

Supplementary Information

February 15, 2019

S1 Additional Dataset Statistics

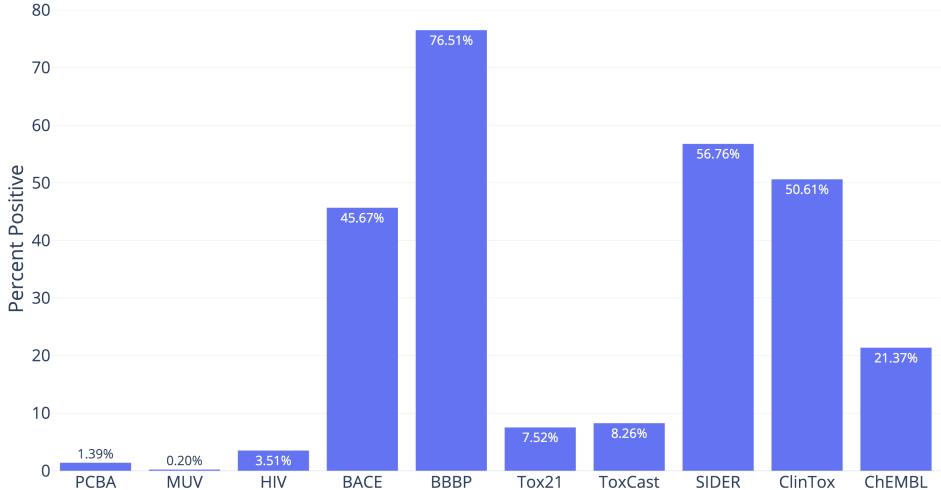


Figure S1: Class balance on the publicly available classification datasets. Values are the weighted average of the percent of positives in each task in a dataset.

S2 Comparison to Baselines

S2.1 Comparison to MoleculeNet

Comparison between our best single model (i.e. optimized hyperparameters and optionally RDKit features but without ensembling) and the best model from MoleculeNet using the splits from MoleculNet[?]. We were unable to reproduce the splits from MoleculeNet on QM7, BACE, and ToxCast, so we leave out those datasets. The QM8, QM9, and PDBbind datasets include 3D coordinates that our model does not use but some MoleculeNet models may use.

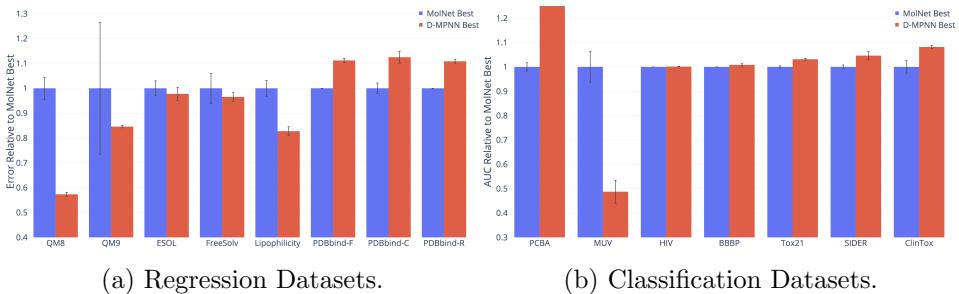


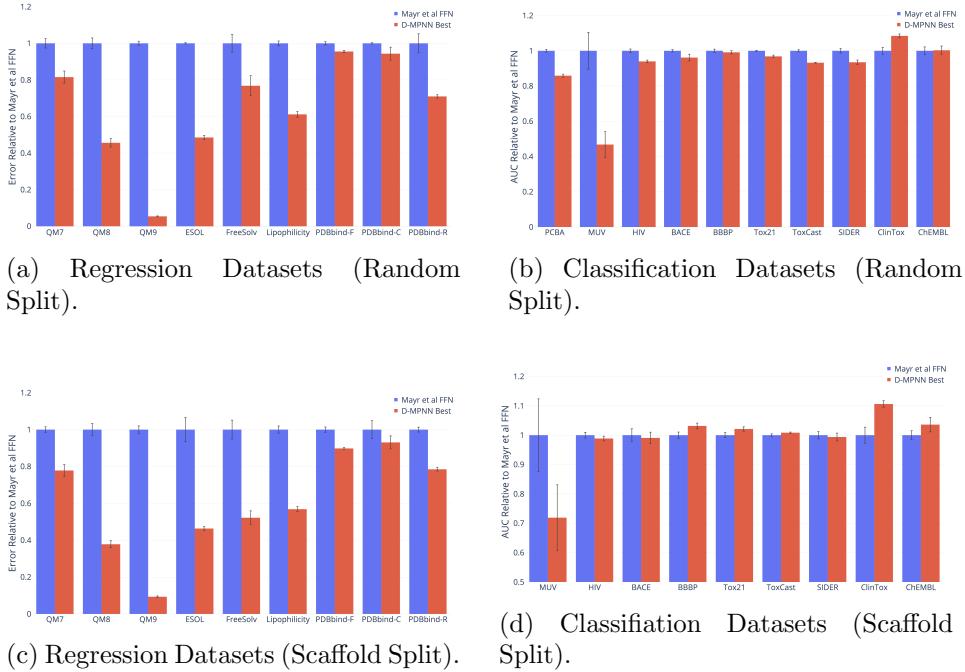
Figure S2: Comparison to MoleculeNet.

Dataset	Metric	Split Type	MolNet Best	D-MPNN Best
QM8	MAE	Random	0.0143 ± 0.0011	0.0082 ± 0.0002 (-42.66%)
QM9	MAE	Random	2.400 ± 1.100	2.030 ± 0.021 (-15.42%)
ESOL	RMSE	Random	0.580 ± 0.030	0.567 ± 0.026 (-2.24%)
FreeSolv	RMSE	Random	1.150 ± 0.120	1.110 ± 0.035 (-3.48%)
Lipophilicity	RMSE	Random	0.655 ± 0.036	0.542 ± 0.020 (-17.25%)
PDBbind-F	RMSE	Time	1.250 ± 0.000	1.390 ± 0.017 (+11.20%)
PDBbind-C	RMSE	Time	1.920 ± 0.070	2.160 ± 0.080 (+12.50%)
PDBbind-R	RMSE	Time	1.380 ± 0.000	1.530 ± 0.018 (+10.87%)
PCBA	PRC-AUC	Random	0.136 ± 0.004	0.397 ± 0.001 (+191.91%)
MUV	PRC-AUC	Random	0.1840 ± 0.0200	0.090 ± 0.015 (-51.25%)
HIV	ROC-AUC	Scaffold	0.792 ± 0.000	0.793 ± 0.001 (+0.13%)
BBBP	ROC-AUC	Scaffold	0.729 ± 0.000	0.735 ± 0.006 (+0.82%)
Tox21	ROC-AUC	Random	0.829 ± 0.006	0.855 ± 0.005 (+3.14%)
SIDER	ROC-AUC	Random	0.648 ± 0.009	0.678 ± 0.019 (+4.63%)
ClinTox	ROC-AUC	Random	0.832 ± 0.037	0.900 ± 0.009 (+8.17%)

Table S1: Comparison to MoleculeNet.

S2.2 Comparison to Mayr et al.

Comparison between our best single model (i.e. optimized hyperparameters and optionally RDKit features but without ensembling) and the feed-forward network (FFN) architecture of Mayr et al[?]. PCBA is left out because the method of mayr2018chembl was unstable on that dataset.



Dataset	Metric	Mayr et al FFN	D-MPNN Best
QM7	11.494	108.100 ± 8.750	60.106 ± 2.250 (-44.40%)
QM8	0.002092	0.029 ± 0.003	0.008 ± 0.000 (-71.38%)
QM9	0.374544	45.400 ± 1.460	1.992 ± 0.017 (-95.61%)
ESOL	0.061	1.700 ± 0.021	0.620 ± 0.041 (-63.53%)
FreeSolv	0.486	2.800 ± 0.430	1.027 ± 0.211 (-63.32%)
Lipophilicity	0.049	1.020 ± 0.041	0.557 ± 0.058 (-45.39%)
PDBbind-F	0.028	1.410 ± 0.041	1.287 ± 0.027 (-8.72%)
PDBbind-C	0.243601	2.170 ± 0.025	1.933 ± 0.143 (-10.92%)
PDBbind-R	0.056603	1.990 ± 0.330	1.325 ± 0.060 (-33.40%)
MUV	0.038943	0.169 ± 0.056	0.126 ± 0.056 (-25.68%)
HIV	0.020414	0.850 ± 0.027	0.827 ± 0.029 (-2.69%)
BACE	0.053	0.886 ± 0.019	0.889 ± 0.024 (+0.34%)
BBBP	0.027	0.927 ± 0.027	0.931 ± 0.016 (+0.43%)
Tox21	0.019	0.853 ± 0.008	0.862 ± 0.008 (+1.06%)
ToxCast	0.005	0.770 ± 0.015	0.769 ± 0.015 (-0.13%)
SIDER	0.027	0.676 ± 0.027	0.658 ± 0.018 (-2.66%)
ClinTox	0.029	0.827 ± 0.050	0.903 ± 0.032 (+9.19%)
ChEMBL	0.031382	0.755 ± 0.029	0.767 ± 0.035 (+1.58%)

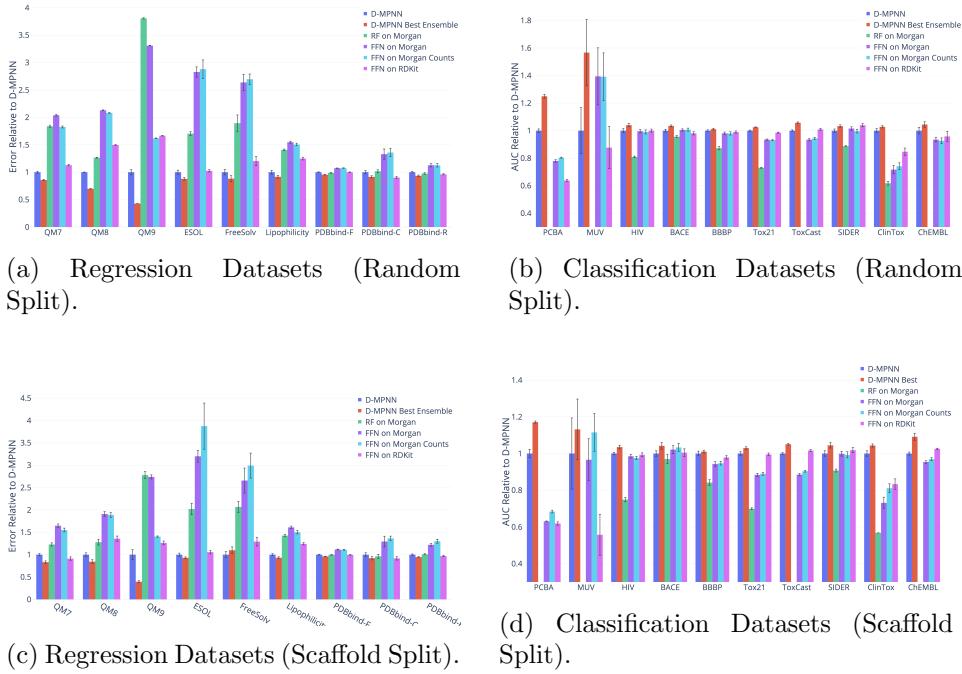
Table S2: Comparison to Mayr et al (Random Split).

Dataset	Metric	Mayr et al FFN	D-MPNN Best
QM7	MAE	113.250 ± 5.440	88.136 ± 11.494 (-22.18%)
QM8	MAE	0.035 ± 0.004	0.013 ± 0.002 (-62.19%)
QM9	MAE	26.400 ± 1.680	2.465 ± 0.375 (-90.66%)
ESOL	RMSE	1.780 ± 0.370	0.825 ± 0.061 (-53.65%)
FreeSolv	RMSE	4.120 ± 0.670	2.151 ± 0.486 (-47.79%)
Lipophilicity	RMSE	1.098 ± 0.068	0.624 ± 0.049 (-43.17%)
PDBbind-F	RMSE	1.500 ± 0.068	1.347 ± 0.028 (-10.20%)
PDBbind-C	RMSE	2.200 ± 0.337	2.047 ± 0.244 (-6.94%)
PDBbind-R	RMSE	1.800 ± 0.076	1.413 ± 0.057 (-21.52%)
MUV	PRC-AUC	0.110 ± 0.043	0.079 ± 0.039 (-28.10%)
HIV	ROC-AUC	0.808 ± 0.025	0.799 ± 0.020 (-1.14%)
BACE	ROC-AUC	0.860 ± 0.059	0.852 ± 0.053 (-0.93%)
BBBP	ROC-AUC	0.891 ± 0.030	0.919 ± 0.027 (+3.14%)
Tox21	ROC-AUC	0.809 ± 0.022	0.826 ± 0.019 (+2.10%)
ToxCast	ROC-AUC	0.712 ± 0.010	0.718 ± 0.005 (+0.84%)
SIDER	ROC-AUC	0.636 ± 0.025	0.632 ± 0.027 (-0.63%)
ClinTox	ROC-AUC	0.811 ± 0.070	0.897 ± 0.029 (+10.60%)
ChEMBL	ROC-AUC	0.731 ± 0.019	0.757 ± 0.031 (+3.59%)

Table S3: Comparison to Mayr et al (Scaffold Split).

S2.3 Comparison to Other Baselines

Comparison to several baselines using feed-forward neural networks on molecular fingerprints/descriptors.



Dataset	Metric	D-MPNN	D-MPNN Best Ensemble
QM7	MAE	67.657 ± 3.647	57.952 ± 2.674 (-14.35%)
QM8	MAE	0.0111 ± 0.0001	0.0077 ± 0.0001 (-30.23%)
QM9	MAE	3.705 ± 0.537	1.585 ± 0.011 (-57.22%)
ESOL	RMSE	0.691 ± 0.080	0.607 ± 0.051 (-12.13%)
FreeSolv	RMSE	1.133 ± 0.176	0.999 ± 0.198 (-11.81%)
Lipophilicity	RMSE	0.598 ± 0.064	0.548 ± 0.050 (-8.38%)
PDBbind-F	RMSE	1.325 ± 0.027	1.264 ± 0.026 (-4.61%)
PDBbind-C	RMSE	2.088 ± 0.194	1.910 ± 0.137 (-8.55%)
PDBbind-R	RMSE	1.411 ± 0.060	1.318 ± 0.063 (-6.62%)
PCBA	PRC-AUC	0.329 ± 0.007	0.411 ± 0.008 (+24.96%)
MUV	PRC-AUC	0.0865 ± 0.0456	0.1356 ± 0.0660 (+56.71%)
HIV	ROC-AUC	0.811 ± 0.037	0.843 ± 0.030 (+3.96%)
BACE	ROC-AUC	0.863 ± 0.028	0.893 ± 0.019 (+3.45%)
BBBP	ROC-AUC	0.920 ± 0.017	0.930 ± 0.017 (+1.07%)
Tox21	ROC-AUC	0.847 ± 0.012	0.868 ± 0.008 (+2.46%)
ToxCast	ROC-AUC	0.735 ± 0.014	0.777 ± 0.014 (+5.73%)
SIDER	ROC-AUC	0.639 ± 0.024	0.660 ± 0.018 (+3.27%)
ClinTox	ROC-AUC	0.892 ± 0.043	0.917 ± 0.023 (+2.77%)

Table S4: Comparison to Baselines, Part I (Random Split).

Dataset	RF on Morgan	FFN on Morgan
QM7	$124.376 \pm 4.278 (+83.83\%)$	$137.932 \pm 3.420 (+103.87\%)$
QM8	$0.0140 \pm 0.0003 (+26.37\%)$	$0.0236 \pm 0.0003 (+112.89\%)$
QM9	$14.104 \pm 0.125 (+280.65\%)$	$12.258 \pm 0.058 (+230.82\%)$
ESOL	$1.176 \pm 0.076 (+70.25\%)$	$1.953 \pm 0.205 (+182.78\%)$
FreeSolv	$2.146 \pm 0.538 (+89.40\%)$	$2.989 \pm 0.533 (+163.76\%)$
Lipophilicity	$0.842 \pm 0.026 (+40.84\%)$	$0.925 \pm 0.035 (+54.74\%)$
PDBbind-F	$1.309 \pm 0.027 (-1.23\%)$	$1.422 \pm 0.021 (+7.28\%)$
PDBbind-C	$2.124 \pm 0.181 (+1.69\%)$	$2.776 \pm 0.617 (+32.92\%)$
PDBbind-R	$1.377 \pm 0.062 (-2.40\%)$	$1.590 \pm 0.138 (+12.63\%)$
PCBA	$0.000 \pm 0.000 (-100.00\%)$	$0.256 \pm 0.005 (-22.02\%)$
MUV	$0.0000 \pm 0.0000 (-100.00\%)$	$0.1207 \pm 0.0567 (+39.43\%)$
HIV	$0.656 \pm 0.015 (-19.07\%)$	$0.807 \pm 0.031 (-0.44\%)$
BACE	$0.825 \pm 0.021 (-4.39\%)$	$0.868 \pm 0.022 (+0.52\%)$
BBBP	$0.803 \pm 0.036 (-12.71\%)$	$0.902 \pm 0.027 (-1.94\%)$
Tox21	$0.619 \pm 0.010 (-26.96\%)$	$0.792 \pm 0.012 (-6.53\%)$
ToxCast	$0.000 \pm 0.000 (-100.00\%)$	$0.687 \pm 0.018 (-6.54\%)$
SIDER	$0.567 \pm 0.008 (-11.29\%)$	$0.649 \pm 0.026 (+1.54\%)$
ClinTox	$0.551 \pm 0.039 (-38.23\%)$	$0.639 \pm 0.085 (-28.33\%)$

Table S5: Comparison to Baselines, Part II (Random Split).

Dataset	FFN on Morgan Counts	FFN on RDKit
QM7	$123.449 \pm 3.806 (+82.46\%)$	$76.300 \pm 2.828 (+12.77\%)$
QM8	$0.0231 \pm 0.0002 (+108.16\%)$	$0.0166 \pm 0.0002 (+49.67\%)$
QM9	$6.002 \pm 0.053 (+61.97\%)$	$6.170 \pm 0.060 (+66.53\%)$
ESOL	$1.988 \pm 0.370 (+187.85\%)$	$0.707 \pm 0.047 (+2.43\%)$
FreeSolv	$3.054 \pm 0.345 (+169.48\%)$	$1.363 \pm 0.296 (+20.32\%)$
Lipophilicity	$0.900 \pm 0.044 (+50.57\%)$	$0.746 \pm 0.046 (+24.83\%)$
PDBbind-F	$1.423 \pm 0.030 (+7.37\%)$	$1.325 \pm 0.024 (-0.05\%)$
PDBbind-C	$2.837 \pm 0.483 (+35.86\%)$	$1.883 \pm 0.124 (-9.81\%)$
PDBbind-R	$1.589 \pm 0.140 (+12.57\%)$	$1.357 \pm 0.048 (-3.87\%)$
PCBA	$0.264 \pm 0.002 (-19.64\%)$	$0.210 \pm 0.004 (-36.24\%)$
MUV	$0.1204 \pm 0.0474 (+39.16\%)$	$0.0759 \pm 0.0418 (-12.34\%)$
HIV	$0.803 \pm 0.038 (-0.96\%)$	$0.811 \pm 0.028 (+0.03\%)$
BACE	$0.870 \pm 0.030 (+0.73\%)$	$0.847 \pm 0.031 (-1.94\%)$
BBBP	$0.901 \pm 0.035 (-2.03\%)$	$0.911 \pm 0.023 (-1.02\%)$
Tox21	$0.789 \pm 0.010 (-6.85\%)$	$0.834 \pm 0.010 (-1.48\%)$
ToxCast	$0.694 \pm 0.017 (-5.63\%)$	$0.742 \pm 0.015 (+0.91\%)$
SIDER	$0.637 \pm 0.027 (-0.39\%)$	$0.664 \pm 0.024 (+3.96\%)$
ClinTox	$0.662 \pm 0.070 (-25.85\%)$	$0.755 \pm 0.075 (-15.33\%)$

Table S6: Comparison to Baselines, Part III (Random Split).

Dataset	Metric	D-MPNN	D-MPNN Best Ensemble
QM7	MAE	102.302 ± 7.532	85.173 ± 9.300 (-16.74%)
QM8	MAE	0.0148 ± 0.0019	0.0125 ± 0.0021 (-15.66%)
QM9	MAE	5.177 ± 1.848	2.043 ± 0.391 (-60.54%)
ESOL	RMSE	0.862 ± 0.094	0.801 ± 0.057 (-7.07%)
FreeSolv	RMSE	1.932 ± 0.412	2.122 ± 0.460 (+9.83%)
Lipophilicity	RMSE	0.656 ± 0.047	0.611 ± 0.045 (-6.92%)
PDBbind-F	RMSE	1.374 ± 0.028	1.319 ± 0.024 (-4.05%)
PDBbind-C	RMSE	2.154 ± 0.290	1.984 ± 0.254 (-7.87%)
PDBbind-R	RMSE	1.468 ± 0.075	1.387 ± 0.057 (-5.51%)
PCBA	PRC-AUC	0.271 ± 0.011	0.317 ± 0.003 (+17.09%)
MUV	PRC-AUC	0.0807 ± 0.0494	0.0913 ± 0.0421 (+13.15%)
HIV	ROC-AUC	0.782 ± 0.016	0.809 ± 0.024 (+3.50%)
BACE	ROC-AUC	0.825 ± 0.039	0.860 ± 0.048 (+4.19%)
BBBP	ROC-AUC	0.915 ± 0.037	0.925 ± 0.023 (+1.03%)
Tox21	ROC-AUC	0.808 ± 0.025	0.832 ± 0.018 (+2.97%)
ToxCast	ROC-AUC	0.690 ± 0.011	0.725 ± 0.009 (+5.03%)
SIDER	ROC-AUC	0.606 ± 0.029	0.633 ± 0.029 (+4.47%)
ClinTox	ROC-AUC	0.874 ± 0.041	0.913 ± 0.024 (+4.37%)
ChEMBL	ROC-AUC	0.709 ± 0.010	0.774 ± 0.024 (+9.12%)

Table S7: Comparison to Baselines, Part I (Scaffold Split).

Dataset	RF on Morgan	FFN on Morgan
QM7	$125.661 \pm 11.322 (+22.83\%)$	$168.496 \pm 12.789 (+64.71\%)$
QM8	$0.0190 \pm 0.0030 (+27.91\%)$	$0.0283 \pm 0.0025 (+90.90\%)$
QM9	$14.392 \pm 1.306 (+178.01\%)$	$14.177 \pm 0.743 (+173.85\%)$
ESOL	$1.741 \pm 0.350 (+102.06\%)$	$2.757 \pm 0.360 (+219.95\%)$
FreeSolv	$3.994 \pm 0.751 (+106.70\%)$	$5.133 \pm 1.706 (+165.68\%)$
Lipophilicity	$0.936 \pm 0.057 (+42.63\%)$	$1.056 \pm 0.061 (+60.99\%)$
PDBbind-F	$1.370 \pm 0.043 (-0.33\%)$	$1.533 \pm 0.035 (+11.52\%)$
PDBbind-C	$2.074 \pm 0.299 (-3.71\%)$	$2.785 \pm 0.816 (+29.33\%)$
PDBbind-R	$1.481 \pm 0.057 (+0.92\%)$	$1.791 \pm 0.149 (+22.02\%)$
PCBA	$0.000 \pm 0.000 (-100.00\%)$	$0.171 \pm 0.001 (-36.96\%)$
MUV	$0.0000 \pm 0.0000 (-100.00\%)$	$0.0780 \pm 0.0293 (-3.40\%)$
HIV	$0.586 \pm 0.029 (-25.12\%)$	$0.770 \pm 0.024 (-1.49\%)$
BACE	$0.801 \pm 0.065 (-3.00\%)$	$0.843 \pm 0.057 (+2.12\%)$
BBBP	$0.770 \pm 0.046 (-15.82\%)$	$0.863 \pm 0.039 (-5.70\%)$
Tox21	$0.565 \pm 0.013 (-30.08\%)$	$0.714 \pm 0.021 (-11.62\%)$
ToxCast	$0.000 \pm 0.000 (-100.00\%)$	$0.611 \pm 0.013 (-11.51\%)$
SIDER	$0.549 \pm 0.015 (-9.27\%)$	$0.605 \pm 0.025 (-0.12\%)$
ClinTox	$0.496 \pm 0.000 (-43.23\%)$	$0.639 \pm 0.085 (-26.87\%)$
ChEMBL	$0.000 \pm 0.000 (-100.00\%)$	$0.677 \pm 0.009 (-4.61\%)$

Table S8: Comparison to Baselines, Part II (Scaffold Split).

Dataset	FFN on Morgan Counts	FFN on RDKit
QM7	$158.701 \pm 13.798 (+55.13\%)$	$93.275 \pm 12.208 (-8.82\%)$
QM8	$0.0280 \pm 0.0024 (+88.68\%)$	$0.0201 \pm 0.0028 (+35.73\%)$
QM9	$7.271 \pm 0.228 (+40.45\%)$	$6.539 \pm 0.694 (+26.30\%)$
ESOL	$3.338 \pm 1.404 (+287.33\%)$	$0.910 \pm 0.106 (+5.63\%)$
FreeSolv	$5.779 \pm 1.707 (+199.11\%)$	$2.495 \pm 0.598 (+29.15\%)$
Lipophilicity	$0.986 \pm 0.080 (+50.23\%)$	$0.815 \pm 0.052 (+24.26\%)$
PDBbind-F	$1.524 \pm 0.034 (+10.91\%)$	$1.368 \pm 0.032 (-0.43\%)$
PDBbind-C	$2.939 \pm 0.341 (+36.45\%)$	$1.976 \pm 0.251 (-8.26\%)$
PDBbind-R	$1.908 \pm 0.205 (+30.01\%)$	$1.422 \pm 0.062 (-3.10\%)$
PCBA	$0.185 \pm 0.004 (-31.62\%)$	$0.168 \pm 0.004 (-38.12\%)$
MUV	$0.0900 \pm 0.0262 (+11.50\%)$	$0.0450 \pm 0.0285 (-44.28\%)$
HIV	$0.763 \pm 0.021 (-2.41\%)$	$0.777 \pm 0.026 (-0.71\%)$
BACE	$0.853 \pm 0.058 (+3.32\%)$	$0.829 \pm 0.054 (+0.48\%)$
BBBP	$0.866 \pm 0.029 (-5.31\%)$	$0.895 \pm 0.028 (-2.14\%)$
Tox21	$0.719 \pm 0.021 (-11.03\%)$	$0.804 \pm 0.019 (-0.48\%)$
ToxCast	$0.624 \pm 0.011 (-9.63\%)$	$0.702 \pm 0.012 (+1.60\%)$
SIDER	$0.601 \pm 0.032 (-0.69\%)$	$0.617 \pm 0.027 (+1.84\%)$
ClinTox	$0.709 \pm 0.066 (-18.89\%)$	$0.729 \pm 0.080 (-16.67\%)$
ChEMBL	$0.687 \pm 0.010 (-3.08\%)$	$0.727 \pm 0.004 (+2.56\%)$

Table S9: Comparison to Baselines, Part III (Scaffold Split).

S3 Proprietary Datasets

S3.1 Amgen

Comparison of our D-MPNN in both its unoptimized and optimized form against baseline models on Amgen internal datasets using a time split of the data. Note: the model was optimized on a random split of the data but evaluated on a time split of the data.

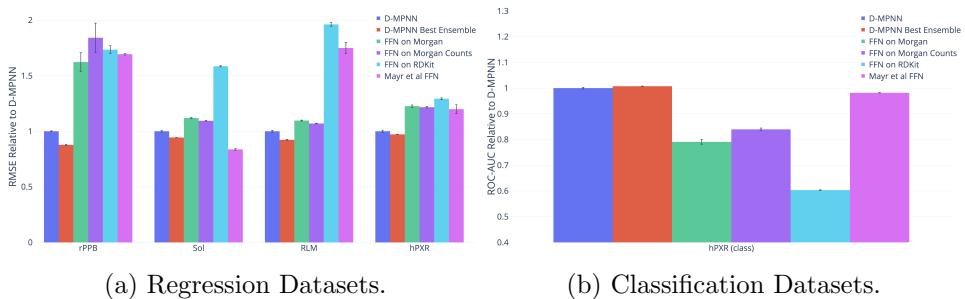


Figure S5: Comparison to Baselines on Amgen.

Dataset	D-MPNN	D-MPNN Best Ensemble	FFN on Morgan
rPPB	1.063 ± 0.014	$0.933 \pm 0.011 (-12.23\%)$	$1.726 \pm 0.282 (+62.33\%)$
Sol	0.710 ± 0.016	$0.670 \pm 0.002 (-5.61\%)$	$0.794 \pm 0.009 (+11.95\%)$
RLM	0.328 ± 0.008	$0.303 \pm 0.003 (-7.70\%)$	$0.360 \pm 0.004 (+9.65\%)$
hPXR	36.789 ± 1.164	$35.761 \pm 0.168 (-2.79\%)$	$45.113 \pm 1.201 (+22.63\%)$
hPXR (class)	0.841 ± 0.008	$0.848 \pm 0.002 (+0.74\%)$	$0.666 \pm 0.024 (-20.90\%)$

Table S10: Comparison to Baselines on Amgen, Part I. Note: hPXR (class) is ROC-AUC, all others are RMSE.

Dataset	FFN on Morgan Counts	FFN on RDKit	Mayr et al
rPPB	$1.957 \pm 0.444 (+84.14\%)$	$1.844 \pm 0.118 (+73.46\%)$	$1.800 \pm 0.023 (+69.33\%)$
Sol	$0.777 \pm 0.005 (+9.45\%)$	$1.125 \pm 0.014 (+58.54\%)$	$0.594 \pm 0.017 (-16.28\%)$
RLM	$0.351 \pm 0.003 (+6.97\%)$	$0.644 \pm 0.017 (+96.20\%)$	$0.574 \pm 0.052 (+74.94\%)$
hPXR	$44.753 \pm 0.774 (+21.65\%)$	$47.574 \pm 0.881 (+29.32\%)$	$44.100 \pm 4.590 (+19.87\%)$
hPXR (class)	$0.706 \pm 0.014 (-16.03\%)$	$0.508 \pm 0.003 (-39.66\%)$	$0.826 \pm 0.002 (-1.82\%)$

Table S11: Comparison to Baselines on Amgen, Part II. Note: hPXR (class) is ROC-AUC, all others are RMSE.

S3.2 Amgen Model Optimizations

Dataset	D-MPNN	D-MPNN Optimized
rPPB	1.063 ± 0.014	$1.104 \pm 0.109 (+3.88\%)$
Sol	0.710 ± 0.016	$0.696 \pm 0.018 (-1.89\%)$
RLM	0.328 ± 0.008	$0.320 \pm 0.006 (-2.40\%)$
hPXR	36.789 ± 1.164	$37.904 \pm 0.920 (+3.03\%)$
hPXR (class)	0.841 ± 0.008	$0.833 \pm 0.007 (-1.00\%)$

Table S12: Optimizations on Amgen, Part I. Note: hPXR (class) is ROC-AUC, all others are RMSE.

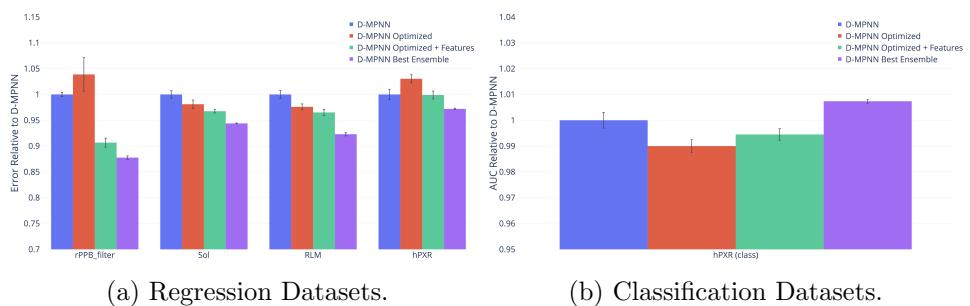


Figure S6: Optimizations on Amgen.

Dataset	D-MPNN Optimized + Features	D-MPNN Best Ensemble
rPPB	0.964 ± 0.029 (-9.33%)	0.933 ± 0.011 (-12.23%)
Sol	0.686 ± 0.008 (-3.25%)	0.670 ± 0.002 (-5.61%)
RLM	0.317 ± 0.006 (-3.48%)	0.303 ± 0.003 (-7.70%)
hPXR	36.751 ± 0.894 (-0.10%)	35.761 ± 0.168 (-2.79%)
hPXR (class)	0.837 ± 0.006 (-0.55%)	0.848 ± 0.002 (+0.74%)

Table S13: Optimizations on Amgen, Part I. Note: hPXR (class) is ROC-AUC, all others are RMSE.

S3.3 BASF

TODO results on random splits; no new information here

S3.4 Novartis

S4 Experimental Error

Comparison of Amgen internal model and our D-MPNN, using a chronological split, to experimental error. The experimental error is not evaluated on the exact same split, but the difference in performance is striking.

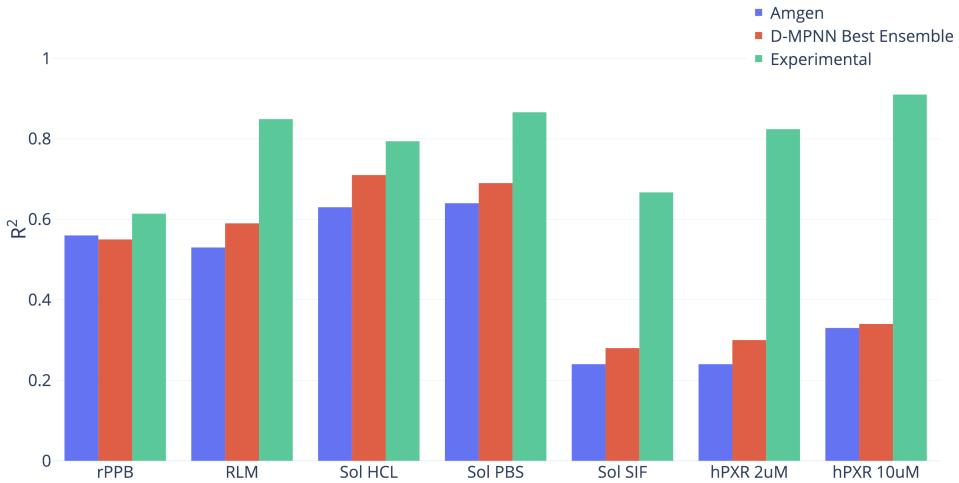


Figure S7: Experimental Error on Amgen.

Dataset	Amgen	D-MPNN Best	Experimental
rPPB	0.56	0.55	0.736
RLM	0.53	0.59	0.848
Sol HCL	0.63	0.71	0.782
Sol PBS	0.64	0.69	0.856
Sol SIF	0.24	0.28	0.614
hPXR 2uM	0.24	0.3	0.579
hPXR 10uM	0.33	0.34	0.668

Table S14: Experimental Error on Amgen. All numbers are R^2 .

S5 Analysis of Split Type

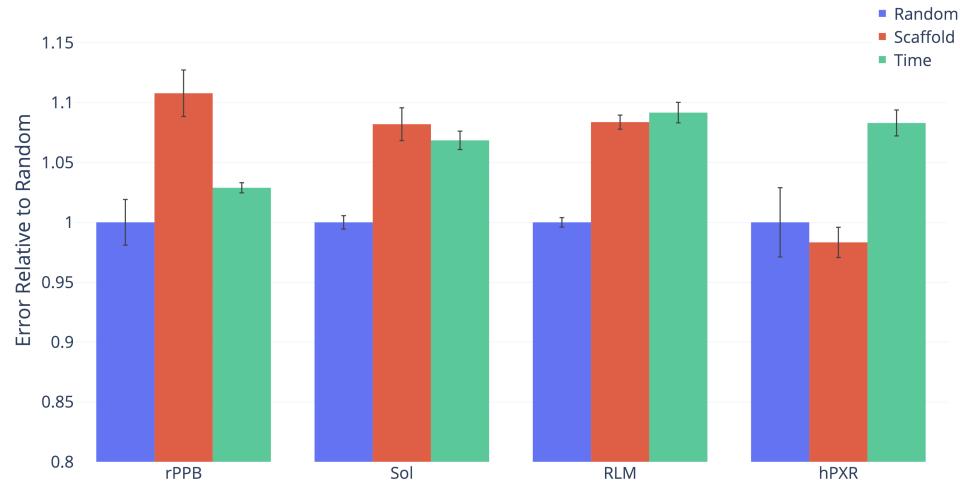


Figure S8: Split Type on Amgen.

Dataset	Metric	Random	Scaffold	Time
rPPB	RMSE	1.033 ± 0.062	$1.145 \pm 0.063 (+10.78\%)$	$1.063 \pm 0.014 (+2.89\%)$
Sol	RMSE	0.664 ± 0.012	$0.718 \pm 0.029 (+8.20\%)$	$0.710 \pm 0.016 (+6.85\%)$
RLM	RMSE	0.301 ± 0.004	$0.326 \pm 0.006 (+8.37\%)$	$0.328 \pm 0.008 (+9.16\%)$
hPXR	RMSE	33.969 ± 3.107	$33.401 \pm 1.354 (-1.67\%)$	$36.789 \pm 1.164 (+8.30\%)$

Table S15: Split Type on Amgen.

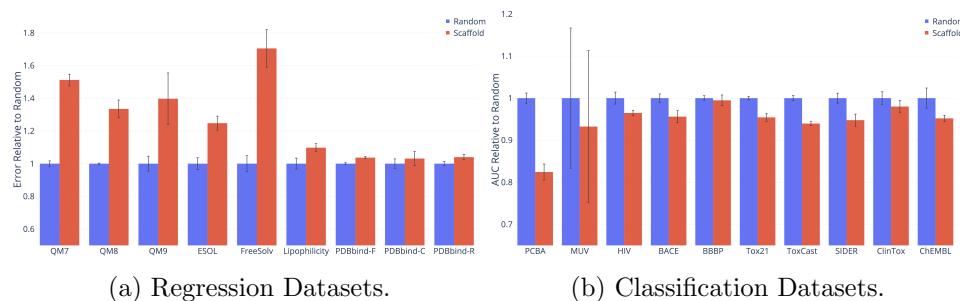


Figure S9: Split Type on Public Datasets.

Dataset	Metric	Random	Scaffold
QM7	MAE	67.657 ± 3.647	$102.302 \pm 7.532 (+51.21\%)$
QM8	MAE	0.0111 ± 0.0001	$0.0148 \pm 0.0019 (+33.49\%)$
QM9	MAE	3.705 ± 0.537	$5.177 \pm 1.848 (+39.72\%)$
ESOL	RMSE	0.691 ± 0.080	$0.862 \pm 0.094 (+24.77\%)$
FreeSolv	RMSE	1.133 ± 0.176	$1.932 \pm 0.412 (+70.49\%)$
Lipophilicity	RMSE	0.598 ± 0.064	$0.656 \pm 0.047 (+9.79\%)$
PDBbind-F	RMSE	1.325 ± 0.027	$1.374 \pm 0.028 (+3.70\%)$
PDBbind-C	RMSE	2.088 ± 0.194	$2.154 \pm 0.290 (+3.12\%)$
PDBbind-R	RMSE	1.411 ± 0.060	$1.468 \pm 0.075 (+4.00\%)$
PCBA	PRC-AUC	0.329 ± 0.007	$0.271 \pm 0.011 (-17.56\%)$
MUV	PRC-AUC	0.0865 ± 0.0456	$0.0807 \pm 0.0494 (-6.76\%)$
HIV	ROC-AUC	0.811 ± 0.037	$0.782 \pm 0.016 (-3.52\%)$
BACE	ROC-AUC	0.863 ± 0.028	$0.825 \pm 0.039 (-4.40\%)$
BBBP	ROC-AUC	0.920 ± 0.017	$0.915 \pm 0.037 (-0.52\%)$
Tox21	ROC-AUC	0.847 ± 0.012	$0.808 \pm 0.025 (-4.57\%)$
ToxCast	ROC-AUC	0.735 ± 0.014	$0.690 \pm 0.011 (-6.04\%)$
SIDER	ROC-AUC	0.639 ± 0.024	$0.606 \pm 0.029 (-5.23\%)$
ClinTox	ROC-AUC	0.892 ± 0.043	$0.874 \pm 0.041 (-2.00\%)$
ChEMBL	ROC-AUC	0.745 ± 0.031	$0.709 \pm 0.010 (-4.81\%)$

Table S16: Split Type on Public Datasets.

S6 Ablations

S6.1 Message Type

Here we describe the implementation and performance of our atom-based and undirected bond-based messages. For the most direct comparison, we implemented these as options in our chemprop repository; the changes are only a few lines of code in each case. Therefore, in each case, we simply detail the differences from our directed bond-based messages.

Atom Messages

We initialize messages based on atom features rather than bond features, according to $h_v^0 = \sigma(W_i x_v)$ rather than $h_{vw}^0 = \sigma(W_i \text{cat}(\mathbf{x}_v, \mathbf{e}_{vw}))$, with matrix dimensions adjusted accordingly.

During message passing, each atom receives messages according to $m_v^{t+1} = \sum_{k \in \{N(v)\}} h_k^t$.

Finally, m_v is the sum of all of the atom hidden states at the end of message passing.

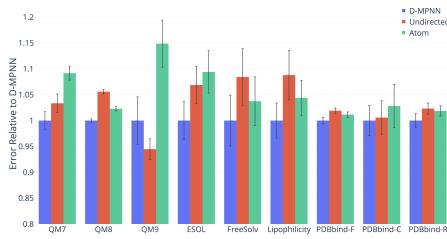
Undirected Bond Messages

The only difference with our D-MPNN is that before each message passing step, for each pair of bonded atoms v and w , we set h_{vw}^t and h_{wv}^t to each be equal to their average. This way, the hidden state for each directed bond is always equal to the hidden state of its reverse bond, resulting in message passing on undirected bonds.

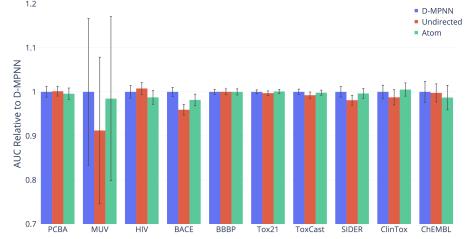
As performance with this variant is on average worse than our D-MPNN, the additional symmetry breaking introduced by directed bonds appears to be helpful.

Comparison of Different Message Types

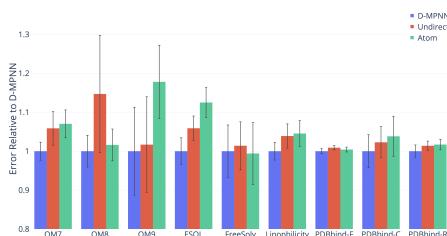
Comparison of performance using different message passing paradigms. Our D-MPNN uses directed messages.



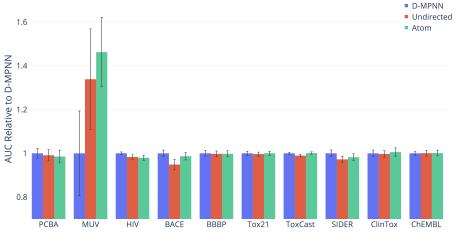
(a) Regression Datasets (Random Split).



(b) Classification Datasets (Random Split).



(c) Regression Datasets (Scaffold Split).



(d) Classification Datasets (Scaffold Split).

Dataset	Metric	D-MPNN	Undirected	Atom
QM7	MAE	67.657 ± 3.647	$69.919 \pm 3.876 (+3.34\%)$	$73.860 \pm 2.849 (+9.17\%)$
QM8	MAE	0.0111 ± 0.0001	$0.0117 \pm 0.0002 (+5.59\%)$	$0.0114 \pm 0.0001 (+2.29\%)$
QM9	MAE	3.705 ± 0.537	$3.501 \pm 0.233 (-5.52\%)$	$4.257 \pm 0.533 (+14.88\%)$
ESOL	RMSE	0.691 ± 0.080	$0.738 \pm 0.079 (+6.88\%)$	$0.756 \pm 0.089 (+9.41\%)$
FreeSolv	RMSE	1.133 ± 0.176	$1.229 \pm 0.196 (+8.42\%)$	$1.176 \pm 0.169 (+3.75\%)$
Lipophilicity	RMSE	0.598 ± 0.064	$0.650 \pm 0.091 (+8.79\%)$	$0.624 \pm 0.064 (+4.40\%)$
PDBbind-F	RMSE	1.325 ± 0.027	$1.351 \pm 0.022 (+1.91\%)$	$1.340 \pm 0.022 (+1.15\%)$
PDBbind-C	RMSE	2.088 ± 0.194	$2.100 \pm 0.216 (+0.58\%)$	$2.147 \pm 0.278 (+2.82\%)$
PDBbind-R	RMSE	1.411 ± 0.060	$1.444 \pm 0.048 (+2.32\%)$	$1.438 \pm 0.044 (+1.87\%)$
PCBA	PRC-AUC	0.329 ± 0.007	$0.329 \pm 0.006 (+0.13\%)$	$0.327 \pm 0.007 (-0.43\%)$
MUV	PRC-AUC	0.0865 ± 0.0456	$0.0790 \pm 0.0455 (-8.77\%)$	$0.0852 \pm 0.0510 (-1.54\%)$
HIV	ROC-AUC	0.811 ± 0.037	$0.817 \pm 0.034 (+0.77\%)$	$0.800 \pm 0.040 (-1.26\%)$
BACE	ROC-AUC	0.863 ± 0.028	$0.828 \pm 0.034 (-4.06\%)$	$0.848 \pm 0.036 (-1.83\%)$
BBBP	ROC-AUC	0.920 ± 0.017	$0.920 \pm 0.020 (+0.01\%)$	$0.920 \pm 0.020 (-0.01\%)$
Tox21	ROC-AUC	0.847 ± 0.012	$0.844 \pm 0.014 (-0.29\%)$	$0.848 \pm 0.013 (+0.10\%)$
ToxCast	ROC-AUC	0.735 ± 0.014	$0.729 \pm 0.017 (-0.76\%)$	$0.734 \pm 0.014 (-0.19\%)$
SIDER	ROC-AUC	0.639 ± 0.024	$0.627 \pm 0.023 (-1.90\%)$	$0.637 \pm 0.022 (-0.36\%)$
ClinTox	ROC-AUC	0.892 ± 0.043	$0.881 \pm 0.049 (-1.24\%)$	$0.897 \pm 0.043 (+0.52\%)$
ChEMBL	ROC-AUC	0.745 ± 0.031	$0.743 \pm 0.026 (-0.24\%)$	$0.735 \pm 0.036 (-1.30\%)$

Table S17: Message Type (Random Split).

Dataset	Metric	D-MPNN	Undirected	Atom
QM7	MAE	102.302 ± 7.532	$108.315 \pm 13.999 (+5.88\%)$	$109.495 \pm 11.577 (+7.03\%)$
QM8	MAE	0.0148 ± 0.0019	$0.0170 \pm 0.0071 (+14.71\%)$	$0.0151 \pm 0.0019 (+1.61\%)$
QM9	MAE	5.177 ± 1.848	$5.264 \pm 2.014 (+1.69\%)$	$6.100 \pm 1.542 (+17.82\%)$
ESOL	RMSE	0.862 ± 0.094	$0.913 \pm 0.086 (+5.89\%)$	$0.970 \pm 0.104 (+12.52\%)$
FreeSolv	RMSE	1.932 ± 0.412	$1.959 \pm 0.376 (+1.40\%)$	$1.921 \pm 0.486 (-0.59\%)$
Lipophilicity	RMSE	0.656 ± 0.047	$0.682 \pm 0.065 (+3.90\%)$	$0.686 \pm 0.069 (+4.54\%)$
PDBbind-F	RMSE	1.374 ± 0.028	$1.386 \pm 0.023 (+0.89\%)$	$1.380 \pm 0.027 (+0.43\%)$
PDBbind-C	RMSE	2.154 ± 0.290	$2.203 \pm 0.273 (+2.29\%)$	$2.236 \pm 0.349 (+3.83\%)$
PDBbind-R	RMSE	1.468 ± 0.075	$1.488 \pm 0.056 (+1.39\%)$	$1.493 \pm 0.062 (+1.71\%)$
PCBA	PRC-AUC	0.271 ± 0.011	$0.269 \pm 0.012 (-0.87\%)$	$0.267 \pm 0.014 (-1.47\%)$
MUV	PRC-AUC	0.0807 ± 0.049	$0.1080 \pm 0.0587 (+33.86\%)$	$0.1180 \pm 0.0405 (+46.24\%)$
HIV	ROC-AUC	0.782 ± 0.016	$0.769 \pm 0.025 (-1.67\%)$	$0.766 \pm 0.029 (-2.02\%)$
BACE	ROC-AUC	0.825 ± 0.039	$0.783 \pm 0.061 (-5.12\%)$	$0.815 \pm 0.044 (-1.30\%)$
BBBP	ROC-AUC	0.915 ± 0.037	$0.913 \pm 0.038 (-0.24\%)$	$0.913 \pm 0.041 (-0.22\%)$
Tox21	ROC-AUC	0.808 ± 0.025	$0.805 \pm 0.023 (-0.34\%)$	$0.808 \pm 0.024 (+0.02\%)$
ToxCast	ROC-AUC	0.690 ± 0.011	$0.683 \pm 0.012 (-1.05\%)$	$0.691 \pm 0.013 (+0.12\%)$
SIDER	ROC-AUC	0.606 ± 0.029	$0.589 \pm 0.026 (-2.75\%)$	$0.595 \pm 0.030 (-1.77\%)$
ClinTox	ROC-AUC	0.874 ± 0.041	$0.872 \pm 0.041 (-0.28\%)$	$0.879 \pm 0.054 (+0.56\%)$
ChEMBL	ROC-AUC	0.709 ± 0.010	$0.710 \pm 0.016 (+0.07\%)$	$0.710 \pm 0.016 (+0.12\%)$

Table S18: Message Type (Scaffold Split).

S6.2 Hyperparameter Optimization

Effect of performing Bayesian hyperparameter optimization on the depth, hidden size, number of fully connect layers, and dropout of our model. Optimization was done on random splits and then the optimized model was applied to scaffold splits to generate these numbers.

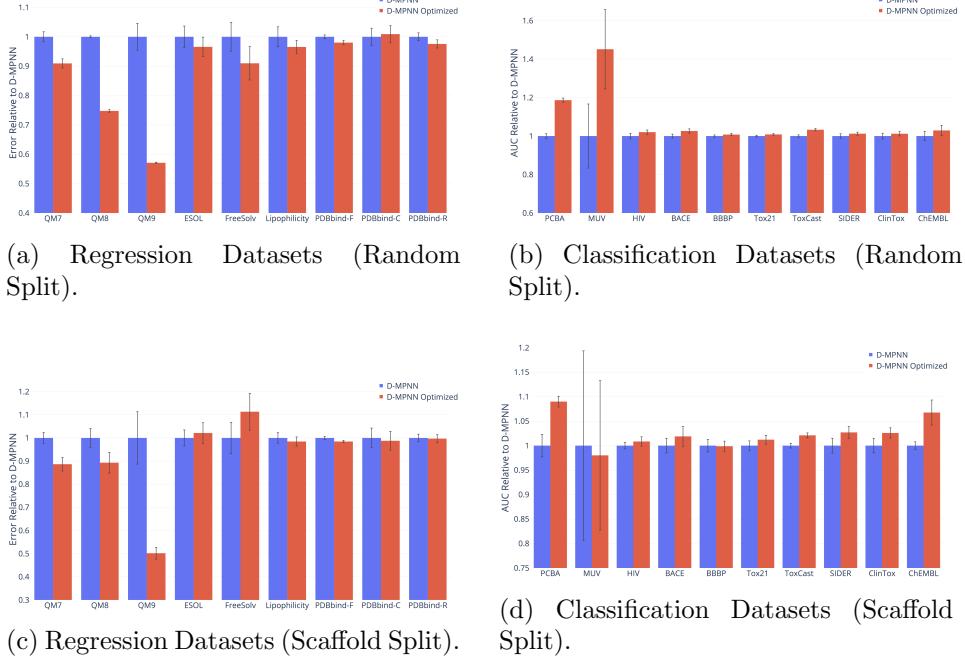


Figure S11: Hyperparameter Optimization.

Dataset	Metric	D-MPNN	D-MPNN Optimized
QM7	MAE	67.657 ± 3.647	61.524 ± 3.337 (-9.07%)
QM8	MAE	0.0111 ± 0.0001	0.0083 ± 0.0002 (-25.23%)
QM9	MAE	3.705 ± 0.537	2.116 ± 0.023 (-42.90%)
ESOL	RMSE	0.691 ± 0.080	0.667 ± 0.072 (-3.43%)
FreeSolv	RMSE	1.133 ± 0.176	1.031 ± 0.204 (-9.02%)
Lipophilicity	RMSE	0.598 ± 0.064	0.577 ± 0.042 (-3.46%)
PDBbind-F	RMSE	1.325 ± 0.027	1.299 ± 0.028 (-1.98%)
PDBbind-C	RMSE	2.088 ± 0.194	2.107 ± 0.191 (+0.91%)
PDBbind-R	RMSE	1.411 ± 0.060	1.377 ± 0.062 (-2.43%)
PCBA	PRC-AUC	0.329 ± 0.007	0.390 ± 0.006 (+18.66%)
MUV	PRC-AUC	0.0865 ± 0.0456	0.126 ± 0.056 (+45.13%)
HIV	ROC-AUC	0.811 ± 0.037	0.827 ± 0.029 (+2.03%)
BACE	ROC-AUC	0.863 ± 0.028	0.886 ± 0.030 (+2.62%)
BBBP	ROC-AUC	0.920 ± 0.017	0.927 ± 0.017 (+0.77%)
Tox21	ROC-AUC	0.847 ± 0.012	0.854 ± 0.011 (+0.84%)
ToxCast	ROC-AUC	0.735 ± 0.014	0.759 ± 0.014 (+3.28%)
SIDER	ROC-AUC	0.639 ± 0.024	0.647 ± 0.016 (+1.24%)
ClinTox	ROC-AUC	0.892 ± 0.043	0.903 ± 0.032 (+1.21%)
ChEMBL	ROC-AUC	0.745 ± 0.031	0.767 ± 0.035 (+2.92%)

Table S19: Hyperparameter Optimization (Random Split).

Dataset	Metric	D-MPNN	D-MPNN Optimized
QM7	MAE	102.302 ± 7.532	90.696 ± 9.463 (-11.34%)
QM8	MAE	0.0148 ± 0.0019	0.0132 ± 0.0021 (-10.69%)
QM9	MAE	5.177 ± 1.848	2.599 ± 0.414 (-49.81%)
ESOL	RMSE	0.862 ± 0.094	0.880 ± 0.121 (+2.12%)
FreeSolv	RMSE	1.932 ± 0.412	2.151 ± 0.486 (+11.33%)
Lipophilicity	RMSE	0.656 ± 0.047	0.646 ± 0.041 (-1.55%)
PDBbind-F	RMSE	1.374 ± 0.028	1.353 ± 0.022 (-1.54%)
PDBbind-C	RMSE	2.154 ± 0.290	2.127 ± 0.277 (-1.24%)
PDBbind-R	RMSE	1.468 ± 0.075	1.463 ± 0.078 (-0.30%)
PCBA	PRC-AUC	0.271 ± 0.011	0.295 ± 0.005 (+9.00%)
MUV	PRC-AUC	0.0807 ± 0.0494	0.0791 ± 0.0389 (-1.99%)
HIV	ROC-AUC	0.782 ± 0.016	0.789 ± 0.024 (+0.85%)
BACE	ROC-AUC	0.825 ± 0.039	0.841 ± 0.053 (+1.89%)
BBBP	ROC-AUC	0.915 ± 0.037	0.914 ± 0.030 (-0.12%)
Tox21	ROC-AUC	0.808 ± 0.025	0.818 ± 0.023 (+1.22%)
ToxCast	ROC-AUC	0.690 ± 0.011	0.705 ± 0.011 (+2.10%)
SIDER	ROC-AUC	0.606 ± 0.029	0.622 ± 0.023 (+2.70%)
ClinTox	ROC-AUC	0.874 ± 0.041	0.897 ± 0.029 (+2.59%)
ChEMBL	ROC-AUC	0.709 ± 0.010	0.757 ± 0.031 (+6.76%)

Table S20: Hyperparameter Optimization (Scaffold Split).

Dataset	Depth	Dropout	# FFN Layers	Hidden Size	# Parameters
QM7	5	0.1	3	700	1,477,601
QM8	5	0	3	2000	9,256,212
QM9	6	0	3	2400	12,918,312
ESOL	6	0.2	2	900	2,143,501
FreeSolv	6	0	3	2100	10,131,001
Lipophilicity	5	0.2	3	1200	3,668,101
PDBbind-F	6	0	2	300	354,901
PDBbind-C	6	0	3	300	445,201
PDBbind-R	4	0.1	2	1800	7,526,401
PCBA	4	0.2	2	2400	12,953,228
MUV	3	0	2	300	359,717
HIV	6	0.3	3	2300	12,007,201
BACE	6	0.1	2	600	1,069,201
BBBP	5	0.15	1	1000	2,282,001
Tox21	6	0.15	2	700	1,390,612
ToxCast	6	0.1	2	600	1,254,617
SIDER	5	0.05	1	1400	4,351,227
ClinTox	6	0.15	3	2200	11,049,402
ChEMBL	6	0.05	2	700	1,781,310

Table S21: Optimal Hyperparameter Settings.

S7 RDKit Features

Effect of adding RDKit features to our optimized D-MPNN.

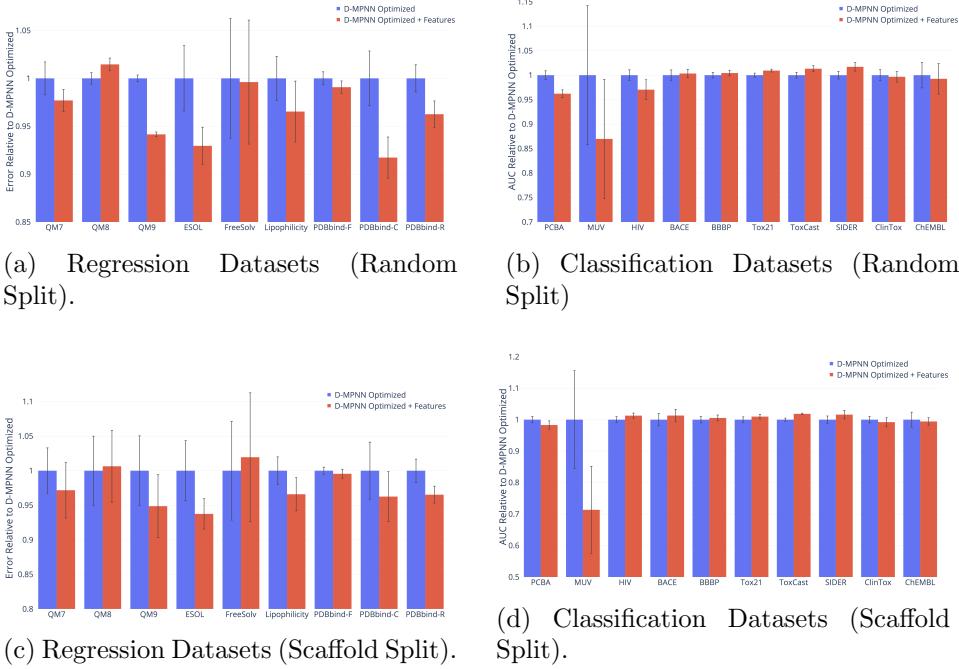


Figure S12: RDKit Features.

Dataset	Metric	D-MPNN Optimized	D-MPNN Optimized + Features
QM7	MAE	61.524 ± 3.337	60.106 ± 2.250 (-2.30%)
QM8	MAE	0.0083 ± 0.0002	0.0084 ± 0.0002 (+1.46%)
QM9	MAE	2.116 ± 0.023	1.992 ± 0.017 (-5.85%)
ESOL	RMSE	0.667 ± 0.072	0.620 ± 0.041 (-7.05%)
FreeSolv	RMSE	1.031 ± 0.204	1.027 ± 0.211 (-0.39%)
Lipophilicity	RMSE	0.577 ± 0.042	0.557 ± 0.058 (-3.47%)
PDBbind-F	RMSE	1.299 ± 0.028	1.287 ± 0.027 (-0.92%)
PDBbind-C	RMSE	2.107 ± 0.191	1.933 ± 0.143 (-8.27%)
PDBbind-R	RMSE	1.377 ± 0.062	1.325 ± 0.060 (-3.74%)
PCBA	PRC-AUC	0.390 ± 0.006	0.375 ± 0.005 (-3.79%)
MUV	PRC-AUC	0.12560 ± 0.056	0.1092 ± 0.0483 (-13.03%)
HIV	ROC-AUC	0.827 ± 0.029	0.803 ± 0.053 (-2.97%)
BACE	ROC-AUC	0.886 ± 0.030	0.889 ± 0.024 (+0.34%)
BBBP	ROC-AUC	0.927 ± 0.017	0.931 ± 0.016 (+0.43%)
Tox21	ROC-AUC	0.854 ± 0.011	0.862 ± 0.008 (+0.94%)
ToxCast	ROC-AUC	0.759 ± 0.014	0.769 ± 0.015 (+1.32%)
SIDER	ROC-AUC	0.647 ± 0.016	0.658 ± 0.018 (+1.70%)
ClinTox	ROC-AUC	0.903 ± 0.032	0.900 ± 0.031 (-0.33%)
ChEMBL	ROC-AUC	0.767 ± 0.035	0.761 ± 0.041 (-0.75%)

Table S22: RDKit Features (Random Split).

Dataset	Metric	D-MPNN Optimized	D-MPNN Optimized + Features
QM7	MAE	90.696 ± 9.463	88.136 ± 11.494 (-2.82%)
QM8	MAE	0.0132 ± 0.0021	0.0133 ± 0.0022 (+0.64%)
QM9	MAE	2.599 ± 0.414	2.465 ± 0.375 (-5.13%)
ESOL	RMSE	0.880 ± 0.121	0.825 ± 0.061 (-6.25%)
FreeSolv	RMSE	2.151 ± 0.486	2.193 ± 0.634 (+1.95%)
Lipophilicity	RMSE	0.646 ± 0.041	0.624 ± 0.049 (-3.41%)
PDBbind-F	RMSE	1.353 ± 0.022	1.347 ± 0.028 (-0.44%)
PDBbind-C	RMSE	2.127 ± 0.277	2.047 ± 0.244 (-3.74%)
PDBbind-R	RMSE	1.463 ± 0.078	1.413 ± 0.057 (-3.47%)
PCBA	PRC-AUC	0.295 ± 0.005	0.290 ± 0.007 (-1.70%)
MUV	PRC-AUC	0.0791 ± 0.0389	0.0564 ± 0.0346 (-28.65%)
HIV	ROC-AUC	0.789 ± 0.024	0.799 ± 0.020 (+1.28%)
BACE	ROC-AUC	0.841 ± 0.053	0.852 ± 0.053 (+1.31%)
BBBP	ROC-AUC	0.914 ± 0.030	0.919 ± 0.027 (+0.55%)
Tox21	ROC-AUC	0.818 ± 0.023	0.826 ± 0.019 (+0.98%)
ToxCast	ROC-AUC	0.705 ± 0.011	0.718 ± 0.005 (+1.84%)
SIDER	ROC-AUC	0.622 ± 0.023	0.632 ± 0.027 (+1.61%)
ClinTox	ROC-AUC	0.897 ± 0.029	0.890 ± 0.040 (-0.78%)
ChEMBL	ROC-AUC	0.757 ± 0.031	0.753 ± 0.016 (-0.57%)

Table S23: RDKit Features (Scaffold Split).

S7.1 Ensembling

Benefit of ensembling five models instead of a single model. All results are using the our best model settings (i.e. optimized hyperparameters and RDKit features, if they improved performance in the single model setting).

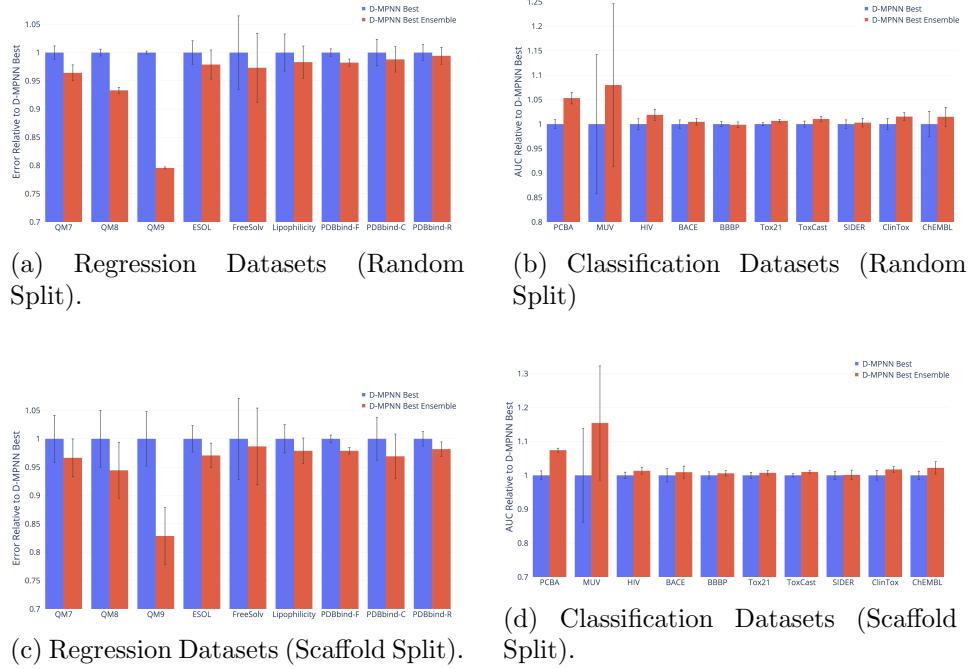


Figure S13: Ensembling.

Dataset	Metric	Features	D-MPNN Best	D-MPNN Best Ensemble
QM7	MAE	Yes	60.106 ± 2.250	57.952 ± 2.674 (-3.58%)
QM8	MAE	No	0.0083 ± 0.0002	0.0077 ± 0.0001 (-6.69%)
QM9	MAE	Yes	1.992 ± 0.017	1.585 ± 0.011 (-20.43%)
ESOL	RMSE	Yes	0.620 ± 0.041	0.607 ± 0.051 (-2.12%)
FreeSolv	RMSE	Yes	1.027 ± 0.211	0.999 ± 0.198 (-2.69%)
Lipophilicity	RMSE	Yes	0.557 ± 0.058	0.548 ± 0.050 (-1.69%)
PDBbind-F	RMSE	Yes	1.287 ± 0.027	1.264 ± 0.026 (-1.78%)
PDBbind-C	RMSE	Yes	1.933 ± 0.143	1.910 ± 0.137 (-1.21%)
PDBbind-R	RMSE	Yes	1.325 ± 0.060	1.318 ± 0.063 (-0.57%)
PCBA	PRC-AUC	No	0.390 ± 0.006	0.411 ± 0.008 (+5.31%)
MUV	PRC-AUC	No	0.1256 ± 0.0565	0.1356 ± 0.0660 (+7.98%)
HIV	ROC-AUC	No	0.827 ± 0.029	0.843 ± 0.030 (+1.89%)
BACE	ROC-AUC	Yes	0.889 ± 0.024	0.893 ± 0.019 (+0.47%)
BBBP	ROC-AUC	Yes	0.931 ± 0.016	0.930 ± 0.017 (-0.14%)
Tox21	ROC-AUC	Yes	0.862 ± 0.008	0.868 ± 0.008 (+0.66%)
ToxCast	ROC-AUC	Yes	0.769 ± 0.015	0.777 ± 0.014 (+1.04%)
SIDER	ROC-AUC	Yes	0.658 ± 0.018	0.660 ± 0.018 (+0.30%)
ClinTox	ROC-AUC	No	0.903 ± 0.032	0.917 ± 0.023 (+1.54%)
ChEMBL	ROC-AUC	No	0.767 ± 0.035	0.778 ± 0.026 (+1.48%)

Table S24: Ensembling (Random Split).

Dataset	Metric	Features	D-MPNN Best	D-MPNN Best Ensemble
QM7	MAE	Yes	88.136 ± 11.494	$85.173 \pm 9.300 (-3.36\%)$
QM8	MAE	No	0.0132 ± 0.0021	$0.0125 \pm 0.0021 (-5.56\%)$
QM9	MAE	Yes	2.465 ± 0.375	$2.043 \pm 0.391 (-17.14\%)$
ESOL	RMSE	Yes	0.825 ± 0.061	$0.801 \pm 0.057 (-2.94\%)$
FreeSolv	RMSE	No	2.151 ± 0.486	$2.122 \pm 0.460 (-1.34\%)$
Lipophilicity	RMSE	Yes	0.624 ± 0.049	$0.611 \pm 0.045 (-2.12\%)$
PDBbind-F	RMSE	Yes	1.347 ± 0.028	$1.319 \pm 0.024 (-2.11\%)$
PDBbind-C	RMSE	Yes	2.047 ± 0.244	$1.984 \pm 0.254 (-3.09\%)$
PDBbind-R	RMSE	Yes	1.413 ± 0.057	$1.387 \pm 0.057 (-1.82\%)$
PCBA	PRC-AUC	No	0.295 ± 0.007	$0.317 \pm 0.003 (+7.43\%)$
MUV	PRC-AUC	No	0.0791 ± 0.0346	$0.0913 \pm 0.0421 (+15.44\%)$
HIV	ROC-AUC	Yes	0.799 ± 0.024	$0.809 \pm 0.024 (+1.33\%)$
BACE	ROC-AUC	Yes	0.852 ± 0.053	$0.860 \pm 0.048 (+0.94\%)$
BBBP	ROC-AUC	Yes	0.919 ± 0.030	$0.925 \pm 0.023 (+0.60\%)$
Tox21	ROC-AUC	Yes	0.826 ± 0.023	$0.832 \pm 0.018 (+0.75\%)$
ToxCast	ROC-AUC	Yes	0.718 ± 0.011	$0.725 \pm 0.009 (+1.01\%)$
SIDER	ROC-AUC	Yes	0.632 ± 0.023	$0.633 \pm 0.029 (+0.11\%)$
ClinTox	ROC-AUC	No	0.897 ± 0.040	$0.913 \pm 0.024 (+1.74\%)$
ChEMBL	ROC-AUC	No	0.757 ± 0.016	$0.774 \pm 0.024 (+2.21\%)$

Table S25: Ensembling (Scaffold Split).

S7.2 Effect of Data Size

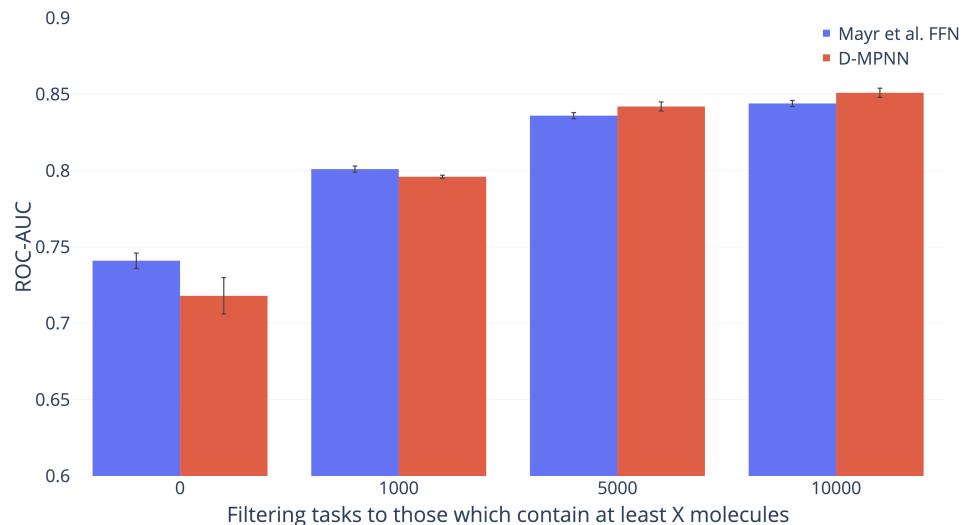


Figure S14: Effect of Data Size on ChEMBL.

Min # of Compounds	Mayr et al FFN	D-MPNN Best
0	0.741	0.741 (-3.10%)
1,000	0.801	0.783 (-0.62%)
5,000	0.836	0.801 (+0.72%)
10,000	0.844	0.836 (+0.83%)

Table S26: Effect of Data Size on ChEMBL. All numbers are ROC-AUC.

S8 RDKit-Calculated Features

We used the following list of RDKit functions to calculate the RDKit features used by our model.

BalabanJ	BertzCT	Chi0
Chi0n	Chi0v	Chi1
Chi1n	Chi1v	Chi2n
Chi2v	Chi3n	Chi3v
Chi4n	Chi4v	EState_VSA1
EState_VSA10	EState_VSA11	EState_VSA2
EState_VSA3	EState_VSA4	EState_VSA5
EState_VSA6	EState_VSA7	EState_VSA8
EState_VSA9	ExactMolWt	FpDensityMorgan1
FpDensityMorgan2	FpDensityMorgan3	FractionCSP3
HallKierAlpha	HeavyAtomCount	HeavyAtomMolWt
Ipc	Kappa1	Kappa2
Kappa3	LabuteASA	MaxAbsEStateIndex
MaxAbsPartialCharge	MaxEStateIndex	MaxPartialCharge
MinAbsEStateIndex	MinAbsPartialCharge	MinEStateIndex
MinPartialCharge	MolLogP	MolMR
MolWt	NHOHCount	NOCount
NumAliphaticCarbocycles	NumAliphaticHeterocycles	NumAliphaticRings
NumAromaticCarbocycles	NumAromaticHeterocycles	NumAromaticRings
NumHAcceptors	NumHDonors	NumHeteroatoms
NumRadicalElectrons	NumRotatableBonds	NumSaturatedCarbocycles
NumSaturatedHeterocycles	NumSaturatedRings	NumValenceElectrons
PEOE_VSA1	PEOE_VSA10	PEOE_VSA11
PEOE_VSA12	PEOE_VSA13	PEOE_VSA14
PEOE_VSA2	PEOE_VSA3	PEOE_VSA4
PEOE_VSA5	PEOE_VSA6	PEOE_VSA7
PEOE_VSA8	PEOE_VSA9	RingCount
SMR_VSA1	SMR_VSA10	SMR_VSA2
SMR_VSA3	SMR_VSA4	SMR_VSA5
SMR_VSA6	SMR_VSA7	SMR_VSA8
SMR_VSA9	SlogP_VSA1	SlogP_VSA10

SlogP_VSA11	SlogP_VSA12	SlogP_VSA2
SlogP_VSA3	SlogP_VSA4	SlogP_VSA5
SlogP_VSA6	SlogP_VSA7	SlogP_VSA8
SlogP_VSA9	TPSA	VSA_EState1
VSA_EState10	VSA_EState2	VSA_EState3
VSA_EState4	VSA_EState5	VSA_EState6
VSA_EState7	VSA_EState8	VSA_EState9
fr_Al_COO	fr_Al_OH	fr_Al_OH_noTert
fr_ArN	fr_Ar_COO	fr_Ar_N
fr_Ar_NH	fr_Ar_OH	fr_COO
fr_COO2	fr_C_O	fr_C_O_noCOO
fr_C_S	fr_HOCCN	fr_Imine
fr_NH0	fr_NH1	fr_NH2
fr_N_O	fr_Ndealkylation1	fr_Ndealkylation2
fr_Nhpyrrole	fr_SH	fr_aldehyde
fr_alkyl_carbamate	fr_alkyl_halide	fr_allylic_oxid
fr_amide	fr_amidine	fr_aniline
fr Aryl_methyl	fr_azide	fr_azo
fr_barbitur	fr_benzene	fr_benzodiazepine
fr_bicyclic	fr_diazo	fr_dihydropyridine
fr_epoxide	fr_esther	fr_ether
fr_furan	fr_guanido	fr_halogen
fr_hdrzine	fr_hdrzone	fr_imidazole
fr_imide	fr_isocyan	fr_isothiocyan
fr_ketone	fr_ketone_Topliss	fr_lactam
fr_lactone	fr_methoxy	fr_morpholine
fr_nitrile	fr_nitro	fr_nitro_arom
fr_nitro_arom_nonortho	fr_nitroso	fr_oxazole
fr_oxime	fr_para_hydroxylation	fr_phenol
fr_phenol_noOrthoHbond	fr_phos_acid	fr_phos_ester
fr_piperdine	fr_piperazine	fr_priamide
fr_prisulfonamid	fr_pyridine	fr_quatN
fr_sulfide	fr_sulfonamid	fr_sulfone
fr_term_acetylene	fr_tetrazole	fr_thiazole
fr_thiocyan	fr_thiophene	fr_unbrch_alkane
fr_urea	qed	