SUPPORTING INFORMATION

PredPotS – A Web Tool for Predicting Standard Reduction Potentials for Organic Molecules

F. B. Németh, A. Hamza, B. Tugyi, M. El-Ali, L. Szegletes, Á. Madarász, I. Pápai*

Contents

S1. The RP-ChEMBL database

All relevant data included in the RP-ChEMBL database are compiled in a comma-separated values (CSV) text file that can be downloaded from the *PredPotS* web application (*Datasets* tab, RP-ChEMBL.csv). The entries of the tabulated dataset include the ChEMBL codes of molecules, SMILES codes, charges and $1e^-$ standard reduction potentials computed with the composite GFN2-xTB/M06-2X protocol. Optimized geometries of the molecules, as well as their reduced forms are also available under the *Datasets* tab (zipped xyz files).

S2. Details of *DeepChem* model training and validation

In our present work we used deep learning tools provided by the open-source Python-based framework *DeepChem*[1] (version 2.8.0) to train predictive models on the RP-ChEMBL molecular database. Model training was performed on a personal laptop equipped with an AMD Ryzen 5 processor, 32 GB RAM, and no dedicated GPU. Depending on the machine learning model used, the training time ranged from 22 to 90 minutes. This timeframe reflects the average duration observed across multiple runs under consistent conditions.

We selected five standard deep learning models provided by *DeepChem* that were all specifically developed for graph property predictions, therefore we think they are suitable for handling larger datasets of organic molecules. These models are: graph convolutional network models *GCN*[2] and *Graph Conv*[3], graph attention network models *GAT*[4] and *Attentive FP*[5], and the directed acyclic graph model *DAG*.[6] The inputs for all these models were the molecules collected in the database RP-ChEMBL represented by their standardized SMILES codes. Conversion of SMILES codes to graph structures was performed using two different featurizers from *DeepChem*. For the *GraphConv* and *DAG* models, we used the built-in ConvMolFeaturizer, which generates a fixed-structure graph where each atom is represented by a feature vector based on local atomic and pairwise information. For the *DAG* model, the resulting graph was further processed using the DAGTransformer to transform it to the model's input requirements. For the *GAT*, *GCN*, and *Attentive FP* models, we used the more flexible MolGraphConvFeaturizer, which constructs graph objects with customizable node and edge features. These featurizers prepare the graph inputs for the graph neural network models enabling them to learn representations that capture both local and global structural information relevant to the chemical property that will be predicted.

For hyperparameter tuning, model training and performance monitoring, we applied a consistent procedure across all five models. The dataset was randomly split into training (train), validation (valid) and test (test) sets using the conventional 0.8:0.1:0.1 ratio. The computed values obtained following the protocol of the composite GFN2-xTB/M06-2X method were used as reference data for the supervised model training. In the following sections, we present the detailed procedures, evaluation methods and performance results for example models, but the same approach was used for all other models.

First, the model's performance was evaluated across multiple random seeds while maintaining the same data split ratio. Table S1 summarizes the results for the *Attentive FP* model. Training was conducted using a dropout rate of 0 and applying the early stopping as regularization strategy. On the test set, the model achieved a mean absolute error (MAE) of 0.134 V. Using the t-distribution (with $t \approx 2.776$) to calculate the 95% confidence interval, the MAE was estimated to lie within ±0.007 V of the mean, yielding an interval of approximately (0.125, 0.142). This means we can be 95% confident that the true test MAE falls within this range.

Table S1. Evaluation metrics results for the *Attentive FP* model using different random seeds for the dataset splitting with the ratio of 0.8:0.1:0.1.

Nr.	split seed	seed	epoch	MAE_ train	MAE_ valid	MAE_ test
1	45	144	98	0.0638	0.1351	0.1263
2	35	135	97	0.0688	0.1335	0.139
3	78	245	101	0.0617	0.1275	0.1261
4	76	123	67	0.0926	0.1387	0.1356
5	74	625	94	0.0643	0.1254	0.1406
Mean				0.070±0.011	0.132±0.005	0.134±0.006

For the five models deployed in the *PredPotS* application a fixed random seed parameter was used during training to ensure reproducibility and consistent dataset partitioning. The default hyperparameters of the neural networks provided by *DeepChem* performed well across all models for our purposes. The most relevant parameters are summarized in Table S2.

Table S2: Most important built-in parameter values for the selected *DeepChem* models.

model	parameter	value
Attentive FP	number of GCN layers	2
	number of atom features	30
	number of bond features	11
	max time steps for updating graph representation	2
GraphConv	graph conv layers	64x64
	dense layer size	128
	number of atom features	75
	batch size	100
GCN	graph conv layers	64x64
	number of atom features	30
	predictor hidden features	128
GAT	graph conv layers	64x64
	graph attention layers	8x8
	n attention heads	8
	number of atom features	30
	predictor hidden features	128
DAG	number of atom features	75
	number of features for atom in the graph	30
	number of features of each molecule	30
	number of hidden layers	100

To avoid overtraining of the models, the series of dropout = 0, 0.1, 0.2 were tested for each model. For this purpose the loss function computed in each epoch for the validation and training set were monitored for trainings consisting of a minimum of 100 epochs and applying the early stopping regularization with a patience of 2. The loss function considered is the squared difference between the true and predicted values as implemented in *DeepChem* (*L2Loss*). To determine the optimal dropout parameter for each model, we examined both parity plots and training/validation loss curves. For the *Graph Conv* model,

results with dropout values of 0, 0.1, and 0.2 are illustrated in Figure S1. While the overall error functions are quite similar across these settings (see Table S3), a notably larger difference between training and validation MAE at the dropout of 0 and 0.1 suggests potential overfitting. In contrast, the model with a dropout of 0.2 exhibits the most balanced performance, indicating it as the most suitable choice. The same routine was used for all models, the final dropout values are listed in Table S4.

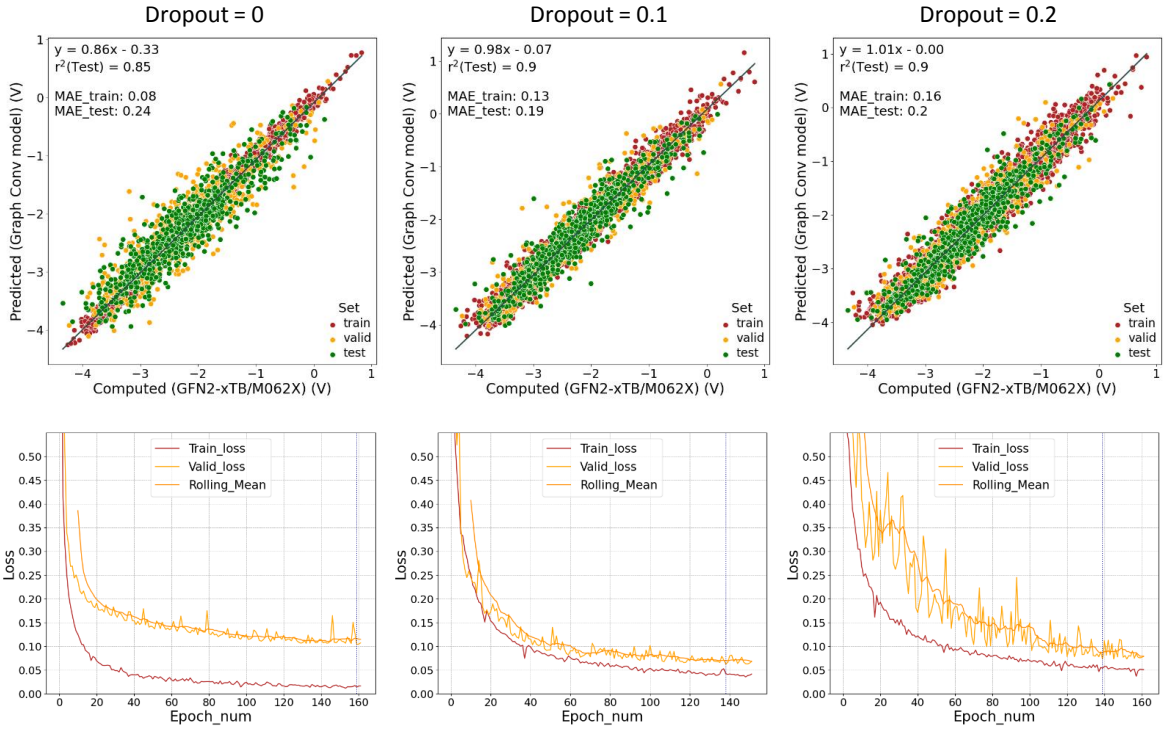


Figure S1: Parity plots for the computed and predicted values (upper) and loss functions for the training and validation sets (lower) for different dropout values: 0.0, 0.1 and 0.2 by using *Graph Conv* model. The computed and predicted reduction potential values are in V.

Table S3: Performance of ML trainings by using the *Graph Conv* model for dropouts 0.0, 0.1, 0.2.

Dropout	Set	MAE	RMSD	R2
0.0	Train	0.0797	0.0999	0.9873
	Valid	0.2419	0.323	0.8345
	Test	0.2356	0.3053	0.851
0.1	Train	0.127	0.1593	0.9708
	Valid	0.1871	0.25	0.9112
	Test	0.19	0.2525	0.9041
0.2	Train	0.1643	0.2067	0.9524
	Valid	0.2066	0.2699	0.9034
	Test	0.204	0.2693	0.8977

Table S4: Dropout parameters used for training the models.

Model	Dropout
Attentive FP	0.1
GraphConv	0.2
GCN	0.1
GAT	0
DAG	0.2

Training machine learning models is inherently stochastic and can exhibit considerable noise, making it important to carefully determine an appropriate stopping point. Since the validation error curves may present several local minima, we extended training to 800 epochs to observe the overall trend in model performance and estimate the minimal necessary training epochs.

S3. Training and evaluation of *DeepChem* models

Herein, we present a detailed analysis of the training strategy applied to the *Attentive FP* model. This includes the investigation of early stopping behavior, and model performance across different training durations. To prevent premature termination due to short-term fluctuations in validation loss, we introduced a parameter specifying the minimum number of epochs (*min_epochs*). This ensures that training proceeds for at least this minimum number of epochs before early stopping is considered. After this threshold is passed, early stopping with a patience of 2 is applied, meaning training stops if no improvement in validation loss is observed for two consecutive epochs. In all cases, the model is restored to the best-performing epoch (*best_epoch*) based on validation loss. This strategy allows flexibility in total training duration, which can extend up to a higher maximum epoch limit if early stopping conditions are not met.

To analyse the model performance over time, we changed the minimum epoch parameter with a step of 100, and evaluated the corresponding error metrics: training and test MAE, training and validation loss at each setting and the overall rolling mean (using a window size of 10) for these functions. This approach allowed us to assess the impact of training duration and early stopping dynamics on model generalization as shown in Figure S2.

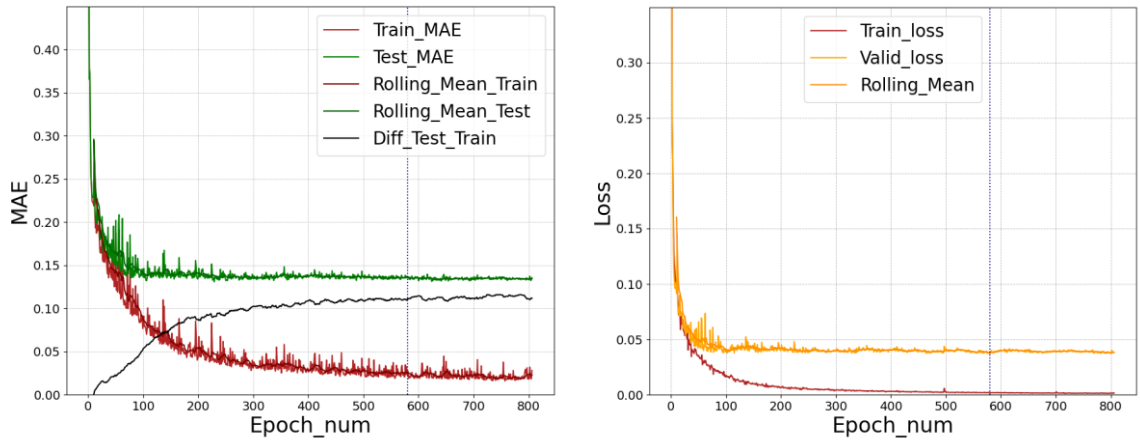


Figure S2: Performance of the model *Attentive FP* recorded for 800 epochs. Left: MAE and rolling mean of MAE for the train and test sets. Right: Loss function and rolling mean for the train and valid sets.

The details of the evaluation are listed in table S5 for the *Attentive FP* model. The 'best epoch' is the epoch number corresponding to the trained model stopped by the early stopping regularization after the respective minimal epochs was reached. It is apparent that after the minimal epochs value of 200 the performance of the trained model doesn't improve indicating that a choice of the minimal epoch leading to the best trained model is between 100 and 200.

Table S5: Performance of model *Attentive FP* evaluated at every 100th epoch.

min epochs	best_epoch	loss valid	loss train	MAE train	MAE valid	MAE test	MAE roll train	MAE roll test
100	95	0.0377	0.0178	0.0837	0.1373	0.1373	0.1008	0.1454
200	102	0.0376	0.0162	0.0789	0.1351	0.1356	0.0926	0.1425
300	102	0.0376	0.0162	0.0789	0.1351	0.1356	0.0926	0.1425
400	102	0.0376	0.0162	0.0789	0.1351	0.1356	0.0926	0.1425
500	464	0.0375	0.0023	0.0232	0.1368	0.1345	0.0289	0.1365
600	580	0.0368	0.0019	0.0224	0.1345	0.1353	0.0255	0.1358
700	580	0.0368	0.0019	0.0224	0.1345	0.1353	0.0255	0.1358

The model weights corresponding to the best epoch determined by the early stopping criterion, were saved and used as the final trained model. In Table S6 we provide a summary of the training parameters for each model.

Table S6: Training parameters used in various deep learning methods.[a]

Model	seed	Dropout	min_epoch	best_epoch
Attentive FP	283	0.1	200	81
GraphConv	768	0.2	250	263
GCN	366	0.1	350	336
GAT	483	0	400	314
DAG	637	0.2	100	86

[a] Notation: Minimum epoch numbers (min_epoch), the epoch number of the best model (best_epoch) according to the early stopping with a patience of 2, and the random seed used for the trainings (seed). For all models the random seed=42 was used for the dataset split into training/validation/test sets.

S4. The performance of *DeepChem* models

The parity plot for the computed (via the protocol) and predicted values of the three sets (train, valid, test), the loss functions for train and valid and the MAE for the test set are shown in Figure S3.

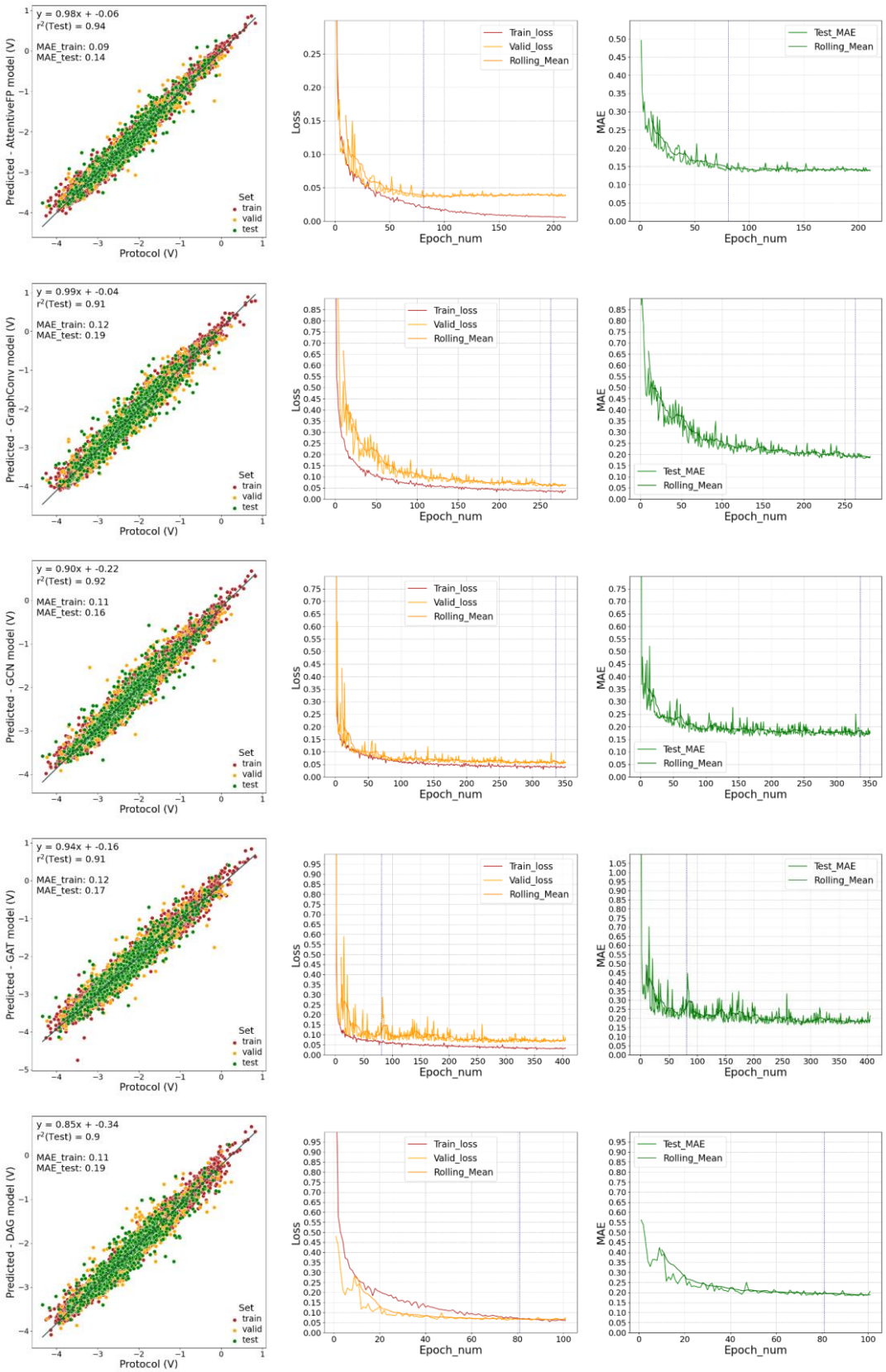


Figure S3. Performance plots for the models used in the *PredPotS* application: *Attentive FP*, *GraphConv*, *GCN*, *GAT*, and *DAG* models.

S5. Analysis of outliers

Despite the satisfying correlation between the predicted and computed reduction potentials, the parity plots display a number of outliers (see Figure S3). To understand the origin of these exceptional discrepancies, we analyzed the structures of these compounds (both original and reduced forms) obtained by the applied GFN2-xTB/M06-2X computational protocol, and compared them to those provided by a more accurate approach, namely by full DFT calculations. In this latter approach, the geometries of both original and reduced forms of molecules were optimized at M06-2X/6-31+G(d,p)/SMD(water) level (with initial geometries from the GFN2-xTB/M06-2X protocol), and the $1e^-$ standard reduction potentials were obtained from Gibbs free energies computed at the same DFT level. The present analysis was carried out for compounds identified as outliers of predictions using the *Attentive FP* model, but we expect similar conclusions for the other deep learning models as well. The molecules from the validation and test datasets, for which the error of *Attentive FP* predictions of reduction potentials is larger than 0.6 V are shown in Figures S4 and S5, and the predicted/computed data are compiled in Table S7.

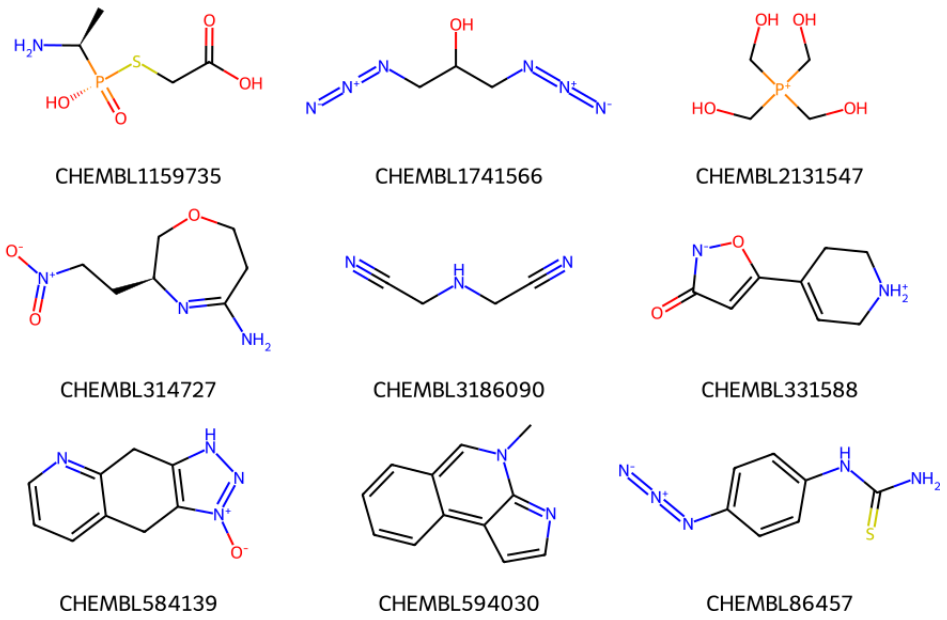


Figure S4: Molecules from the validation set of the RP-ChEMBL database for which the *Attentive FP* model gives exceptionally erroneous reduction potential predictions (discrepancy with respect to values computed with the GFN2-xTB/M06-2X protocol is larger than 0.6 V).

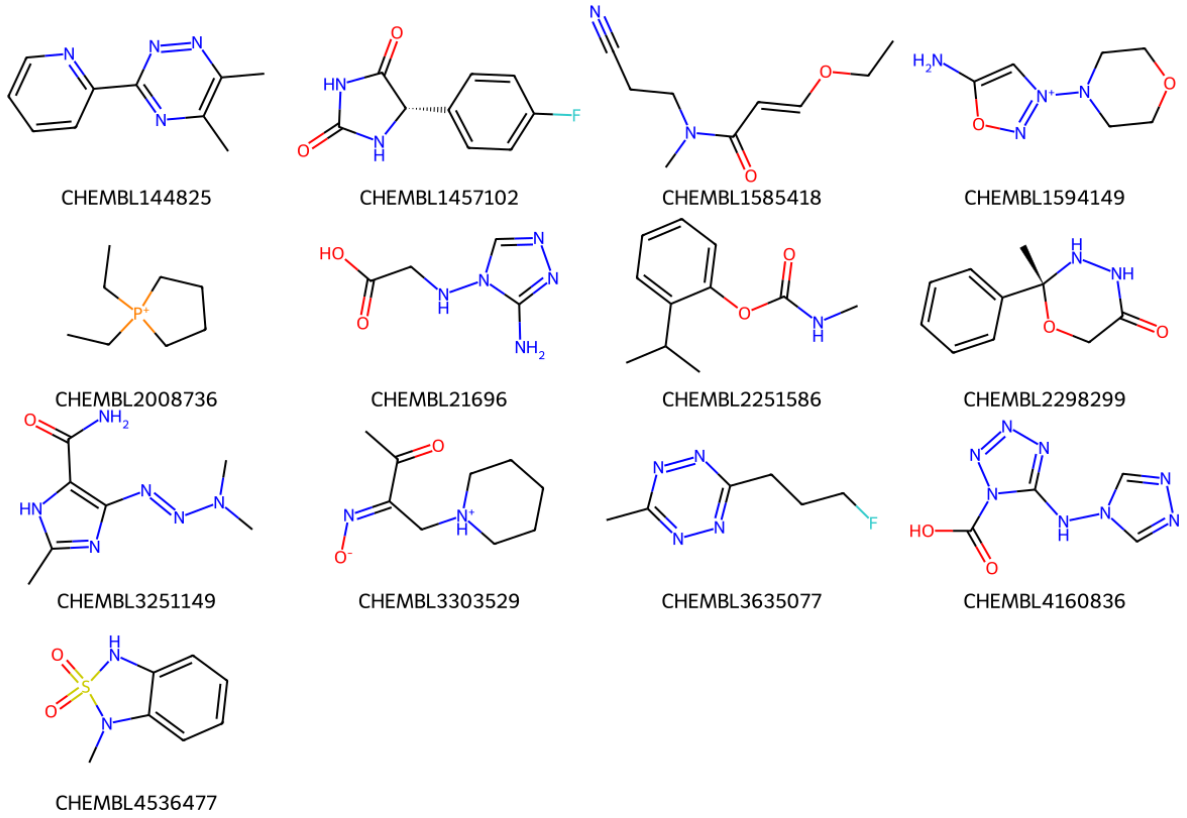


Figure S5: Molecules from the test set of the RP-ChEMBL database for which the *Attentive FP* model gives exceptionally erroneous reduction potential predictions (discrepancy with respect to values computed with the GFN2-xTB/M06-2X protocol is larger than 0.6 V).

Our structural analysis indicated that some of the exceptional discrepancies between the predicted and computed reduction potentials clearly originate from the inaccurate structures obtained with the GFN2-xTB method, as evidenced by large differences between the GFN2-xTB/M06-2X and full DFT results (E_o^2 data in Table S7**).** The structural artefacts include bond weakenings, partial bond formations between substituents, amplified inaccuracies for particular structural motives (e.g. amide, N=N bonds, C≡N bonds, hydrogen bonds). In two cases, we found that the conformational search of the GFN2-xTB/M06-2X protocol (*crest* utility) did not find the most stable forms of the molecules. In all other cases, where the E_o^2 differences are insignificant, the large E_o^1 deviations are likely related to the deep limited capacity of deep learning models to generalize across diverse molecular structures. This also depends on the sensitivity of the methods to the unique properties of certain molecules.

Table S7: Computed and predicted reduction potentials of molecules from the validation and test set of the RP-ChEMBL database giving large errors in the predictions with the *Attentive FP* model.[a]

ChEMBL code	CHG	SMILES	Set	E_o^{prot}	E_o^{pred}	E_o^1	E_o^{DFT}	E_o^2	Origin of error
CHEMBL1159735	0	[P@](=O)(SCC(=O)O)([C@@H](N)C)O	valid	-2.46	-1.79	0.66	-1.95	**0.51**	xTB
CHEMBL1741566	0	OC(CN=[N+]=[N-])CN=[N+]=[N-]	valid	-1.99	-0.99	0.99	-1.49	**0.50**	xTB
CHEMBL2131547	1	[P+](CO)(CO)(CO)CO	valid	-2.08	-3.07	-0.99	-2.10	-0.02	ML + xTB
CHEMBL314727	0	[N+](=O)([O-])CC[C@@H]1N=C(N)CCOC1	valid	-0.17	-1.24	-1.07	-0.95	**-0.78**	xTB
CHEMBL3186090	0	N(CC#N)CC#N	valid	-3.70	-3.04	0.67	-3.68	0.02	xTB
CHEMBL331588	0	c1c(o[n-]c1=O)C1=CC[NH2+]CC1	valid	-2.26	-2.96	-0.70	-2.34	-0.07	ML
CHEMBL584139	0	c1cnc2c(c1)Cc1c(C2)[nH]n[n+]1[O-]	valid	-2.76	-2.13	0.63	-2.53	0.23	xTB + ML
CHEMBL594030	0	c1ccc2c(c1)c1c(n(c2)C)ncc1	valid	-1.59	-2.27	-0.68	-1.65	-0.06	ML
CHEMBL86457	0	[N+](=[N-])=Nc1ccc(NC(=S)N)cc1	valid	-2.01	-1.38	0.64	-1.11	**0.90**	xTB
CHEMBL144825	0	c1(nc(c(nn1)C)C)c1ncccc1	test	-1.75	-1.10	0.65	-1.72	0.03	ML
CHEMBL1457102	0	Fc1ccc([C@@H]2NC(=O)NC2=O)cc1	test	-3.39	-2.52	0.87	-2.75	**0.63**	xTB
CHEMBL1585418	0	O=C(N(CCC#N)C)/C=C/OCC	test	-3.11	-2.45	0.66	-2.40	**0.71**	xTB
CHEMBL1594149	1	o1n[n+](N2CCOCC2)cc1N	test	-0.69	-1.50	-0.81	-0.78	-0.09	ML
CHEMBL2008736	1	[P+]1(CCCC1)(CC)CC	test	-2.87	-3.51	-0.64	-2.79	0.08	ML
CHEMBL21696	0	n1(c(nnc1)N)NCC(=O)O	test	-2.57	-3.18	-0.60	-2.48	0.09	ML
CHEMBL2251586	0	c1(ccccc1OC(=O)NC)C(C)C	test	-4.01	-3.39	0.62	-3.08	**0.93**	xTB
CHEMBL2298299	0	c1ccc(cc1)[C@]1(OCC(=O)NN1)C	test	-3.69	-2.92	0.77	-2.89	**0.80**	xTB
CHEMBL3251149	0	c1(nc(c([nH]1)C(=O)N)/N=N/N(C)C)C	test	-1.32	-2.02	-0.70	-1.41	-0.09	crest
CHEMBL3303529	0	O=C(/C(=N/[O-])/C[NH+]1CCCCC1)C	test	-1.41	-2.04	-0.64	-1.41	-0.01	ML
CHEMBL3635077	0	n1c(nnc(n1)CCCF)C	test	-0.59	0.08	0.68	-0.79	-0.20	ML
CHEMBL4160836	0	c1(n(nnn1)C(=O)O)Nn1cnnc1	test	-1.66	-2.47	-0.80	-1.30	**0.37**	crest
CHEMBL4536477	0	CN1c2ccccc2NS1(=O)=O	test	-1.15	-1.76	-0.61	-0.48	**0.67**	xTB

[a]Notation: E_o^{prot}, E_o^{pred} and E_o^{DFT} refer to 1e⁻ standard reduction potentials obtained via the GFN2-xTB/M06-2X computational protocol, via predictions of the *Attentive FP* model, and via full DFT calculations, respectively; $E_o^1 = E_o^{pred} - E_o^{prot}$; $E_o^2 = E_o^{DFT} - E_o^{prot}$. The last column of the table specifies the origin of the error: xTB – structural uncertainty inherited from the semiempirical GFN2-xTB method, ML – uncertainty of machine learning training, crest – uncertainty from crest conformational analysis.

To further examine the uncertainty level of the GFN2-xTB/M06-2X protocol, we carried out additional test calculations for a set of compounds from the RP-ChEMBL database. We selected 50 compounds from the database randomly, but also ensuring that the reduction potential of these compounds covers the entire potential range of RP-ChEMBL compounds. We then carried out geometry optimizations for the original and reduced forms at the M06-2X/6-31+G(d,p)/SMD(water) level of DFT and computed the $1e^-$ standard reduction potentials at the same DFT level. The results are compiled in Table S8, which includes the ChEMBL and SMILES codes of the selected compounds and all data relevant to the test calculations.

The parity plot of mean predicted versus the GFN2-xTB/M06-2X potentials shows an excellent correlation for the molecules of this test set (R^2 = 0.99, MAE = 0.13 V, see Figure S6). The results of full DFT calculations reveal that the GFN2-xTB/M06-2X protocol gives quite accurate data for the majority of the molecules, and there are only a few exceptions in the highly negative potential region as illustrated in Figure S7. Larger discrepancies (>0.4 V) are obtained only for four compounds, which include structural units that appear problematic for the GFN2-xTB method. Inaccuracies associated with the amide and the C≡N units have already been mentioned in the previous section, but the structural analysis indicates that the reduced form of CHEMBL1631455 (i.e. the pyramidality of the carbonyl group) is not accurately described either by the GFN2-xTB method.

The present analysis provides further support for the efficiency of the composite GFN2-xTB/M06-2X protocol in the calculation or redox potentials, but it also highlights its limitations.

Table S8: Computed and predicted reduction potentials of molecules selected from the RP-ChEMBL database for test calculations.[a]

ChEMBL code	CHG	SMILES	E_o^{pred}	E_o^{prot}	E_o^{DFT}	$E_o^{prot-DFT}$
CHEMBL1999493	0	O=C(N[C@H](CNC(=O)N)C)N	-3.79	-4.23	-3.75	-0.47
CHEMBL3558608	1	C1[C@H](CC=C1)C[N+](C)(C)C	-3.77	-4.20	-3.81	-0.39
CHEMBL1191543	0	C(=O)([C@@H]1C2(C[NH2+]CC1)CCCCC2)[O-]	-3.75	-4.14	-3.88	-0.26
CHEMBL431307	0	C1(=O)N([C@@H](Cc2n(cnc2)C)CO1)C	-3.44	-3.93	-3.48	-0.45
CHEMBL1978310	0	P(=O)(C([C@H](C(C)C)C)(C)C)(C)C	-3.79	-3.83	-3.65	-0.18
CHEMBL1741675	0	N(CC(=C)C)(CC(=C)C)CC#N	-3.53	-3.73	-2.88	-0.85
CHEMBL1489129	0	N12C(=NCCC1)CCCCC2	-3.57	-3.63	-3.67	0.05
CHEMBL1621372	0	n1(nc(c(c1N)C)CC)C	-3.61	-3.53	-3.56	0.04
CHEMBL4517591	0	N[C@H]1CCn2ccnc12	-3.54	-3.42	-3.47	0.05
CHEMBL3899142	0	NC/C=C/CN	-3.39	-3.32	-3.36	0.04
CHEMBL343650	0	C(=O)([C@@H](N)[C@@H](CC)C)N(CCC)C	-3.15	-3.22	-3.18	-0.04
CHEMBL2109969	0	C(C#C[C@H]1CCCN1)N1CCCC1	-3.31	-3.12	-3.10	-0.02
CHEMBL1631455	0	C(SCCC(=O)C)CO	-2.66	-3.02	-2.52	-0.50
CHEMBL4549622	0	O=C1CNc2cccc(F)c2N1	-2.91	-2.92	-2.71	-0.20
CHEMBL1432925	0	S=C(NC(C)(C)C)NC(C)(C)C	-3.02	-2.82	-2.84	0.02
CHEMBL2111859	0	[C@@]123[C@H](c4c(C1)ccnc4)N(CC2)CC3	-2.66	-2.72	-2.76	0.04
CHEMBL125783	0	c1(c2ccc(cc2)O)c(F)cccc1	-2.51	-2.61	-2.62	0.00
CHEMBL1732601	0	[nH]1ncc2c1ncnc2NCC	-2.58	-2.51	-2.50	-0.02
CHEMBL92537	0	c1(nc(no1)N)[C@H]1N(CCN(C1)C)C	-2.63	-2.41	-2.38	-0.04
CHEMBL1999762	0	O=c1n2c([nH]c3c1CCCC3)ncc2	-2.51	-2.31	-2.32	0.01
CHEMBL12849	0	c12c(cc(s2)N)[nH]c(=O)[nH]c1=O	-2.25	-2.21	-2.14	-0.07
CHEMBL1213612	0	O=C(O)c1ccsc1NC(=O)C	-2.13	-2.11	-2.05	-0.05
CHEMBL1313117	0	o1c2c(c(N)c(N)c1=O)cccc2	-2.01	-2.01	-2.02	0.02
CHEMBL195075	0	N(=C\c1ccccc1)/Oc1ccccc1	-1.91	-1.91	-1.95	0.04
CHEMBL1896392	0	O=C(N1[C@H](C1)C)c1c(cccc1)C	-2.11	-1.80	-1.97	0.17
CHEMBL2132927	0	O=C1C(=C(c2ccccc2)CC1)C	-1.64	-1.70	-1.67	-0.04
CHEMBL4583850	0	c1c(ccc(c1)/C=C(/C(=O)O)\O)Cl	-1.60	-1.60	-1.63	0.03
CHEMBL3185350	0	N#Cc1c(cccc1)C#N	-1.53	-1.50	-1.50	0.00
CHEMBL1977616	0	O=C1C(=O)C=C1O	-1.19	-1.40	-1.28	-0.12
CHEMBL1990202	0	S1(=O)(=O)N(C(=O)[C@H](N1)C)C	-1.43	-1.30	-1.20	-0.10
CHEMBL1548859	0	s1c(C(=O)c2ncc[nH]2)ccc1	-1.19	-1.20	-1.18	-0.01

CHEMBL29207	0	[N+](=[N-])=CC(=O)C=C=CCCCCCC	-0.99	-1.10	-1.05	-0.05
CHEMBL357682	0	S(=O)(=O)(O/C=C\1/C(=O)OCC1)C	-1.09	-0.99	-0.67	-0.33
CHEMBL479228	0	c1c[n+](cc(c1)C(=O)O)CC(=O)[O-]	-0.80	-0.89	-1.07	0.18
CHEMBL1651797	0	c1c(cccc1)NC(=O)/C(=N/O)/C#N	-0.77	-0.79	-0.79	-0.01
CHEMBL489894	0	C1[C@H](COc2n1c(cn2)[N+](=O)[O-])O	-0.76	-0.69	-0.71	0.02
CHEMBL164925	0	[N+](=O)(c1cc2c(NN(C2)C)cc1)[O-]	-0.71	-0.59	-0.64	0.05
CHEMBL231045	0	c1ccc(c(c1)/C=C/[N+](=O)[O-])O	-0.47	-0.49	-0.47	-0.02
CHEMBL241419	0	[n+]1(nc([n+](c2c1cccc2)[O-])C)[O-]	-0.52	-0.39	-0.44	0.06
CHEMBL1269722	0	c1ccc(cc1)/C=C(/[N+](=O)[O-])\O	-0.39	-0.28	-0.27	-0.01
CHEMBL253356	0	C1(=C(C(=CC(=O)C1=O)OC)O)C	-0.16	-0.18	-0.20	0.02
CHEMBL2296053	0	C1(=O)C(=C(C(=O)C(=C1O)CC)O)CC	-0.15	-0.08	-0.08	0.00
CHEMBL1559303	0	O=C1C(=C(O)C(=O)C(=C1O)O)O	-0.12	0.02	0.07	-0.05
CHEMBL1985450	0	s1cc2c(c(=O)c(=O)c2=O)c1	-0.33	0.13	0.14	-0.01
CHEMBL1971808	0	O=C1c2c(C(=O)C=C1)ccc(=O)cc2	0.14	0.22	0.23	-0.01
CHEMBL1743218	0	C1=CC(=CC(=O)C1=O)C[C@H](C(=O)[O-])[NH3+]	0.33	0.34	0.25	0.09
CHEMBL174035	0	C1(=O)C(=O)[C@H](OC1=O)[C@H](O)CO	0.37	0.46	0.40	0.06
CHEMBL2361875	1	O=[N+]1C=C(/N=N/N(CC)CC)C=CC1	0.35	0.55	0.17	0.38
CHEMBL1189149	2	[N+]1(=CC=[N+]([C@@H]1C)C)C	0.64	0.65	0.62	0.03
CHEMBL2134077	0	O=C1[C@H]([N+](=O)[O-])C(=O)NC(=O)C1=O	0.76	0.74	0.68	0.05

[a]Notation: E_o^{pred}, E_o^{prot} and E_o^{DFT} refer to 1e⁻ standard reduction potentials obtained via predictions (mean values with Z-score filtering switched on), the GFN2-xTB/M06-2X computational protocol, and via full DFT calculations, respectively; $E_o^{prot-DFT} = E_o^{prot} - E_o^{DFT}$.

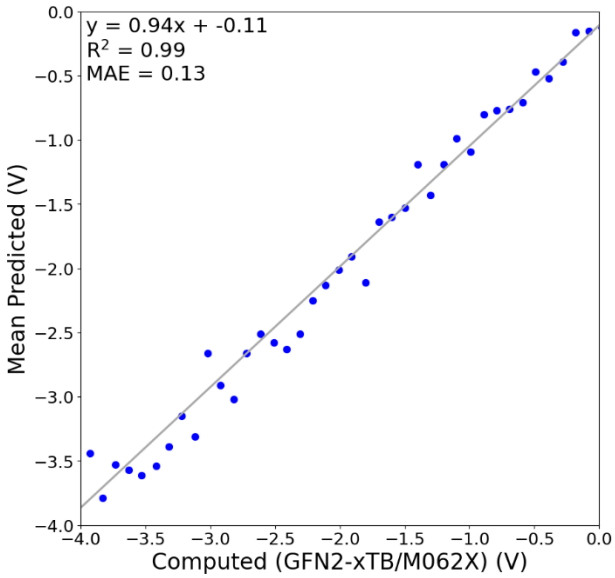


Figure S6: Parity plot of predicted 1e⁻ reduction potentials (mean predictions) vs. the values computed with the GFN2-xTB/M06-2X protocol for a set molecules selected from the RP-ChEMBL database (see Table S8).

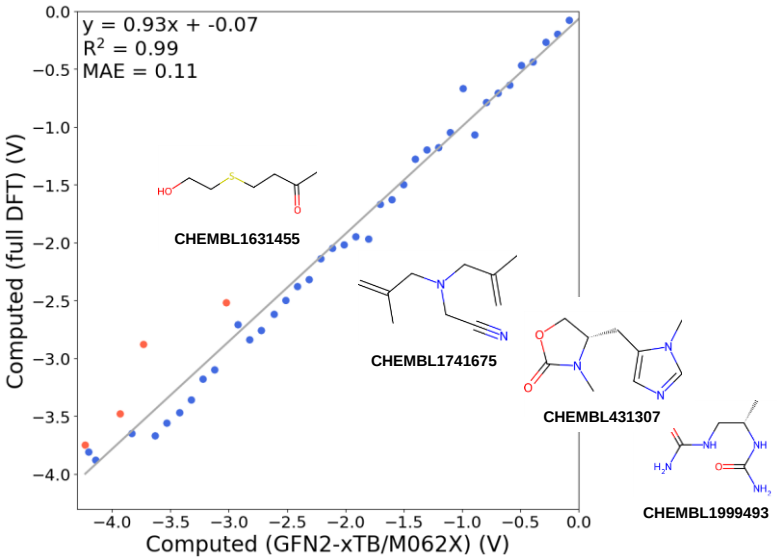


Figure S7: Parity plot of 1e⁻ reduction potentials computed with the full DFT and GFN2-xTB/M06-2X methods for a set molecules selected from the RP-ChEMBL database (see Table S8). Outlier data with discrepancies >0.4 V are highlighted in orange.

S6. Distribution of signed errors

The distribution of signed errors for predictions using the five deep learning models are reported in Figure S8.

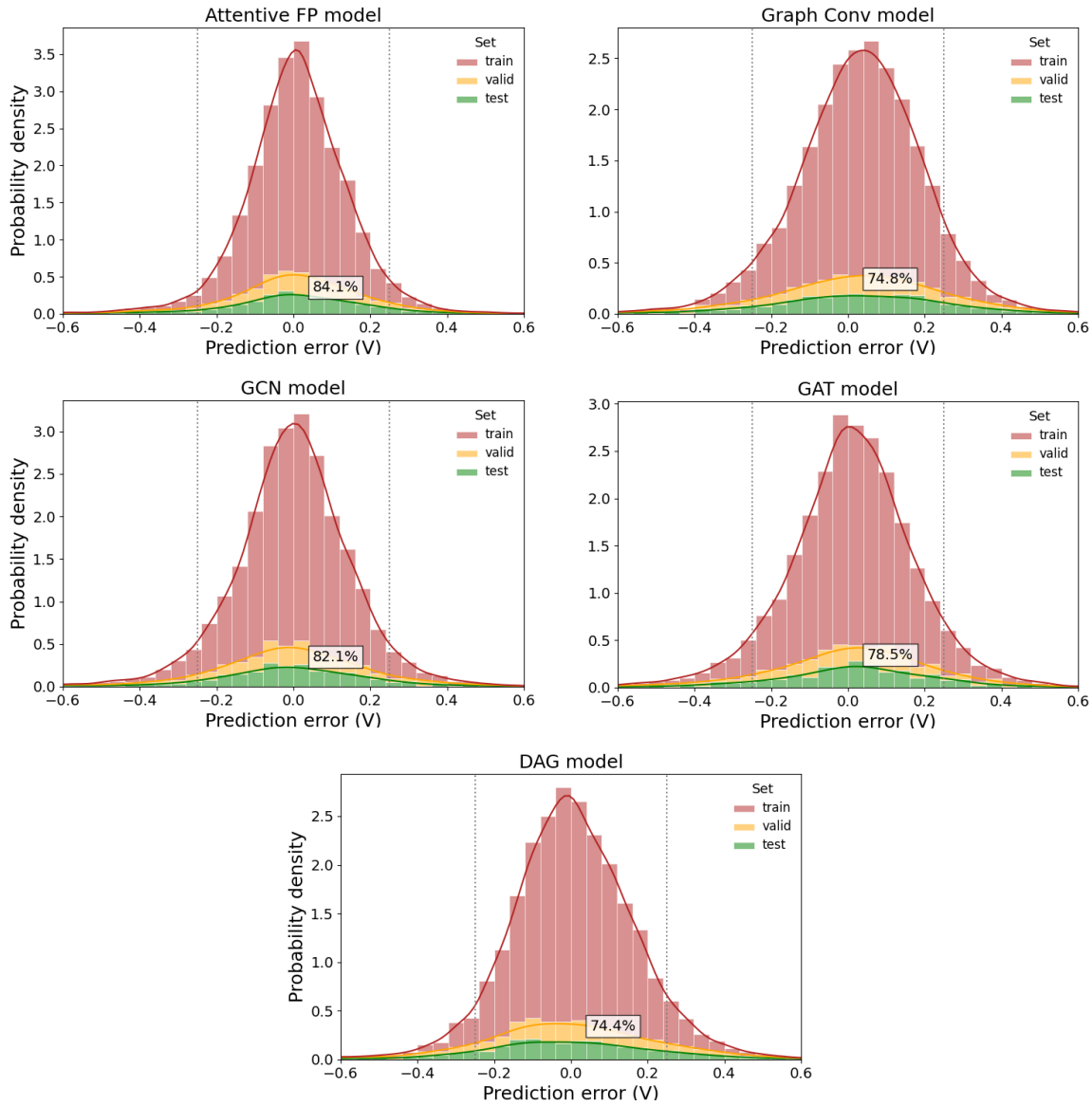


Figure S8: Distribution of signed errors for predictions using the models: *Attentive FP*, *Graph Conv*, *GCN*, *GAT*, and *DAG* with bars colored by data split (train, valid and test). Overlaid kerned density estimate curves provide a smoothed view of the error distribution for each set. Percentage value indicates the fraction of test molecules with errors within the -0.25 to 0.25 V range.

S7. The *PredPotS* web tool

The online application was implemented by using the dash (version 2.17.0) and plotly (version 5.22.0) python program packages. The trained models are loaded in the background and used to make predictions for the SMILES code(s) uploaded by the user. The program is hosted on a local webserver at http://predpots.ttk.hu/.

Herein we provide a detailed description of the usage of *PredPotS* according to the tabs displayed on the main page.

Predict

The predicted one-electron reduction potentials for molecules of interest can be calculated directly from the main page of the application. Multiple input options are available: users can paste a single SMILES

code or a semicolon-separated list of SMILES codes into the first input box. Alternatively, a CSV file containing a list of SMILES codes can be uploaded using the second input box. If needed, SMILES codes can be generated via the link provided in the third box of the application, where users can draw the molecule and obtain the corresponding SMILES code as input for the prediction (Figure S9).

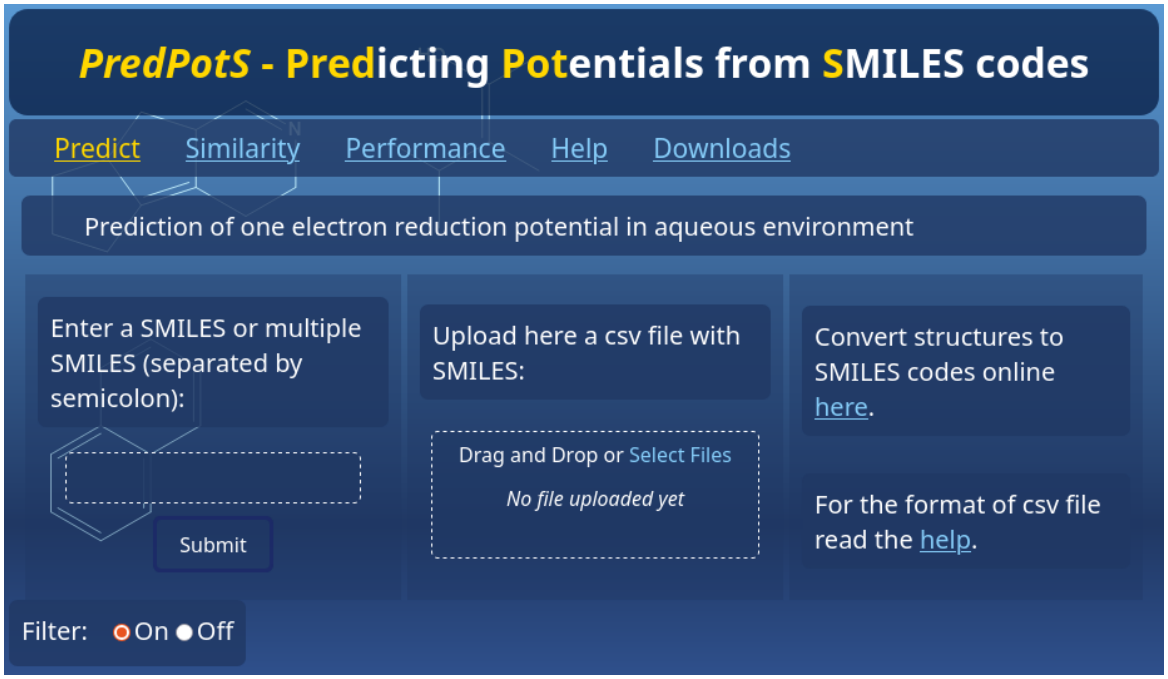


Figure S9. The main page of *PredPotS*.

Upon submitting an input, the web application displays the following information in a table:

- The ***smiles*** of the compound
- The ***mean prediction*** of the reduction potential
- The associated ***prediction confidence***
- The individual predictions from each of the models: *Attentive FP*, *GraphConv*, *GCN*, *GAT* and *DAG*

The ***mean prediction*** is calculated using a default *Z-score filter* set to *On*, which corresponds to a Z-score threshold of 1.5. This means that only predictions from models whose values lie within 1.5 standard deviations (±1.5 V) of the mean of all five predicted potentials are included in the final averaged value. This filtering helps to reduce the impact of potential outliers among the model outputs, however it can be switched O*ff* by the user.

The ***prediction confidence*** is derived from the 95% confidence interval (CI) around the filtered predictions, calculated using the t-distribution. To make this information more user-friendly, the width of the CI is mapped onto a ***star scale*** ranging from 1-5, where a narrower CI (i.e., higher agreement between models) results in a higher confidence rating. The ratings are illustrated in Table S9. In addition to the star rating, up to four types of warnings may appear in the prediction confidence column to highlight potential issues with the input molecule which might lead to poor prediction performances:

- *w1* indicates that the molecule contains fewer than six heavy atoms
- *w2* flags molecules that are not a single covalently bonded entity
- *w3* denotes molecules containing four or more aromatic rings
- *w4* denotes molecules with molecular weight M > 300 g/mol

Table S9. The relation between the width of the confidence interval and the star scale ratings.

CI values	Rating[a]
CI < 0.2	⋆⋆⋆⋆⋆
0.2 < CI < 0.3	⋆⋆⋆⋆
0.3 < CI < 0.4	⋆⋆⋆
0.4 < CI < 0.6	⋆⋆
CI > 0.6	⋆

[a] Additionally, if the number of outlier predictions is equal or greater than two the rating automatically drops to ⋆.

The table displayed by *PredPotS* can be exported from the application to a comma-separated values CSV type of file for further analysis and documentation. By using the radio button in the first column of the table the Lewis structure of the molecule is displayed below the table. An illustrated example is shown in Figure S10.

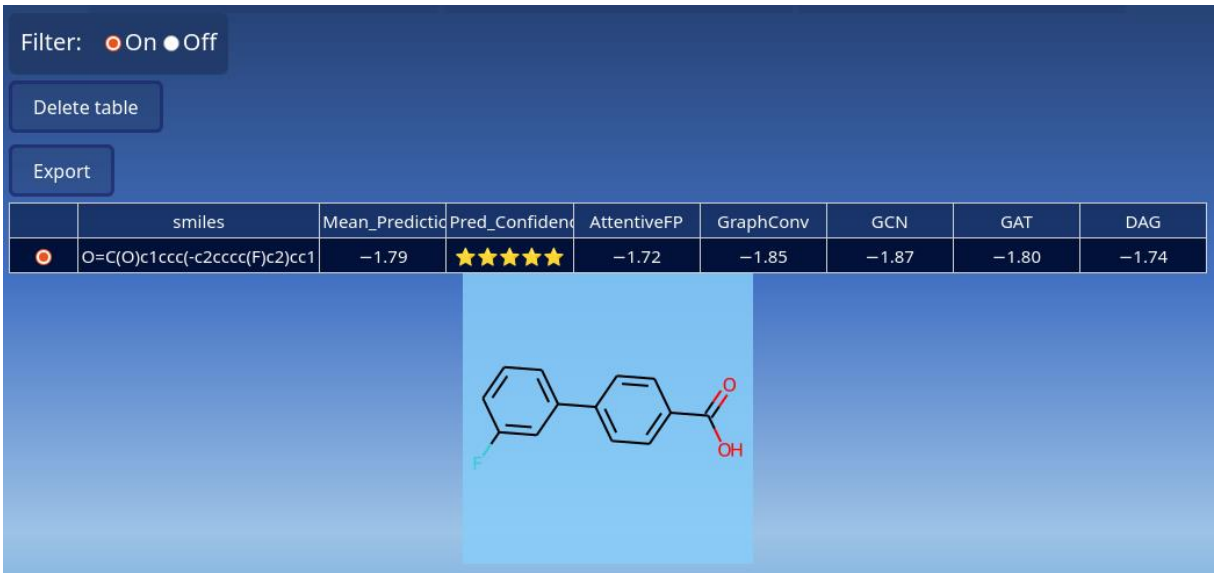


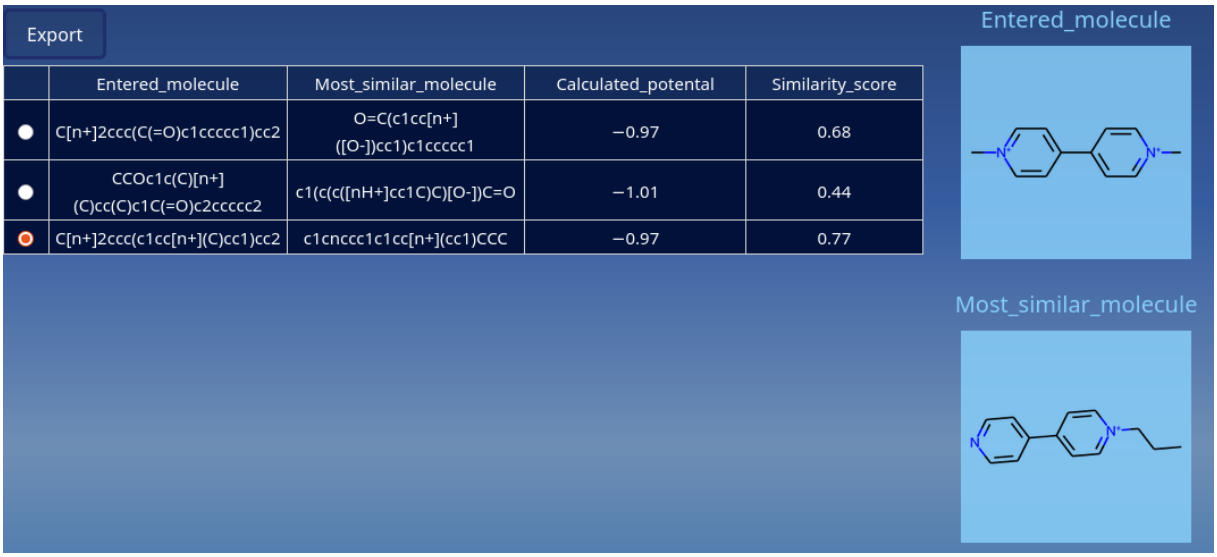
Figure S10. The table for the predicted reduction potential of the molecule represented by the SMILES code: O=C(O)c1ccc(-c2cccc(F)c2)cc1 and the Lewis structure displayed.

Similarity

The *Similarity* tab displays the most similar molecule(s) in the underlying computed Rp-ChEMBL database to the inserted or uploaded SMILES. The result is displayed in a table that contains the SMILES codes of the entered and the most similar molecule, the potential from the RP-ChEMBL database, and the similarity score. The latter is calculated by the RDKit[7] (version 2022.03.5) program package, according to the Tanimoto coefficient formula:

$$T(Q,S) = \frac{|Q \cap S|}{|Q| + |S| - |Q \cap S|}$$

where Q and S are the unfolded fingerprint vectors computed by the RDKit topological algorithm for the query and the most similar molecule from the dataset, respectively. For the visualization of the structures use the radio button from the first column (Figure S11).

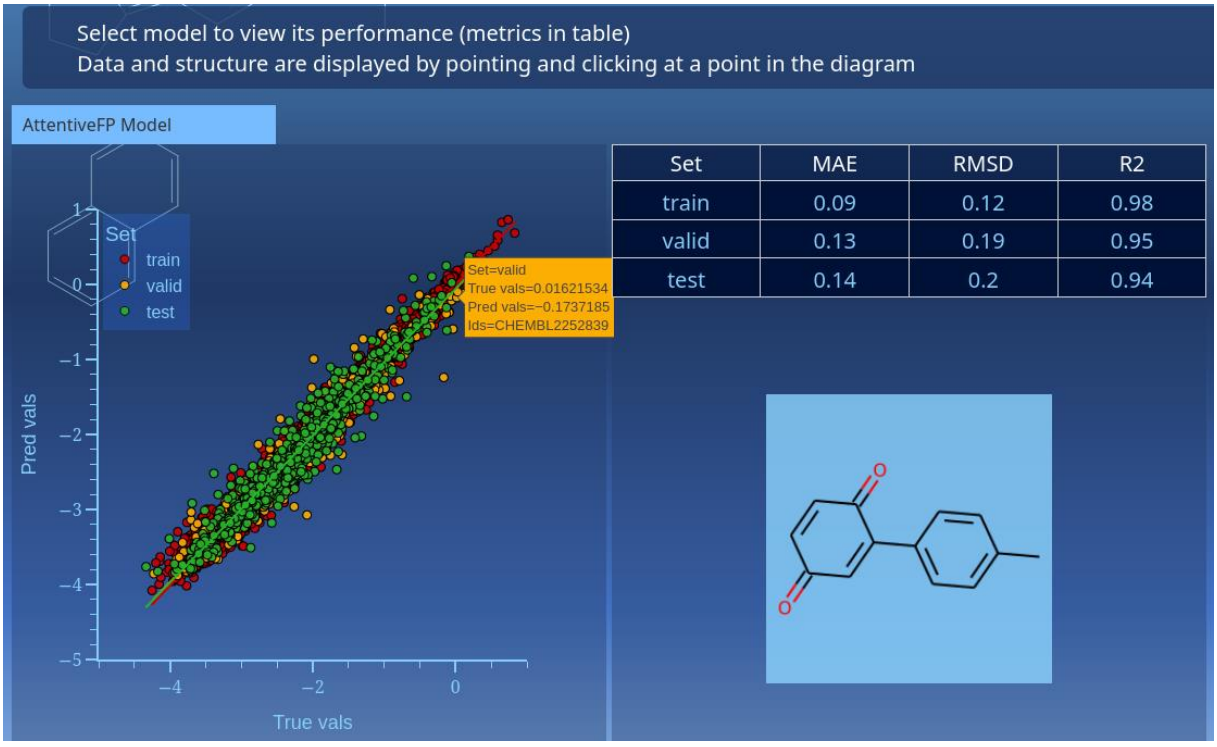


	Entered_molecule	Most_similar_molecule	Calculated_potental	Similarity_score
○	C[n+]2ccc(C(=O)c1ccccc1)cc2	O=C(c1cc[n+]([O-])cc1)c1ccccc1	−0.97	0.68
○	CCOc1c(C)[n+](C)cc(C)c1C(=O)c2ccccc2	c1(c(c([nH+]cc1C)C)[O-])C=O	−1.01	0.44
◉	C[n+]2ccc(c1cc[n+](C)cc1)cc2	c1cnccc1c1cc[n+](cc1)CCC	−0.97	0.77

Figure S11. Similarity search in the RP-ChEMBL dataset for the submitted SMILES codes: C[n+]2ccc(C(=O)c1ccccc1)cc2 (**1**$^{+}$); CCOc1c(C)[n+](C)cc(C)c1C(=O)c2ccccc2 (**2**$^{+}$); C[n+]2ccc(c1cc[n+](C)cc1)cc2 (**3**$^{2+}$). The Lewis structure of **3**$^{2+}$ and the most similar molecule is shown at right.

Performance

The accuracy and error scores for the different models used for predicting the reduction potential can be found under the *Performance* Tab. The performance of the different models can be visualized by using the dropdown menu in the upper left corner. The parity plot of the computed against the predicted potential values of the trained models is shown in an interactive manner. Hovering with the mouse over the points of the diagram will show the detailed data for the corresponding molecule and clicking on the selected point reveals the structure of the molecule (see Figure S12).



Set	MAE	RMSD	R2
train	0.09	0.12	0.98
valid	0.13	0.19	0.95
test	0.14	0.2	0.94

Figure S12. The performance of *Attentive FP* model is displayed.

Help

A short presentation, citation and funding of *PredPotS* and the usage of the web-tool with point-by-point guidance is described in the *Help* menu.

Dataset

A curated list of datasets used and referenced in this web application and the related paper (ref) is available for download. These datasets can be accessed directly from the *Datasets* section of the interface. The available datasets include:

- RP-ChEMBL dataset – ids, SMILES, computed and predicted potentials and distribution into train-valid-test sets
- RP-ChEMBL-ext extended dataset – ids, SMILES and computed potentials
- Wardman experimental dataset – ids, SMILES and experimental potentials

Each file is provided in CSV format and are freely available.

S8. Wardman compilation

Parity plots of predicted $1e^-$ reduction potentials (mean values an predictions using the five deep learning methods) vs. the experimental data from the Wardman compilation ($M < 200$ set) are shown in Figure S13.

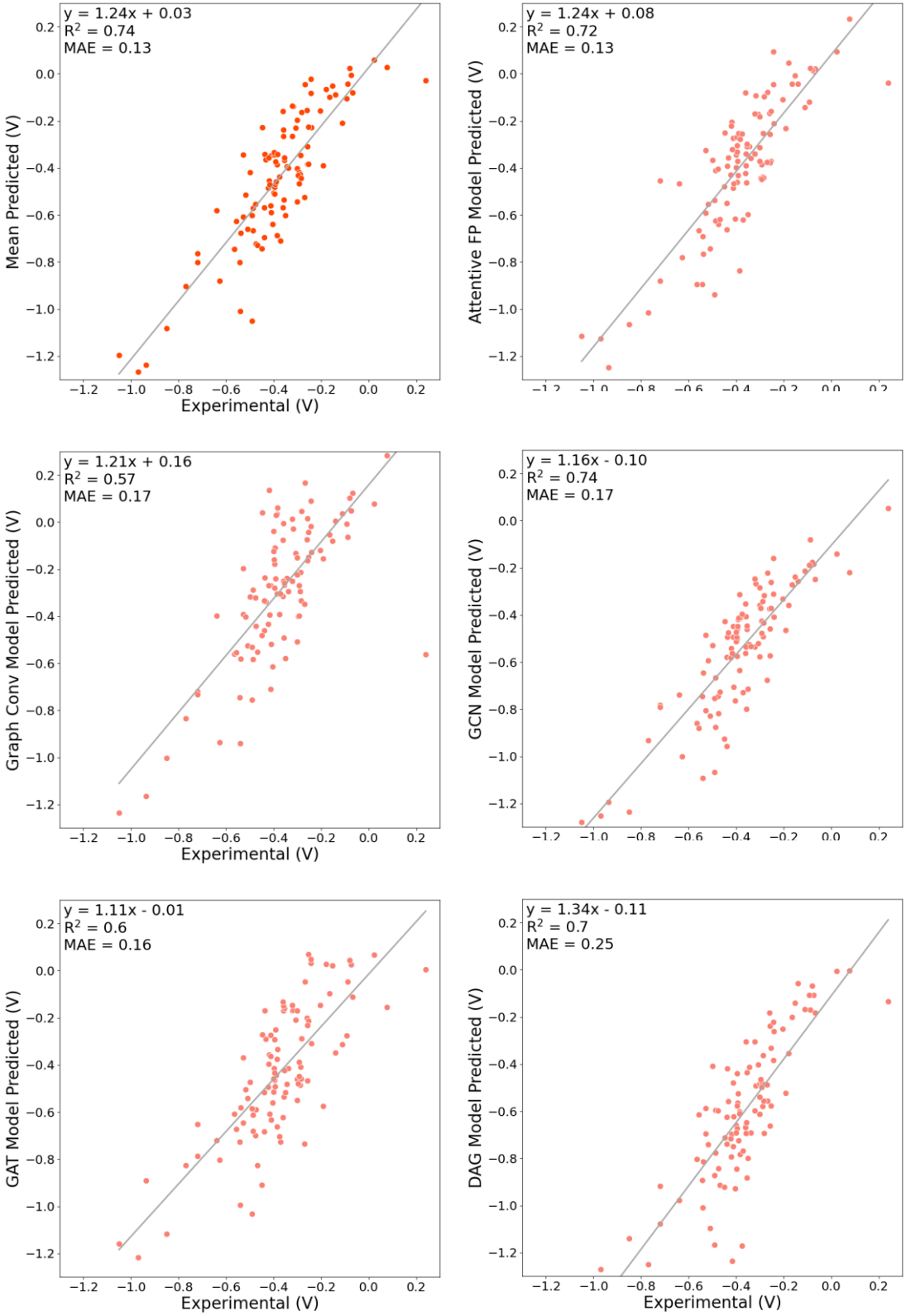


Figure S13. Parity plots of predicted vs. experimental 1e⁻ reduction potentials for the $M < 200$ set of Wardman compilation. Top-left plot corresponds to mean predictions, the others to those from particular model predictions as indicated on the y axis.

S9. Application to larger molecules

To evaluate the applicability of the *PredPotS* to larger molecules, we downloaded about 300k structures from the ChEMBL library with molecular weight in the $200 < M < 300$ range and used the *PredPotS* application to predict their $1e^-$ reduction potentials. From this set a subset of 2000 molecules was selected such that the distribution of the mean predicted potentials matched the shape of the distribution observed for the RP-ChEMBL dataset. Subsequently, the $1e^-$ reduction potentials of these selected molecules were computed using the GFN2-xTB/M06-2X protocol. Following data filtering, the resulting RP-ChEMBL-ext dataset comprises 1719 entries, each including the ChEMBL ID of the molecule, SMILES code, charge and $1e^-$ standard reduction potential computed with the protocol.

Parity plots of predicted $1e^-$ reduction potentials using the different deep learning methods vs. the values computed with the GFN2-xTB/M06-2X protocol for the RP-ChEMBL-ext dataset are shown in Figure S14.

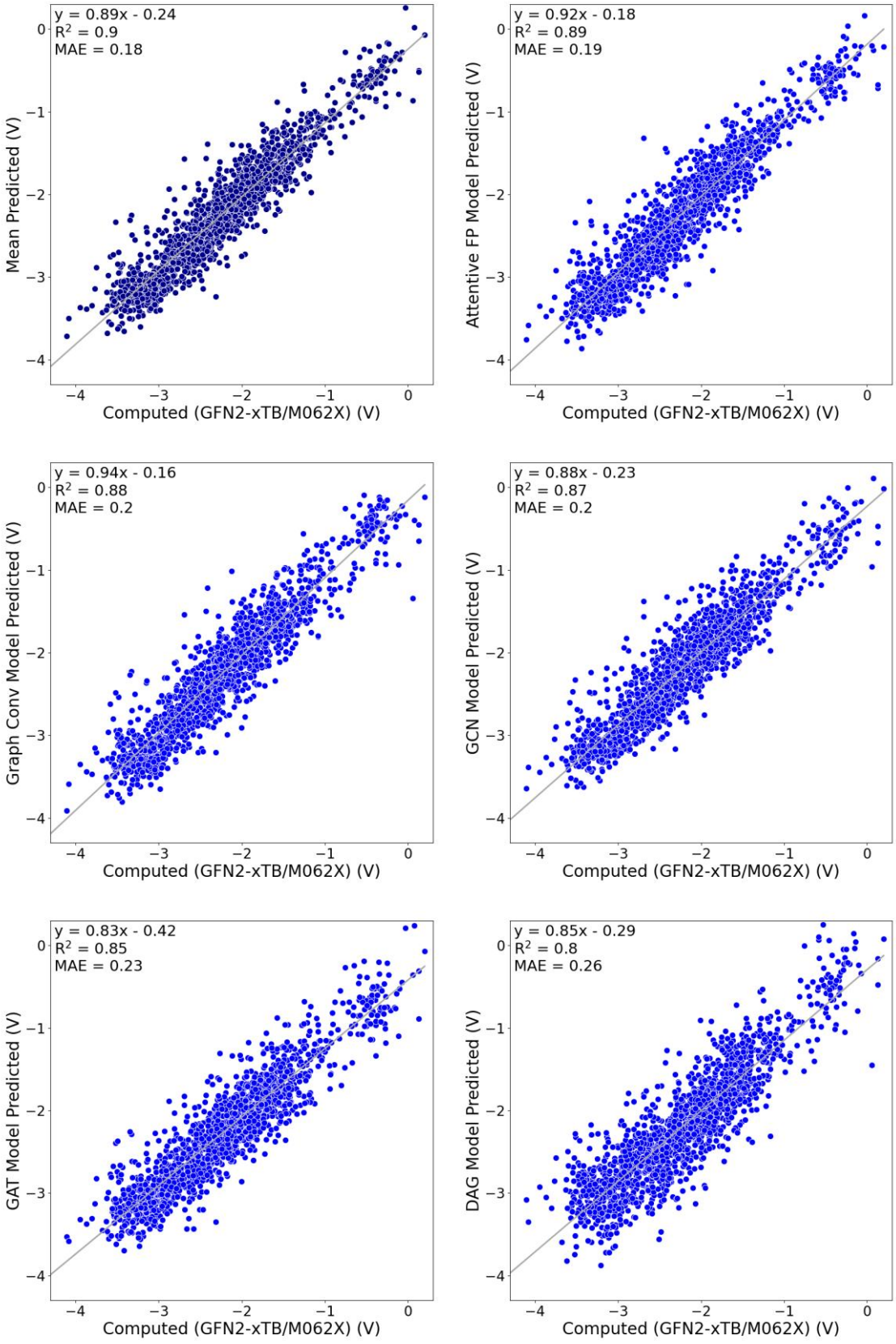


Figure S14. Parity plots of predicted 1e⁻ reduction potentials vs. the values computed with the GFN2-xTB/M06-2X protocol for a set molecules from the RP-ChEMBL-ext (molecular weight 200 < *M* < 300 g/mol). Top-left plot corresponds to mean predictions, the others to those from particular model predictions as indicated on the y axis.

References

[1] *DeepChem:* a) https://deepchem.io/; b) B. Ramsundar, P. Eastman, P, Walters, V. Pande, *Deep Learning for the Life Sciences*, O'Reilly, Sebastopol, 2019.

[2] Thomas N. Kipf and Max Welling. "Semi-Supervised Classification with Graph Convolutional Networks." ICLR 2017.

[3] Duvenaud, David K., et al. "Convolutional networks on graphs for learning molecular fingerprints." Advances in neural information processing systems. 2015.

[4] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. "Graph Attention Networks." ICLR 2018.

[5] Zhaoping Xiong, Dingyan Wang, Xiaohong Liu, Feisheng Zhong, Xiaozhe Wan, Xutong Li, Zhaojun Li, Xiaomin Luo, Kaixian Chen, Hualiang Jiang, and Mingyue Zheng. "Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism." Journal of Medicinal Chemistry. 2020, 63, 16, 8749–8760.

[6] Lusci, Alessandro, Gianluca Pollastri, and Pierre Baldi. "Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules." Journal of chemical information and modeling 53.7 (2013): 1563-1575.

[7] RDKit: Open-source cheminformatics. https://www.rdkit.org Release version: https://github.com/rdkit/rdkit/tree/Release_2022_03_5