noncovalent) is termed a “bond-critical point”. Figure inspired by ref. 21 and created using ParaView.<sup>22</sup>

neural networks. Previous studies on BCP-based QSAR models mainly relied on aggregated information or scalar descriptors of structural information.<sup>41–43,60,61</sup> For example, a recent study showcased a strong correlation ( $r = 0.891$ ) between the sum of the electron density at BCPs and experimental protein–ligand binding affinity for 34 D2-dopamine receptor inhibitors.<sup>43</sup> The authors argued that this relationship depends on *e.g.*, the number of rotatable bonds in the series of ligands, as a low number of rotatable bonds may indicate that entropic effects (which cannot be captured by the static picture of a QTAIM analysis) can be neglected or are very similar across the set of ligands.<sup>43</sup>

This current study offers two key contributions: First, we investigated whether specific quantum mechanical (QM) properties and their spatial distribution can be utilized to predict binding affinity. We employed 3D message-passing neural networks (MPNNs) for this analysis. An automated Python-based pipeline was created to prepare protein structures, perform QM calculations, and conduct electron density analysis for this research. We conducted a comprehensive evaluation of the suggested method and discovered that, for two extensive data sets, the selected representation did not appear to offer any discernible advantages over established methods in the specific cases we examined. Our intention in sharing these findings is to offer valuable insights for future research directions that concentrate on the utilization of electron density-based descriptors for predicting protein–ligand binding affinity. Second, we analyzed the correlation between the sum of the electron density at BCPs and the measured binding affinities across two large-scale datasets of protein–ligand binding complexes (*i.e.*, PDBBind<sup>17,18</sup> and PDE10A<sup>62</sup>).

### Data compilation & processing

We performed our analysis using two datasets of protein–ligand complex structures: the commonly used PDBbind (version 2019) dataset<sup>17,18</sup> (as prepared in ref. 2) and a recently released collection of PDE10A inhibitors<sup>62</sup> originating from a former discovery project at Roche. In addition to consistently measured binding affinities, crystal structures, and expert-curated docking poses, the PDE10A dataset contributes dataset splits inspired by real-world drug discovery programs. These splitting strategies include temporal splits and splits according to binding mode, enabling a thorough investigation of a model’s ability to extrapolate to unseen types of interactions.<sup>62</sup> The PDE10A dataset may help address some of the shortcomings and biases previously identified in using the PDBbind dataset to assess model performance. These shortcomings include similar performance when using ligand-only or protein-only representations and a failure of PDBbind-trained models to meaningfully capture physical interactions.<sup>2</sup> In keeping with previous work<sup>63</sup> and to reduce computational cost, we removed protein residues in which all atoms are farther away than 6 Å from the ligand (see details in Section S1†).

### QM & QTAIM calculations

While previous QTAIM studies of biomolecular systems have used density functional theory (DFT) to obtain electron

Table 1 Quantum-mechanical properties at bond-critical points (BCPs) and their chemical interpretation

Property	Formula	Interpretation	Ref.
Electron density	$\rho$	Bond order/strength of interaction	33 and 34
Laplacian	$\nabla^2\rho$	Covalent character of interaction	34
Electron localization function	ELF	Degree of electron localization	35 and 36
Localized orbital locator	$t_\sigma$	Degree of electron localization	37
Reduced density gradient	RDG	Homogeneity of electron density & type of interactions	38
sign ( $\lambda_2$ ) $\rho$	sign ( $\lambda_2$ ) $\rho$	Strength & type of interactions <sup>a</sup>	38
Gradient norm	$ \nabla\rho $	Magnitude of the gradient	39
Bond ellipticity	$\epsilon$	Cylindrical symmetry and $\pi$ -character of interaction	34
ETA index	$\eta$	Type of interaction	40

<sup>a</sup>  $\lambda_2$  denotes the second-largest eigenvalue of the electron density's Hessian matrix.

densities for a few dozen structures,<sup>43,66</sup> such an approach is challenging to apply to tens of thousands of protein–ligand complexes, each consisting of hundreds to thousands of atoms. Accordingly, the semiempirical method GFN2-xTB<sup>67–70</sup> (version 6.4.1) was used for QM calculations using Analytical Linearized Poisson–Boltzmann (ALPB) implicit water solvent.<sup>71,72</sup> GFN2-xTB has been successfully used in several biological applications,<sup>63,73</sup> though the considerable speed-up with respect to DFT comes at the expense of lower accuracy.<sup>74</sup> However, it has been observed that the electron density shows little sensitivity to the electronic structure method being used for computation and that key features are well preserved between different methods.<sup>38,75,76</sup> In a preliminary study, we benchmarked GFN2-xTB against two DFT approaches,  $\omega$ B97X-D/def2-QZVP<sup>77,78</sup> and B3LYP-D3/6-31G\*,<sup>79,80</sup> calculated using Psi4 (ref. 81) (version 1.7). While we observed some discrepancies (Section S2†), the overall acceptable performance and dramatic speedup (5–6 or 3 orders of magnitude, respectively [Table S1†]) of GFN2-xTB *versus*  $\omega$ B97X-D/def2-QZVP or B3LYP-D3/6-31G\*, respectively, suggested GFN2-xTB as a suitable choice given the need for computational efficiency in such a large-scale investigation.

Following QM calculations, Multiwfn<sup>39</sup> (version 3.8(dev), 03/2023) was used to find BCPs and compute their QM properties (Table 1) based on the wavefunction files obtained from the GFN2-xTB calculation. RDKit<sup>82</sup> (version 2021.09.4) was used for general molecular processing tasks. At the end of this pipeline, 14 181 (out of 14 215) and 1162 (out of 1162) successfully processed graphs were obtained for PDBbind and PDE10A, respectively (Section S1†).

### Molecular representation

While we primarily investigated the usefulness of BCP-based, interaction-centric graphs (Fig. 2), we additionally investigated a related graph setup based on nucleus-critical points (NCPs). Although this approach deviates slightly from the recommendation of Rognan and coworkers<sup>2</sup> to use interaction-centric molecular representations, it appears as a natural alternative to BCP-based graphs for the use of MPNNs in the context of the QTAIM. Both graphs consist of nodes, edges, and node positional information ( $\mathcal{G} = (\nu, \mathcal{E}, \mathcal{R})$ ) and were constructed as follows.

**BCP graphs.** The nodes  $\nu$  in the graph  $\mathcal{G}$  correspond to BCPs, and edges  $\mathcal{E}$  were added between any pair of nodes within 6 Å (similar to previous studies<sup>83,84</sup>). Initial node features  $\mathbf{v}_i^0$  were the QM properties at the respective BCP (see details in Section S3†). Optionally, the identities of atoms connected by the BCP were added. Node-to-node distances  $d_{ij}$  described *via* a sinusoidal and cosinusoidal encoding (similar to other 3D-tasks<sup>85,86</sup>) were used as edge features  $e_{ij}$ .

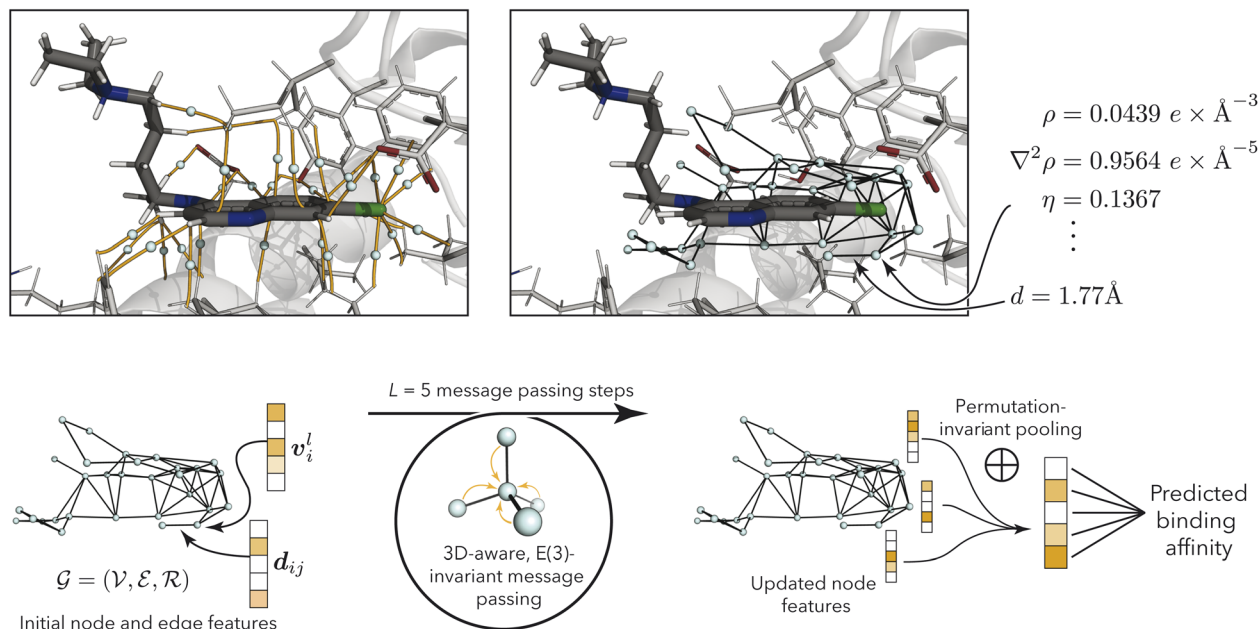
**NCP graphs.** Nodes  $\nu$  were placed at NCPs (atomic nuclei) that participate in interactions between the protein and ligand (as identified by the presence of intermolecular BCPs). Initial node features  $\mathbf{v}_i^0$  for NCPs were obtained from their QM properties (Table 1) and (optionally) atomic identities. Intermolecular edges were added between interacting NCPs (one on the protein side and one on the ligand side) and featurized using Fourier-like encoded distances and QM properties of the corresponding BCP. Additional intramolecular edges were introduced between all NCPs of the ligand and featurized using their respective BCPs for covalent bonds (if present). See Section S3† for full details.

Extreme QM values of individual BCPs or NCPs rendered common input normalization strategies unsuitable, prompting the use of a custom scaling method (Section S3†).

### Model architecture & training

3D-MPNNs based on the EGNN architecture,<sup>85,91</sup> which had previously shown good performance in several other 3D-based prediction tasks,<sup>55,86</sup> were used to operate on the BCP/NCP-based graphs. Initial node features were based on QM properties at the BCPs (for BCP-based graphs) or at the NCPs (for NCP-based graphs), respectively, and optionally combined with atom types. QM properties were transformed using multi-layer perceptrons (MLPs). Atom types were embedded using MLPs. For BCP-based graphs with atom types, the atom embeddings of both connected atoms were summed to achieve permutation invariance to the neighbor ordering. The transformed QM properties and atom embeddings were concatenated (in cases where both were present) to obtain initial node features  $\mathbf{v}_0^l$ . Node features  $\mathbf{v}_i^l$  in layer  $l$  were iteratively updated *via* the message-passing scheme

$$\mathbf{m}_{ij} = \phi_e(\mathbf{v}_i^l, \mathbf{v}_j^l, e_{ij}^l) \quad (1)$$



**Fig. 2** (Top left) bond-critical points (teal) in the binding pocket (PDB ID: 1CET<sup>64</sup>), connecting protein and ligand atoms *via* bond-critical paths (orange). Atoms are shown in black (carbon in ligand), gray (carbon in protein), white (hydrogen), blue (nitrogen), red (oxygen), and green (chlorine). The figure was generated using PyMol.<sup>65</sup> (Top right) Resulting bond-critical point graph with example annotations using quantum-mechanical properties (see Table 1) and node-to-node distances. For clarity, only edges between nodes within 3 Å are shown, while the discussed models use a 6 Å cutoff. (Bottom) model overview, processing node and edge features *via* geometry-aware E(3)-invariant message-passing to predict protein–ligand binding affinity.

$$\mathbf{m}_i = \bigoplus_{j \in \mathcal{N}_i} \mathbf{m}_{ij} \quad (2)$$

$$\mathbf{v}_i^{l+1} = \phi_h(\mathbf{v}_i^l, \mathbf{m}_i) \quad (3)$$

where the non-linear transformations on edge and node features,  $\phi_e$  and  $\phi_h$ , were described with SiLU-activated MLPs.<sup>92</sup> Edge messages  $\mathbf{m}_{ij}$  were obtained from the features of a pair of connected nodes and their (Fourier-encoded) distance  $e_{ij}$  (eqn (1)), achieving E(3)-invariance (to translation, rotation, and inversion of the input). Incoming edges to one node were mean-aggregated (eqn (2)), and the node's features were updated based on the aggregated message  $\mathbf{m}_i$  and the previous node features  $\mathbf{v}_i^l$  (eqn (3)). After five message-passing steps, the node features from each step were concatenated and transformed again using a SiLU-activated MLP  $\phi_f$  to obtain final node-level features  $\mathbf{V}_i$ :

$$\mathbf{V}_i = \phi_f(\text{concat}_{l=0}^{l=5}(\mathbf{v}_i^l)). \quad (4)$$

The final node-level features  $\mathbf{V}_i$  were mean-pooled (achieving permutation invariance) as preliminary experiments did not indicate a benefit from using sum or multi-headed attention pooling (not shown). The predicted binding affinity was obtained using another MLP. In line with previous work,<sup>86</sup> models were trained for 1000 epochs to minimize the mean squared error (MSE) loss using the Adam optimizer<sup>93</sup> with an initial learning rate of  $10^{-4}$ , a learning rate decay factor of 0.7, and a patience of 20 epochs. Combinations of the hyperparameters

batch size ( $\in \{16, 32, 64, 128\}$ ), kernel dimension ( $\in \{16, 32, 64, 128\}$ ), and MLP dimension ( $\in \{128, 256, 512\}$ ) were screened, and the configuration with the lowest root mean squared error (RMSE) on the validation set was used for testing. Models were built and trained using PyTorch<sup>94</sup> (version 1.9.1), and PyTorch Geometric<sup>95</sup> (version 2.0.3). Code for structure preparation, QM/QTAIM calculations, and model training is available at <https://github.com/ETHmodlab/bcpaff> (resp. <https://doi.org/10.5281/zenodo.8097403>).

## Benchmarks

One recently published benchmark model was selected for each dataset. For the PDBbind dataset, we used the “merged protein, ligand and interaction graph”,<sup>2</sup> a 3D-aware MPNN model that has shown strong performance on the PDBbind dataset and has undergone careful evaluation. This model uses pseudoatoms to represent protein–ligand interactions. For the PDE10A dataset, we used the 2D3D hybrid model which has previously performed well on this dataset.<sup>62</sup> The 2D3D hybrid constitutes an ensemble model combining predictions from the RF-PLP<sup>88,89</sup> (3D, protein–ligand structure-based) and AttentiveFP<sup>90</sup> (2D, ligand-based) approaches.

## Electron density-based models can predict binding affinity but show greater errors than benchmark models

Initial experiments suggested that using all the available QM properties (Table 1) as node features might be beneficial for model performance (see Section S4.4† for statistical analysis).

Unless otherwise specified, all discussed models used the entire range of QM properties given in Table 1. Model performance is shown in Fig. 3 (Section S4†). These models were trained using only QM properties without access to atomic identities, providing the desired “interaction-centric view”. These electron density-based models (both BCP-/NCP-based graphs) achieved root-mean-squared errors (RMSEs) of 1.4–1.8 log units on the PDBbind dataset, and 1.0–1.7 log units on the PDE10A dataset, not outperforming the benchmark models. This finding suggests that the proposed electron density-based representation did not improve on existing methods for the prediction of absolute protein–ligand binding affinities. Particularly for the non-random splits of the PDE10A dataset, these deep-learning models even failed to outperform the mean absolute deviation (MAD) baseline model, which predicts the arithmetic mean of the training and validation sets for all compounds in the test set. This apparent failure to outperform the MAD highlights the more challenging extrapolation task in these cases. The errors on the random split of the PDE10A dataset were generally lower than those on the PDBbind dataset ( $\sim 1$  log unit vs.  $\sim 1.5$ – $1.7$  log units), which is potentially due to consistently measured binding affinities and the presence of only a single protein target across the entire dataset. However, also the MAD was lower for the PDE10A dataset than for the PDBbind dataset, indicating a higher bar for non-trivial performance. While the PDBbind-trained BCP/NCP models achieved Pearson correlation coefficients  $r$  in the range of 0.45–0.76, the PDE10A-trained models achieved  $r > 0.5$  only for the random and the temporal 2013 (NCP model only) splits, while showing very poor or no correlation to experimental values for other dataset splits (Section S4†). For the BCP models trained on binding mode splits, the predicted values were contained within a small range of values around the mean (Fig. S6†), mirroring previous observations.<sup>2</sup> While the NCP-based models showed significantly lower test set errors than the BCP-based models for the PDBbind dataset (Section S4.3† for statistical analysis), this trend was less pronounced or even reversed for the individual

splits of the PDE10A dataset. Specifically, when considering the splits by binding mode, which necessitate more complex generalization across various interaction types, the incorporation of NCPs resulted in a rise in test set error. This suggests that this approach struggles with poor generalizability in such cases.

In order to more closely assess the predictive performance of the proposed representation, we additionally analyzed BCP/NCP-based models trained on the PDBbind dataset to make predictions for the CASF-2013 (ref. 96 and 97) and CASF-2016 (ref. 98) challenges, respectively. For model training, only structures that were not part of the test sets (CASF-2013 and CASF-2016) were incorporated, in adherence to an established splitting strategy.<sup>99</sup> (see Section S1.3† for details). With Pearson correlation coefficients of 0.552 and 0.591, and RMSE values of  $\sim 1.9$  and  $\sim 1.8$  for the CASF-2013 and CASF-2016 benchmark sets respectively, the BCP-based models did not exhibit any advantages in comparison to other benchmark models when considering these metrics (see details in Section S4.1.3†). A somewhat different scenario emerged for the models based on NCP descriptors, as they ranked center-field among the benchmark methods in terms of Pearson correlation coefficients and RMSE (see details in Section S4.1.3†). The finding that the NCP-based models showed higher correlations and lower RMSEs than the BCP-based models is in line with the findings for the test and core sets of PDBbind. This outcome reinforces the notion that, for the PDBbind dataset, details about the atomic characteristics of interacting atoms offer advantages in conjunction with interaction-centric information provided by the BCPs. However, no such pattern was identified for the PDE10A dataset splits, underscoring the possibility of a dataset bias. Overall, no substantial advantages were identified concerning scoring performance when compared to the best benchmark model that was examined (GraphscoreDTA<sup>99</sup>).

Furthermore, we explored the approach of training the PDBbind-based models exclusively on the refined set and then making predictions for the test and core sets. In both BCP- and

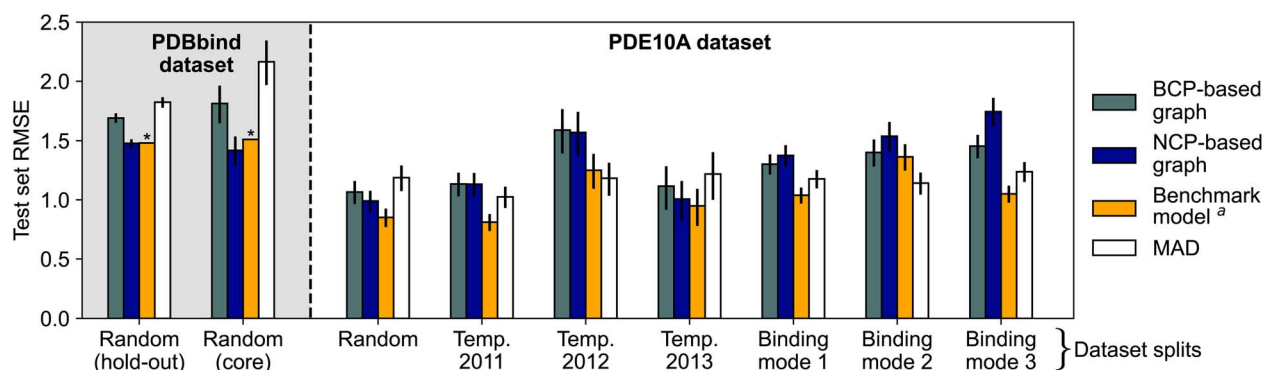


Fig. 3 Test set root-mean-squared error (RMSE) of electron density-based deep learning models for PDBbind (shaded background) and PDE10A (white background) with different data splitting strategies. Binding modes 1, 2, and 3 refer to the aminohetaryl-C1-amide, C1-hetaryl-alkyl-C2-hetaryl, and aryl-C1-amide-C2-hetaryl splits, respectively. Models trained using QM properties as initial node features. Error bars show 95% confidence intervals.<sup>87</sup> Model performance for PDBbind is reported for the 2019 hold-out test set and the 2016 core set (ignoring complexes for which processing failed, see Section S1†).<sup>a</sup> Benchmark models are the “merged protein, ligand and interaction graph<sup>2</sup>” model for PDBbind and the 2D3D hybrid model (ensemble of RF-PLP<sup>88,89</sup> and AttentiveFP<sup>90</sup>) for PDE10A.<sup>62</sup> \* No 95% confidence interval available for benchmark model.<sup>2</sup>

NCP-based models, there were moderate increases in prediction errors observed on the core set (RMSEs increasing from  $\sim 1.8$  to  $\sim 2.1$  log units, and from  $\sim 1.4$  to  $\sim 1.6$  log units, respectively [Section S4.1.2†]), suggesting that the larger amount of training data available in the PDBbind general set had a small positive impact on modeling performance. Similar trends were observed for the test set predictions (see details in Section S4.1.2†).

Based on these results, our initial hypothesis regarding the use of electron density-based graphs for binding affinity prediction *via* MPNNs was rejected. There are several potential explanations for this outcome, offering insights into why our deep models trained on electron density-based graphs did not provide more accurate binding affinity predictions than benchmark models.

Firstly, the (calculated) electron density was obtained by inputting atomic coordinates into the QM method of choice (*i.e.*, GFN2-xTB). The output, in terms of BCP-/NCP-based graphs annotated with QM properties, was thus mapped 1 : 1 to the initial atomic coordinates. Accordingly, much of the information contained in the BCP-centric view was already implicitly contained in the atomic coordinates, lacking only the information contributed by the data GFN2-xTB was fitted to. Owing to this lack of additional information, a BCP-centric view may not exhibit an advantage over more traditional atom position-based views, and the results of this study do not indicate that this alternative molecular representation renders the prediction task more feasible.

Secondly, although a generally acceptable agreement with more accurate QM calculations was confirmed (Section S2†), the lower accuracy of GFN2-xTB (compared to *e.g.*, DFT) may still contribute to the limited predictive accuracy of the deep-learning models. A similar effect was observed when using electron densities calculated with DFTB<sup>+100</sup> as an alternative semiempirical method (Section S4.2.1† for details). Given the substantial computational cost associated with running higher-accuracy QM calculations for thousands of protein–ligand interactions (Table S1†), one might consider turning towards recently developed ML-based methods that predict the electron density at a fraction of the cost of first-principles methods.<sup>101–109</sup> In addition to the epistemological problem associated with stacking multiple models on top of one another, making existing approaches compatible with commonly used BCP/NCP calculation software<sup>39,110,111</sup> is not straight-forward.

Thirdly, when considering BCP-based representations (without access to ligand coordinates), evaluating the inherent strain in the ligand within a bound conformation may pose greater challenges than with alternative methods. Ligand strain, which refers to the energy penalty associated with a ligand potentially having to deviate from a more stable solution-phase conformation to fit into the binding pocket, can considerably impact the measured binding affinity.<sup>112,113</sup> This hypothesis was tested by visualizing absolute per-structure model errors against ligand strain energies calculated for structures from the PDBbind dataset<sup>113</sup> (Fig. S6 and S7†). Based on the lack of correlation (slope =  $-0.028 \pm 0.010$  log units kcal<sup>-1</sup> mol and  $-0.008 \pm 0.022$  log units kcal<sup>-1</sup> mol for the test and core sets, BCP-based models) between ligand strain and model error, we

rejected the hypothesis that poorly capturing ligand strain is a key driver behind the unsatisfactory model performance.

Fourthly, “a more precise chemical description of the protein–ligand complex does not generally lead to a more accurate prediction of binding affinity”.<sup>114</sup> Potential challenges associated with the use of a more precise description lie in the difficulty of extracting information from such a representation, and in additional modeling assumptions required to produce the representation, such as protonation states and the choice of method for electron density calculation.<sup>114</sup>

Finally, an inherent limitation of many current deep learning-based binding affinity prediction methods lies in their focus on fixed atom positions (respectively fixed BCP/NCP positions, as in the present study). This view fails to capture the relevance of entropic contributions and the dynamic process of protein–ligand binding.<sup>115–118</sup> It has previously been argued that descriptors that remain relatively constant during the binding event (such as atomic identities) may alleviate the restrictions from this focus on the bound state to an extent.<sup>114</sup> Accordingly, the strong dependence of BCP-based graphs on the underlying molecular structure<sup>42</sup> may contribute to their inability to capture the dynamic binding process. This strong dependence of BCPs on the molecular structure may also amplify the effects of *e.g.*, incorrectly assigned atom positions or protonation states.<sup>52</sup> Other potentially contributing factors include the free energy of protein–ligand proton transfer or solvation effects.<sup>119</sup>

### Using atom identities instead of quantum-mechanical properties achieves similar performance

To assess the impact of using atom identities in the initial node features (as is typically done in structure-based deep learning methods for binding affinity prediction<sup>3,4</sup>), we compared the test set errors of models trained using only QM properties to those of models trained using only atom types or a combination of both. This approach moves away from a purely interaction-focused view, as it directly includes information about the chemical composition of the protein and ligand, respectively. Fig. 4 shows the effect of varying node features on the performance of deep learning models. The effect of modifying the node features was limited, resulting in mostly overlapping 95% confidence intervals for RMSE estimates. For the “binding mode 1” and “binding mode 3” splits of the PDE10A dataset, the models trained on BCP-based graphs with QM features (Fig. 4A) performed significantly better (Section S4.3† for statistical analysis) than models trained using only atom types or a combination of QM features and atom types, with RMSE values differing by up to 0.6 log units. While one might hypothesize that using atom types instead of QM properties could be detrimental for BCP-based graphs (essentially placing pseudo-atoms at the BCP locations but not using their respective QM properties), this effect was not observed in other dataset splits. Because the binding mode splits require the model to make more challenging extrapolations than a random split, the increasing error when going from QM properties to atom types to a combination of both could be related to the models

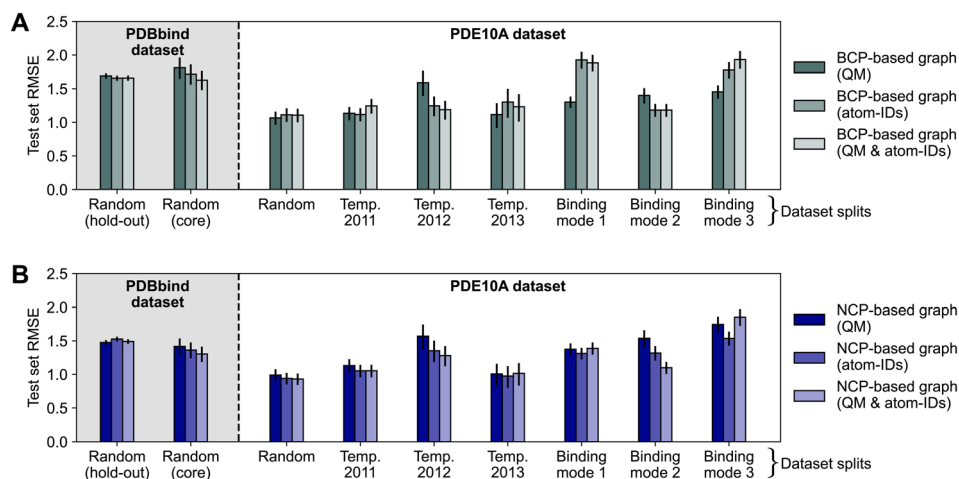


Fig. 4 Impact of varying node features on model performance of message-passing deep learning models trained on BCP-based graphs (A) and NCP-based graphs (B), respectively. Results reported as test set root-mean-squared error (RMSE) for the PDBbind (shaded background) and PDE10A (white background) datasets with different data splitting strategies. Error bars show 95% confidence intervals.<sup>87</sup> Model performance for PDBbind is reported for the 2019 hold-out test set and the 2016 core set.

overfitting to this information. For NCP-based graphs (Fig. 4B), for which an augmentation with atom types appears more natural than for BCP-based graphs (as NCPs correspond to positions of atomic nuclei), minor improvements were observed for individual data splitting strategies (binding modes 2 and 3), though no decisive impact was measured.

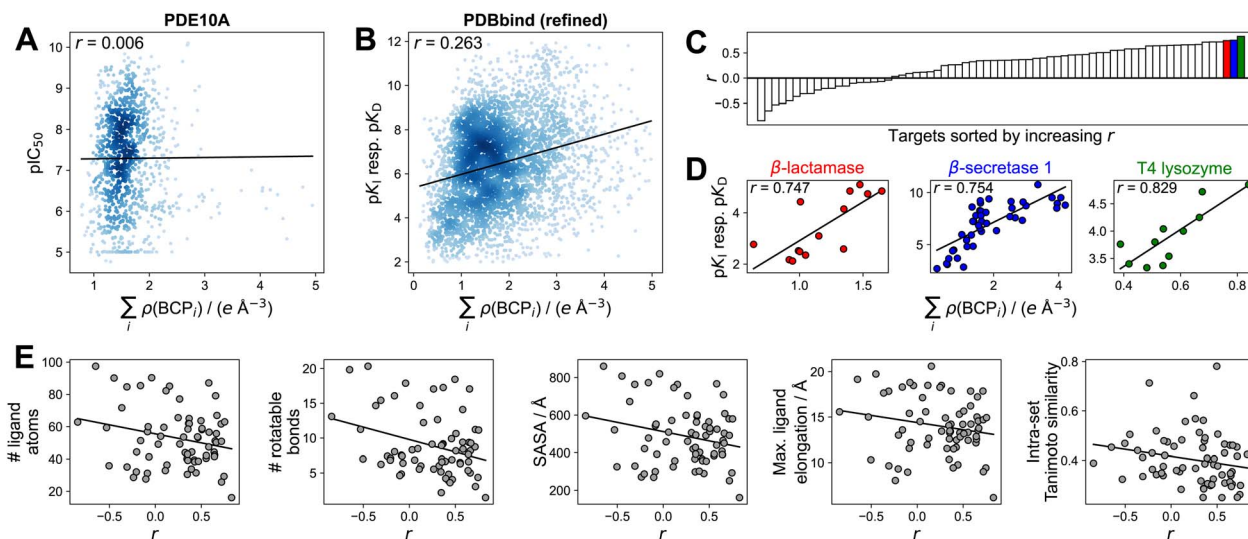
#### Intermolecular electron density correlates with binding affinity for some, but not all, protein targets

To gain additional insights into the utility of electron density for binding affinity prediction, we turned to a simpler method, foregoing ML techniques. We pursued a previously used approach<sup>43,48</sup> of assessing the correlation between the sum of electron density at intermolecular BCPs on the one hand, and the binding affinity on the other hand (Section S5†). Focusing this analysis on the PDBbind refined set and the PDE10A dataset was motivated by the goal of using high-quality structures and binding affinity measurements and ensuring comparability between individual binding affinity measurements (e.g., not comparing IC<sub>50</sub> values to pK<sub>i</sub> or pK<sub>d</sub> values as present in the PDBbind general set). While a previous study<sup>43</sup> used molecular dynamics-refined docking structures, we directly used crystal structures for the PDBbind dataset. We found no correlation ( $r = 0.006$ ) or extremely limited correlation ( $r = 0.263$ ) between the sum of electron density at intermolecular BCPs and binding affinity when looking at the PDE10A and PDBbind datasets, respectively (Fig. 5A and B), indicating that no such general trend exists when comparing these large sets of different protein–ligand interactions. This observation may be related to very different entropic contributions to protein–ligand binding across the datasets. When analyzing protein-specific correlations in the PDBbind dataset (Fig. 5C), large differences in correlations were observed for different proteins. In this more detailed view, individual proteins emerged for which good correlations ( $r > 0.7$ ) were observed (Fig. 5D). The three top-scoring protein targets featured sets of ligands with

both a wide range of affinity values (8 log units for  $\beta$ -secretase 1) and with narrow ranges (1.5 log units for T4 lysozyme, 3 log units for  $\beta$ -lactamase). These ranges were also reflected in the respective ranges of electron densities. An additional analysis using a very simple measure, such as the number of ligand atoms instead of the sum of electron density, is shown in Table S15.† For T4 lysozyme, the intermolecular electron density and number of intermolecular BCPs showed a substantially better correlation with binding affinity than the number of ligand atoms, though less pronounced trends were observed for the other protein targets. This result suggests that for these targets, the connection between intermolecular electron density and binding affinity is at least partly driven by ligand size.

Aiming to analyze to what degree the observed trends were affected by the accuracy of the semiempirical GFN2-xTB method, electron densities at the B3LYP-D3/6-31G\* level of theory<sup>79,80</sup> (analogous to a previous study<sup>43</sup>) were computed for three sets of ligands binding to different protein targets (Section S5†). These targets included a set of two high-correlation protein targets with medium-to-high affinity ranges ( $\beta$ -lactamase and  $\beta$ -secretase 1) and one low-correlation target (acetylcholine-binding protein) from the PDBbind refined set. This analysis revealed that the correlations identified using GFN2-xTB were also identified using B3LYP-D3/6-31G\*, showing only minor deviations ( $\Delta r < 0.05$ ).

To understand the impact of previously suggested<sup>43</sup> structural factors that may contribute to the differing correlations between sets of ligands, we assessed the average number of atoms, number of rotatable bonds, solvent-accessible surface area (SASA), spatial dimensions, and mean pairwise ligand similarity for each set of ligands binding to one protein target (Fig. 5E). This analysis indicated a rather weak trend of smaller and more rigid ligands showing better correlations. Groups of ligands with higher mean pairwise similarity between them showed a slight tendency to have poorer correlation between electron density and binding affinity. While such weak trends were observed, none of the



**Fig. 5** Correlation between the sum of electron density at intermolecular BCPs and binding affinity, expressed as Pearson's correlation coefficient  $r$ . (A) PDE10A dataset. (B) PDBbind (refined) dataset. Complexes with  $\sum_i \rho(\text{BCP}_i) > 5 \text{ e } \text{Å}^{-3}$  not shown in (A and B). (C) Protein targets (grouped by Uniprot-ID and protein name) sorted by increasing correlation. Only protein targets with at least 10 ligands shown. See Fig. S9† for complete labels. (D) Correlation analysis for the three highest-correlated protein targets ( $\beta$ -lactamase (Uniprot-ID P00811),  $\beta$ -secretase 1 (Uniprot-ID P56817), T4 lysozyme (Uniprot-ID P00720)). (E) Average structural properties of ligands per protein target from the PDBbind (refined) set, indicating limited predictive value for the correlation  $r$  between  $\sum_i \rho(\text{BCP}_i)$  and binding affinity. SASA = solvent-accessible surface area. Intra-set Tanimoto similarity measured by mean pairwise Tanimoto similarity of Morgan fingerprints (radius 2, 2048 bit).<sup>120</sup> Solid black lines in panels (A, B, D, and E) indicate least-squares fit.

investigated properties fully rationalized the vastly different correlations observed between the electron density at intermolecular BCPs and the measured binding affinity, suggesting that more detailed studies on the applicability of simple correlation-based QTAIM approaches are warranted.

## Conclusion

Herein, we explored the use of a QM-based, interaction-centric descriptor of protein–ligand complexes, focusing on the electron density as a fundamental physical description of molecular systems. To this end, a computational pipeline was introduced that enables the large-scale computation of QM properties and the extraction of BCPs for protein–ligand complexes. Training geometric deep learning models on BCP-based descriptors does not seem to alleviate the “frustration to predict binding affinities from protein–ligand structures with deep neural networks”.<sup>2</sup> The lack of competitive predictive performance of electron density-based deep neural networks may be driven by insufficient accuracy of the chosen QM method, the already implicitly-contained information in atom center-based representations, and the inability to account for entropic contributions in protein–ligand binding.

The findings regarding the correlation of the electron density at intermolecular BCPs with the binding affinity may help focus future QTAIM analyses on protein targets for which informative results can be obtained. Certain groups of ligands that bind to the same target show better correlations than others. However, the specific structural features of ligands and protein pockets that drive the utility of electron density as a predictor of binding affinity remain currently unclear.

To complement the incomplete picture emerging from the static view provided by a QTAIM analysis, a molecular dynamics (MD) approach might be suitable. Such an approach could use the ensemble of BCPs sampled from an MD trajectory of a protein–ligand complex, aiming to capture a physically more meaningful representation of the binding event. Specifics of the graph-construction process (“through-time” edges or other aggregation strategies) and whether such a strategy potentially provides benefits over a corresponding atom-centered view remain as questions for future work. An opportunity lies in the exploration of predicted (rather than calculated) electron densities using deep models trained on *ab initio* calculations.

## Author contributions

C. I.: conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, writing – original draft. K. A.: formal analysis, investigation, methodology, writing – review & editing. S. R.: formal analysis, investigation, methodology, supervision, writing – review & editing. G. S.: formal analysis, investigation, methodology, supervision, funding acquisition, project administration, writing – review & editing.

## Conflicts of interest

G. S. declares a potential financial conflict of interest as co-founder of <https://insili.com/> GmbH, Zurich, and in his role as a scientific consultant to the pharmaceutical industry.



## Acknowledgements

We thank Dr Andreas Tosstorff for providing raw model predictions for the 2D3D hybrid model from ref. 62. This work was financially supported by the Swiss National Science Foundation (grant no. 205321\_182176). C. I. acknowledges support from the Scholarship Fund of the Swiss Chemical Industry.

## Notes and references

- 1 T. Steinbrecher, in *Free Energy Calculations in Drug Lead Optimization*, John Wiley & Sons, Ltd, 2012, ch. 11, pp. 207–236.
- 2 M. Volkov, J.-A. Turk, N. Drizard, N. Martin, B. Hoffmann, Y. Gaston-Mathé and D. Rognan, *J. Med. Chem.*, 2022, **65**, 7946–7958.
- 3 S. Li, J. Zhou, T. Xu, L. Huang, F. Wang, H. Xiong, W. Huang, D. Dou and H. Xiong, *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2021, pp. 975–985.
- 4 S. Moon, W. Zhung, S. Yang, J. Lim and W. Y. Kim, *Chem. Sci.*, 2022, **13**, 3661–3673.
- 5 Z. Yang, W. Zhong, Q. Lv, T. Dong and C. Yu-Chian Chen, *J. Phys. Chem. Lett.*, 2023, **14**, 2020–2033.
- 6 J. Jiménez, S. Doerr, G. Martínez-Rosell, A. S. Rose and G. De Fabritiis, *Bioinformatics*, 2017, **33**, 3036–3042.
- 7 L. Möller, L. Guerci, C. Isert, K. Atz and G. Schneider, *Mol. Inf.*, 2022, **41**, 2200059.
- 8 G. Corso, H. Stärk, B. Jing, R. Barzilay and T. Jaakkola, *arXiv*, 2022, preprint, arXiv:2210.01776, DOI: [10.48550/arXiv.2210.01776](https://doi.org/10.48550/arXiv.2210.01776).
- 9 M. A. Ketata, C. Laue, R. Mammadov, H. Stärk, M. Wu, G. Corso, C. Marquet, R. Barzilay and T. S. Jaakkola, *arXiv*, 2023, preprint, arXiv:2304.03889, DOI: [10.48550/arXiv.2304.03889](https://doi.org/10.48550/arXiv.2304.03889).
- 10 K. Atz, L. C. Muñoz, C. Isert, M. Håkansson, D. Focht, D. F. Nippa, M. Hilleke, M. Iff, J. Ledergerber, C. C. Schiebroek, J. A. Hiss, D. Merk, P. Schneider, B. Kuhn, U. Grether and G. Schneider, *ChemRxiv*, 2023, preprint, DOI: [10.26434/chemrxiv-2023-cbq9k](https://doi.org/10.26434/chemrxiv-2023-cbq9k).
- 11 C. Shen, J. Ding, Z. Wang, D. Cao, X. Ding and T. Hou, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2020, **10**, e1429.
- 12 R. Singh, S. Sledzieski, B. Bryson, L. Cowen and B. Berger, *Proc. Natl. Acad. Sci. U. S. A.*, 2023, **120**, e2220778120.
- 13 J. Sieg, F. Flachsenberg and M. Rarey, *J. Chem. Inf. Model.*, 2019, **59**, 947–961.
- 14 L. Chen, A. Cruz, S. Ramsey, C. J. Dickson, J. S. Duca, V. Hornak, D. R. Koes and T. Kurtzman, *PLoS One*, 2019, **14**, e0220113.
- 15 J. Scantlebury, N. Brown, F. Von Delft and C. M. Deane, *J. Chem. Inf. Model.*, 2020, **60**, 3722–3730.
- 16 T. Janela and J. Bajorath, *Nat. Mach. Intell.*, 2022, **4**, 1246–1255.
- 17 Z. Liu, Y. Li, L. Han, J. Li, J. Liu, Z. Zhao, W. Nie, Y. Liu and R. Wang, *Bioinformatics*, 2015, **31**, 405–412.
- 18 R. Wang, X. Fang, Y. Lu and S. Wang, *J. Med. Chem.*, 2004, **47**, 2977–2980.
- 19 J. Yang, C. Shen and N. Huang, *Front. Pharmacol.*, 2020, **11**, 69.
- 20 G. C. Kanakala, R. Aggarwal, D. Nayar and U. D. Priyakumar, *ACS Omega*, 2023, **8**, 2389–2397.
- 21 C. F. Matta and A. A. Arabi, *Future Med. Chem.*, 2011, **3**, 969–994.
- 22 J. Ahrens, B. Geveci, C. Law, C. Hansen and C. Johnson, in *ParaView: An end-user tool for large-data visualization*, Citeseer, 2005, vol. 717, pp. 50038–50041.
- 23 V. R. Somnath, C. Bunne and A. Krause, *Adv. Neural Inf. Process.*, 2021, **34**, 25244–25255.
- 24 M. A. Moesser, D. Klein, F. Boyles, C. M. Deane, A. Baxter and G. M. Morris, *bioRxiv*, 2022, preprint, bioRxiv:2022.03.04.483012, DOI: [10.1101/2022.03.04.483012](https://doi.org/10.1101/2022.03.04.483012).
- 25 C. Isert, K. Atz and G. Schneider, *Curr. Opin. Struct. Biol.*, 2023, **79**, 102548.
- 26 K. Atz, F. Grisoni and G. Schneider, *Nat. Mach. Intell.*, 2021, **3**, 1023–1032.
- 27 L. Wang, J. Chambers and R. Abel, in *Protein–Ligand Binding Free Energy Calculations with FEP<sup>+</sup>*, ed. M. Bonomi and C. Camilloni, Springer New York, New York, NY, 2019, pp. 201–232.
- 28 T. Steinbrecher, R. Abel, A. Clark and R. Friesner, *J. Mol. Biol.*, 2017, **429**, 923–929.
- 29 K. Huang, S. Luo, Y. Cong, S. Zhong, J. Z. Zhang and L. Duan, *Nanoscale*, 2020, **12**, 10737–10750.
- 30 B. Kuhn and P. A. Kollman, *J. Med. Chem.*, 2000, **43**, 3786–3791.
- 31 B. Kuhn, P. Gerber, T. Schulz-Gasch and M. Stahl, *J. Med. Chem.*, 2005, **48**, 4040–4048.
- 32 R. F. Bader, *Acc. Chem. Res.*, 1985, **18**, 9–15.
- 33 R. F. Bader, *Atoms in Molecules: A Quantum Theory*, Clarendon Press, 1990.
- 34 C. F. Matta and R. J. Boyd, in *An Introduction to the Quantum Theory of Atoms in Molecules*, ed. C. F. Matta and R. J. Boyd, Wiley, 2007, pp. 1–34.
- 35 A. D. Becke and K. E. Edgecombe, *J. Chem. Phys.*, 1990, **92**, 5397–5403.
- 36 T. Lu and F.-W. Chen, *Acta Phys.-Chim. Sin.*, 2011, **27**, 2786–2792.
- 37 H. Schmider and A. Becke, *J. Mol. Struct.: THEOCHEM*, 2000, **527**, 51–61.
- 38 E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, *J. Am. Chem. Soc.*, 2010, **132**, 6498–6506.
- 39 T. Lu and F. Chen, *J. Comput. Chem.*, 2012, **33**, 580–592.
- 40 B. Niepötter, R. Herbst-Irmer, D. Kratzert, P. P. Samuel, K. C. Mondal, H. W. Roesky, P. Jerabek, G. Frenking and D. Stalke, *Angew. Chem., Int. Ed.*, 2014, **53**, 2766–2770.
- 41 N. Sukumar and C. M. Breneman, in *QTAIM in Drug Discovery and Protein Modeling*, ed. C. F. Matta and R. J. Boyd, Wiley, 2007, pp. 473–498.
- 42 R. D. Tosso, M. Vettorazzi, S. A. Andujar, L. J. Gutierrez, J. C. Garro, E. Angelina, R. Rodríguez, F. D. Suvire, M. Noguera, J. Cobo and R. D. Enriz, *J. Mol. Struct.*, 2017, **1134**, 464–474.

- 43 S. Rojas, O. Parravicini, M. Vettorazzi, R. Tosso, A. Garro, L. Gutiérrez, S. Andújar and R. Enriz, *Eur. J. Med. Chem.*, 2020, **208**, 112792.
- 44 R. D. Tosso, S. A. Andujar, L. Gutierrez, E. Angelina, R. Rodriguez, M. Nogueras, H. Baldoni, F. D. Suvire, J. Cobo and R. D. Enriz, *J. Chem. Inf. Model.*, 2013, **53**, 2018–2032.
- 45 M. Vettorazzi, E. Angelina, S. Lima, T. Gonec, J. Otevrel, P. Marvanova, T. Padrtova, P. Mokry, P. Bobal, L. M. Acosta, A. Palma, J. Cobo, J. Bobalova, J. Csollei, I. Malik, S. Alvarez, S. Spiegel, J. Jampilek and R. D. Enriz, *Eur. J. Med. Chem.*, 2017, **139**, 461–481.
- 46 C. L. Firme, N. K. Monteiro and S. R. Silva, *Comput. Theor. Chem.*, 2017, **1111**, 40–49.
- 47 A. M. Luchi, R. N. Villafaña, J. L. Gomez Chavez, M. L. Bogado, E. L. Angelina and N. M. Peruchena, *ACS Omega*, 2019, **4**, 19582–19594.
- 48 L. J. Gutiérrez, O. Parravicini, E. Sánchez, R. Rodríguez, J. Cobo and R. D. Enriz, *J. Biomol. Struct.*, 2019, **37**, 229–246.
- 49 O. A. von Lilienfeld, K.-R. Müller and A. Tkatchenko, *Nat. Rev. Chem.*, 2020, **4**, 347–358.
- 50 O. T. Unke, S. Chmiela, H. E. Saucedo, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K.-R. Müller, *Chem. Rev.*, 2021, **121**, 10142–10186.
- 51 D. Lemm, G. F. von Rudorff and O. A. von Lilienfeld, *Nat. Commun.*, 2021, **12**, 4468.
- 52 B. K. Rai, V. Sresht, Q. Yang, R. Unwalla, M. Tu, A. M. Mathiowetz and G. A. Bakken, *J. Chem. Inf. Model.*, 2022, **62**, 785–800.
- 53 A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth and B. Kozinsky, *Nat. Commun.*, 2023, **14**, 579.
- 54 C. Isert, J. C. Kromann, N. Stiefl, G. Schneider and R. A. Lewis, *ACS Omega*, 2023, **8**, 2046–2056.
- 55 D. F. Nippa, K. Atz, R. Hohler, A. T. Müller, A. Marx, C. Bartelmus, G. Wuitschik, I. Marzuoli, V. Jost, J. Wolfard, M. Binder, A. F. Stepan, D. B. Konrad, U. Grether, R. E. Martin and G. Schneider, *Nat. Chem.*, 2023, DOI: [10.1038/s41557-023-01360-5](https://doi.org/10.1038/s41557-023-01360-5).
- 56 T. Stuyver and C. W. Coley, *J. Chem. Phys.*, 2022, **156**, 084104.
- 57 C. Isert, K. Atz, J. Jiménez-Luna and G. Schneider, *Sci. Data*, 2022, **9**, 273.
- 58 R. M. Neeser, C. Isert, T. Stuyver, G. Schneider and C. W. Coley, *Chem. Data Collect.*, 2023, **46**, 101040.
- 59 D. F. Nippa, K. Atz, A. T. Müller, J. Wolfard, C. Isert, M. Binder, O. Scheidegger, D. B. Konrad, U. Grether, R. E. Martin, *et al.*, *Commun. Chem.*, 2023, **6**, 256.
- 60 D. V. Eldred, C. L. Weikel, P. C. Jurs and K. L. Kaiser, *Chem. Res. Toxicol.*, 1999, **12**, 670–678.
- 61 C. M. Breneman, C. M. Sundling, N. Sukumar, L. Shen, W. P. Katt and M. J. Embrechts, *J. Comput.-Aided Mol. Des.*, 2003, **17**, 231–240.
- 62 A. Tosstorff, M. G. Rudolph, J. C. Cole, M. Reutlinger, C. Kramer, H. Schaffhauser, A. Nilly, A. Flohr and B. Kuhn, *J. Comput.-Aided Mol. Des.*, 2022, **36**, 753–765.
- 63 Y.-q. Chen, Y.-j. Sheng, Y.-q. Ma and H.-m. Ding, *Phys. Chem. Chem. Phys.*, 2022, **24**, 14339–14347.
- 64 J. A. Read, K. W. Wilkinson, R. Tranter, R. B. Sessions and R. L. Brady, *J. Biol. Chem.*, 1999, **274**, 10213–10218.
- 65 Schrödinger, LLC, *PyMOL. The PyMOL Molecular Graphics System*, Version 2.5.2, Schrödinger, LLC.
- 66 E. L. Angelina, S. A. Andujar, R. D. Tosso, R. D. Enriz and N. M. Peruchena, *J. Phys. Org. Chem.*, 2014, **27**, 128–134.
- 67 S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2017, **13**, 1989–2009.
- 68 C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.
- 69 S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 2847–2862.
- 70 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.
- 71 G. Sigalov, A. Fenley and A. Onufriev, *J. Chem. Phys.*, 2006, **124**, 124902.
- 72 S. Ehlert, M. Stahn, S. Spicher and S. Grimme, *J. Chem. Theory Comput.*, 2021, **17**, 4250–4261.
- 73 S. Schmitz, J. Seibert, K. Ostermeir, A. Hansen, A. H. Göller and S. Grimme, *J. Phys. Chem. B*, 2020, **124**, 3636–3646.
- 74 L. Gundelach, T. Fox, C. S. Tautermann and C.-K. Skylaris, *Phys. Chem. Chem. Phys.*, 2021, **23**, 9381–9393.
- 75 C. F. Matta, *J. Comput. Chem.*, 2010, **31**, 1297–1311.
- 76 C. F. Matta, *J. Comput. Chem.*, 2014, **35**, 1165–1198.
- 77 J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.
- 78 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
- 79 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 80 W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.
- 81 D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. Di Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O'Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, I. Schaefer, F. Henry, A. Y. Sokolov, K. Patkowski, I. DePrince, A. Eugene, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford and C. D. Sherrill, *J. Chem. Phys.*, 2020, **152**, 184108.
- 82 G. Landrum, *RDKit: Open-source cheminformatics and machine learning*, <https://www.rdkit.org/docs/index.html>, accessed 19.06.23.
- 83 J. Lim, S. Ryu, K. Park, Y. J. Choe, J. Ham and W. Y. Kim, *J. Chem. Inf. Model.*, 2019, **59**, 3981–3988.
- 84 W. Torng and R. B. Altman, *J. Chem. Inf. Model.*, 2019, **59**, 4131–4149.
- 85 V. G. Satorras, E. Hoogeboom and M. Welling, *ICML*, 2021, pp. 9323–9332.
- 86 K. Atz, C. Isert, M. N. Böcker, J. Jiménez-Luna and G. Schneider, *Phys. Chem. Chem. Phys.*, 2022, **24**, 10775–10783.

- 87 J. H. Jensen, *PeerJ Prepr.*, 2017, **5**, e2693v1.
- 88 A. Tosstorff, J. C. Cole, R. Taylor, S. F. Harris and B. Kuhn, *J. Chem. Inf. Model.*, 2020, **60**, 6595–6611.
- 89 A. Tosstorff, J. C. Cole, R. Bartelt and B. Kuhn, *ChemMedChem*, 2021, **16**, 3428–3438.
- 90 Z. Xiong, D. Wang, X. Liu, F. Zhong, X. Wan, X. Li, Z. Li, X. Luo, K. Chen, H. Jiang and M. Zheng, *J. Med. Chem.*, 2019, **63**, 8749–8760.
- 91 *EGNN-PyTorch*, <https://github.com/lucidrains/egnn-pytorch>, accessed 19.06.23.
- 92 S. Elfving, E. Uchibe and K. Doya, *Neural Networks*, 2018, **107**, 3–11.
- 93 D. P. Kingma and J. Ba, *arXiv*, 2014, preprint, arXiv:1412.6980, DOI: **10.48550/arXiv.1412.6980**.
- 94 A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala, *Adv. Neural Inf. Process*, 2019, **32**, 8026–8037.
- 95 M. Fey and J. E. Lenssen, *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.
- 96 Y. Li, Z. Liu, J. Li, L. Han, J. Liu, Z. Zhao and R. Wang, *J. Chem. Inf. Model.*, 2014, **54**, 1700–1716.
- 97 Y. Li, L. Han, Z. Liu and R. Wang, *J. Chem. Inf. Model.*, 2014, **54**, 1717–1736.
- 98 M. Su, Q. Yang, Y. Du, G. Feng, Z. Liu, Y. Li and R. Wang, *J. Chem. Inf. Model.*, 2018, **59**, 895–913.
- 99 K. Wang, R. Zhou, J. Tang and M. Li, *Bioinform.*, 2023, **39**, btad340.
- 100 B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshayé, T. Dumitrică, A. Dominguez, S. Ehlert, M. Elstner, T. van der Heide, J. Hermann, S. Irle, J. J. Kranz, C. Köhler, T. Kowalczyk, T. Kubař, I. S. Lee, V. Lutsker, R. J. Maurer, S. K. Min, I. Mitchell, C. Negre, T. A. Niehaus, A. M. N. Niklasson, A. J. Page, A. Pecchia, G. Penazzi, M. P. Persson, J. Řezáč, C. G. Sánchez, M. Sternberg, M. Stöhr, F. Stuckenberg, A. Tkatchenko, V. W.-z. Yu and T. Frauenheim, *J. Chem. Phys.*, 2020, **152**, 124101.
- 101 A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen and R. Ramprasad, *npj Comput. Mater.*, 2019, **5**, 22.
- 102 A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf and M. Ceriotti, *ACS Cent. Sci.*, 2018, **5**, 57–64.
- 103 B. Cuevas-Zuiviría and L. F. Pacios, *J. Chem. Inf. Model.*, 2020, **60**, 3831–3842.
- 104 Z. Qiao, A. S. Christensen, M. Welborn, F. R. Manby, A. Anandkumar and T. F. Miller III, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2205221119.
- 105 O. Unke, M. Bogojeski, M. Gastegger, M. Geiger, T. Smidt and K.-R. Müller, *Adv. Neural Inf. Process*, 2021, **34**, 14434–14447.
- 106 P. B. Jørgensen and A. Bhowmik, *npj Comput. Mater.*, 2022, **8**, 183.
- 107 B. Cuevas-Zuiviría and L. F. Pacios, *J. Chem. Inf. Model.*, 2021, **61**, 2658–2666.
- 108 J. A. Rackers, L. Tecot, M. Geiger and T. E. Smidt, *arXiv*, 2022, preprint, arXiv:2201.03726, DOI: **10.48550/arXiv.2201.03726**.
- 109 A. J. Lee, J. A. Rackers and W. P. Bricker, *Biophys. J.*, 2022, **121**, 3883–3895.
- 110 T. A. Keith, *AIMAll (Version 19.10.12)*, TK Gristmill Software, Overland Park KS, USA, 2019, <https://aim.tkgristmill.com/>, accessed 29.06.23.
- 111 J. Cheeseman and T. A. Keith and R. F. W. Bader, *AIMPAC Program Package*, McMaster University, Hamilton, Ontario, 1992, <https://www.chemistry.mcmaster.ca/aimpac/imagemap/imagemap.htm>, accessed 29.06.23.
- 112 S. Gu, M. S. Smith, Y. Yang, J. J. Irwin and B. K. Shoichet, *J. Chem. Inf. Model.*, 2021, **61**, 4331–4341.
- 113 A. N. Jain, A. C. Brueckner, A. E. Cleves, M. Reibarkh and E. C. Sherer, *J. Med. Chem.*, 2023, **66**, 1955–1971.
- 114 P. J. Ballester, A. Schreyer and T. L. Blundell, *J. Chem. Inf. Model.*, 2014, **54**, 944–955.
- 115 G. Schneider, *Nat. Rev. Drug Discovery*, 2010, **9**, 273–276.
- 116 J. D. Chodera and D. L. Mobley, *Annu. Rev. Biophys.*, 2013, **42**, 121–142.
- 117 S. Wu, W. Zhang, W. Li, W. Huang, Q. Kong, Z. Chen, W. Wei and S. Yan, *Biochemistry*, 2022, **61**, 433–445.
- 118 T. Siebenmorgen, F. Menezes, S. Benassou, E. Merdivan, S. Kesselheim, M. Piraud, F. J. Theis, M. Sattler and G. M. Popowicz, *bioRxiv*, 2023, preprint, DOI: **10.1101/2023.05.24.542082**.
- 119 A. Pecina, J. Fanfrlík, M. Lepšík and J. Řezáč, *ChemRxiv*, 2023, preprint, DOI: **10.26434/chemrxiv-2023-zh03k**.
- 120 H. L. Morgan, *J. Chem. Doc.*, 1965, **5**, 107–113.