

# Supporting Information for "Exploring protein-ligand binding affinity prediction with electron density-based geometric deep learning"

Clemens Isert<sup>1</sup>, Kenneth Atz<sup>1</sup>, Sereina Riniker<sup>1</sup>, and Gisbert Schneider\*<sup>1</sup>

<sup>1</sup>ETH Zurich, Department of Chemistry and Applied Biosciences, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland.

## Contents

<b>S1 Structure preparation</b>	<b>S2</b>
S1.1 PDBbind (core/test set) . . . . .	S2
S1.2 PDE10A . . . . .	S2
S1.3 CASF-2013 and CASF-2016 . . . . .	S2
<b>S2 Assessment of GFN2-xTB-calculated electron density</b>	<b>S3</b>
<b>S3 Node features</b>	<b>S6</b>
<b>S4 Model performance</b>	<b>S7</b>
S4.1 PDBbind . . . . .	S7
S4.1.1 PDBbind - core and test set . . . . .	S7
S4.1.2 Training on the refined set . . . . .	S7
S4.1.3 CASF-2013 and CASF-2016 . . . . .	S8
S4.1.4 Impact of ligand strain on model performance . . . . .	S8
S4.2 PDE10A . . . . .	S9
S4.2.1 Models trained on DFTB+-calculated electron density . . . . .	S10
S4.3 Comparison of model performances . . . . .	S10
S4.4 Feature ablation . . . . .	S16
<b>S5 Correlation of sum of electron density at BCPs with binding affinity</b>	<b>S16</b>

---

\*Correspondence: [gisbert@ethz.ch](mailto:gisbert@ethz.ch)

## S1 Structure preparation

During the extraction of BCPs with Multiwfn<sup>[1]</sup>, the interacting atoms could not be identified in a small number of instances, in which case the corresponding atom type was set to "unknown".

### S1.1 PDBbind (core/test set)

Mol2 files for the PDBbind (v. 2019) dataset<sup>[2,3]</sup> were downloaded as prepared by Volkov *et al.*<sup>[4]</sup> (incl. protonation). After fixing issues related to formal charge and bond order annotation in *e.g.*, phosphoric acid and carboxylic acid groups, protein residues were removed if no atom of the residue was within 6 Å of any ligand atom. The removal was performed by cutting between the  $\alpha$ - and  $\beta$ -carbons (to cut between side chains and the backbone) resp. between the  $\alpha$ -carbon and the adjacent carboxy carbon (to cut between adjacent residues in the backbone) and cut positions were padded with hydrogens.

The following complexes from PDBbind failed during structure preparation, QM calculation, and BCP/NCP extraction: 1a7x, 1ao8, 1nlt, 1qpf, 1w9u, 2fou, 2fov, 2foy, 2whp, 2ycf, 2yir, 2ylc, 3d45, 3i4y, 3kr4, 3l0k, 3t8w, 3v7t, 3w54, 4acu, 4djv, 4ew2, 4ew3, 4r6t, 4s3e, 4wef, 4wvs, 4zla, 5adq, 5adt, 5fpp, 5mrp, 6bt0, 6nao.

### S1.2 PDE10A

All complexes from PDE10A could successfully be processed.

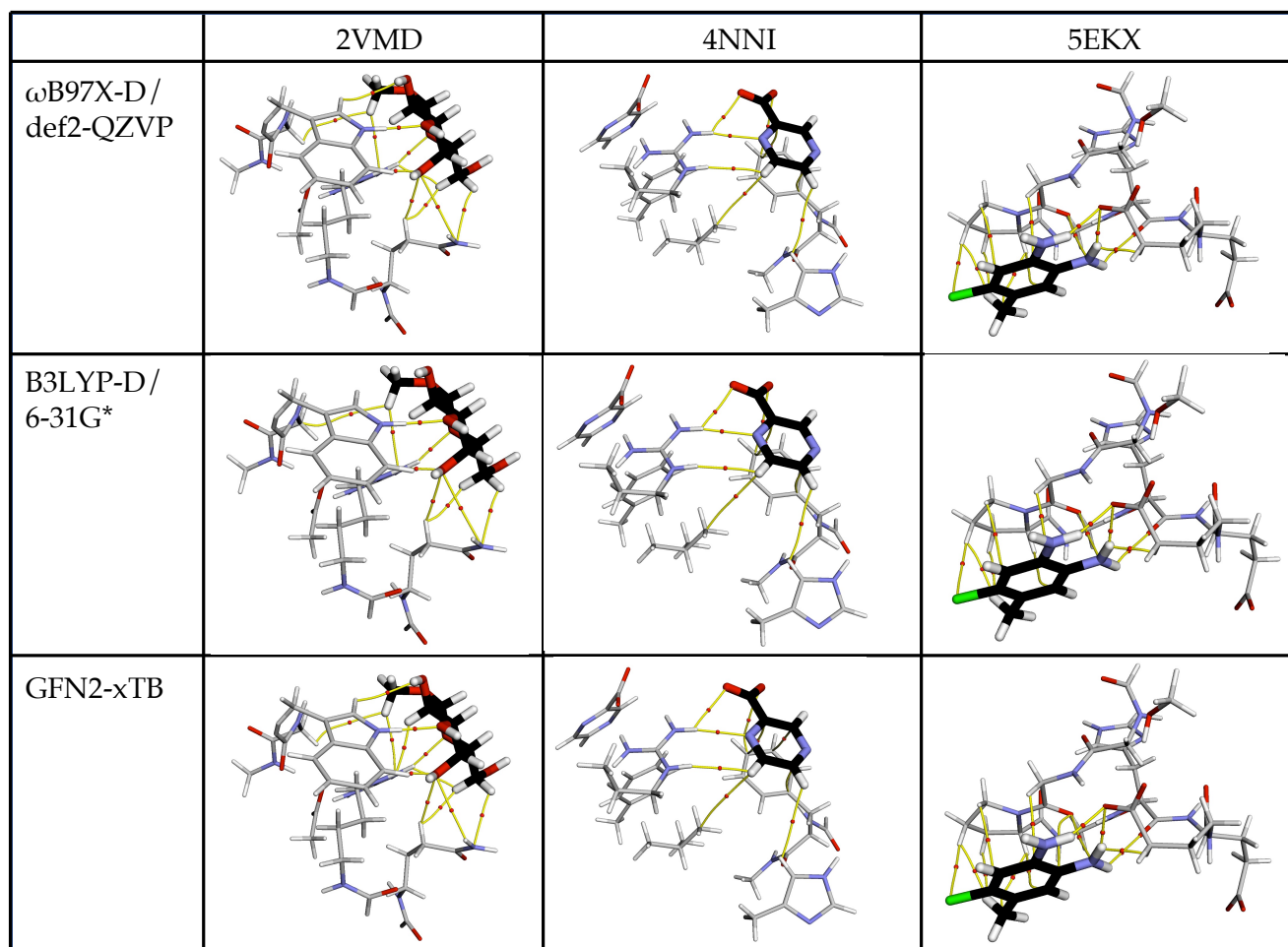
### S1.3 CASF-2013 and CASF-2016

For the models trained for the CASF-2013 and CASF-2016 challenges, structure files were used directly from the PDBbind dataset<sup>[2,3]</sup> (v. 2019). Water molecules were protonated at physiological pH (7.4) using OpenBabel<sup>[5]</sup> (version 3.1.1), and otherwise processed as described in Section S1.1.

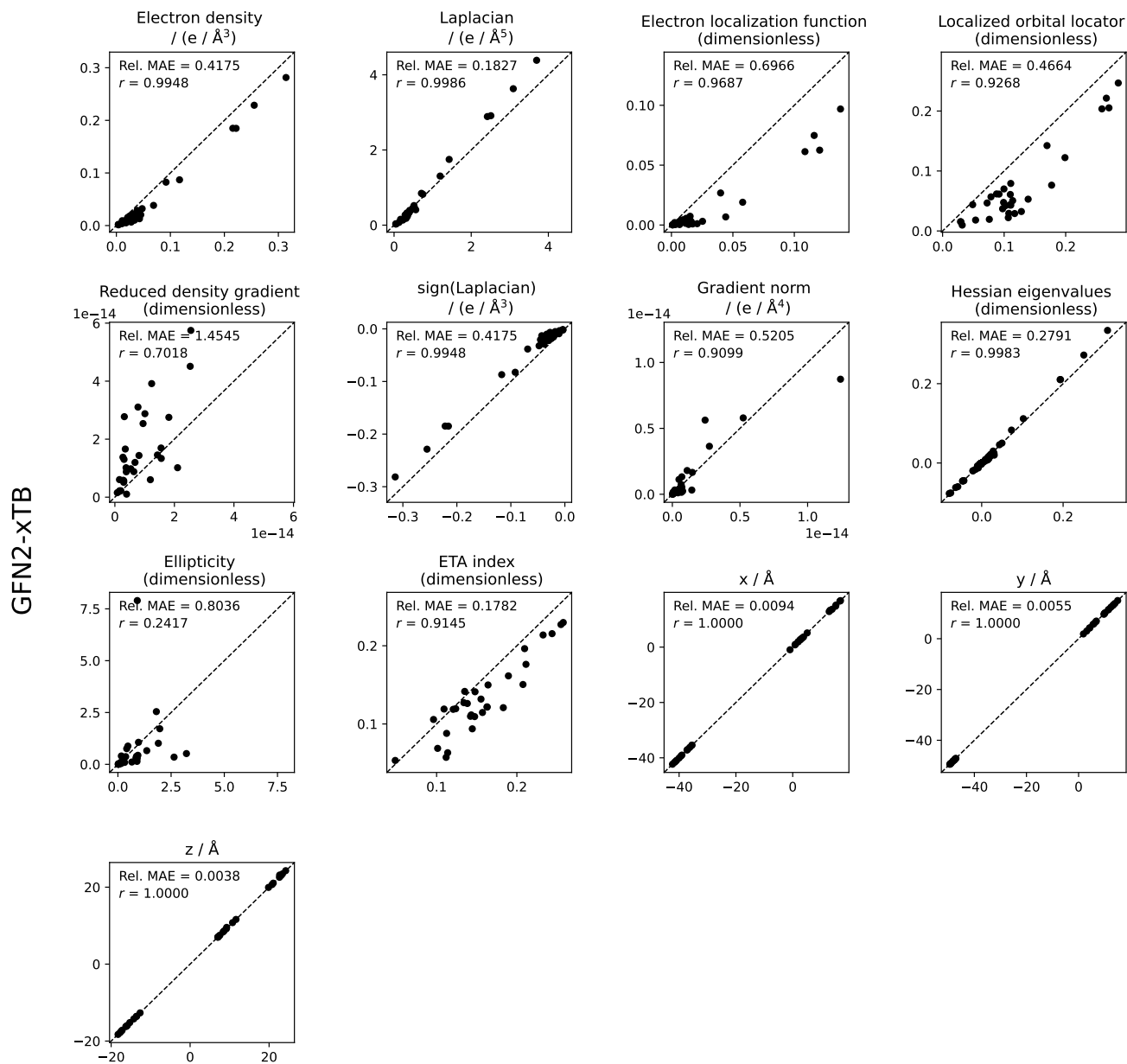
The following complexes from PDBbind failed during structure preparation, QM calculation, and BCP/NCP extraction (max. 24 h wall time): 1a7x, 1a8t, 1ao8, 1b4d, 1bnn, 1dl7, 1h07, 1i9o, 1qq9, 1uyi, 1w8l, 1x9d, 1zsr, 2bo4, 2eep, 2fou, 2fov, 2foy, 2hrm, 2hs1, 2iyf, 2j83, 2oc4, 2pll, 2qu3, 2vxa, 2x9f, 2xv1, 2xvd, 2ycf, 2yir, 2z3z, 3arg, 3bho, 3cwk, 3d1v, 3eyd, 3hf6, 3hfb, 3jyj, 3kdt, 3kdu, 3le8, 3tmk, 3tu1, 3u8h, 3ump, 3wv3, 4acu, 4anu, 4bps, 4c1e, 4cp7, 4cps, 4djv, 4ew2, 4ew3, 4hdb, 4he9, 4hww, 4hxq, 4hze, 4i06, 4ie2, 4ie3, 4iu0, 4ixu, 4j8t, 4jpa, 4kai, 4kb7, 4kbi, 4kne, 4l6q, 4lv1, 4m3b, 4m3d, 4m3e, 4m3f, 4m3p, 4mdq, 4mm4, 4nal, 4ngm, 4ngr, 4no1, 4ob0, 4ob1, 4ob2, 4oc3, 4oc4, 4p44, 4q3r, 4qrh, 4tkj, 4u0s, 4u0u, 4u0x, 4uyd, 4wri, 4wyp, 4z7f, 4zum, 4zun, 4zuo, 4zup, 4zuq, 4zur, 5a5v, 5aan, 5ad1, 5ans, 5ar0, 5b5p, 5ezh, 5f94, 5fun, 5hip, 5js3, 5kdf, 5l87, 5lsh, 5lwn, 5myl, 5myn, 5myr, 5mys, 5ng9, 5ny1, 5oah, 5tcy, 5tq7, 5uwl, 5uwm, 5v1r, 6bt0, 6c0t, 6ela, 6eog, 6epa, 6epy, 6mu1, 6nao.

## S2 Assessment of GFN2-xTB-calculated electron density

In order to assess whether GFN2-xTB is suitable for calculating electron densities in the analyzed protein-ligand complexes, three structures from the PDBbind dataset (PDB-IDs 2VMD<sup>[6]</sup>, 4NNI<sup>[7]</sup>, 5EKX<sup>[8]</sup>) were computed using both GFN2-xTB<sup>[9-12]</sup> and DFT ( $\omega$ B97X-D/def2-QZVP<sup>[13,14]</sup>). Choosing these three relatively small structures was motivated by the need to constrain computational cost for the DFT calculations. No substantial effects of molecular size on benchmark results are expected given the local character of electron density ("nearsightedness of electronic matter"<sup>[15]</sup>). Psi4<sup>[16]</sup> (version 1.7) was used for DFT calculations and Multiwfn<sup>[1]</sup> (version 3.8(dev), 03/2023) was used to determine BCPs and to extract QM properties (see Table 1 in the main paper) at the intermolecular BCPs. BCPs from each method were matched by finding the spatially closest GFN2-xTB-calculated BCP to each DFT-calculated BCP. Figure S1 shows the resulting intermolecular BCPs found using both QM methods, and Figures S2 and S3 show correlations between the respective QM methods for the different properties of interest. While generally good agreement is observed, individual properties (particularly reduced density gradient and ellipticity) show poorer correlation. Table S1 show that speedups of 5-6 orders of magnitude can be achieved by using GFN2-xTB instead of  $\omega$ B97X-D/def2-QZVP, respectively speedups of 3 orders of magnitude when using GFN2-xTB instead of B3LYP-D3/6-31G\*.

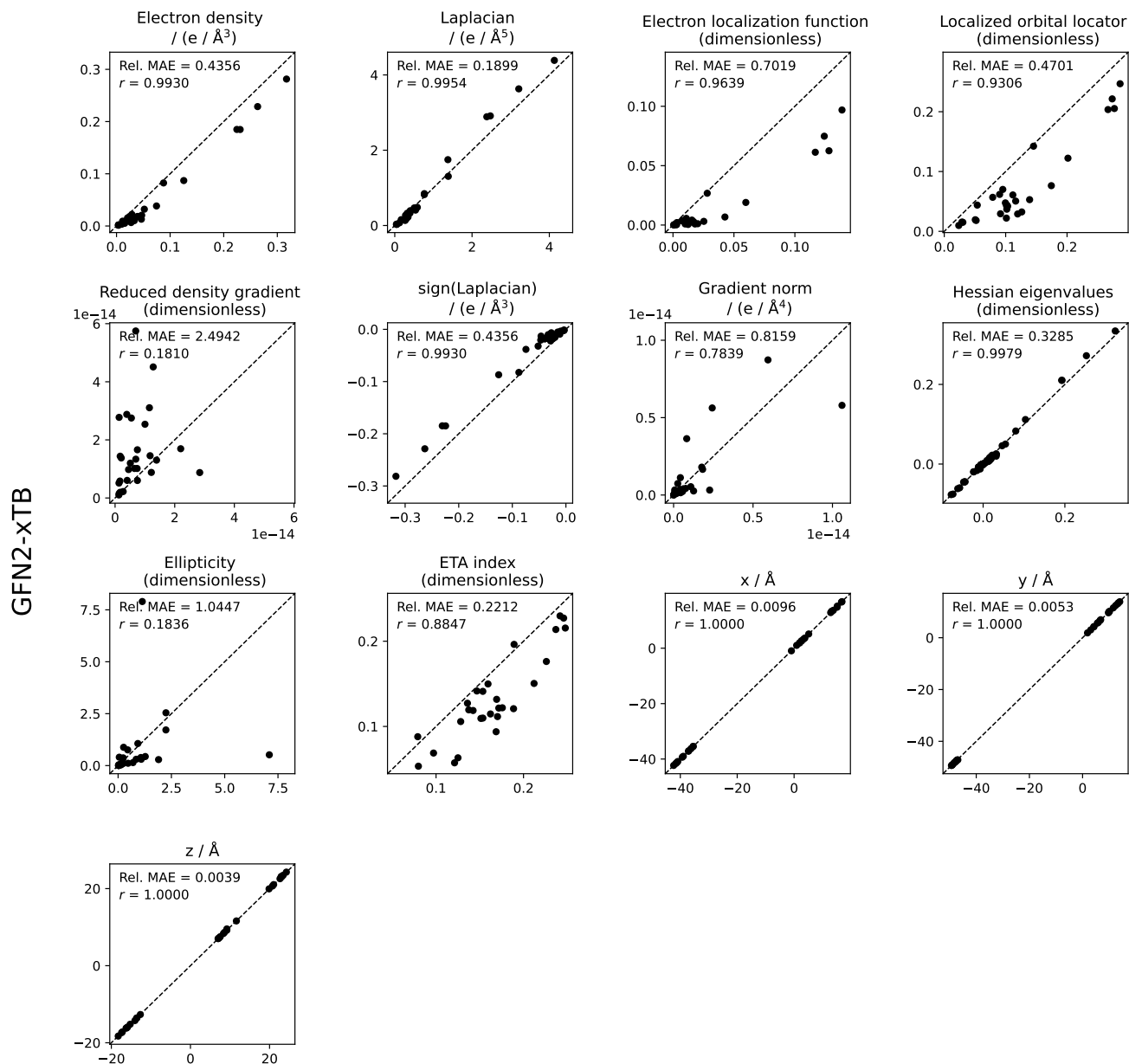


**Figure S1.** Three example structures with bond-critical points calculated using DFT ( $\omega$ B97X-D/def2-QZVP & B3LYP-D3/6-31G\*) and semiempirical GFN2-xTB. Bond-critical points show as red spheres, bond paths as yellow lines. Atoms are shown in black (carbon, ligand), gray (carbon, protein), white (hydrogen), blue (nitrogen), red (oxygen), green (chlorine).



$\omega$ B97X-D/def2-QZVP

**Figure S2.** Correlation between QM properties calculated using semi-empirical GFN2-xTB and DFT ( $\omega$ B97X-D/def2-QZVP) for three structures (PDB IDs 2VMD, 4NNI, 5EKX).



### B3LYP-D3/6-31G\*

**Figure S3.** Correlation between QM properties calculated using semi-empirical GFN2-xTB and DFT (B3LYP-D3/6-31G\*) for three structures (PDB IDs 2VMD, 4NNI, 5EKX).

**Table S1.** Compute times for QM calculations of three complexes from the PDBbind dataset.  $\omega$ B97X-D/def2-QZVP and B3LYP-D3/6-31G\* computed using 8 CPU cores, GFN2-xTB computed using 1 CPU core.

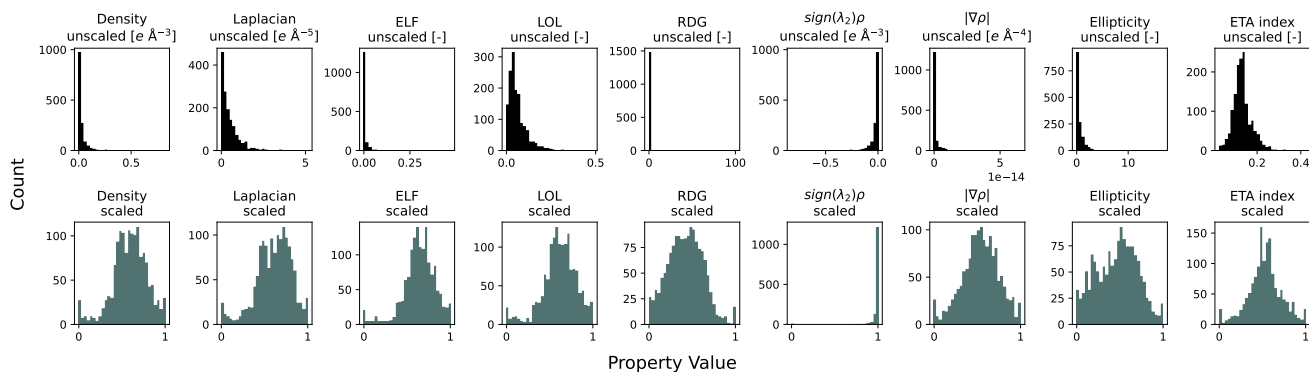
PDB ID	$\omega$ B97X-D/def2-QZVP CPU time (s)	B3LYP-D3/6-31G* CPU time (s)	GFN2-xTB CPU time (s)	Speedup GFN2-xTB vs $\omega$ B97X-D/def2-QZVP (-)	Speedup GFN2-xTB vs B3LYP-D3/6-31G* (-)
2VMD	136,627	4,036	1.333	102,496	3,028
4NNI	136,192	4,623	1.462	93,155	3,162
5EKX	143,417	4,411	1.028	139,511	4,291

### S3 Node features

The unprocessed node features (for BCPs and NCPs) contained some extreme values, and typical scaling methods (incl. simple percentile-based methods) did not appear to yield satisfactory results. Accordingly, a custom scaling function for the QM properties was designed in order to achieve a more informative representation.

$$X_{\text{scaler}} = \text{MinMaxScaler}(\log_{10}[\text{clip}(X, \min = X_{1\%}, \max = X_{99\%}) + (1 + \epsilon) \cdot \text{abs}(X_{1\%})]) \quad (1)$$

where  $\epsilon = 0.001$  is a small constant to push the minima just over 0 (to permit  $\log_{10}$  computation) and an additional small offset (0.00001) is added for ellipticity values. The clipping operation is used to set very low or very high values to the 1<sup>st</sup> and 99<sup>th</sup> percentile, respectively. The required scaling factors are computed using the respective training sets, and are applied to all features in the training, validation, and test sets. Figure S4 shows the distribution of QM properties before and after scaling.



**Figure S4.** QM properties before and after scaling BCP-based graphs of the PDE10A training set using a random split.

Computed electron density-based graphs for the datasets used in this work are available at <https://doi.org/10.3929/ethz-b-000653723>.

## S4 Model performance

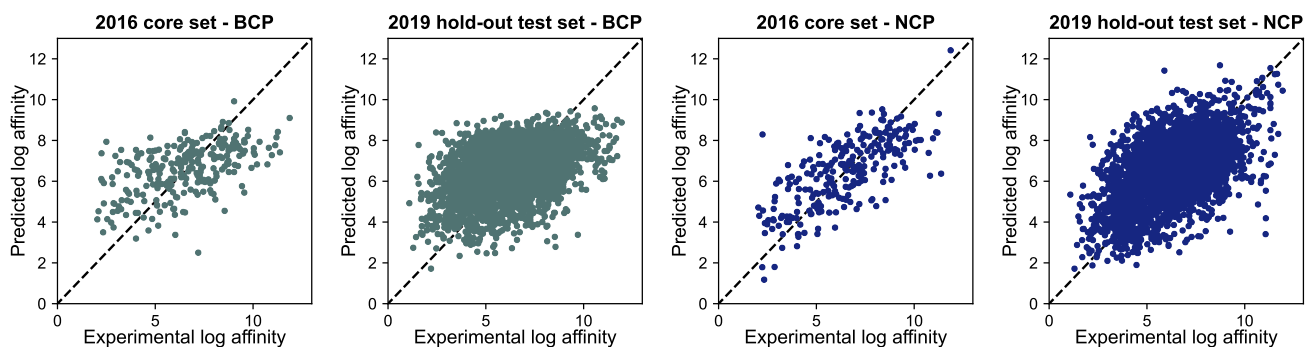
### S4.1 PDBbind

#### S4.1.1 PDBbind - core and test set

Table S2 and Figure S5 show model performance on the core and test sets after models were trained on the PDBbind training set as provided in reference 4.

**Table S2.** Predictive performance of deep-learning models for the PDBbind dataset measured using Pearson’s correlation coefficient  $R$  and root-mean-squared error RMSE. Values in sub- and superscript denote lower and upper bounds of 95%-confidence intervals, respectively.<sup>[17]</sup> BCP/NCP models trained using QM properties (ignoring complexes for which processing failed, see Section S1.1). PLI: "merged protein, ligand and interaction graph" model<sup>[4]</sup>. \* No 95%-confidence intervals available for benchmark model.

Model	PDBbind 2016 core set		PDBbind 2019 test set	
	$r$ ( $\uparrow$ )	RMSE ( $\downarrow$ )	$r$ ( $\uparrow$ )	RMSE ( $\downarrow$ )
BCP-based graph	0.552 <sub>0.461</sub> <sup>0.631</sup>	1.812 <sub>1.647</sub> <sup>1.963</sup>	0.446 <sub>0.419</sub> <sup>0.473</sup>	1.692 <sub>1.651</sub> <sup>1.732</sup>
NCP-based graph	0.764 <sub>0.709</sub> <sup>0.811</sup>	1.418 <sub>1.289</sub> <sup>1.536</sup>	0.607 <sub>0.585</sub> <sup>0.628</sup>	1.478 <sub>1.442</sub> <sup>1.512</sup>
PLI <sup>[4]</sup> *	0.813	1.511	0.652	1.481
MAD	-	2.165 <sub>1.968</sub> <sup>2.345</sup>	-	1.823 <sub>1.779</sub> <sup>1.866</sup>



**Figure S5.** Predicted vs calculated plots for the PDBbind dataset. Models trained with QM properties on the random split for the PDBbind dataset and tested on the 2016 core set and the 2019 hold-out test set. The dashed  $y = x$  line indicates a perfect prediction.

#### S4.1.2 Training on the refined set

To investigate the impact of training set choice, we trained BCP/NCP-based models on the refined set of the PDBbind dataset. To this end, the structures in the training set provided by Volkov *et al.*<sup>[4]</sup> (see Section S1 for details on structure preparation) were filtered to include only those that are part of the refined set in PDBbind<sup>[2,3]</sup> (version 2019). The sets of optimal hyperparameters identified for the PDBbind dataset trained on the general set (Section S4.1.1) were used and models were retrained on the refined set. Table S3 shows model performance on the core and test sets.

**Table S3.** Predictive performance of deep-learning models trained on the refined set of the PDBbind dataset measured using Pearson’s correlation coefficient  $R$  and root-mean-squared error RMSE. Values in sub- and superscript denote lower and upper bounds of 95%-confidence intervals, respectively.<sup>[17]</sup> BCP/NCP models trained using QM properties (ignoring complexes for which processing failed, see Section S1.1).

Model	PDBbind 2016 core set		PDBbind 2019 test set	
	$r$ ( $\uparrow$ )	RMSE ( $\downarrow$ )	$r$ ( $\uparrow$ )	RMSE ( $\downarrow$ )
BCP-based graph	0.435 <sub>0.327</sub> <sup>0.531</sup>	2.080 <sub>1.883</sub> <sup>2.259</sup>	0.406 <sub>0.352</sub> <sup>0.456</sup>	2.020 <sub>1.929</sub> <sup>2.107</sup>
NCP-based graph	0.666 <sub>0.589</sub> <sup>0.730</sup>	1.620 <sub>1.467</sub> <sup>1.760</sup>	0.594 <sub>0.552</sub> <sup>0.633</sup>	1.500 <sub>1.432</sub> <sup>1.564</sup>
MAD	-	2.148 <sub>1.946</sub> <sup>2.333</sup>	-	1.821 <sub>1.739</sub> <sup>1.899</sup>

### S4.1.3 CASF-2013 and CASF-2016

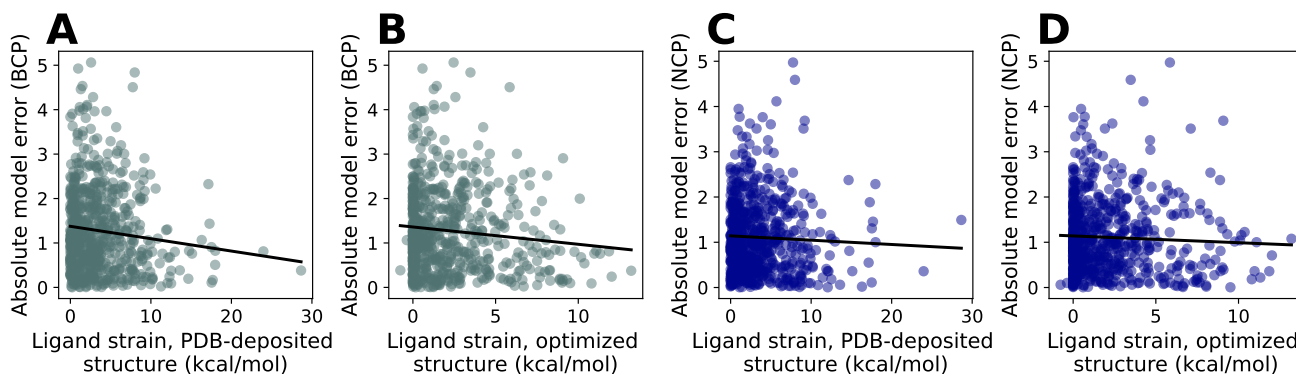
We report results for the CASF-2013<sup>[18,19]</sup> and CASF-2016<sup>[20]</sup> challenges. We use the dataset splits suggested in reference 21 (compound similarity-based clustering). The sets of optimal hyperparameters identified for the PDBbind dataset (Section S4.1.1) were used for the BCP/NCP-based models and models were retrained using the training set from reference 21. Table S4 shows model performance on the two test sets.

**Table S4.** Predictive performance of deep-learning models trained on the training set<sup>[21]</sup> of the PDBbind dataset measured using Pearson’s correlation coefficient  $R$  and root-mean-squared error RMSE. Values in sub- and superscript denote lower and upper bounds of 95%-confidence intervals, respectively.<sup>[17]</sup> BCP/NCP models trained using QM properties (ignoring complexes for which processing failed, see Section S1.3). Performance metrics for other models from reference 21

Model	CASF-2013		CASF-2016	
	$r$ ( $\uparrow$ )	RMSE ( $\downarrow$ )	$r$ ( $\uparrow$ )	RMSE ( $\downarrow$ )
BCP-based graph	0.552	1.948	0.591	1.809
NCP-based graph	0.728	1.590	0.776	1.435
MAD	-	2.265	-	2.181
DeepDTA <sup>[22]</sup>	0.710	1.608	0.745	1.458
DeepDTAF <sup>[23]</sup>	0.616	2.010	0.765	1.426
Pafnucy <sup>[24]</sup>	0.662	1.716	0.754	1.527
BAPA <sup>[25]</sup>	0.620	2.014	0.615	1.936
Fusion model <sup>[26]</sup>	0.662	1.720	0.746	1.513
GraphscoreDTA <sup>[21]</sup>	0.757	1.486	0.831	1.249

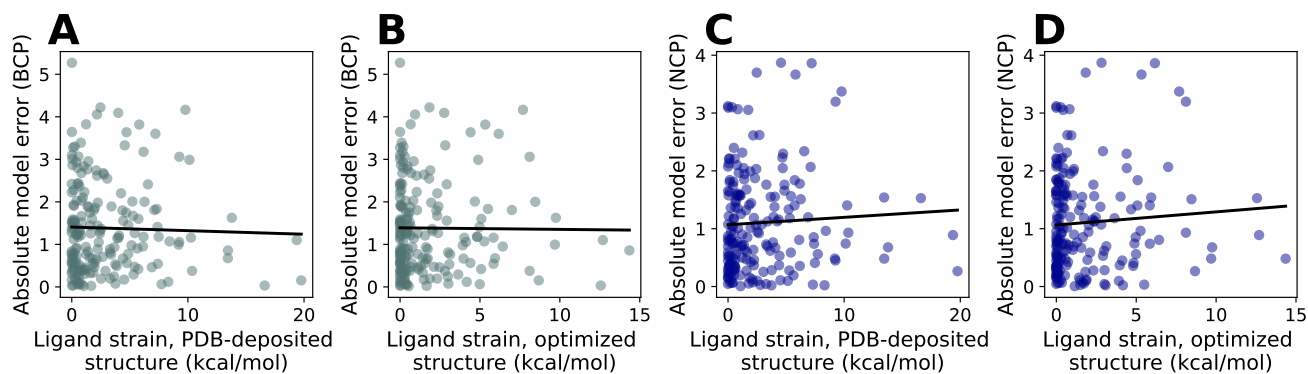
### S4.1.4 Impact of ligand strain on model performance

To investigate the impact of ligand strain on model performance, we used calculated ligand strains for structures in the PDBbind dataset as provided in reference 27. In this work, both ligand strain energies for ligands as deposited in the PDB database were calculated, as well as ligand strains for strain-reduced ligand poses that still fit the experimentally measured electron density well.<sup>[27]</sup> The calculated ligand strains were matched against the PDBbind core and test sets, respectively. Figures S6 and S7 show how the absolute model errors of BCP/NCP-based models for the test and core sets (models trained using QM properties) depend on the calculated ligand strain.



**Figure S6.** Absolute model errors on the PDBbind test set vs ligand strain. **A & B:** BCP-based models trained using QM properties. **C & D:** NCP-based models trained using QM properties. Solid black lines indicate least-squares fit.





**Figure S7.** Absolute model errors on the PDBbind core set vs ligand strain. **A & B:** BCP-based models trained using QM properties. **C & D:** NCP-based models trained using QM properties. Solid black lines indicate least-squares fit.

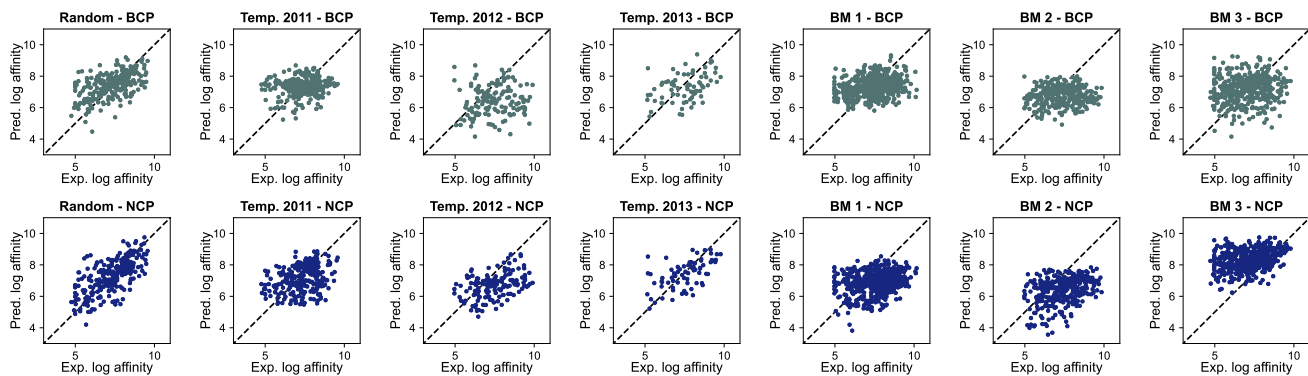
## S4.2 PDE10A

**Table S5.** Predictive performance of deep-learning models for the PDE10A dataset measured using Pearson’s correlation coefficient  $r$  and root-mean-squared error RMSE. Values in sub- and superscript denote lower and upper bounds of 95%-confidence intervals, respectively.<sup>[17]</sup> BCP/NCP models trained using QM properties. <sup>a</sup> 2D3D hybrid model<sup>[28]</sup> (ensemble of RF-PLP<sup>[29,30]</sup> and AttentiveFP<sup>[31]</sup>).

Model	Random		Temp. 2011		Temp. 2012		Temp. 2013	
	$r$ ( $\uparrow$ )	RMSE ( $\downarrow$ )	$R$ ( $\uparrow$ )	RMSE ( $\downarrow$ )	$r$ ( $\uparrow$ )	RMSE ( $\downarrow$ )	$r$ ( $\uparrow$ )	RMSE ( $\downarrow$ )
BCP-based graph	0.525 <sup>0.611</sup> <sub>0.426</sub>	1.066 <sup>1.159</sup> <sub>0.965</sub>	0.246 <sup>0.358</sup> <sub>0.127</sub>	1.134 <sup>1.230</sup> <sub>1.030</sub>	-0.009 <sup>0.154</sup> <sub>-0.171</sub>	1.591 <sup>1.767</sup> <sub>1.392</sub>	0.480 <sup>0.635</sup> <sub>0.288</sub>	1.116 <sup>1.285</sup> <sub>0.915</sub>
NCP-based graph	0.601 <sup>0.676</sup> <sub>0.513</sub>	0.990 <sup>1.077</sup> <sub>0.896</sub>	0.300 <sup>0.408</sup> <sub>0.184</sub>	1.132 <sup>1.228</sup> <sub>1.028</sub>	0.331 <sup>0.469</sup> <sub>0.178</sub>	1.570 <sup>1.744</sup> <sub>1.374</sub>	0.559 <sup>0.695</sup> <sub>0.384</sub>	1.007 <sup>1.160</sup> <sub>0.826</sub>
2D3D hybrid <sup>a</sup>	0.712 <sup>0.770</sup> <sub>0.643</sub>	0.852 <sup>0.926</sup> <sub>0.770</sub>	0.608 <sup>0.680</sup> <sub>0.524</sub>	0.812 <sup>0.880</sup> <sub>0.737</sub>	0.559 <sup>0.662</sup> <sub>0.436</sub>	1.250 <sup>1.389</sup> <sub>1.094</sub>	0.637 <sup>0.753</sup> <sub>0.483</sub>	0.949 <sup>1.093</sup> <sub>0.779</sub>
MAD	-	1.187 <sup>1.291</sup> <sub>1.074</sub>	-	1.025 <sup>1.112</sup> <sub>0.931</sub>	-	1.183 <sup>1.314</sup> <sub>1.035</sub>	-	1.219 <sup>1.404</sup> <sub>1.000</sub>

**Table S6.** Predictive performance of deep-learning models for the PDE10A dataset measured using Pearson’s correlation coefficient  $r$  and root-mean-squared error RMSE. Values in sub- and superscript denote lower and upper bounds of 95%-confidence intervals, respectively.<sup>[17]</sup> BCP/NCP models trained using QM properties. <sup>a</sup> 2D3D hybrid model<sup>[28]</sup> (ensemble of RF-PLP<sup>[29,30]</sup> and AttentiveFP<sup>[31]</sup>). 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.

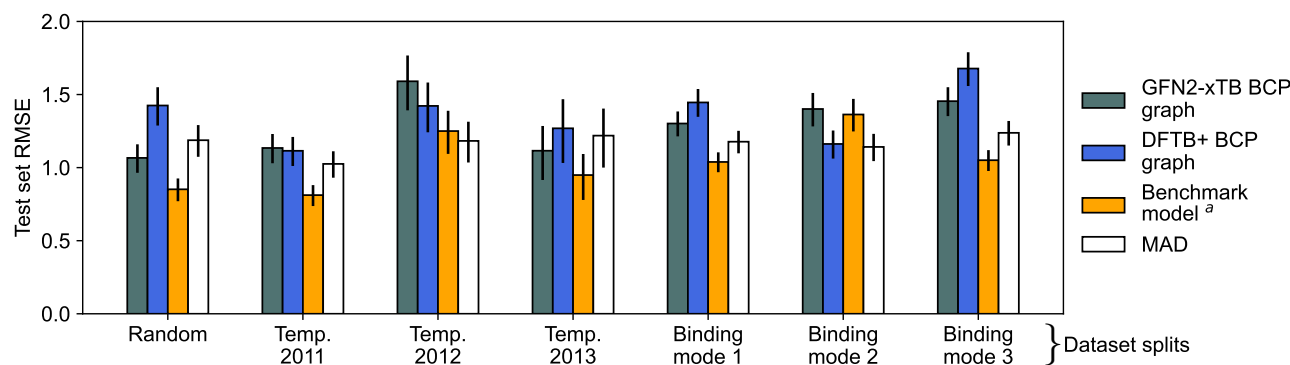
Model	Binding mode 1		Binding mode 2		Binding mode 3	
	$r$ ( $\uparrow$ )	RMSE ( $\downarrow$ )	$r$ ( $\uparrow$ )	RMSE ( $\downarrow$ )	$r$ ( $\uparrow$ )	RMSE ( $\downarrow$ )
BCP-based graph	0.207 <sup>0.293</sup> <sub>0.118</sub>	1.302 <sup>1.384</sup> <sub>1.214</sub>	0.064 <sup>0.177</sup> <sub>-0.050</sub>	1.401 <sup>1.511</sup> <sub>1.282</sub>	0.139 <sup>0.232</sup> <sub>0.045</sub>	1.455 <sup>1.550</sup> <sub>1.352</sub>
NCP-based graph	0.299 <sup>0.381</sup> <sub>0.213</sub>	1.376 <sup>1.463</sup> <sub>1.283</sub>	0.355 <sup>0.451</sup> <sub>0.251</sub>	1.538 <sup>1.659</sup> <sub>1.407</sub>	0.328 <sup>0.411</sup> <sub>0.240</sub>	1.746 <sup>1.860</sup> <sub>1.623</sub>
2D3D hybrid <sup>a</sup>	0.453 <sup>0.523</sup> <sub>0.377</sub>	1.039 <sup>1.104</sup> <sub>0.969</sub>	0.333 <sup>0.430</sup> <sub>0.227</sub>	1.363 <sup>1.470</sup> <sub>1.248</sub>	0.494 <sup>0.563</sup> <sub>0.419</sub>	1.051 <sup>1.120</sup> <sub>0.977</sub>
MAD	-	1.177 <sup>1.252</sup> <sub>1.098</sub>	-	1.142 <sup>1.231</sup> <sub>1.044</sub>	-	1.238 <sup>1.319</sup> <sub>1.151</sub>



**Figure S8.** Predicted vs calculated plots for the PDE10A dataset. Models trained with QM properties and tested on the respective test sets (see subpanel titles). The dashed  $y = x$  line indicates a perfect prediction.

#### S4.2.1 Models trained on DFTB+-calculated electron density

To assess if using DFTB+<sup>[32]</sup> (version 22.2) as an alternative semi-empirical method to GFN2-xTB for electron density calculations, models were trained on DFTB+-calculated electron densities. Electron densities were calculated using the DFTB3 method<sup>[33–36]</sup> (ThirdOrderFull = Yes) and GBSA implicit water. BCPs and QM properties were extracted using critic2<sup>[37,38]</sup>. Calculations could successfully be completed for 1139 structures (given 24 h wall time, 1 core, 8 GB RAM) and were used for model training/testing according to the predefined splits. The DFTB+-based models use the following QM properties: Electron density, Laplacian of the electron density, and gradient norm of the electron density. For each dataset split, hyperparameters were used as optimized for the QM-property experiments. Figure S9 shows model performance on the PDE10A dataset.



**Figure S9.** Test set root-mean-squared error (RMSE) of electron density-based deep learning models for PDE10A with different data splitting strategies, using BCP-based models trained using electron densities either calculated with GFN2-xTB or DFTB+. 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>[17]</sup> <sup>a</sup> Benchmark model is the 2D3D hybrid model (ensemble of RF-PLP<sup>[29,30]</sup> and AttentiveFP<sup>[31]</sup>) for PDE10A<sup>[28]</sup>.

### S4.3 Comparison of model performances

Model performance, as measured by RMSE, is compared using Equation 2 to assess statistical significance (95%) using composite errors<sup>[17,39]</sup> computed using the statsig Python package<sup>[40]</sup>:

$$error_A - error_B > \sqrt{L_A^2 + U_B^2 - 2r_{AB}L_AU_B} \quad (2)$$

where  $L_A$  and  $U_B$  are the upper and lower bounds of the confidence intervals for methods  $A$  and  $B$  respectively. With method  $A$  being the more accurate one, if the inequality in Equation 2 holds, method  $A$  is more accurate than  $B$  on the 95% significance level.<sup>[17]</sup> Tables S8 - S15 show pairwise model comparisons.

**Table S7.** Pairwise comparison of different model setups for the PDBbind test set (random split) and test whether errors are significantly different on the 95% confidence level.<sup>[17,39,40]</sup> Difference in RMSE values, composite errors, and whether or not the null hypothesis that the errors are the same holds are shown.

Method <i>A</i>	Method <i>B</i>	RMSE <sub><i>A</i></sub>	RMSE <sub><i>B</i></sub>	RMSE <sub><i>A</i></sub> - RMSE <sub><i>B</i></sub>	Comp. error	Same error?
BCP QM	BCP atom-IDs	1.692	1.658	0.034	0.035	True
BCP atom-IDs & QM	BCP atom-IDs	1.659	1.658	0.001	0.029	True
BCP atom-IDs	NCP atom-IDs	1.658	1.528	0.131	0.035	False
BCP atom-IDs	NCP QM	1.658	1.478	0.181	0.034	False
BCP atom-IDs	NCP atom-IDs & QM	1.658	1.491	0.167	0.033	False
BCP QM	BCP atom-IDs & QM	1.692	1.659	0.033	0.036	True
BCP QM	NCP atom-IDs	1.692	1.528	0.164	0.037	False
BCP QM	NCP QM	1.692	1.478	0.214	0.036	False
BCP QM	NCP atom-IDs & QM	1.692	1.491	0.201	0.037	False
BCP atom-IDs & QM	NCP atom-IDs	1.659	1.528	0.131	0.035	False
BCP atom-IDs & QM	NCP QM	1.659	1.478	0.182	0.034	False
BCP atom-IDs & QM	NCP atom-IDs & QM	1.659	1.491	0.168	0.034	False
NCP atom-IDs	NCP QM	1.528	1.478	0.050	0.021	False
NCP atom-IDs	NCP atom-IDs & QM	1.528	1.491	0.037	0.018	False
NCP atom-IDs & QM	NCP QM	1.491	1.478	0.013	0.020	True

**Table S8.** Pairwise comparison of different model setups for the PDBbind core set (random split) and test whether errors are significantly different on the 95% confidence level.<sup>[17,39,40]</sup> Difference in RMSE values, composite errors, and whether or not the null hypothesis that the errors are the same holds are shown.

Method <i>A</i>	Method <i>B</i>	RMSE <sub><i>A</i></sub>	RMSE <sub><i>B</i></sub>	RMSE <sub><i>A</i></sub> - RMSE <sub><i>B</i></sub>	Comp. error	Same error?
BCP QM	BCP atom-IDs	1.812	1.718	0.095	0.122	True
BCP atom-IDs	BCP atom-IDs & QM	1.718	1.628	0.090	0.102	True
BCP atom-IDs	NCP atom-IDs	1.718	1.363	0.354	0.118	False
BCP atom-IDs	NCP QM	1.718	1.418	0.299	0.120	False
BCP atom-IDs	NCP atom-IDs & QM	1.718	1.305	0.413	0.113	False
BCP QM	BCP atom-IDs & QM	1.812	1.628	0.185	0.132	False
BCP QM	NCP atom-IDs	1.812	1.363	0.449	0.132	False
BCP QM	NCP QM	1.812	1.418	0.394	0.135	False
BCP QM	NCP atom-IDs & QM	1.812	1.305	0.507	0.126	False
BCP atom-IDs & QM	NCP atom-IDs	1.628	1.363	0.264	0.108	False
BCP atom-IDs & QM	NCP QM	1.628	1.418	0.210	0.110	False
BCP atom-IDs & QM	NCP atom-IDs & QM	1.628	1.305	0.323	0.103	False
NCP QM	NCP atom-IDs	1.418	1.363	0.055	0.065	True
NCP atom-IDs	NCP atom-IDs & QM	1.363	1.305	0.058	0.055	False
NCP QM	NCP atom-IDs & QM	1.418	1.305	0.113	0.062	False

**Table S9.** Pairwise comparison of different model setups for the PDE10A (random split) and test whether errors are significantly different on the 95% confidence level.<sup>[17,39,40]</sup> Difference in RMSE values, composite errors, and whether or not the null hypothesis that the errors are the same holds are shown.

Method $A$	Method $B$	RMSE $_A$	RMSE $_B$	RMSE $_A$ - RMSE $_B$	Comp. error	Same error?
BCP atom-IDs	BCP QM	1.112	1.066	0.046	0.097	True
BCP atom-IDs	BCP atom-IDs & QM	1.112	1.107	0.006	0.072	True
BCP atom-IDs	NCP atom-IDs	1.112	0.941	0.172	0.081	False
BCP atom-IDs	NCP QM	1.112	0.990	0.122	0.080	False
BCP atom-IDs	NCP atom-IDs & QM	1.112	0.932	0.181	0.081	False
BCP atom-IDs & QM	BCP QM	1.107	1.066	0.040	0.097	True
BCP QM	NCP atom-IDs	1.066	0.941	0.126	0.082	False
BCP QM	NCP QM	1.066	0.990	0.076	0.077	True
BCP QM	NCP atom-IDs & QM	1.066	0.932	0.135	0.072	False
BCP atom-IDs & QM	NCP atom-IDs	1.107	0.941	0.166	0.081	False
BCP atom-IDs & QM	NCP QM	1.107	0.990	0.116	0.081	False
BCP atom-IDs & QM	NCP atom-IDs & QM	1.107	0.932	0.175	0.082	False
NCP QM	NCP atom-IDs	0.990	0.941	0.050	0.055	True
NCP atom-IDs	NCP atom-IDs & QM	0.941	0.932	0.009	0.046	True
NCP QM	NCP atom-IDs & QM	0.990	0.932	0.059	0.052	False

**Table S10.** Pairwise comparison of different model setups for the PDE10A (temporal 2011 split) and test whether errors are significantly different on the 95% confidence level.<sup>[17,39,40]</sup> Difference in RMSE values, composite errors, and whether or not the null hypothesis that the errors are the same holds are shown.

Method $A$	Method $B$	RMSE $_A$	RMSE $_B$	RMSE $_A$ - RMSE $_B$	Comp. error	Same error?
BCP QM	BCP atom-IDs	1.134	1.117	0.017	0.101	True
BCP atom-IDs & QM	BCP atom-IDs	1.245	1.117	0.127	0.100	False
BCP atom-IDs	NCP atom-IDs	1.117	1.054	0.063	0.099	True
NCP QM	BCP atom-IDs	1.132	1.117	0.015	0.119	True
BCP atom-IDs	NCP atom-IDs & QM	1.117	1.055	0.062	0.095	True
BCP atom-IDs & QM	BCP QM	1.245	1.134	0.110	0.107	False
BCP QM	NCP atom-IDs	1.134	1.054	0.080	0.104	True
BCP QM	NCP QM	1.134	1.132	0.002	0.120	True
BCP QM	NCP atom-IDs & QM	1.134	1.055	0.079	0.104	True
BCP atom-IDs & QM	NCP atom-IDs	1.245	1.054	0.190	0.104	False
BCP atom-IDs & QM	NCP QM	1.245	1.132	0.112	0.115	True
BCP atom-IDs & QM	NCP atom-IDs & QM	1.245	1.055	0.190	0.101	False
NCP QM	NCP atom-IDs	1.132	1.054	0.078	0.096	True
NCP atom-IDs & QM	NCP atom-IDs	1.055	1.054	0.001	0.035	True
NCP QM	NCP atom-IDs & QM	1.132	1.055	0.077	0.093	True

**Table S11.** Pairwise comparison of different model setups for the PDE10A (temporal 2012 split) and test whether errors are significantly different on the 95% confidence level.<sup>[17,39,40]</sup> Difference in RMSE values, composite errors, and whether or not the null hypothesis that the errors are the same holds are shown.

Method $A$	Method $B$	RMSE $_A$	RMSE $_B$	RMSE $_A$ - RMSE $_B$	Comp. error	Same error?
BCP QM	BCP atom-IDs	1.591	1.247	0.344	0.219	False
BCP atom-IDs	BCP atom-IDs & QM	1.247	1.188	0.058	0.157	True
NCP atom-IDs	BCP atom-IDs	1.353	1.247	0.106	0.186	True
NCP QM	BCP atom-IDs	1.570	1.247	0.323	0.233	False
NCP atom-IDs & QM	BCP atom-IDs	1.281	1.247	0.035	0.173	True
BCP QM	BCP atom-IDs & QM	1.591	1.188	0.402	0.240	False
BCP QM	NCP atom-IDs	1.591	1.353	0.237	0.245	True
BCP QM	NCP QM	1.591	1.570	0.020	0.262	True
BCP QM	NCP atom-IDs & QM	1.591	1.281	0.309	0.234	False
NCP atom-IDs	BCP atom-IDs & QM	1.353	1.188	0.165	0.213	True
NCP QM	BCP atom-IDs & QM	1.570	1.188	0.382	0.249	False
NCP atom-IDs & QM	BCP atom-IDs & QM	1.281	1.188	0.093	0.198	True
NCP QM	NCP atom-IDs	1.570	1.353	0.217	0.155	False
NCP atom-IDs	NCP atom-IDs & QM	1.353	1.281	0.072	0.085	True
NCP QM	NCP atom-IDs & QM	1.570	1.281	0.289	0.152	False

**Table S12.** Pairwise comparison of different model setups for the PDE10A (temporal 2013 split) and test whether errors are significantly different on the 95% confidence level.<sup>[17,39,40]</sup> Difference in RMSE values, composite errors, and whether or not the null hypothesis that the errors are the same holds are shown.

Method $A$	Method $B$	RMSE $_A$	RMSE $_B$	RMSE $_A$ - RMSE $_B$	Comp. error	Same error?
BCP atom-IDs	BCP QM	1.302	1.116	0.187	0.179	False
BCP atom-IDs	BCP atom-IDs & QM	1.302	1.233	0.069	0.163	True
BCP atom-IDs	NCP atom-IDs	1.302	0.976	0.326	0.173	False
BCP atom-IDs	NCP QM	1.302	1.007	0.295	0.172	False
BCP atom-IDs	NCP atom-IDs & QM	1.302	1.016	0.286	0.157	False
BCP atom-IDs & QM	BCP QM	1.233	1.116	0.117	0.189	True
BCP QM	NCP atom-IDs	1.116	0.976	0.140	0.141	True
BCP QM	NCP QM	1.116	1.007	0.109	0.150	True
BCP QM	NCP atom-IDs & QM	1.116	1.016	0.100	0.142	True
BCP atom-IDs & QM	NCP atom-IDs	1.233	0.976	0.257	0.182	False
BCP atom-IDs & QM	NCP QM	1.233	1.007	0.226	0.193	False
BCP atom-IDs & QM	NCP atom-IDs & QM	1.233	1.016	0.217	0.175	False
NCP QM	NCP atom-IDs	1.007	0.976	0.031	0.094	True
NCP atom-IDs & QM	NCP atom-IDs	1.016	0.976	0.040	0.065	True
NCP atom-IDs & QM	NCP QM	1.016	1.007	0.009	0.091	True

**Table S13.** Pairwise comparison of different model setups for the PDE10A (aminohetaryl-C1-amide split) and test whether errors are significantly different on the 95% confidence level.<sup>[17,39,40]</sup> Difference in RMSE values, composite errors, and whether or not the null hypothesis that the errors are the same holds are shown.

Method <i>A</i>	Method <i>B</i>	RMSE <sub><i>A</i></sub>	RMSE <sub><i>B</i></sub>	RMSE <sub><i>A</i></sub> - RMSE <sub><i>B</i></sub>	Comp. error	Same error?
BCP atom-IDs	BCP QM	1.926	1.302	0.624	0.138	False
BCP atom-IDs	BCP atom-IDs & QM	1.926	1.883	0.043	0.114	True
BCP atom-IDs	NCP atom-IDs	1.926	1.313	0.613	0.130	False
BCP atom-IDs	NCP QM	1.926	1.376	0.551	0.132	False
BCP atom-IDs	NCP atom-IDs & QM	1.926	1.389	0.538	0.142	False
BCP atom-IDs & QM	BCP QM	1.883	1.302	0.581	0.137	False
NCP atom-IDs	BCP QM	1.313	1.302	0.012	0.103	True
NCP QM	BCP QM	1.376	1.302	0.074	0.105	True
NCP atom-IDs & QM	BCP QM	1.389	1.302	0.087	0.110	True
BCP atom-IDs & QM	NCP atom-IDs	1.883	1.313	0.569	0.114	False
BCP atom-IDs & QM	NCP QM	1.883	1.376	0.507	0.114	False
BCP atom-IDs & QM	NCP atom-IDs & QM	1.883	1.389	0.494	0.120	False
NCP QM	NCP atom-IDs	1.376	1.313	0.062	0.066	True
NCP atom-IDs & QM	NCP atom-IDs	1.389	1.313	0.075	0.050	False
NCP atom-IDs & QM	NCP QM	1.389	1.376	0.013	0.073	True

**Table S14.** Pairwise comparison of different model setups for the PDE10A (C1-hetaryl-alkyl-C2-hetaryl split) and test whether errors are significantly different on the 95% confidence level.<sup>[17,39,40]</sup> Difference in RMSE values, composite errors, and whether or not the null hypothesis that the errors are the same holds are shown.

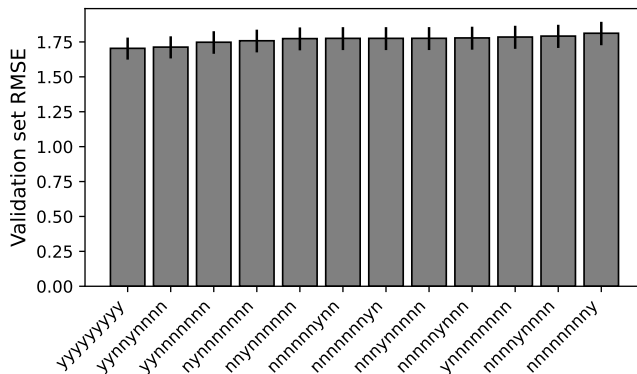
Method <i>A</i>	Method <i>B</i>	RMSE <sub><i>A</i></sub>	RMSE <sub><i>B</i></sub>	RMSE <sub><i>A</i></sub> - RMSE <sub><i>B</i></sub>	Comp. error	Same error?
BCP QM	BCP atom-IDs	1.401	1.183	0.218	0.144	False
BCP atom-IDs	BCP atom-IDs & QM	1.183	1.182	0.001	0.072	True
NCP atom-IDs	BCP atom-IDs	1.319	1.183	0.136	0.140	True
NCP QM	BCP atom-IDs	1.538	1.183	0.355	0.155	False
BCP atom-IDs	NCP atom-IDs & QM	1.183	1.100	0.083	0.126	True
BCP QM	BCP atom-IDs & QM	1.401	1.182	0.219	0.160	False
BCP QM	NCP atom-IDs	1.401	1.319	0.082	0.140	True
NCP QM	BCP QM	1.538	1.401	0.137	0.144	True
BCP QM	NCP atom-IDs & QM	1.401	1.100	0.301	0.126	False
NCP atom-IDs	BCP atom-IDs & QM	1.319	1.182	0.137	0.146	True
NCP QM	BCP atom-IDs & QM	1.538	1.182	0.356	0.149	False
BCP atom-IDs & QM	NCP atom-IDs & QM	1.182	1.100	0.082	0.128	True
NCP QM	NCP atom-IDs	1.538	1.319	0.219	0.091	False
NCP atom-IDs	NCP atom-IDs & QM	1.319	1.100	0.219	0.060	False
NCP QM	NCP atom-IDs & QM	1.538	1.100	0.438	0.090	False

**Table S15.** Pairwise comparison of different model setups for the PDE10A (aryl-C1-amide-C2-hetaryl split) and test whether errors are significantly different on the 95% confidence level.<sup>[17,39,40]</sup> Difference in RMSE values, composite errors, and whether or not the null hypothesis that the errors are the same holds are shown.

Method <i>A</i>	Method <i>B</i>	RMSE <sub><i>A</i></sub>	RMSE <sub><i>B</i></sub>	RMSE <sub><i>A</i></sub> - RMSE <sub><i>B</i></sub>	Comp. error	Same error?
BCP atom-IDs	BCP QM	1.777	1.455	0.322	0.140	False
BCP atom-IDs & QM	BCP atom-IDs	1.932	1.777	0.155	0.098	False
BCP atom-IDs	NCP atom-IDs	1.777	1.537	0.240	0.130	False
BCP atom-IDs	NCP QM	1.777	1.746	0.031	0.138	True
NCP atom-IDs & QM	BCP atom-IDs	1.853	1.777	0.077	0.144	True
BCP atom-IDs & QM	BCP QM	1.932	1.455	0.477	0.147	False
NCP atom-IDs	BCP QM	1.537	1.455	0.082	0.124	True
NCP QM	BCP QM	1.746	1.455	0.291	0.144	False
NCP atom-IDs & QM	BCP QM	1.853	1.455	0.399	0.141	False
BCP atom-IDs & QM	NCP atom-IDs	1.932	1.537	0.395	0.131	False
BCP atom-IDs & QM	NCP QM	1.932	1.746	0.186	0.137	False
BCP atom-IDs & QM	NCP atom-IDs & QM	1.932	1.853	0.078	0.130	True
NCP QM	NCP atom-IDs	1.746	1.537	0.209	0.100	False
NCP atom-IDs & QM	NCP atom-IDs	1.853	1.537	0.316	0.079	False
NCP atom-IDs & QM	NCP QM	1.853	1.746	0.108	0.120	True

## S4.4 Feature ablation

To assess whether using all available QM properties (Table 1 in the main paper) is useful, the set of hyperparameters was chosen which showed the lowest validation-set RMSE during initial hyperparameter screening. Models were trained using these hyperparameters but with varying subsets of QM properties or only on individual properties. The PDBbind dataset was used in this experiment, and results are shown in Figure S10 and Table S16. While the best-performing model uses all available QM properties, the drop in performance is limited when some individual properties are left out.



**Figure S10.** Model performance when using only a subset of QM properties. The labels "y" and "n" refer to whether a specific property (see Table 1) is used as a feature.

**Table S16.** Pairwise comparison of using all QM properties vs only subsets for the PDBbind validation set (random split) and test whether errors are significantly different on the 95% confidence level.<sup>[17,39,40]</sup> Difference in RMSE values, composite errors, and whether or not the null hypothesis that the errors are the same holds are shown.

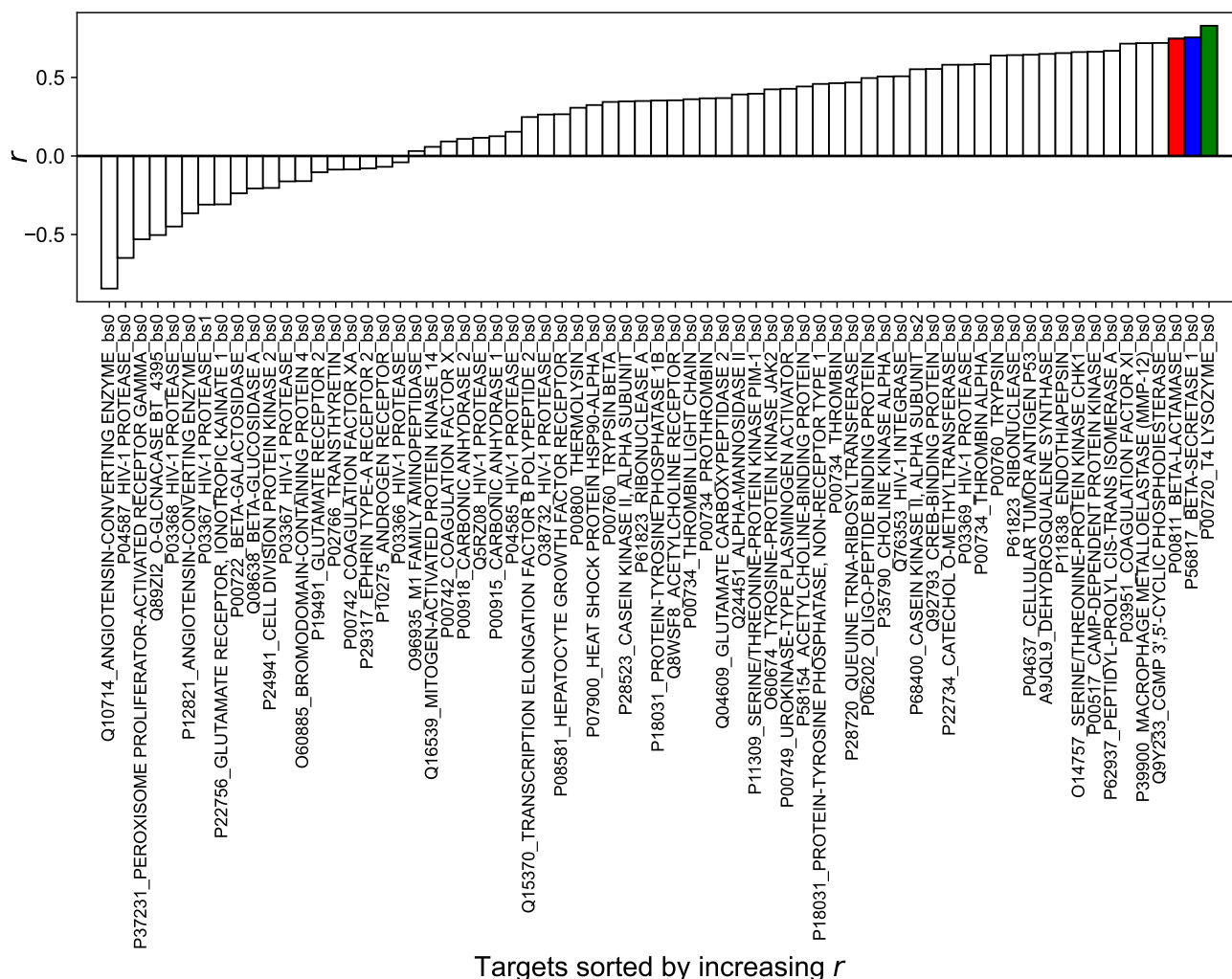
Method A	Method B	RMSE <sub>A</sub>	RMSE <sub>B</sub>	RMSE <sub>A</sub> - RMSE <sub>B</sub>	Comp. error	Same error?
ynnnnnnnn	yyyyyyyyy	1.785	1.704	0.081	0.044	False
nynnnnnnn	yyyyyyyyy	1.758	1.704	0.054	0.045	False
nnynnnnnn	yyyyyyyyy	1.774	1.704	0.070	0.043	False
nnnynnnnn	yyyyyyyyy	1.776	1.704	0.072	0.043	False
nnnnynnnn	yyyyyyyyy	1.792	1.704	0.088	0.054	False
nnnnnyynn	yyyyyyyyy	1.779	1.704	0.075	0.052	False
nnnnnnyyn	yyyyyyyyy	1.776	1.704	0.071	0.049	False
nnnnnnnyn	yyyyyyyyy	1.776	1.704	0.072	0.054	False
nnnnnnnny	yyyyyyyyy	1.812	1.704	0.108	0.050	False
yyynnnnnn	yyyyyyyyy	1.748	1.704	0.044	0.044	True
yyynyynn	yyyyyyyyy	1.713	1.704	0.009	0.042	True

## S5 Correlation of sum of electron density at BCPs with binding affinity

For the three best-correlating protein targets ( $\beta$ -lactamase (Uniprot-ID P00811),  $\beta$ -secretase 1 (Uniprot-ID P56817), T4 lysozyme (Uniprot-ID P00720)), we further analyzed the correlation between binding affinity and other descriptors, specifically the number of ligand atoms and number of intermolecular BCPs. The resulting Pearson correlation coefficients are shown in Table S17.

Protein-ligand complexes from the PDBbind<sup>[2,3]</sup> refined set were grouped by their Uniprot-ID and annotated protein name. Complexes that were annotated as allosteric binding sites in PDBbind were removed. To further separate groups of ligands binding to the same target but in different binding sites, protein chains bound to a ligand were aligned to one another using PyMol's<sup>[41]</sup> `align` functionality. The number of binding sites per target was manually annotated, and clusters of ligands were assigned using  $k$ -means clustering (with  $k$  equal to the number of binding sites) based on the atom coordinates of the first atom in each ligand. Due to the typically relatively large spatial separation of individual binding sites, no complications from this arbitrary choice of atom were expected, but the clustering was manually verified for all protein targets. The targets were (in addition to their Uniprot-ID





**Figure S11.** Protein targets (grouped by Uniprot-ID and protein name) sorted by increasing Pearson correlation coefficient  $r$ . Only protein targets with at least 10 ligands shown.

and protein name) also sub-divided by binding site (see <https://doi.org/10.3929/ethz-b-000619200>). Maximum ligand elongation was computed as the largest distance between any two ligand atoms in the binding site. Number of rotatable bonds and solvent-accessible surface area (SASA) were computed using RDKit.<sup>[42]</sup>

To analyze whether observed correlation trends between the sum of the electron density at intermolecular BCPs and the binding affinity were caused by the accuracy of the GFN2-xTB method, we used an analogous DFT approach (B3LYP-D3/6-31G\*<sup>[43,44]</sup>) to a previous study<sup>[45]</sup> to identify correlation between the two quantities. Psi4<sup>[16]</sup> (version 1.7) was used for DFT calculation and the methodology was applied to three sets of ligands:

- **Set 1:** 15 ligands for the acetylcholine-binding protein (Uniprot-ID P58154, PDB-IDs 1UV6, 1UW6, 3U8J, 3U8K, 3U8L, 3U8N, 3WTJ, 3WTL, 3WTM, 3WTN, 3WTO, 3ZDG, 3ZDH, 4ALX, 4QAC)
- **Set 2:** 14  $\beta$ -lactamase ligands (Uniprot-ID P00811, PDB-IDs 1XGI, 2PU2, 2R9W, 2R9X, 3GQZ, 3GR2, 3GTC, 3GV9, 3GVB, 4JXS, 4KZ3, 4KZ4, 4KZ6, 4KZ7)
- **Set 3:** 43  $\beta$ -secretase 1 ligands (Uniprot-ID P56817, PDB-IDs 1FKN, 1M4H, 2FDP, 2G94, 2P4J, 2QMG, 2VKM, 3BRA, 3BUF, 3BUG, 3BUH, 3CKP, 3I25, 3KMX, 3KMY, 3L59, 3LPI, 3LPK, 3RSX, 3RU1, 3UDH, 4DJU, 4DJW, 4DJX, 4DJY, 4FM7, 4FM8, 4FS4, 4FSL, 4GID, 4H3F, 4H3G, 4H3J, 4HA5, 5DQC, 5HTZ, 5I3V, 5I3W, 5I3X, 5I3Y, 5IE1, 6EJ2, 6EJ3)

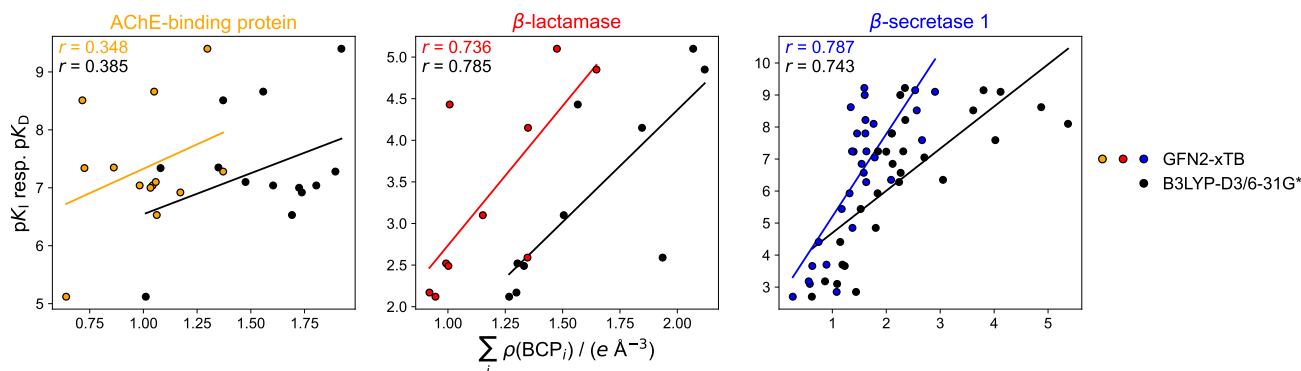
Computations were run with 8 cores, 128 GB RAM (resp. 256 GB for Set 3), 120 h maximum wall time, and 1,920 GB scratch space. The number of successfully computed complexes and Pearson correlation coefficients between the sum of electron density at intermolecular BCPs and binding affinity is reported in Table S18 and shown in Figure S12. While GFN2-xTB tends to underestimate the sum of the electron density at intermolecular BCPs, the overall correlation trends appear to be preserved, with only small differences in Pearson correlation coefficients between electron density computed at the two levels of theory and binding affinity.

**Table S17.** Pearson correlation coefficients  $r$  between binding affinity and the sum of electron density at intermolecular BCPs, the number of intermolecular BCPs, and the number of ligand atoms, respectively.

Protein target	$r(\sum \rho)$	$r(\# \text{ BCPs})$	$r(\# \text{ ligand atoms})$
$\beta$ -lactamase	0.7467	0.6319	0.7055
$\beta$ -secretase 1	0.7538	0.7419	0.7549
T4 lysozyme	0.8291	0.8793	0.5627

**Table S18.** Pearson correlation coefficients  $r$  between the sum of intermolecular BCPs and binding affinity for three sets of ligands binding to different protein targets: Set 1 (acetylcholine-binding protein), Set 2 ( $\beta$ -lactamase), Set 3 ( $\beta$ -secretase 1). Correlation coefficients computed only on complexes for which both DFT and GFN2-xTB calculations were successful.

Dataset	Set 1	Set 2	Set 3
Number of successfully computed complexes	13	10	29
DFT-correlation $r_{\text{B3LYP-D3/6-31G}^*}$	0.385	0.785	0.743
Semiempirical correlation $r_{\text{GFN2-xTB}}$	0.348	0.736	0.787



**Figure S12.** Correlation between sum of electron density at intermolecular BCPs (calculated using GFN2-xTB resp. B3LYP-D3/6-31G\*) and binding affinity. Only complexes shown for which electron densities from both GFN2-xTB and B3LYP-D3/6-31G\* could successfully be computed.

## References

- [1] T. Lu, F. Chen, *J. Comput. Chem.* **2012**, *33*, 580–592.
- [2] Z. Liu, Y. Li, L. Han, J. Li, J. Liu, Z. Zhao, W. Nie, Y. Liu, R. Wang, *Bioinformatics* **2015**, *31*, 405–412.
- [3] R. Wang, X. Fang, Y. Lu, S. Wang, *J. Med. Chem.* **2004**, *47*, 2977–2980.
- [4] M. Volkov, J.-A. Turk, N. Drizard, N. Martin, B. Hoffmann, Y. Gaston-Mathé, D. Rognan, *J. Med. Chem.* **2022**, *65*, 7946–7958.
- [5] The Open Babel Package, version 3.1.1, <http://openbabel.org>.
- [6] K. S. Aragão, M. Satre, A. Imberty, A. Varrot, *Proteins: Struct. Funct. Bioinf.* **2008**, *73*, 43–52.
- [7] J. Yang, Y. Liu, J. Bi, Q. Cai, X. Liao, W. Li, C. Guo, Q. Zhang, T. Lin, Y. Zhao, H. Wang, J. Liu, X. Zhang, D. Lin, *Mol. Microbiol.* **2015**, *95*, 791–803.
- [8] F. Benmansour, I. Trist, B. Coutard, E. Decroly, G. Querat, A. Brancale, K. Barral, *Eur. J. Med. Chem.* **2017**, *125*, 865–880.
- [9] S. Grimme, C. Bannwarth, P. Shushkov, *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.
- [10] C. Bannwarth, S. Ehlert, S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671.
- [11] S. Grimme, *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862.
- [12] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, S. Grimme, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, e01493.

- [13] J.-D. Chai, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.
- [14] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.
- [15] E. Prodan, W. Kohn, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 11635–11638.
- [16] 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, Henry F., A. Y. Sokolov, K. Patkowski, I. DePrince, A. Eugene, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford, C. D. Sherrill, *J. Chem. Phys.* **2020**, *152*, 184108.
- [17] J. H. Jensen, *PeerJ Preprints* **2017**, *5*, e2693v1.
- [18] Y. Li, Z. Liu, J. Li, L. Han, J. Liu, Z. Zhao, R. Wang, *J. Chem. Inf. Model.* **2014**, *54*, 1700–1716.
- [19] Y. Li, L. Han, Z. Liu, R. Wang, *J. Chem. Inf. Model.* **2014**, *54*, 1717–1736.
- [20] M. Su, Q. Yang, Y. Du, G. Feng, Z. Liu, Y. Li, R. Wang, *J. Chem. Inf. Model.* **2018**, *59*, 895–913.
- [21] K. Wang, R. Zhou, J. Tang, M. Li, *Bioinform.* **2023**, *39*, btad340.
- [22] H. Öztürk, A. Özgür, E. Ozkirimli, *Bioinform.* **2018**, *34*, i821–i829.
- [23] K. Wang, R. Zhou, Y. Li, M. Li, *Brief. Bioinform.* **2021**, *22*, bbab072.
- [24] M. M. Stepniewska-Dziubinska, P. Zielenkiewicz, P. Siedlecki, *Bioinform.* **2018**, *34*, 3666–3674.
- [25] S. Seo, J. Choi, S. Park, J. Ahn, *BMC Bioinform.* **2021**, *22*, 542.
- [26] D. Jones, H. Kim, X. Zhang, A. Zemla, G. Stevenson, W. D. Bennett, D. Kirshner, S. E. Wong, F. C. Lightstone, J. E. Allen, *J. Chem. Inf. Model.* **2021**, *61*, 1583–1592.
- [27] A. N. Jain, A. C. Brueckner, A. E. Cleves, M. Reibarkh, E. C. Sherer, *J. Med. Chem.* **2023**, *66*, 1955–1971.
- [28] A. Tosstorff, M. G. Rudolph, J. C. Cole, M. Reutlinger, C. Kramer, H. Schaffhauser, A. Nilly, A. Flohr, B. Kuhn, *J. Comput. Aided Mol. Des.* **2022**, *36*, 753–765.
- [29] A. Tosstorff, J. C. Cole, R. Taylor, S. F. Harris, B. Kuhn, *J. Chem. Inf. Model.* **2020**, *60*, 6595–6611.
- [30] A. Tosstorff, J. C. Cole, R. Bartelt, B. Kuhn, *ChemMedChem* **2021**, *16*, 3428–3438.
- [31] Z. Xiong, D. Wang, X. Liu, F. Zhong, X. Wan, X. Li, Z. Li, X. Luo, K. Chen, H. Jiang, M. Zheng, *J. Med. Chem.* **2019**, *63*, 8749–8760.
- [32] 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. Irlé, 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, T. Frauenheim, *J. Chem. Phys.* **2020**, *152*, 124101.
- [33] M. Gaus, A. Goez, M. Elstner, *J. Chem. Theory Comput.* **2013**, *9*, 338–354.
- [34] M. Gaus, X. Lu, M. Elstner, Q. Cui, *J. Chem. Theory Comput.* **2014**, *10*, 1518–1537.
- [35] X. Lu, M. Gaus, M. Elstner, Q. Cui, *J. Phys. Chem. B* **2015**, *119*, 1062–1082.
- [36] M. Kubillus, T. Kubar, M. Gaus, J. Rezac, M. Elstner, *J. Chem. Theory Comput.* **2015**, *11*, 332–342.
- [37] A. Otero-de-la-Roza, M. Blanco, A. M. Pendás, V. Luaña, *Comput. Phys. Commun.* **2009**, *180*, 157–166.
- [38] A. Otero-de-la-Roza, E. R. Johnson, V. Luaña, *Comput. Phys. Commun.* **2014**, *185*, 1007–1018.
- [39] A. Nicholls, *J. Comput. Aided Mol. Des.* **2016**, *30*, 103–126.
- [40] Jensen, Jan H. and Kromann, Jimmy C. and Christensen, Anders S., statsig - Determination of statistical significance using composite errors, <https://github.com/jensengroup/statsig> (accessed 19.06.23).
- [41] Schrödinger, LLC, “The PyMOL Molecular Graphics System, Version 2.5.2”.
- [42] Landrum, Greg, RDKit: Open-source cheminformatics and machine learning, <https://www.rdkit.org/docs/index.html> (accessed 19.06.23).
- [43] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
- [44] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257–2261.
- [45] S. Rojas, O. Parravicini, M. Vettorazzi, R. Tosso, A. Garro, L. Gutiérrez, S. Andújar, R. Enriz, *Eur. J. Med. Chem.* **2020**, *208*, 112792.