Supplementary Information for

ACES-GNN: Can Graph Neural Network Learn to Explain Activity Cliffs?

Xu Chen,[†] Dazhou Yu,[‡] Liang Zhao,[‡] Fang Liu[†]

[†]*Department of Chemistry, Emory University, Atlanta, Georgia, 30322, United States*

[‡]*Department of Computer Science, Emory University, Atlanta, Georgia, 30322, United States*

Table of Contents

Table S1. Atomic encoding features. The initial atomic encoding for each atom in a graph neural network input. For continuous values, the values are directly used as the features. For categorical features, they are one-hot encoded.

Atom feature	Description
Atom type	Element symbols
Implicit valence	0, 1, 2, 3, 4, >4
Degree	1, 2, 3, 4, >4
Formal charge	-3, -2, -1, 0, 1, 2, 3, other
Is in ring	Whether the atom is in a ring
Is aromatic	Whether the atom is part of an aromatic system
Hybridizations	s, sp, sp2, sp3, sp3d, dp3d2, other
Atomic mass	Scaled atomic mass
Vdw radius	Scaled van der Waals radius
Covalent radius	Scaled covalent radius

Table S2. Bond encoding features. The initial bond encoding for each atom in a graph neural network. All bond features are one-hot encoded.

Bond feature	Description
Type	Single, double, triple, aromatic
Conjugated	Whether the bond is conjugated
In ring	Whether the bond in part of an aromatic system

Table S3. Tuning range of hyperparameters. Hyperparameter tuning was performed for all graph neural network models using Hyperopt.[1] Adaptive Tree Parzen Estimators (TPE) was used as the searching algorithm with a maximum of 50 trials and an early stop of 10 trials. Explanation loss factor was chosen based on grid-search for each target dataset. The first column indicates the hyperparameters tuned in each model. The second column shows the discrete values considered when doing hyperparameter tuning.

Hyperparameters	Search space
Hidden dimension	64, 128
Global pooling	mean, add
Number of convolutional layers	2, 3, 5
Weight decay	0., 0.001, 0.0001
Dropout rate for MLP	0., 0.2, 0.5
Learning rate	0.001, 0.0003, 0.0001
Batch size	32, 64
Explanation loss factor	0.1, 0.05, 0.045, 0.04, 0.035, 0.03, 0.025, 0.02, 0.015, 0.01, 0.005, 0.001

Table S4. GradCAM global direction accuracy for MPNN and ACES-MPNN across datasets. Each value represents the mean ± standard deviation (SD) across 10 independent runs. We report both sources of variation—seed SD reflecting stochasticity in model initialization and training, and split SD reflecting variation due to different data partitioning. The global direction accuracy measures alignment between the attribution direction (from GradCAM) and the true activity change direction across molecular perturbations. Higher values indicate better explainability performance with respect to activity-related feature attribution.

dataset	MPNN global direction (seed SD)	ACES-MPNN global direction (seed SD)	MPNN global direction (split SD)	MPNN global direction (split SD)
CHEMBL4616_EC50	0.876±0.022	**0.886±0.023**	0.813 ± 0.065	**0.856 ± 0.052**
CHEMBL4792_Ki	0.805±0.018	**0.840±0.020**	0.795 ± 0.066	**0.845 ± 0.047**
CHEMBL1871_Ki	0.629±0.033	**0.726±0.033**	0.657 ± 0.115	**0.681 ± 0.089**
CHEMBL2971_Ki	0.555±0.071	**0.669±0.128**	0.603 ± 0.117	**0.659 ± 0.097**
CHEMBL239_EC50	0.563±0.030	**0.645±0.022**	0.696 ± 0.080	**0.747 ± 0.061**
CHEMBL233_Ki	0.692±0.033	**0.718±0.028**	0.733 ± 0.034	**0.746 ± 0.034**
CHEMBL235_EC50	0.683±0.031	**0.692±0.016**	0.683 ± 0.047	**0.741 ± 0.049**
CHEMBL231_Ki	0.796±0.033	**0.844±0.041**	0.738 ± 0.105	**0.774 ± 0.085**
CHEMBL218_EC50	0.813±0.041	0.794±0.042	0.675 ± 0.080	**0.732 ± 0.094**
CHEMBL244_Ki	0.765±0.029	**0.844±0.016**	0.754 ± 0.030	**0.807 ± 0.038**
CHEMBL234_Ki	0.817±0.022	**0.833±0.028**	0.844 ± 0.034	**0.862 ± 0.031**
CHEMBL237_Ki	0.748±0.025	**0.771±0.031**	0.746 ± 0.030	**0.772 ± 0.043**
CHEMBL1862_Ki	0.681±0.072	**0.789±0.041**	0.708 ± 0.062	**0.785 ± 0.053**
CHEMBL4203_Ki	0.525±0.075	0.500±0.079	0.564 ± 0.117	0.554 ± 0.149
CHEMBL2047_EC50	0.689±0.042	**0.700±0.022**	0.713 ± 0.079	**0.800 ± 0.081**
CHEMBL219_Ki	0.759±0.037	**0.796±0.032**	0.780 ± 0.053	**0.813 ± 0.062**
CHEMBL236_Ki	0.713±0.017	**0.773±0.026**	0.715 ± 0.048	**0.768 ± 0.031**
CHEMBL228_Ki	0.676±0.033	**0.745±0.033**	0.726 ± 0.055	**0.751 ± 0.050**
CHEMBL2147_Ki	0.795±0.051	**0.823±0.030**	0.875 ± 0.022	**0.881 ± 0.046**
CHEMBL204_Ki	0.766±0.026	**0.845±0.012**	0.739 ± 0.040	**0.812 ± 0.041**
CHEMBL262_Ki	0.563±0.074	**0.600±0.061**	0.656 ± 0.100	**0.657 ± 0.094**
CHEMBL287_Ki	0.782±0.043	**0.816±0.043**	0.779 ± 0.037	**0.781 ± 0.051**
CHEMBL2034_Ki	0.654±0.058	**0.716±0.035**	0.667 ± 0.119	**0.754 ± 0.130**
CHEMBL3979_EC50	0.618±0.041	**0.682±0.022**	0.699 ± 0.078	**0.740 ± 0.053**
CHEMBL238_Ki	0.822±0.029	0.817±0.060	0.804 ± 0.092	0.787 ± 0.091
CHEMBL2835_Ki	0.640±0.150	**0.780±0.108**	0.627 ± 0.271	**0.696 ± 0.246**
CHEMBL4005_Ki	0.656±0.056	**0.670±0.056**	0.692 ± 0.065	**0.734 ± 0.071**
CHEMBL237_EC50	0.668±0.048	**0.703±0.054**	0.691 ± 0.069	**0.772 ± 0.047**
CHEMBL264_Ki	0.828±0.025	**0.848±0.027**	0.756 ± 0.048	**0.820 ± 0.028**
CHEMBL214_Ki	0.726±0.023	**0.773±0.021**	0.749 ± 0.032	**0.811 ± 0.034**
average	0.710	0.755	0.723	0.765

Table S5. Test set prediction performance of MPNN and ACES-MPNN across datasets. Each reported value represents the mean ± standard deviation (SD) over 10 independent runs. For each dataset, two types of variation are reported: variation due to model random initialization (seed SD) and variation due to different data splits (split SD). Lower values of RMSE and $RMSE_{cliff}$ indicate better performance, with $RMSE_{cliff}$ specifically measuring model sensitivity to activity cliffs.

dataset	MPNN RMSE (seed SD)	ACES-MPNN RMSE (seed SD)	MPNN RMSE (split SD)	ACES-MPNN RMSE (split SD)	MPNN $RMSE_{cliff}$ (seed SD)	ACES-MPNN $RMSE_{cliff}$ (split SD)	MPNN $RMSE_{cliff}$ (seed SD)	ACES-MPNN $RMSE_{cliff}$ (split SD)
CHEMBL4616_EC50	0.608±0.030	0.615±0.020	0.704 ± 0.058	**0.691 ± 0.063**	0.656±0.038	0.667±0.018	0.767 ± 0.087	0.775 ± 0.068
CHEMBL4792_Ki	0.665±0.016	**0.656±0.021**	0.651 ± 0.035	0.676 ± 0.052	0.685±0.024	**0.660±0.032**	0.686 ± 0.052	0.696 ± 0.057
CHEMBL1871_Ki	0.729±0.024	0.738±0.015	0.655 ± 0.075	0.658 ± 0.074	0.874±0.038	0.880±0.038	0.727 ± 0.129	**0.698 ± 0.118**
CHEMBL2971_Ki	0.663±0.030	0.677±0.041	0.768 ± 0.079	**0.744 ± 0.048**	0.854±0.052	**0.802±0.064**	0.994 ± 0.152	**0.889 ± 0.103**
CHEMBL239_EC50	0.844±0.016	**0.820±0.028**	0.750 ± 0.053	**0.737 ± 0.046**	0.921±0.026	**0.895±0.029**	0.890 ± 0.091	**0.863 ± 0.076**
CHEMBL233_Ki	0.844±0.019	**0.828±0.014**	0.793 ± 0.045	**0.785 ± 0.034**	0.940±0.028	**0.932±0.017**	0.833 ± 0.068	0.838 ± 0.061
CHEMBL235_EC50	0.761±0.027	**0.720±0.015**	0.733 ± 0.057	**0.731 ± 0.046**	0.877±0.031	**0.853±0.025**	0.857 ± 0.063	**0.849 ± 0.062**
CHEMBL231_Ki	0.782±0.051	**0.740±0.038**	0.731 ± 0.100	**0.713 ± 0.088**	0.823±0.088	**0.764±0.067**	0.802 ± 0.095	**0.781 ± 0.108**
CHEMBL218_EC50	0.741±0.028	**0.734±0.024**	0.737 ± 0.065	**0.721 ± 0.063**	0.748±0.026	**0.745±0.034**	0.835 ± 0.096	**0.802 ± 0.097**
CHEMBL244_Ki	0.737±0.019	0.743±0.012	0.817 ± 0.059	**0.805 ± 0.079**	0.798±0.021	0.801±0.017	0.879 ± 0.053	**0.862 ± 0.074**
CHEMBL234_Ki	0.734±0.023	**0.732±0.040**	0.727 ± 0.155	0.739 ± 0.142	0.630±0.025	**0.617±0.036**	0.675 ± 0.156	0.678 ± 0.170
CHEMBL237_Ki	0.715±0.022	**0.708±0.017**	0.770 ± 0.050	**0.767 ± 0.064**	0.793±0.022	**0.788±0.018**	0.810 ± 0.065	0.815 ± 0.060
CHEMBL1862_Ki	0.875±0.023	**0.863±0.032**	0.967 ± 0.355	**0.964 ± 0.359**	0.915±0.034	**0.879±0.045**	0.834 ± 0.141	0.834 ± 0.155
CHEMBL4203_Ki	1.026±0.024	**1.008±0.032**	0.934 ± 0.099	0.951 ± 0.084	1.185±0.095	**1.177±0.097**	1.123 ± 0.288	1.132 ± 0.230
CHEMBL2047_EC50	0.636±0.019	0.640±0.024	0.635 ± 0.050	0.637 ± 0.030	0.696±0.029	**0.706±0.026**	0.727 ± 0.079	**0.713 ± 0.059**
CHEMBL219_Ki	0.822±0.034	**0.786±0.025**	0.890 ± 0.273	**0.828 ± 0.206**	0.793±0.058	**0.766±0.031**	0.913 ± 0.231	**0.862 ± 0.168**
CHEMBL236_Ki	0.709±0.013	**0.705±0.020**	0.743 ± 0.053	0.754 ± 0.057	0.797±0.027	0.799±0.025	0.818 ± 0.060	0.840 ± 0.058
CHEMBL228_Ki	0.742±0.022	0.748±0.022	0.821 ± 0.136	**0.798 ± 0.110**	0.800±0.024	**0.791±0.023**	0.847 ± 0.114	**0.819 ± 0.120**
CHEMBL2147_Ki	0.746±0.048	0.782±0.033	0.733 ± 0.062	0.770 ± 0.076	0.672±0.033	0.708±0.064	0.725 ± 0.078	0.744 ± 0.089
CHEMBL204_Ki	0.703±0.024	0.707±0.019	0.777 ± 0.049	0.778 ± 0.041	0.835±0.034	0.839±0.023	0.867 ± 0.072	**0.864 ± 0.047**
CHEMBL262_Ki	0.810±0.037	**0.784±0.033**	0.822 ± 0.076	0.834 ± 0.068	0.921±0.045	**0.912±0.041**	0.942 ± 0.127	**0.915 ± 0.119**
CHEMBL287_Ki	0.797±0.028	**0.755±0.036**	0.870 ± 0.240	**0.839 ± 0.112**	0.852±0.035	**0.830±0.058**	0.884 ± 0.199	**0.874 ± 0.120**
CHEMBL2034_Ki	0.763±0.042	0.805±0.024	0.695 ± 0.148	**0.686 ± 0.132**	0.754±0.032	0.754±0.034	0.722 ± 0.156	**0.673 ± 0.124**
CHEMBL3979_EC50	0.735±0.033	**0.716±0.027**	0.689 ± 0.052	0.721 ± 0.051	0.830±0.046	**0.812±0.014**	0.752 ± 0.094	0.756 ± 0.097
CHEMBL238_Ki	0.632±0.039	0.636±0.046	0.685 ± 0.050	0.689 ± 0.054	0.717±0.032	0.736±0.046	0.788 ± 0.111	0.789 ± 0.120
CHEMBL2835_Ki	0.589±0.046	**0.515±0.060**	0.497 ± 0.104	0.523 ± 0.097	1.172±0.092	**0.976±0.103**	0.833 ± 0.274	**0.823 ± 0.265**
CHEMBL4005_Ki	0.760±0.030	**0.755±0.022**	0.679 ± 0.077	**0.672 ± 0.068**	0.820±0.036	0.826±0.021	0.752 ± 0.089	**0.721 ± 0.096**
CHEMBL237_EC50	0.925±0.029	**0.916±0.023**	0.977 ± 0.369	**0.962 ± 0.273**	1.034±0.040	**1.030±0.043**	1.029 ± 0.296	**0.983 ± 0.207**
CHEMBL264_Ki	0.669±0.037	0.679±0.028	0.759 ± 0.229	**0.733 ± 0.148**	0.674±0.026	0.680±0.023	0.821 ± 0.268	**0.783 ± 0.173**
CHEMBL214_Ki	0.663±0.017	**0.650±0.009**	0.670 ± 0.042	**0.665 ± 0.037**	0.740±0.022	**0.733±0.018**	0.745 ± 0.043	0.756 ± 0.048
average	0.748	0.739	0.756	0.752	0.827	0.812	0.829	0.814

Table S6. Ablation study: ACES-MPNN without the uncommon part of the explanation loss. Each reported value represents the mean ± standard deviation (SD) over 10 independent runs. For each dataset, two types of variation are reported: variation due to model random initialization (seed SD) and variation due to different data splits (split SD). Lower values of RMSE and RMSEcliff indicate better performance, with $RMSE_{cliff}$ specifically measuring model sensitivity to activity cliffs. The performance of averaged test set RMSE, $RMSE_{cliff}$ and global direction accuracy are not as good as when considering both common and uncommon part of explanation loss.

dataset	RMSE (seed SD)↓	RMSE (split SD)↓	$RMSE_{cliff}$ (seed SD)↓	$RMSE_{cliff}$ (split SD)↓	global direction (seed SD) ↑	global direction (split SD) ↑
CHEMBL4616_EC50	0.611±0.026	0.703 ± 0.057	0.668±0.034	0.785 ± 0.093	0.873±0.025	0.831 ± 0.041
CHEMBL4792_Ki	0.652±0.015	0.680 ± 0.050	0.659±0.022	0.692 ± 0.078	0.815±0.019	0.811 ± 0.078
CHEMBL1871_Ki	0.728±0.019	0.654 ± 0.091	0.912±0.034	0.736 ± 0.144	0.629±0.041	0.648 ± 0.107
CHEMBL2971_Ki	0.705±0.039	0.752 ± 0.062	0.817±0.087	0.940 ± 0.115	0.510±0.065	0.545 ± 0.084
CHEMBL239_EC50	0.828±0.021	0.749 ± 0.056	0.906±0.029	0.889 ± 0.069	0.618±0.029	0.695 ± 0.075
CHEMBL233_Ki	0.836±0.024	0.789 ± 0.035	0.929±0.017	0.838 ± 0.059	0.695±0.028	0.720 ± 0.044
CHEMBL235_EC50	0.738±0.008	0.734 ± 0.060	0.860±0.014	0.860 ± 0.048	0.689±0.031	0.691 ± 0.043
CHEMBL231_Ki	0.746±0.030	0.900 ± 0.452	0.765±0.037	0.974 ± 0.394	0.822±0.083	0.716 ± 0.099
CHEMBL218_EC50	0.735±0.023	0.715 ± 0.058	0.751±0.029	0.823 ± 0.100	0.774±0.032	0.708 ± 0.094
CHEMBL244_Ki	0.763±0.015	0.798 ± 0.061	0.818±0.027	0.863 ± 0.071	0.796±0.019	0.782 ± 0.022
CHEMBL234_Ki	0.711±0.025	0.716 ± 0.162	0.604±0.020	0.674 ± 0.205	0.835±0.019	0.847 ± 0.035
CHEMBL237_Ki	0.709±0.016	0.748 ± 0.036	0.786±0.020	0.794 ± 0.039	0.741±0.011	0.742 ± 0.039
CHEMBL1862_Ki	0.867±0.030	0.952 ± 0.337	0.891±0.037	0.825 ± 0.136	0.724±0.047	0.721 ± 0.090
CHEMBL4203_Ki	0.988±0.057	0.950 ± 0.088	1.195±0.062	1.140 ± 0.254	0.488±0.088	0.498 ± 0.225
CHEMBL2047_EC50	0.638±0.030	0.647 ± 0.093	0.698±0.036	0.739 ± 0.130	0.705±0.034	0.730 ± 0.084
CHEMBL219_Ki	0.780±0.021	0.843 ± 0.240	0.775±0.027	0.857 ± 0.207	0.788±0.031	0.811 ± 0.061
CHEMBL236_Ki	0.717±0.026	0.818 ± 0.228	0.810±0.028	0.870 ± 0.164	0.725±0.031	0.748 ± 0.040
CHEMBL228_Ki	0.760±0.024	0.808 ± 0.114	0.794±0.030	0.839 ± 0.108	0.724±0.041	0.699 ± 0.056
CHEMBL2147_Ki	0.786±0.043	0.770 ± 0.071	0.691±0.033	0.776 ± 0.104	0.815±0.043	0.874 ± 0.039
CHEMBL204_Ki	0.705±0.017	0.771 ± 0.045	0.836±0.022	0.856 ± 0.042	0.825±0.020	0.790 ± 0.047
CHEMBL262_Ki	0.822±0.052	0.813 ± 0.065	0.924±0.035	0.931 ± 0.145	0.538±0.097	0.632 ± 0.063
CHEMBL287_Ki	0.758±0.043	0.841 ± 0.166	0.835±0.047	0.868 ± 0.145	0.807±0.030	0.745 ± 0.085
CHEMBL2034_Ki	0.810±0.033	0.718 ± 0.115	0.752±0.023	0.727 ± 0.124	0.663±0.047	0.645 ± 0.087
CHEMBL3979_EC50	0.738±0.031	0.766 ± 0.173	0.804±0.023	0.862 ± 0.322	0.663±0.021	0.708 ± 0.073
CHEMBL238_Ki	0.627±0.041	0.679 ± 0.055	0.723±0.062	0.797 ± 0.113	0.839±0.070	0.781 ± 0.091
CHEMBL2835_Ki	0.502±0.056	0.579 ± 0.208	0.997±0.076	0.936 ± 0.391	0.820±0.108	0.639 ± 0.259
CHEMBL4005_Ki	0.760±0.024	0.682 ± 0.085	0.823±0.033	0.755 ± 0.118	0.646±0.054	0.716 ± 0.076
CHEMBL237_EC50	0.922±0.023	1.005 ± 0.321	1.040±0.012	1.022 ± 0.258	0.681±0.035	0.724 ± 0.087
CHEMBL264_Ki	0.702±0.057	0.758 ± 0.182	0.688±0.020	0.805 ± 0.213	0.829±0.017	0.792 ± 0.028
CHEMBL214_Ki	0.662±0.017	0.679 ± 0.033	0.737±0.016	0.749 ± 0.037	0.735±0.027	0.787 ± 0.041
average	0.744	0.767	0.816	0.841	0.727	0.726

Table S7. Ablation study: ACES-MPNN without the common part of the explanation loss. Each reported value represents the mean ± standard deviation (SD) over 10 independent runs. For each dataset, two types of variation are reported: variation due to model random initialization (seed SD) and variation due to different data splits (split SD). The performance of averaged test set RMSE, $RMSE_{cliff}$ and global direction accuracy are not as good as when considering both common and uncommon part of explanation loss.

dataset	RMSE (seed SD)↓	RMSE (split SD)↓	$RMSE_{cliff}$ (seed SD)↓	$RMSE_{cliff}$ (split SD)↓	global direction (seed SD) ↑	global direction (split SD) ↑
CHEMBL4616_EC50	0.606±0.020	0.680 ± 0.057	0.661±0.023	0.752 ± 0.091	0.896±0.022	0.870 ± 0.062
CHEMBL4792_Ki	0.665±0.024	0.680 ± 0.045	0.674±0.017	0.691 ± 0.065	0.833±0.015	0.836 ± 0.043
CHEMBL1871_Ki	0.736±0.018	0.675 ± 0.094	0.872±0.033	0.722 ± 0.116	0.761±0.044	0.719 ± 0.099
CHEMBL2971_Ki	0.682±0.031	0.762 ± 0.068	0.792±0.060	0.886 ± 0.200	0.717±0.077	0.737 ± 0.120
CHEMBL239_EC50	0.855±0.026	0.744 ± 0.067	0.904±0.024	0.861 ± 0.074	0.615±0.032	0.757 ± 0.072
CHEMBL233_Ki	0.860±0.025	0.786 ± 0.032	0.954±0.020	0.830 ± 0.053	0.735±0.032	0.746 ± 0.051
CHEMBL235_EC50	0.741±0.028	0.729 ± 0.044	0.880±0.025	0.831 ± 0.053	0.738±0.030	0.746 ± 0.047
CHEMBL231_Ki	0.776±0.054	0.890 ± 0.484	0.788±0.063	0.939 ± 0.434	0.831±0.023	0.763 ± 0.045
CHEMBL218_EC50	0.744±0.023	0.738 ± 0.065	0.775±0.042	0.824 ± 0.104	0.804±0.069	0.732 ± 0.106
CHEMBL244_Ki	0.738±0.027	0.799 ± 0.051	0.790±0.041	0.863 ± 0.068	0.852±0.022	0.807 ± 0.026
CHEMBL234_Ki	0.720±0.038	0.733 ± 0.151	0.615±0.028	0.679 ± 0.163	0.828±0.018	0.857 ± 0.031
CHEMBL237_Ki	0.718±0.031	0.757 ± 0.045	0.800±0.025	0.794 ± 0.041	0.768±0.033	0.770 ± 0.034
CHEMBL1862_Ki	0.868±0.041	0.938 ± 0.348	0.907±0.043	0.834 ± 0.142	0.734±0.035	0.750 ± 0.053
CHEMBL4203_Ki	1.012±0.025	0.937 ± 0.062	1.173±0.051	1.119 ± 0.299	0.513±0.067	0.623 ± 0.166
CHEMBL2047_EC50	0.643±0.028	0.621 ± 0.068	0.702±0.032	0.682 ± 0.067	0.701±0.032	0.757 ± 0.073
CHEMBL219_Ki	0.806±0.025	0.864 ± 0.269	0.781±0.031	0.893 ± 0.233	0.778±0.031	0.825 ± 0.044
CHEMBL236_Ki	0.712±0.018	0.821 ± 0.213	0.811±0.021	0.870 ± 0.156	0.733±0.026	0.768 ± 0.046
CHEMBL228_Ki	0.780±0.032	0.823 ± 0.115	0.813±0.045	0.847 ± 0.129	0.718±0.062	0.766 ± 0.044
CHEMBL2147_Ki	0.721±0.043	0.729 ± 0.065	0.662±0.041	0.727 ± 0.085	0.829±0.024	0.893 ± 0.029
CHEMBL204_Ki	0.695±0.029	0.780 ± 0.058	0.802±0.041	0.865 ± 0.062	0.851±0.027	0.818 ± 0.038
CHEMBL262_Ki	0.816±0.032	0.867 ± 0.111	0.929±0.029	0.955 ± 0.109	0.591±0.081	0.660 ± 0.101
CHEMBL287_Ki	0.810±0.032	0.864 ± 0.199	0.842±0.036	0.884 ± 0.157	0.821±0.038	0.810 ± 0.038
CHEMBL2034_Ki	0.786±0.034	0.672 ± 0.126	0.761±0.030	0.707 ± 0.109	0.728±0.066	0.727 ± 0.079
CHEMBL3979_EC50	0.709±0.034	0.701 ± 0.057	0.799±0.023	0.742 ± 0.097	0.682±0.030	0.741 ± 0.072
CHEMBL238_Ki	0.655±0.051	0.697 ± 0.080	0.724±0.044	0.787 ± 0.130	0.856±0.052	0.794 ± 0.125
CHEMBL2835_Ki	0.515±0.060	0.543 ± 0.202	0.976±0.103	0.917 ± 0.362	0.780±0.108	0.664 ± 0.165
CHEMBL4005_Ki	0.761±0.019	0.687 ± 0.097	0.830±0.025	0.759 ± 0.120	0.678±0.044	0.777 ± 0.066
CHEMBL237_EC50	0.910±0.020	1.013 ± 0.422	1.038±0.028	1.033 ± 0.351	0.685±0.055	0.734 ± 0.068
CHEMBL264_Ki	0.701±0.073	0.753 ± 0.183	0.679±0.015	0.811 ± 0.219	0.843±0.030	0.798 ± 0.036
CHEMBL214_Ki	0.656±0.018	0.671 ± 0.022	0.740±0.024	0.750 ± 0.032	0.754±0.029	0.797 ± 0.039
average	0.747	0.765	0.816	0.828	0.755	0.768

Table S8. Test set prediction and global direction accuracy of GIN. For each dataset, the mean ± std of 10 individual runs has been reported.

dataset	Prediction ↓				Global direction ↑			
	$RMSE_{all}$	$RMSE_{cliff}$	MAE_{all}	MAE_{cliff}	GradInput	SmoothGrad	GradCAM	IG
CHEMBL4616_EC50	1.312 ± 0.299	1.208 ± 0.303	1.084 ± 0.271	1.029 ± 0.257	0.656 ± 0.093	0.646 ± 0.081	0.756 ± 0.081	0.738 ± 0.065
CHEMBL4792_Ki	1.849 ± 0.599	1.230 ± 0.294	1.629 ± 0.569	1.605 ± 0.565	0.566 ± 0.085	0.560 ± 0.060	0.698 ± 0.069	0.664 ± 0.079
CHEMBL1871_Ki	1.254 ± 0.671	1.252 ± 0.545	1.029 ± 0.635	1.082 ± 0.555	0.595 ± 0.117	0.618 ± 0.089	0.613 ± 0.050	0.638 ± 0.132
CHEMBL2971_Ki	0.986 ± 0.342	1.074 ± 0.283	0.685 ± 0.261	0.839 ± 0.296	0.658 ± 0.099	0.551 ± 0.111	0.693 ± 0.121	0.631 ± 0.080
CHEMBL239_EC50	1.148 ± 0.648	1.005 ± 0.158	0.943 ± 0.625	0.948 ± 0.430	0.573 ± 0.079	0.548 ± 0.065	0.672 ± 0.061	0.613 ± 0.060
CHEMBL233_Ki	1.023 ± 0.365	0.960 ± 0.202	0.812 ± 0.335	0.840 ± 0.266	0.571 ± 0.074	0.568 ± 0.055	0.660 ± 0.039	0.594 ± 0.059
CHEMBL235_EC50	1.068 ± 0.414	1.029 ± 0.205	0.854 ± 0.407	0.927 ± 0.441	0.571 ± 0.040	0.559 ± 0.076	0.700 ± 0.038	0.635 ± 0.067
CHEMBL231_Ki	1.050 ± 0.608	1.018 ± 0.484	0.813 ± 0.548	0.837 ± 0.434	0.602 ± 0.111	0.557 ± 0.117	0.640 ± 0.102	0.682 ± 0.097
CHEMBL218_EC50	1.005 ± 0.400	0.952 ± 0.195	0.822 ± 0.355	0.818 ± 0.250	0.583 ± 0.071	0.595 ± 0.081	0.653 ± 0.064	0.651 ± 0.039
CHEMBL244_Ki	1.391 ± 0.479	1.327 ± 0.374	1.135 ± 0.430	1.116 ± 0.381	0.608 ± 0.035	0.567 ± 0.035	0.686 ± 0.075	0.658 ± 0.040
CHEMBL234_Ki	1.016 ± 0.303	0.921 ± 0.259	0.810 ± 0.308	0.809 ± 0.365	0.627 ± 0.092	0.588 ± 0.087	0.795 ± 0.031	0.739 ± 0.042
CHEMBL237_Ki	0.974 ± 0.261	0.993 ± 0.268	0.759 ± 0.244	0.793 ± 0.241	0.566 ± 0.047	0.556 ± 0.032	0.675 ± 0.054	0.604 ± 0.069
CHEMBL1862_Ki	1.051 ± 0.453	0.914 ± 0.254	0.827 ± 0.378	0.705 ± 0.181	0.598 ± 0.069	0.569 ± 0.091	0.617 ± 0.095	0.582 ± 0.069
CHEMBL4203_Ki	1.211 ± 0.568	1.258 ± 0.395	0.997 ± 0.552	1.057 ± 0.342	0.565 ± 0.192	0.525 ± 0.163	0.549 ± 0.065	0.668 ± 0.142
CHEMBL2047_EC50	1.235 ± 0.937	1.206 ± 0.693	1.072 ± 0.928	1.088 ± 0.781	0.606 ± 0.120	0.598 ± 0.101	0.606 ± 0.101	0.691 ± 0.096
CHEMBL219_Ki	1.085 ± 0.357	0.976 ± 0.146	0.845 ± 0.326	0.873 ± 0.274	0.632 ± 0.072	0.586 ± 0.068	0.737 ± 0.063	0.609 ± 0.093
CHEMBL236_Ki	1.078 ± 0.363	1.073 ± 0.255	0.846 ± 0.328	0.851 ± 0.231	0.575 ± 0.082	0.558 ± 0.078	0.689 ± 0.033	0.600 ± 0.074
CHEMBL228_Ki	1.116 ± 0.372	1.019 ± 0.247	0.836 ± 0.300	0.814 ± 0.214	0.544 ± 0.061	0.565 ± 0.067	0.631 ± 0.059	0.585 ± 0.037
CHEMBL2147_Ki	0.859 ± 0.092	0.875 ± 0.102	0.645 ± 0.058	0.667 ± 0.086	0.624 ± 0.078	0.600 ± 0.078	0.759 ± 0.081	0.731 ± 0.071
CHEMBL204_Ki	1.648 ± 0.499	1.487 ± 0.315	1.319 ± 0.417	1.378 ± 0.468	0.608 ± 0.060	0.566 ± 0.045	0.669 ± 0.057	0.628 ± 0.054
CHEMBL262_Ki	0.958 ± 0.399	1.041 ± 0.353	0.777 ± 0.336	0.820 ± 0.215	0.473 ± 0.122	0.467 ± 0.087	0.591 ± 0.135	0.503 ± 0.112
CHEMBL287_Ki	0.999 ± 0.344	0.924 ± 0.194	0.786 ± 0.288	0.776 ± 0.224	0.719 ± 0.053	0.635 ± 0.089	0.751 ± 0.046	0.739 ± 0.034
CHEMBL2034_Ki	1.033 ± 0.355	0.889 ± 0.327	0.755 ± 0.278	0.766 ± 0.262	0.572 ± 0.116	0.534 ± 0.088	0.632 ± 0.102	0.582 ± 0.100
CHEMBL3979_EC50	0.719 ± 0.043	0.758 ± 0.063	0.561 ± 0.031	0.645 ± 0.081	0.620 ± 0.055	0.596 ± 0.087	0.700 ± 0.081	0.685 ± 0.072
CHEMBL238_Ki	0.999 ± 0.586	0.978 ± 0.343	0.806 ± 0.534	0.830 ± 0.376	0.592 ± 0.135	0.624 ± 0.072	0.698 ± 0.081	0.644 ± 0.115
CHEMBL2835_Ki	0.734 ± 0.228	1.081 ± 0.433	0.487 ± 0.151	0.925 ± 0.390	0.625 ± 0.209	0.586 ± 0.234	0.475 ± 0.121	0.533 ± 0.193
CHEMBL4005_Ki	1.328 ± 0.361	1.171 ± 0.280	1.066 ± 0.306	1.104 ± 0.306	0.603 ± 0.065	0.552 ± 0.079	0.650 ± 0.090	0.599 ± 0.086
CHEMBL237_EC50	1.806 ± 0.335	1.481 ± 0.296	1.488 ± 0.292	1.415 ± 0.247	0.583 ± 0.060	0.526 ± 0.058	0.645 ± 0.063	0.608 ± 0.049
CHEMBL264_Ki	1.129 ± 0.194	1.146 ± 0.183	0.912 ± 0.156	0.923 ± 0.156	0.595 ± 0.042	0.530 ± 0.043	0.704 ± 0.045	0.640 ± 0.060
CHEMBL214_Ki	1.011 ± 0.337	0.965 ± 0.235	0.797 ± 0.291	0.806 ± 0.246	0.569 ± 0.033	0.570 ± 0.045	0.677 ± 0.036	0.624 ± 0.045
average	1.136	1.146	0.907	0.936	0.596	0.570	0.667	0.637

Table S9. Test set prediction and global direction accuracy of ACES-GIN. For each dataset, the mean ± std of 10 individual runs has been reported.

dataset	Prediction ↓				Global direction ↑			
	$RMSE_{all}$	$RMSE_{cliff}$	MAE_{all}	MAE_{cliff}	GradInput	SmoothGrad	GradCAM	IG
CHEMBL4616_EC50	1.270 ± 0.308	1.208 ± 0.303	1.057 ± 0.279	0.996 ± 0.272	0.663 ± 0.126	0.608 ± 0.164	0.807 ± 0.046	0.775 ± 0.054
CHEMBL4792_Ki	1.256 ± 0.309	1.230 ± 0.294	1.058 ± 0.302	1.030 ± 0.277	0.574 ± 0.048	0.543 ± 0.046	0.783 ± 0.053	0.671 ± 0.086
CHEMBL1871_Ki	1.178 ± 0.615	1.252 ± 0.545	0.955 ± 0.553	1.020 ± 0.461	0.568 ± 0.128	0.615 ± 0.107	0.607 ± 0.112	0.641 ± 0.128
CHEMBL2971_Ki	0.973 ± 0.350	1.074 ± 0.283	0.681 ± 0.266	0.856 ± 0.251	0.579 ± 0.063	0.543 ± 0.052	0.716 ± 0.081	0.591 ± 0.103
CHEMBL239_EC50	0.956 ± 0.297	1.005 ± 0.158	0.763 ± 0.281	0.821 ± 0.135	0.593 ± 0.073	0.539 ± 0.077	0.712 ± 0.079	0.637 ± 0.058
CHEMBL233_Ki	0.942 ± 0.234	0.960 ± 0.202	0.739 ± 0.221	0.781 ± 0.184	0.588 ± 0.058	0.567 ± 0.051	0.703 ± 0.043	0.620 ± 0.063
CHEMBL235_EC50	0.937 ± 0.195	1.029 ± 0.205	0.727 ± 0.174	0.821 ± 0.157	0.563 ± 0.057	0.538 ± 0.042	0.700 ± 0.075	0.624 ± 0.067
CHEMBL231_Ki	1.019 ± 0.579	1.018 ± 0.484	0.814 ± 0.519	0.832 ± 0.429	0.595 ± 0.117	0.592 ± 0.119	0.660 ± 0.065	0.675 ± 0.074
CHEMBL218_EC50	0.932 ± 0.270	0.952 ± 0.195	0.744 ± 0.243	0.768 ± 0.161	0.642 ± 0.068	0.631 ± 0.066	0.698 ± 0.056	0.689 ± 0.060
CHEMBL244_Ki	1.346 ± 0.407	1.327 ± 0.374	1.108 ± 0.391	1.087 ± 0.334	0.617 ± 0.055	0.579 ± 0.055	0.739 ± 0.066	0.679 ± 0.053
CHEMBL234_Ki	0.962 ± 0.235	0.921 ± 0.259	0.761 ± 0.240	0.741 ± 0.240	0.669 ± 0.081	0.589 ± 0.098	0.823 ± 0.030	0.737 ± 0.044
CHEMBL237_Ki	0.934 ± 0.263	0.993 ± 0.268	0.738 ± 0.241	0.796 ± 0.237	0.589 ± 0.055	0.552 ± 0.085	0.690 ± 0.071	0.597 ± 0.077
CHEMBL1862_Ki	1.042 ± 0.451	0.914 ± 0.254	0.836 ± 0.364	0.715 ± 0.185	0.602 ± 0.095	0.572 ± 0.083	0.671 ± 0.104	0.611 ± 0.090
CHEMBL4203_Ki	1.206 ± 0.542	1.258 ± 0.395	0.991 ± 0.528	1.065 ± 0.306	0.489 ± 0.168	0.539 ± 0.159	0.550 ± 0.163	0.512 ± 0.163
CHEMBL2047_EC50	1.202 ± 0.822	1.206 ± 0.693	1.012 ± 0.801	1.013 ± 0.674	0.634 ± 0.126	0.632 ± 0.138	0.716 ± 0.061	0.740 ± 0.078
CHEMBL219_Ki	0.958 ± 0.168	0.976 ± 0.146	0.738 ± 0.157	0.784 ± 0.130	0.610 ± 0.102	0.594 ± 0.101	0.776 ± 0.039	0.689 ± 0.066
CHEMBL236_Ki	1.023 ± 0.280	1.073 ± 0.255	0.798 ± 0.267	0.860 ± 0.219	0.533 ± 0.085	0.552 ± 0.054	0.691 ± 0.060	0.591 ± 0.072
CHEMBL228_Ki	1.056 ± 0.337	1.019 ± 0.247	0.801 ± 0.271	0.791 ± 0.177	0.587 ± 0.080	0.573 ± 0.073	0.679 ± 0.095	0.626 ± 0.092
CHEMBL2147_Ki	0.870 ± 0.089	0.875 ± 0.102	0.670 ± 0.059	0.689 ± 0.076	0.648 ± 0.088	0.551 ± 0.120	0.809 ± 0.054	0.744 ± 0.061
CHEMBL204_Ki	1.437 ± 0.316	1.487 ± 0.315	1.181 ± 0.291	1.232 ± 0.286	0.600 ± 0.056	0.553 ± 0.076	0.682 ± 0.098	0.628 ± 0.067
CHEMBL262_Ki	0.990 ± 0.405	1.041 ± 0.353	0.758 ± 0.242	0.859 ± 0.178	0.572 ± 0.093	0.531 ± 0.140	0.562 ± 0.107	0.566 ± 0.141
CHEMBL287_Ki	0.963 ± 0.248	0.924 ± 0.194	0.740 ± 0.187	0.723 ± 0.172	0.706 ± 0.070	0.668 ± 0.108	0.753 ± 0.059	0.702 ± 0.098
CHEMBL2034_Ki	0.967 ± 0.319	0.889 ± 0.327	0.704 ± 0.235	0.703 ± 0.226	0.645 ± 0.111	0.579 ± 0.124	0.652 ± 0.134	0.583 ± 0.084
CHEMBL3979_EC50	0.714 ± 0.067	0.758 ± 0.063	0.552 ± 0.053	0.612 ± 0.067	0.632 ± 0.083	0.549 ± 0.067	0.697 ± 0.107	0.700 ± 0.088
CHEMBL238_Ki	0.925 ± 0.477	0.978 ± 0.343	0.746 ± 0.433	0.805 ± 0.293	0.633 ± 0.085	0.644 ± 0.066	0.717 ± 0.147	0.652 ± 0.104
CHEMBL2835_Ki	0.723 ± 0.242	1.081 ± 0.433	0.491 ± 0.159	0.903 ± 0.369	0.521 ± 0.192	0.592 ± 0.241	0.569 ± 0.230	0.595 ± 0.308
CHEMBL4005_Ki	1.157 ± 0.303	1.171 ± 0.280	0.928 ± 0.244	0.953 ± 0.237	0.560 ± 0.083	0.575 ± 0.097	0.679 ± 0.102	0.633 ± 0.051
CHEMBL237_EC50	1.589 ± 0.336	1.481 ± 0.296	1.303 ± 0.300	1.231 ± 0.272	0.591 ± 0.091	0.585 ± 0.102	0.726 ± 0.060	0.658 ± 0.069
CHEMBL264_Ki	1.158 ± 0.192	1.146 ± 0.183	0.945 ± 0.172	0.932 ± 0.141	0.625 ± 0.035	0.524 ± 0.066	0.723 ± 0.046	0.669 ± 0.079
CHEMBL214_Ki	0.962 ± 0.282	0.965 ± 0.235	0.757 ± 0.238	0.776 ± 0.206	0.577 ± 0.053	0.550 ± 0.038	0.733 ± 0.063	0.650 ± 0.063
average	1.055	1.074	0.837	0.873	0.600	0.575	0.700	0.649

Table S10. Test set prediction and global direction accuracy of GAT. For each dataset, the mean ± std of 10 individual runs has been reported.

dataset	Prediction ↓				Global direction ↑			
	$RMSE_{all}$	$RMSE_{cliff}$	MAE_{all}	MAE_{cliff}	GradInput	SmoothGrad	GradCAM	IG
CHEMBL4616_EC50	1.070 ± 0.300	1.050 ± 0.275	0.866 ± 0.264	0.849 ± 0.231	0.598 ± 0.127	0.560 ± 0.142	0.780 ± 0.047	0.759 ± 0.051
CHEMBL4792_Ki	0.947 ± 0.400	0.960 ± 0.400	0.750 ± 0.387	0.764 ± 0.380	0.537 ± 0.040	0.550 ± 0.060	0.665 ± 0.091	0.606 ± 0.064
CHEMBL1871_Ki	1.484 ± 0.573	1.579 ± 0.624	1.265 ± 0.541	1.369 ± 0.567	0.576 ± 0.089	0.542 ± 0.066	0.544 ± 0.145	0.600 ± 0.134
CHEMBL2971_Ki	1.127 ± 0.306	1.191 ± 0.322	0.920 ± 0.334	0.968 ± 0.350	0.487 ± 0.151	0.471 ± 0.116	0.529 ± 0.180	0.536 ± 0.196
CHEMBL239_EC50	1.193 ± 0.488	1.232 ± 0.454	1.002 ± 0.468	1.048 ± 0.431	0.595 ± 0.064	0.516 ± 0.062	0.641 ± 0.057	0.628 ± 0.067
CHEMBL233_Ki	1.115 ± 0.248	1.119 ± 0.213	0.889 ± 0.239	0.905 ± 0.194	0.557 ± 0.059	0.529 ± 0.041	0.628 ± 0.071	0.582 ± 0.056
CHEMBL235_EC50	1.142 ± 0.504	1.169 ± 0.369	0.937 ± 0.496	0.947 ± 0.346	0.550 ± 0.048	0.536 ± 0.072	0.594 ± 0.055	0.598 ± 0.047
CHEMBL231_Ki	1.131 ± 0.410	1.103 ± 0.378	0.854 ± 0.371	0.881 ± 0.328	0.546 ± 0.132	0.530 ± 0.107	0.556 ± 0.054	0.565 ± 0.074
CHEMBL218_EC50	1.257 ± 0.385	1.195 ± 0.315	1.025 ± 0.332	0.965 ± 0.237	0.544 ± 0.080	0.531 ± 0.084	0.538 ± 0.128	0.585 ± 0.087
CHEMBL244_Ki	1.302 ± 0.338	1.286 ± 0.280	1.067 ± 0.321	1.050 ± 0.248	0.512 ± 0.034	0.499 ± 0.043	0.605 ± 0.104	0.557 ± 0.058
CHEMBL234_Ki	1.618 ± 0.621	1.643 ± 0.738	1.393 ± 0.619	1.443 ± 0.746	0.563 ± 0.091	0.517 ± 0.070	0.686 ± 0.085	0.644 ± 0.081
CHEMBL237_Ki	1.104 ± 0.258	1.103 ± 0.274	0.863 ± 0.229	0.877 ± 0.223	0.539 ± 0.068	0.507 ± 0.066	0.619 ± 0.044	0.578 ± 0.047
CHEMBL1862_Ki	1.292 ± 0.389	1.035 ± 0.157	1.022 ± 0.316	0.800 ± 0.121	0.546 ± 0.112	0.555 ± 0.087	0.616 ± 0.074	0.599 ± 0.069
CHEMBL4203_Ki	1.054 ± 0.194	1.110 ± 0.224	0.903 ± 0.217	0.950 ± 0.189	0.549 ± 0.237	0.448 ± 0.249	0.535 ± 0.155	0.537 ± 0.214
CHEMBL2047_EC50	1.356 ± 0.889	1.348 ± 0.802	1.162 ± 0.878	1.177 ± 0.798	0.633 ± 0.094	0.526 ± 0.086	0.581 ± 0.138	0.651 ± 0.135
CHEMBL219_Ki	1.138 ± 0.387	1.149 ± 0.370	0.913 ± 0.346	0.928 ± 0.339	0.580 ± 0.066	0.555 ± 0.077	0.704 ± 0.068	0.679 ± 0.066
CHEMBL236_Ki	1.201 ± 0.448	1.248 ± 0.484	0.956 ± 0.406	1.018 ± 0.434	0.581 ± 0.066	0.553 ± 0.054	0.638 ± 0.072	0.595 ± 0.063
CHEMBL228_Ki	1.163 ± 0.276	1.085 ± 0.257	0.898 ± 0.248	0.864 ± 0.220	0.532 ± 0.050	0.550 ± 0.054	0.632 ± 0.051	0.608 ± 0.059
CHEMBL2147_Ki	1.312 ± 0.401	1.299 ± 0.477	1.071 ± 0.406	1.090 ± 0.465	0.592 ± 0.086	0.518 ± 0.089	0.644 ± 0.103	0.646 ± 0.096
CHEMBL204_Ki	1.504 ± 0.325	1.627 ± 0.380	1.224 ± 0.289	1.347 ± 0.340	0.549 ± 0.041	0.504 ± 0.043	0.647 ± 0.054	0.599 ± 0.030
CHEMBL262_Ki	1.155 ± 0.385	1.089 ± 0.300	0.995 ± 0.423	0.904 ± 0.312	0.513 ± 0.126	0.546 ± 0.147	0.531 ± 0.097	0.541 ± 0.143
CHEMBL287_Ki	0.966 ± 0.250	0.983 ± 0.163	0.751 ± 0.208	0.762 ± 0.141	0.641 ± 0.064	0.622 ± 0.076	0.719 ± 0.047	0.719 ± 0.056
CHEMBL2034_Ki	0.993 ± 0.260	0.944 ± 0.242	0.721 ± 0.212	0.727 ± 0.175	0.493 ± 0.085	0.507 ± 0.117	0.580 ± 0.086	0.539 ± 0.031
CHEMBL3979_EC50	1.312 ± 0.490	1.219 ± 0.416	1.068 ± 0.420	0.988 ± 0.338	0.544 ± 0.088	0.520 ± 0.068	0.569 ± 0.056	0.581 ± 0.084
CHEMBL238_Ki	0.942 ± 0.361	0.990 ± 0.293	0.750 ± 0.333	0.792 ± 0.235	0.612 ± 0.091	0.602 ± 0.073	0.705 ± 0.120	0.681 ± 0.106
CHEMBL2835_Ki	0.776 ± 0.254	1.080 ± 0.452	0.538 ± 0.198	0.935 ± 0.385	0.600 ± 0.362	0.584 ± 0.311	0.633 ± 0.268	0.558 ± 0.298
CHEMBL4005_Ki	1.105 ± 0.278	1.126 ± 0.274	0.899 ± 0.292	0.929 ± 0.267	0.564 ± 0.107	0.514 ± 0.081	0.558 ± 0.102	0.601 ± 0.103
CHEMBL237_EC50	1.487 ± 0.482	1.442 ± 0.382	1.208 ± 0.410	1.182 ± 0.321	0.563 ± 0.107	0.515 ± 0.109	0.651 ± 0.088	0.588 ± 0.089
CHEMBL264_Ki	1.069 ± 0.259	1.077 ± 0.253	0.872 ± 0.286	0.890 ± 0.263	0.591 ± 0.064	0.532 ± 0.042	0.652 ± 0.058	0.640 ± 0.049
CHEMBL214_Ki	1.028 ± 0.263	1.007 ± 0.234	0.806 ± 0.233	0.811 ± 0.207	0.544 ± 0.053	0.523 ± 0.054	0.660 ± 0.056	0.603 ± 0.062
average	1.178	1.183	0.953	0.972	0.561	0.532	0.621	0.607

Table S11. Test set prediction and global direction accuracy of ACES-GAT. For each dataset, the mean ± std of 10 individual runs has been reported.

dataset	Prediction ↓				Global direction ↑			
	$RMSE_{all}$	$RMSE_{cliff}$	MAE_{all}	MAE_{cliff}	GradInput	SmoothGrad	GradCAM	IG
CHEMBL4616_EC50	0.925 ± 0.168	0.916 ± 0.149	0.763 ± 0.151	0.759 ± 0.126	0.550 ± 0.115	0.561 ± 0.067	0.806 ± 0.066	0.744 ± 0.059
CHEMBL4792_Ki	0.828 ± 0.165	0.836 ± 0.149	0.648 ± 0.149	0.655 ± 0.129	0.591 ± 0.051	0.563 ± 0.052	0.743 ± 0.059	0.637 ± 0.049
CHEMBL1871_Ki	1.234 ± 0.372	1.325 ± 0.387	1.004 ± 0.343	1.080 ± 0.351	0.487 ± 0.119	0.554 ± 0.094	0.596 ± 0.117	0.641 ± 0.136
CHEMBL2971_Ki	1.213 ± 0.301	1.195 ± 0.289	0.999 ± 0.335	0.984 ± 0.276	0.509 ± 0.155	0.498 ± 0.150	0.614 ± 0.186	0.547 ± 0.099
CHEMBL239_EC50	1.128 ± 0.497	1.195 ± 0.548	0.917 ± 0.426	0.983 ± 0.485	0.536 ± 0.058	0.534 ± 0.049	0.643 ± 0.108	0.619 ± 0.071
CHEMBL233_Ki	1.064 ± 0.212	1.073 ± 0.165	0.861 ± 0.210	0.879 ± 0.151	0.569 ± 0.054	0.539 ± 0.048	0.680 ± 0.049	0.608 ± 0.045
CHEMBL235_EC50	1.023 ± 0.231	1.039 ± 0.136	0.838 ± 0.233	0.846 ± 0.120	0.549 ± 0.050	0.509 ± 0.056	0.644 ± 0.057	0.620 ± 0.063
CHEMBL231_Ki	1.112 ± 0.474	1.133 ± 0.420	0.873 ± 0.443	0.917 ± 0.367	0.543 ± 0.095	0.569 ± 0.089	0.574 ± 0.116	0.637 ± 0.100
CHEMBL218_EC50	0.974 ± 0.189	0.984 ± 0.112	0.783 ± 0.172	0.782 ± 0.089	0.571 ± 0.110	0.549 ± 0.090	0.636 ± 0.096	0.638 ± 0.112
CHEMBL244_Ki	1.343 ± 0.375	1.344 ± 0.318	1.116 ± 0.354	1.112 ± 0.282	0.528 ± 0.045	0.499 ± 0.031	0.639 ± 0.115	0.577 ± 0.073
CHEMBL234_Ki	1.274 ± 0.221	1.260 ± 0.308	1.070 ± 0.217	1.070 ± 0.319	0.657 ± 0.056	0.581 ± 0.061	0.787 ± 0.052	0.733 ± 0.067
CHEMBL237_Ki	1.137 ± 0.288	1.157 ± 0.270	0.917 ± 0.251	0.934 ± 0.218	0.545 ± 0.065	0.505 ± 0.049	0.667 ± 0.069	0.596 ± 0.066
CHEMBL1862_Ki	1.283 ± 0.427	1.028 ± 0.193	1.035 ± 0.350	0.796 ± 0.157	0.615 ± 0.090	0.565 ± 0.056	0.650 ± 0.091	0.628 ± 0.066
CHEMBL4203_Ki	1.053 ± 0.134	1.117 ± 0.172	0.862 ± 0.149	0.918 ± 0.170	0.508 ± 0.239	0.452 ± 0.128	0.455 ± 0.176	0.589 ± 0.182
CHEMBL2047_EC50	0.974 ± 0.431	0.948 ± 0.367	0.804 ± 0.454	0.797 ± 0.378	0.603 ± 0.113	0.560 ± 0.130	0.731 ± 0.159	0.737 ± 0.099
CHEMBL219_Ki	0.962 ± 0.181	1.000 ± 0.178	0.751 ± 0.158	0.806 ± 0.159	0.568 ± 0.064	0.552 ± 0.061	0.720 ± 0.065	0.682 ± 0.062
CHEMBL236_Ki	1.125 ± 0.350	1.150 ± 0.357	0.894 ± 0.310	0.943 ± 0.318	0.562 ± 0.060	0.533 ± 0.046	0.674 ± 0.055	0.610 ± 0.071
CHEMBL228_Ki	1.026 ± 0.164	0.974 ± 0.139	0.795 ± 0.164	0.783 ± 0.143	0.597 ± 0.039	0.573 ± 0.064	0.712 ± 0.067	0.633 ± 0.060
CHEMBL2147_Ki	1.343 ± 0.405	1.144 ± 0.285	1.106 ± 0.414	0.929 ± 0.249	0.611 ± 0.096	0.554 ± 0.104	0.718 ± 0.087	0.692 ± 0.067
CHEMBL204_Ki	1.453 ± 0.320	1.574 ± 0.400	1.182 ± 0.276	1.297 ± 0.347	0.542 ± 0.064	0.530 ± 0.026	0.693 ± 0.064	0.631 ± 0.073
CHEMBL262_Ki	1.033 ± 0.233	1.019 ± 0.230	0.842 ± 0.183	0.810 ± 0.184	0.513 ± 0.122	0.532 ± 0.102	0.585 ± 0.094	0.579 ± 0.127
CHEMBL287_Ki	0.925 ± 0.224	0.935 ± 0.171	0.712 ± 0.169	0.731 ± 0.136	0.680 ± 0.066	0.626 ± 0.071	0.749 ± 0.092	0.722 ± 0.067
CHEMBL2034_Ki	0.916 ± 0.235	0.858 ± 0.211	0.689 ± 0.194	0.697 ± 0.151	0.536 ± 0.130	0.520 ± 0.110	0.645 ± 0.126	0.510 ± 0.094
CHEMBL3979_EC50	1.199 ± 0.387	1.085 ± 0.310	0.974 ± 0.317	0.885 ± 0.236	0.546 ± 0.102	0.512 ± 0.078	0.656 ± 0.091	0.647 ± 0.060
CHEMBL238_Ki	0.896 ± 0.281	0.933 ± 0.200	0.728 ± 0.254	0.746 ± 0.174	0.557 ± 0.072	0.544 ± 0.076	0.675 ± 0.141	0.624 ± 0.107
CHEMBL2835_Ki	0.773 ± 0.260	1.091 ± 0.418	0.539 ± 0.218	0.935 ± 0.358	0.689 ± 0.139	0.601 ± 0.239	0.758 ± 0.185	0.681 ± 0.192
CHEMBL4005_Ki	0.960 ± 0.152	0.947 ± 0.178	0.764 ± 0.136	0.765 ± 0.147	0.561 ± 0.102	0.541 ± 0.097	0.666 ± 0.098	0.628 ± 0.090
CHEMBL237_EC50	1.327 ± 0.312	1.286 ± 0.244	1.067 ± 0.235	1.050 ± 0.168	0.543 ± 0.075	0.546 ± 0.077	0.679 ± 0.059	0.586 ± 0.082
CHEMBL264_Ki	0.985 ± 0.157	0.992 ± 0.144	0.787 ± 0.153	0.799 ± 0.127	0.559 ± 0.097	0.516 ± 0.082	0.691 ± 0.061	0.651 ± 0.057
CHEMBL214_Ki	0.963 ± 0.204	0.990 ± 0.182	0.765 ± 0.193	0.804 ± 0.166	0.552 ± 0.059	0.510 ± 0.050	0.681 ± 0.090	0.599 ± 0.067
average	1.082	1.084	0.870	0.883	0.566	0.541	0.672	0.634

Table S12. Test set prediction and global direction accuracy of RF combined with ECFP. Each reported value represents the mean ± standard deviation (SD) over 10 independent runs. For each dataset, two types of variation are reported: variation due to model random initialization (seed SD) and variation due to different data splits (split SD). The attribution value for each atom is obtained via the atomic masking method.[2]

dataset	Prediction ↓						Global direction ↑	
	$RMSE_{all}$ (seed SD)	$RMSE_{all}$ (split SD)	$RMSE_{cliff}$ (seed SD)	$RMSE_{cliff}$ (split SD)	MAE_{all} (split SD)	MAE_{cliff}(split SD)	Mask (seed SD)	mask (split SD)
CHEMBL4616_EC50	0.611 ± 0.008	0.644 ± 0.038	0.640 ± 0.006	0.731 ± 0.076	0.494 ± 0.036	0.574 ± 0.081	0.940 ± 0.011	0.913 ± 0.041
CHEMBL4792_Ki	0.682 ± 0.006	0.680 ± 0.042	0.682 ± 0.007	0.715 ± 0.055	0.511 ± 0.023	0.553 ± 0.030	0.891 ± 0.013	0.885 ± 0.035
CHEMBL1871_Ki	0.654 ± 0.007	0.634 ± 0.078	0.810 ± 0.011	0.707 ± 0.152	0.491 ± 0.055	0.562 ± 0.126	0.729 ± 0.065	0.734 ± 0.086
CHEMBL2971_Ki	0.663 ± 0.010	0.662 ± 0.078	0.607 ± 0.014	0.775 ± 0.140	0.432 ± 0.040	0.612 ± 0.100	0.955 ± 0.022	0.862 ± 0.115
CHEMBL239_EC50	0.768 ± 0.006	0.685 ± 0.045	0.880 ± 0.009	0.823 ± 0.064	0.520 ± 0.039	0.665 ± 0.046	0.717 ± 0.012	0.828 ± 0.051
CHEMBL233_Ki	0.810 ± 0.004	0.724 ± 0.045	0.902 ± 0.007	0.779 ± 0.064	0.539 ± 0.030	0.614 ± 0.049	0.827 ± 0.009	0.811 ± 0.047
CHEMBL235_EC50	0.672 ± 0.007	0.672 ± 0.041	0.846 ± 0.010	0.805 ± 0.047	0.490 ± 0.035	0.640 ± 0.047	0.846 ± 0.016	0.835 ± 0.037
CHEMBL231_Ki	0.798 ± 0.010	0.736 ± 0.089	0.690 ± 0.016	0.808 ± 0.133	0.535 ± 0.057	0.641 ± 0.110	0.718 ± 0.046	0.807 ± 0.093
CHEMBL218_EC50	0.729 ± 0.009	0.693 ± 0.058	0.808 ± 0.010	0.802 ± 0.123	0.530 ± 0.037	0.643 ± 0.087	0.886 ± 0.022	0.840 ± 0.056
CHEMBL244_Ki	0.727 ± 0.006	0.783 ± 0.040	0.834 ± 0.007	0.868 ± 0.038	0.579 ± 0.029	0.679 ± 0.038	0.817 ± 0.013	0.813 ± 0.023
CHEMBL234_Ki	0.642 ± 0.003	0.642 ± 0.017	0.603 ± 0.006	0.651 ± 0.038	0.479 ± 0.012	0.500 ± 0.032	0.904 ± 0.007	0.895 ± 0.027
CHEMBL237_Ki	0.742 ± 0.004	0.711 ± 0.034	0.853 ± 0.008	0.786 ± 0.059	0.537 ± 0.022	0.621 ± 0.040	0.789 ± 0.017	0.808 ± 0.049
CHEMBL1862_Ki	0.754 ± 0.010	0.759 ± 0.063	0.752 ± 0.007	0.719 ± 0.065	0.573 ± 0.055	0.566 ± 0.038	0.811 ± 0.017	0.860 ± 0.071
CHEMBL4203_Ki	0.787 ± 0.012	0.806 ± 0.069	1.030 ± 0.029	0.982 ± 0.250	0.622 ± 0.045	0.852 ± 0.233	0.512 ± 0.067	0.776 ± 0.119
CHEMBL2047_EC50	0.557 ± 0.011	0.590 ± 0.054	0.607 ± 0.020	0.683 ± 0.083	0.435 ± 0.042	0.526 ± 0.058	0.864 ± 0.012	0.874 ± 0.042
CHEMBL219_Ki	0.715 ± 0.008	0.712 ± 0.037	0.699 ± 0.011	0.789 ± 0.066	0.534 ± 0.031	0.626 ± 0.059	0.906 ± 0.008	0.846 ± 0.048
CHEMBL236_Ki	0.648 ± 0.005	0.690 ± 0.041	0.783 ± 0.007	0.801 ± 0.058	0.509 ± 0.026	0.627 ± 0.050	0.829 ± 0.014	0.822 ± 0.017
CHEMBL228_Ki	0.761 ± 0.003	0.751 ± 0.037	0.876 ± 0.007	0.815 ± 0.079	0.550 ± 0.029	0.641 ± 0.073	0.829 ± 0.015	0.829 ± 0.040
CHEMBL2147_Ki	0.698 ± 0.008	0.695 ± 0.060	0.630 ± 0.017	0.705 ± 0.110	0.512 ± 0.041	0.536 ± 0.076	0.847 ± 0.015	0.832 ± 0.056
CHEMBL204_Ki	0.714 ± 0.009	0.758 ± 0.037	0.838 ± 0.012	0.866 ± 0.069	0.552 ± 0.027	0.675 ± 0.053	0.836 ± 0.013	0.855 ± 0.025
CHEMBL262_Ki	0.747 ± 0.009	0.695 ± 0.080	0.735 ± 0.009	0.751 ± 0.132	0.545 ± 0.061	0.632 ± 0.128	0.772 ± 0.028	0.869 ± 0.090
CHEMBL287_Ki	0.747 ± 0.007	0.760 ± 0.063	0.777 ± 0.006	0.788 ± 0.105	0.576 ± 0.046	0.641 ± 0.079	0.824 ± 0.019	0.848 ± 0.064
CHEMBL2034_Ki	0.776 ± 0.014	0.646 ± 0.134	0.770 ± 0.010	0.664 ± 0.127	0.457 ± 0.071	0.525 ± 0.094	0.860 ± 0.014	0.836 ± 0.064
CHEMBL3979_EC50	0.662 ± 0.006	0.658 ± 0.041	0.770 ± 0.005	0.747 ± 0.083	0.514 ± 0.028	0.609 ± 0.060	0.855 ± 0.021	0.796 ± 0.062
CHEMBL238_Ki	0.604 ± 0.006	0.649 ± 0.064	0.626 ± 0.007	0.774 ± 0.117	0.499 ± 0.048	0.646 ± 0.118	0.893 ± 0.038	0.838 ± 0.060
CHEMBL2835_Ki	0.357 ± 0.009	0.414 ± 0.056	0.968 ± 0.036	0.800 ± 0.242	0.275 ± 0.041	0.714 ± 0.235	0.860 ± 0.092	0.865 ± 0.125
CHEMBL4005_Ki	0.659 ± 0.010	0.613 ± 0.049	0.720 ± 0.007	0.692 ± 0.057	0.459 ± 0.032	0.539 ± 0.051	0.822 ± 0.025	0.880 ± 0.059
CHEMBL237_EC50	0.892 ± 0.008	0.838 ± 0.054	0.906 ± 0.012	0.875 ± 0.097	0.654 ± 0.035	0.711 ± 0.076	0.896 ± 0.011	0.803 ± 0.041
CHEMBL264_Ki	0.678 ± 0.003	0.635 ± 0.030	0.702 ± 0.004	0.711 ± 0.061	0.471 ± 0.023	0.558 ± 0.039	0.897 ± 0.007	0.866 ± 0.026
CHEMBL214_Ki	0.656 ± 0.005	0.672 ± 0.030	0.740 ± 0.007	0.766 ± 0.048	0.502 ± 0.023	0.598 ± 0.045	0.800 ± 0.015	0.829 ± 0.024
average	0.697	0.687	0.770	0.773	0.513	0.618	0.831	0.838

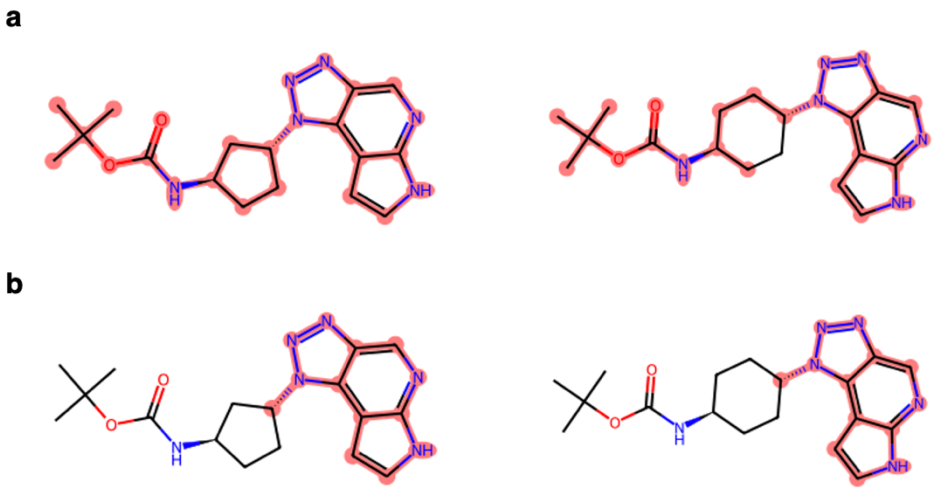

Figure S1. Exemplary illustration of maximum common substructure with different FindMCS configuration settings in rdkit. (a) 'matchValences', 'ringMatchesRingOnly', and 'completeRingOnly' are set to False. (b) 'matchValences', 'ringMatchesRingOnly', and 'completeRingOnly' are set to True.

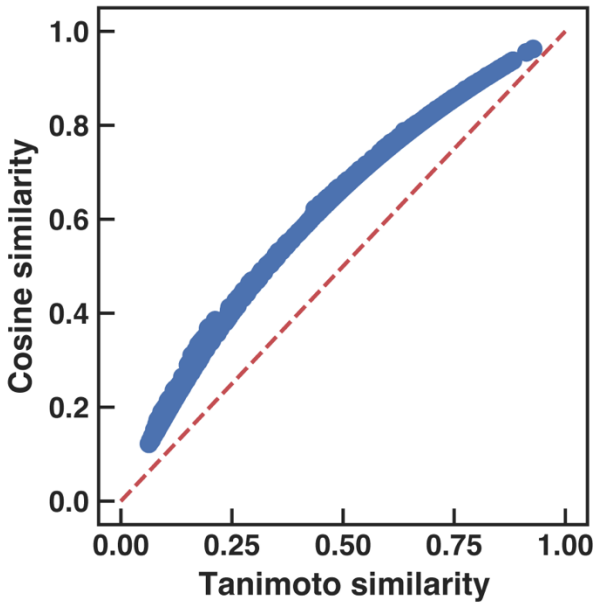

Figure S2. Comparison of cosine similarity and Tanimoto similarity to measure the similarity between fingerprints. The figure shows all pairwise ECFP similarity using either cosine or Tanimoto similarity metrics for molecules in CHEMBL2835 dataset. This two metrics exhibit similar values of similarity when two molecules are very similar (>0.90) or dissimilar (<0.10).

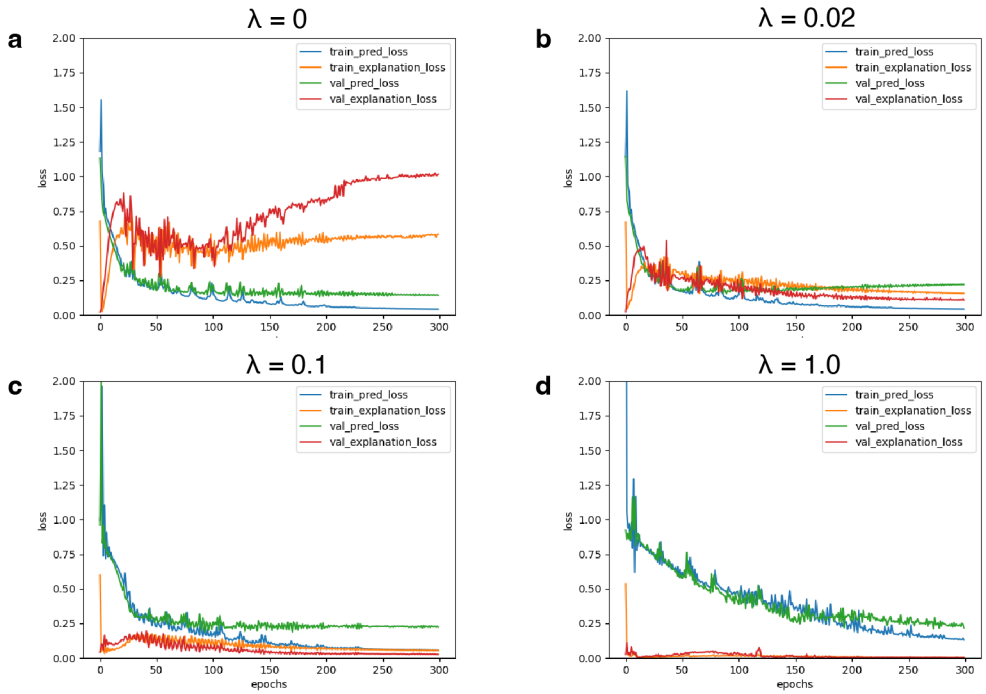

Figure S3. Comparison of training dynamics with various explanation loss scaling factors. The figures show the training dynamics (epoch vs. loss) at various explanation loss scaling factors (λ). The blue, orange, green, red color lines represent training prediction loss, training explanation loss, validation prediction loss, validation explanation loss, respectively.

Text S1 Impact of scaling factor on the training dynamics

We observed an interesting phenomenon while selecting the optimal explanation loss scaling factor. When this scaling factor was chosen to be larger values, the explanation performance can almost always improve, however, the prediction performance may be sufficiently jeopardized. By inspecting the training dynamics (Figure S3), we found that the introduction of a large explanation loss can sufficiently affect the optimization of training loss, thus leading to an irreversible prediction performance degradation. By grid-searching a relatively small scaling factor based on the performance of validation prediction performance, we found most of the target data set can improve both the explanation accuracy and the prediction performance.

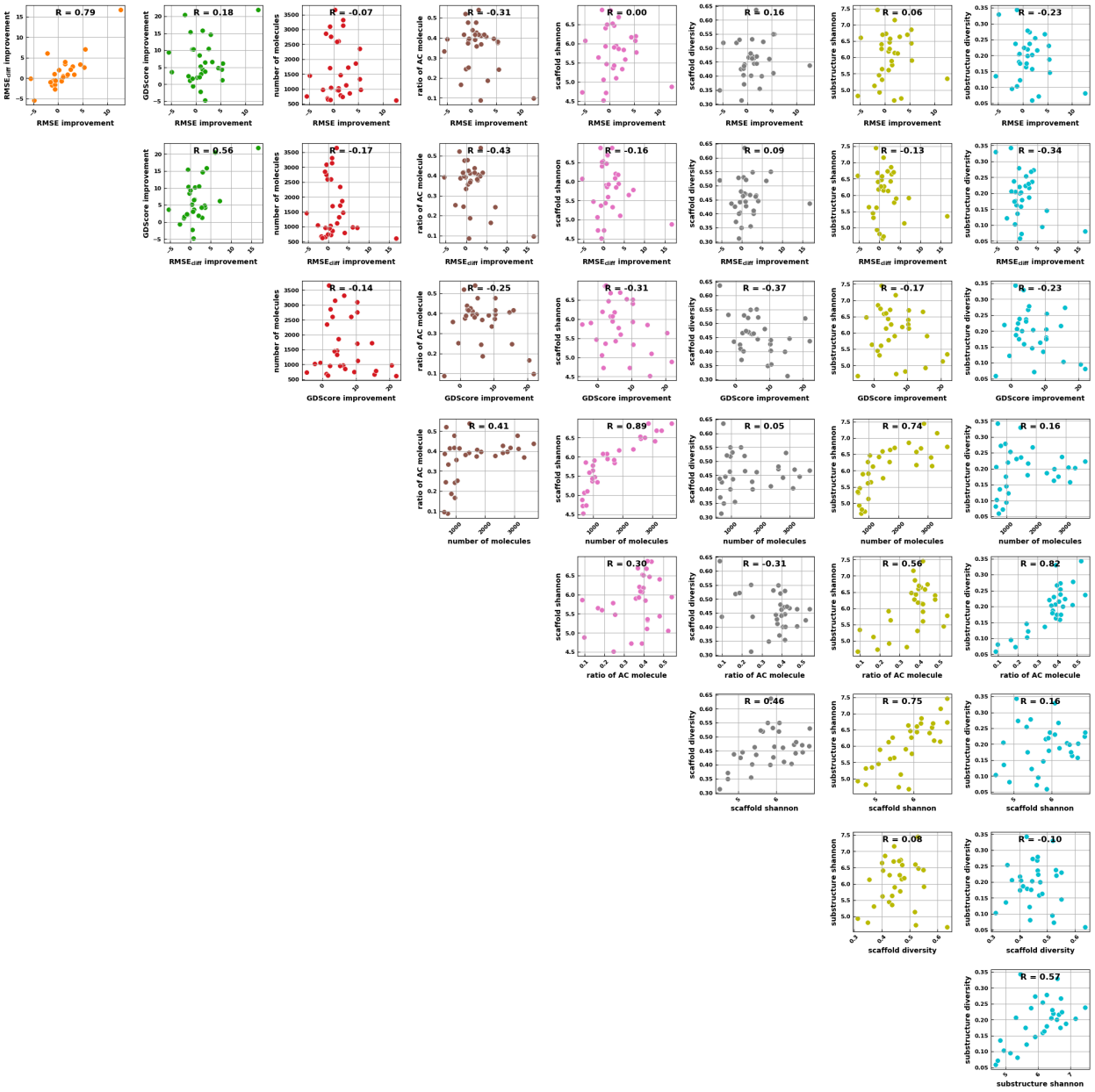

Figure S4. Correlations between various performance metrics and possible contributing factors. The scatter plots along with Pearson correlation coefficient (R) between the improvement of test set RMSE, $RMSE_{cliff}$ and global direction score (GDScore), number of molecules in the dataset, ratio of AC molecules in the dataset, scaffold diversity, scaffold Shannon entropy, ground-truth substructure diversity and ground-truth substructure Shannon entropy for the 30 target datasets.

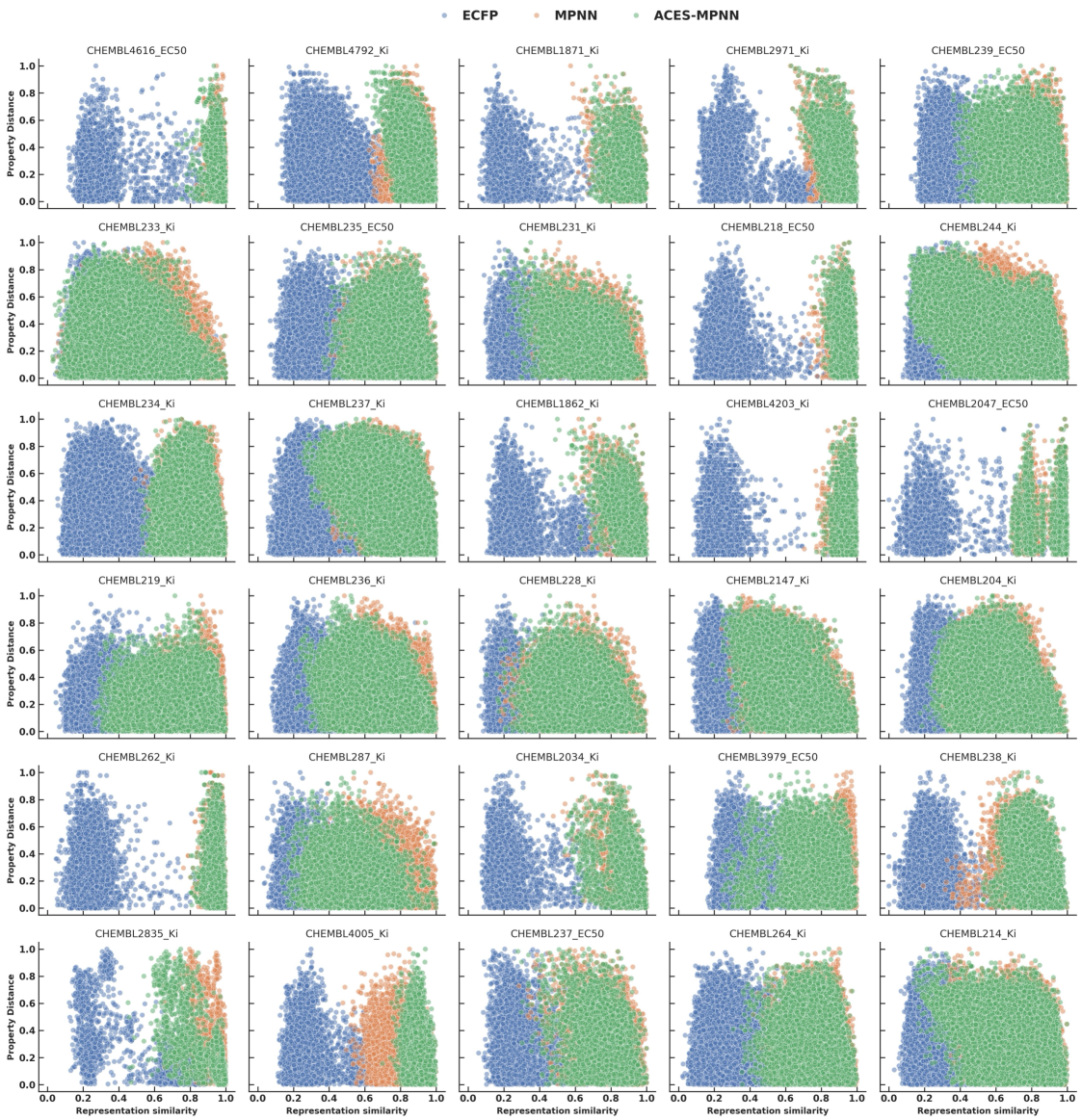

Figure S5. RPSMap of all 30 target datasets. The blue, orange, green dots represent the distribution of the representations generated from ECFP, MPNN and ACES-MPNN, respectively. All the data points are from the respective test set molecules. For many datasets (e.g., CHEMBL 4792, CHEMBL219, CHEMBL3979, all with improved RMSE and $RMSE_{cliff}$), we can observe the orange dots (MPNN) occupying the upper right corner of the RPSMap while green dots (ACES-MPNN) do not, indicating that the MPNN-generated representations are improved after activity-cliff explanation supervision. ECFP representations (blue) always locate at the left side of the RMSMap, occupying distinct region compared to GNN generated representations.

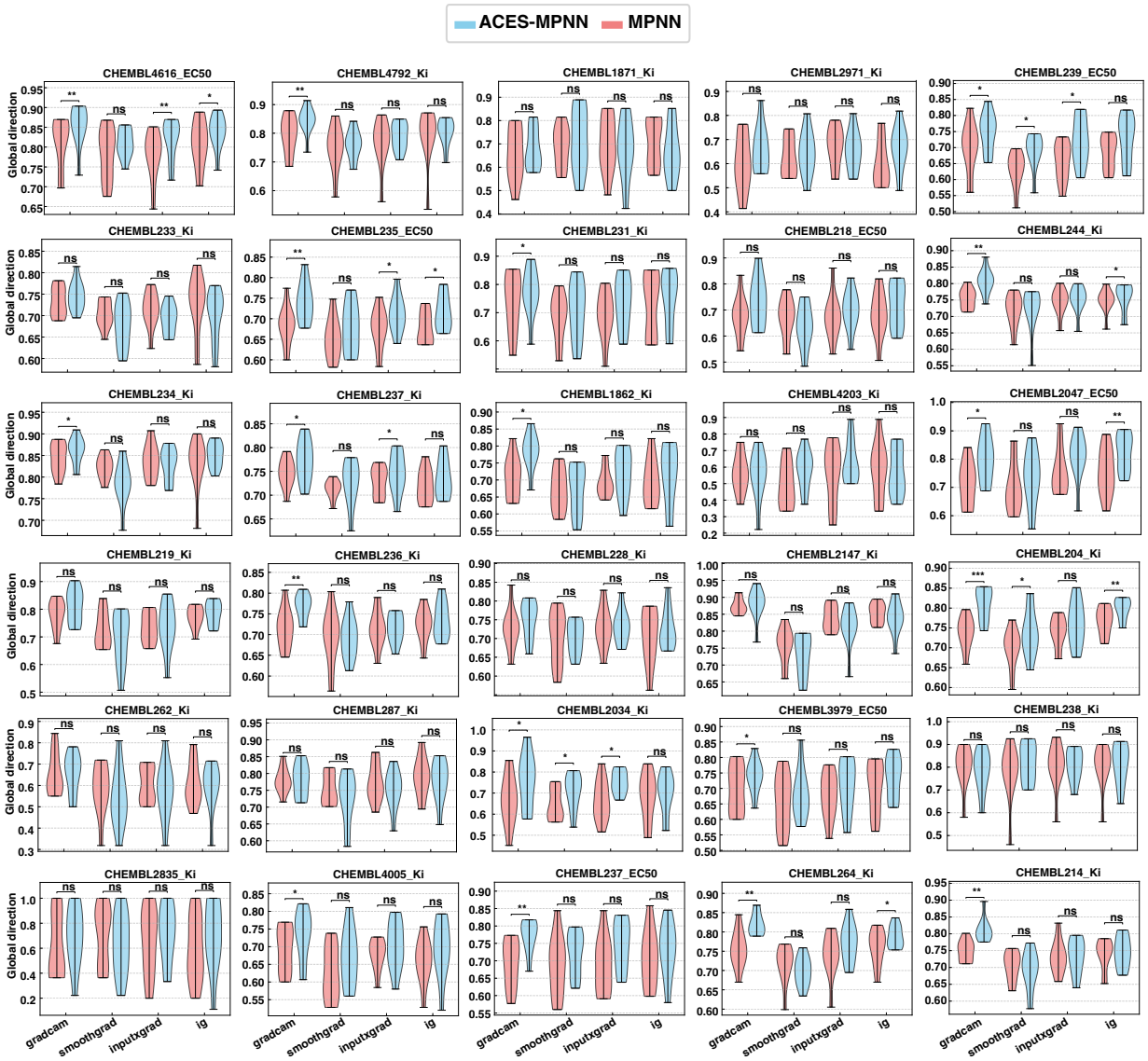

Figure S6. Dataset-wise global-direction performance for various gradient-based attribution methods. Violin plots show the distribution of global-direction scores for ACES-MPNN (blue) and MPNN (red) on each dataset, evaluated using four different attribution methods (GradCAM, SmoothGrad, GradInput and IG). Each violin represents the distribution of scores obtained from 10 independent runs. Asterisks indicate the statistical significance of performance differences between the two models based on the Wilcoxon signed-rank test: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), and $p < 0.0001$ (****); 'ns' denotes non-significant differences ($p \geq 0.05$).

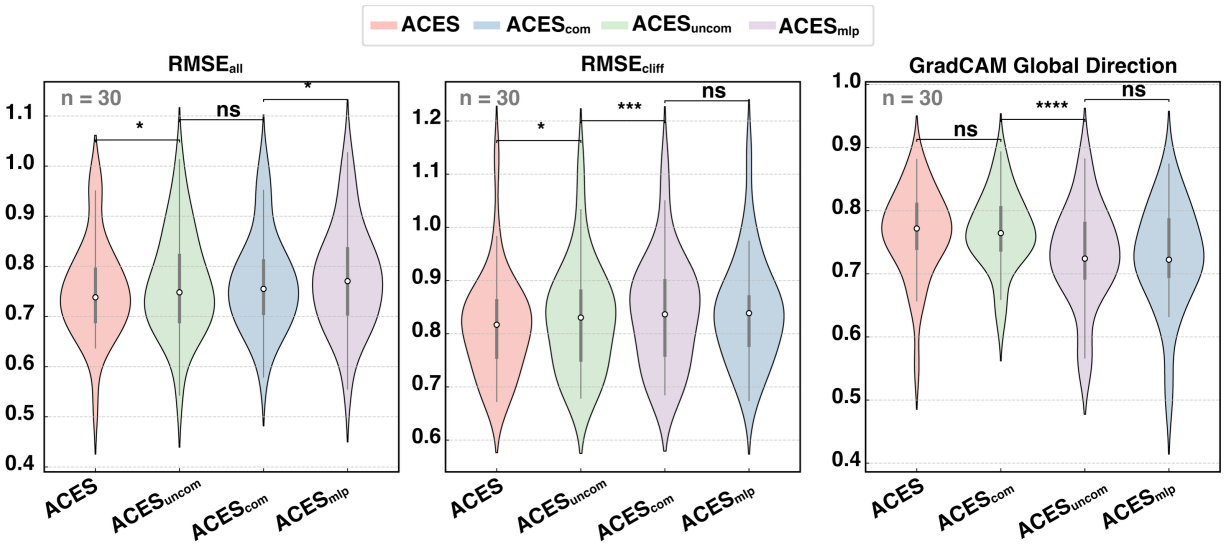

Figure S7. Comparison of predictive and attribution performance across ACES model variants. Violin plots summarize the distribution of (left) RMSE on all test samples ($RMSE_{all}$), (middle) RMSE on activity cliff compounds ($RMSE_{cliff}$), and (right) GradCAM global direction scores across 30 benchmark datasets. Each value represents the average score over 10 random stratified scaffold splits per target. The figure compares four ACES variants: standard ACES, ACES with common part explanation supervision only ($ACES_{com}$), ACES with uncommon part explanations supervision ($ACES_{uncom}$), and ACES with both part of explanation supervision but GradCAM is replaced with a multi-layer perceptron (MLP) ($ACES_{mlp}$). Statistical comparisons were conducted using the Wilcoxon signed-rank test. Asterisks denote significance levels: $p < 0.05$ (*), $p < 0.001$ (***), $p < 0.0001$ (****); 'ns' indicates non-significant differences ($p \geq 0.05$). Each violin represents the distribution over datasets; white circles denote the median values. The width of each violin represents the density of data points, with wider sections indicating a higher concentration of values.

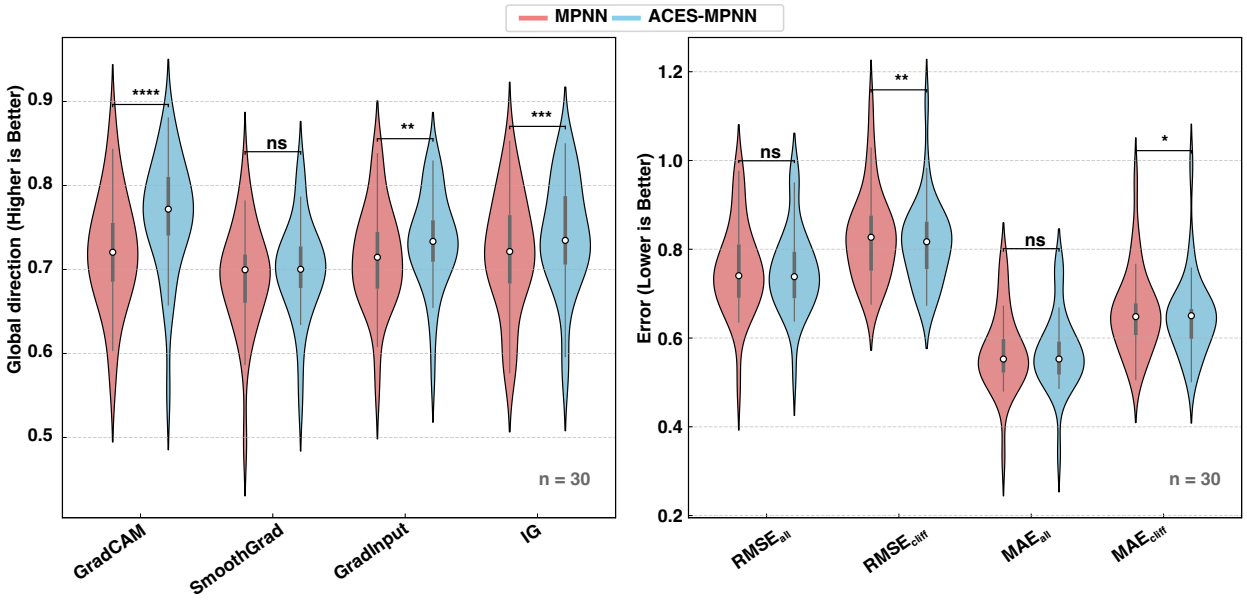

Figure S8. Comparison of attribution quality and predictive performance between baseline MPNN and ACES-MPNN models across 30 benchmark targets. (Left) Violin plots show the distribution of global direction scores for four gradient-based methods (GradCAM, SmoothGrad, GradInput, and Integrated Gradients). Higher scores indicate better alignment. (Right) Violin plots show the distribution of predictive errors, including RMSE and MAE for all test compounds (subscript all) and for activity cliff compounds (subscript cliff). Lower values indicate better predictive accuracy. In both cases, each value represents the average score across 10 random stratified splits. White circles denote the median values across the 30 targets. Asterisks denote statistical significance based on Wilcoxon signed-rank tests: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), $p < 0.0001$ (****); 'ns' denotes non-significant differences ($p \geq 0.05$).

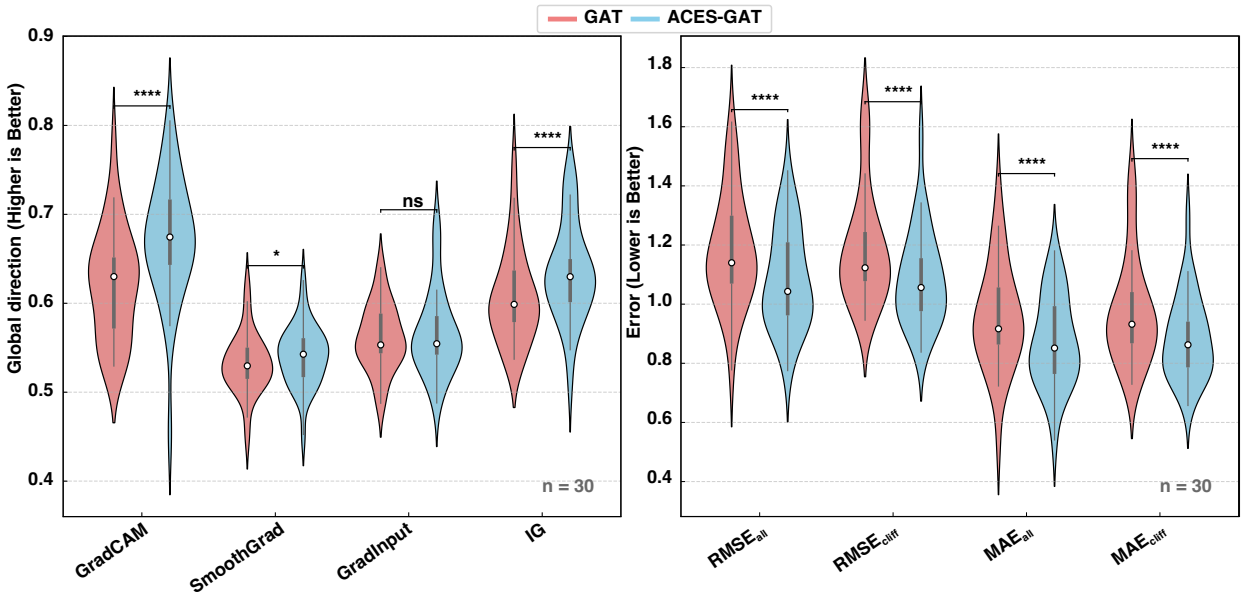

Figure S9. Comparison of attribution quality and predictive performance between baseline GAT and ACES-GAT models across 30 benchmark targets. (Left) Violin plots show the distribution of global direction scores for four gradient-based methods (GradCAM, SmoothGrad, GradInput, and Integrated Gradients). Higher scores indicate better alignment. (Right) Violin plots show the distribution of predictive errors, including RMSE and MAE for all test compounds (subscript all) and for activity cliff compounds (subscript cliff). Lower values indicate better predictive accuracy. In both cases, each value represents the average score across 10 random stratified splits. White circles denote the median values across the 30 targets. Asterisks denote statistical significance based on Wilcoxon signed-rank tests: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), $p < 0.0001$ (****); 'ns' denotes non-significant differences ($p \geq 0.05$).

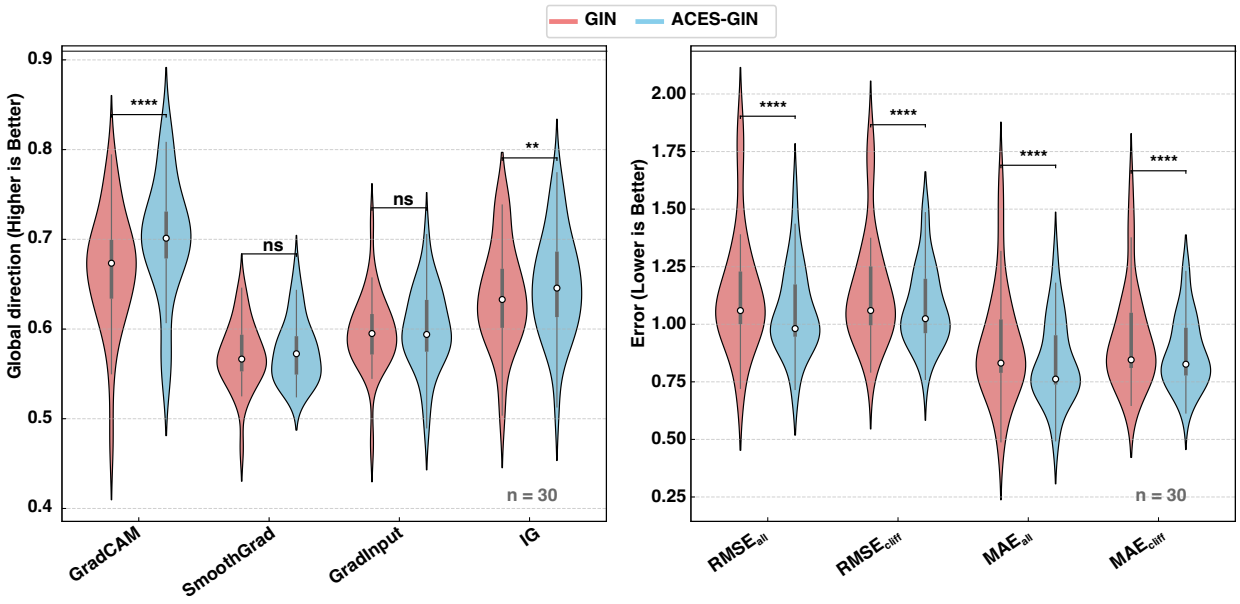

Figure S10. Comparison of attribution quality and predictive performance between baseline GIN and ACES-GIN models across 30 benchmark targets. (Left) Violin plots show the distribution of global direction scores for four gradient-based methods (GradCAM, SmoothGrad, GradInput, and Integrated Gradients). Higher scores indicate better alignment. (Right) Violin plots show the distribution of predictive errors, including RMSE and MAE for all test compounds (subscript all) and for activity cliff compounds (subscript cliff). Lower values indicate better predictive accuracy. In both cases, each value represents the average score across 10 random stratified splits. White circles denote the median values across the 30 targets. Asterisks denote statistical significance based on Wilcoxon signed-rank tests: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), $p < 0.0001$ (****); 'ns' denotes non-significant differences ($p \geq 0.05$).

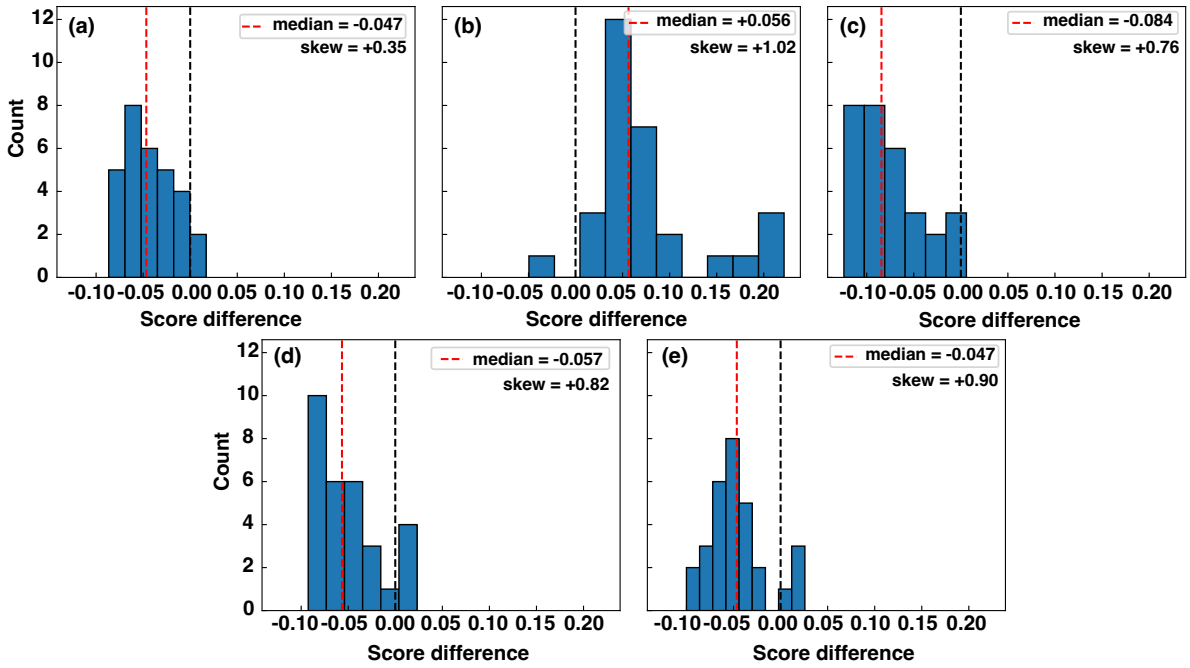

Figure S11. Paired-difference histograms (30 targets) comparing each attribution method to the ACES-MPNN combined with GradCAM. For every target, the metric value of our method is subtracted from that of the alternative method, i.e., (a) GradCAM - ours; (b) RF Mask - ours; (c) SmoothGrad - ours; (d) GradInput - ours; (e) IG - ours; positive bars therefore indicate superior performance of the alternative. Dashed red lines mark the sample medians, dashed black lines denote zero (no difference). Reported skewness coefficients confirm substantial asymmetry in all distributions, justifying the use of the Wilcoxon signed-rank test instead of a paired *t*-test.

Text S2. Paired-difference analysis for attribution methods Histograms of the paired score differences (Figure S11) reveal that the distributions deviate markedly from normality: four of the five comparison groups are left-skewed (median < 0), whereas the RF-Mask differences are right-skewed. Sample skew coefficients range from +0.35 to +1.02, and Shapiro–Wilk tests reject normality for every set of differences (all $p < 0.01$). Because the critical assumption of a symmetric, approximately Gaussian distribution required by the paired *t*-test is violated, we adopted the Wilcoxon signed-rank test. This non-parametric procedure operates on the ranks of the within-pair differences and therefore provides valid control of the type I error rate under the observed skewness while incurring only a modest loss of power relative to the parametric alternative.

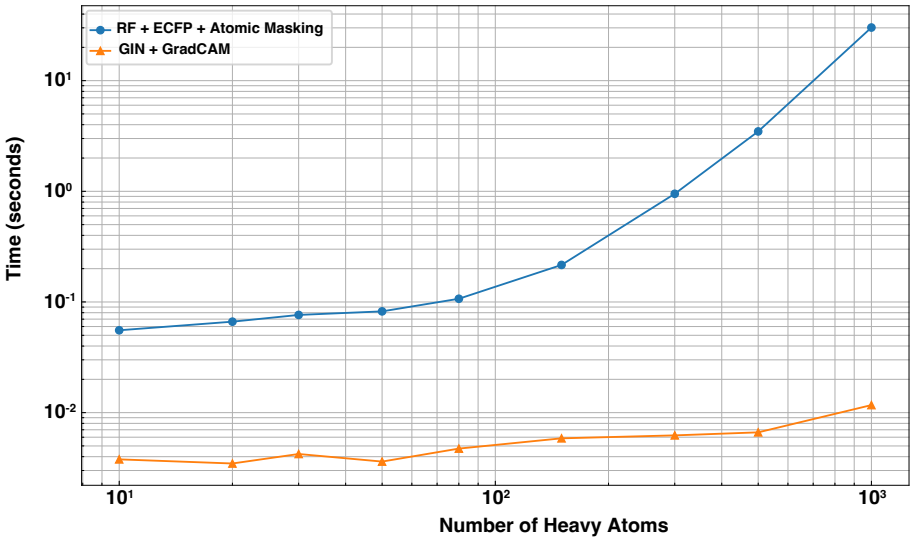

Figure S12. Computational time scaling for atomic attribution methods. Comparison of the time required to generate atomic attributions as a function of molecule size (number of heavy atoms). The methods evaluated are: (i) Random Forest (n_estimators=500) with ECFP4 fingerprints and atomic masking (RF + ECFP + Atomic Masking; blue circles), and (ii) a Graph Isomorphism Network (5 layers with 128 hidden dimension each) with GradCAM (GIN + GradCAM; orange triangles). Runtimes are reported in seconds. Both the x-axis (Number of Heavy Atoms) and y-axis (Time (seconds)) are presented on a logarithmic scale. The GIN + GradCAM method exhibits substantially lower computational cost and more favorable scaling with increasing molecular size compared to the RF-based approach.

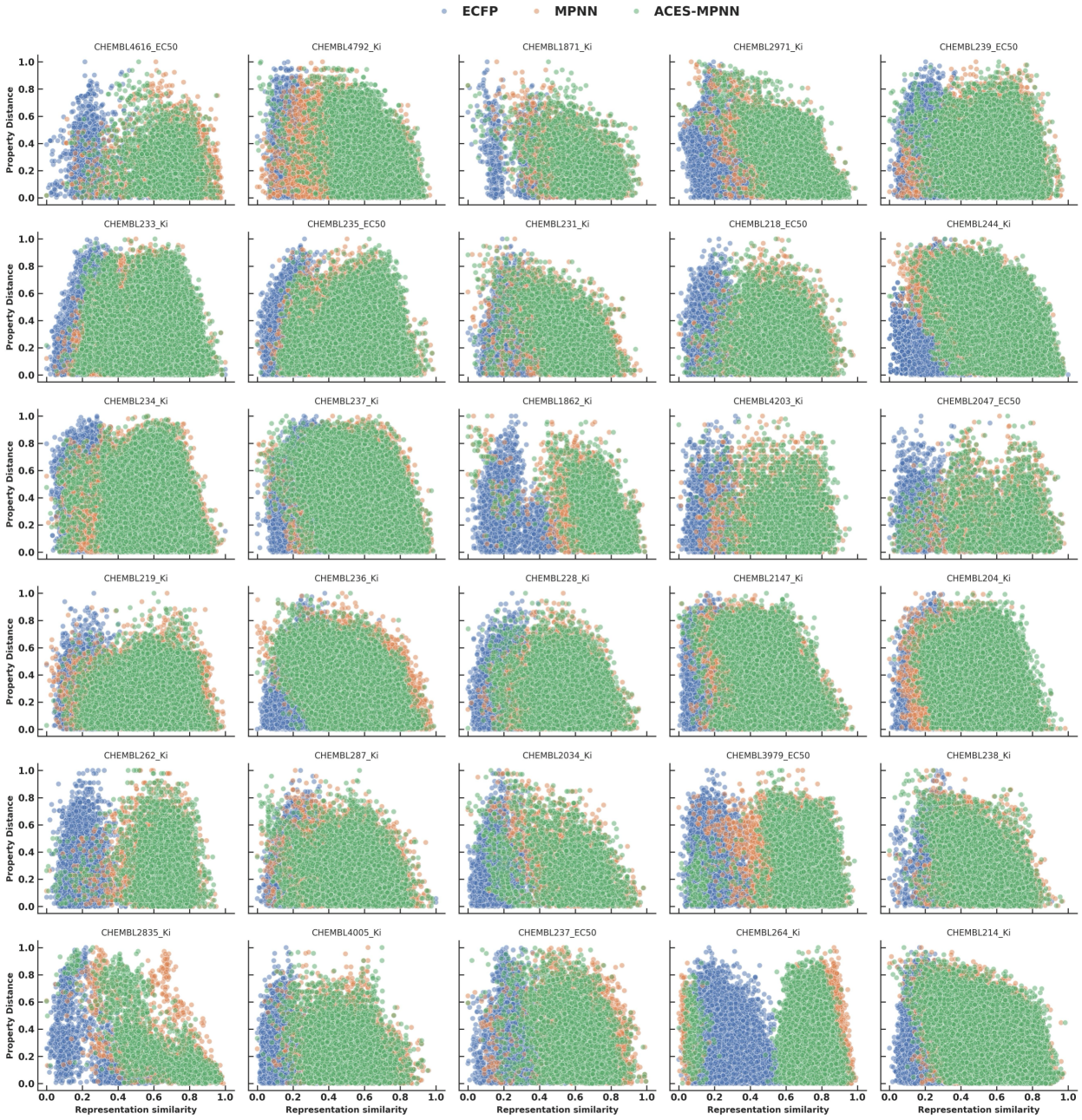

Figure S13. RPSMap of all 30 target datasets using normalized Euclidean distance as the representation similarity metric. The blue, orange, green dots represent the distribution of the representations generated from ECFP, MPNN and ACES-MPNN, respectively. All the data points are from the respective test set molecules. The normalized Euclidean distance is used as the representation similarity function.

Text S3. Normalized Euclidean similarity metric function

As discussed in the previous study,[3] Euclidean distance cannot be directly exploited in the similarity measure as the scale of the representations for different molecules may be different. As a result, the MinMax normalized Euclidean distance is transformed to similarity for the

visualization in Figure S13. Specifically, if the distance between two molecular representations ($\boldsymbol{z}_i$ and $\boldsymbol{z}_j$) is denoted as $d_{ij} = \left\|\boldsymbol{z}_i - \boldsymbol{z}_j\right\|$, then defined as $d_z\left(\boldsymbol{z}_i, \boldsymbol{z}_j\right) = \left(d_{ij} - \min_{i,j}(d_{ij})\right) / \left(\max_{i,j}(d_{ij}) - \min_{i,j}(d_{ij})\right)$. The similarity between two representations is defined as $sim_z\left(\boldsymbol{z}_i, \boldsymbol{z}_j\right) = 1 - d_z\left(\boldsymbol{z}_i, \boldsymbol{z}_j\right)$.

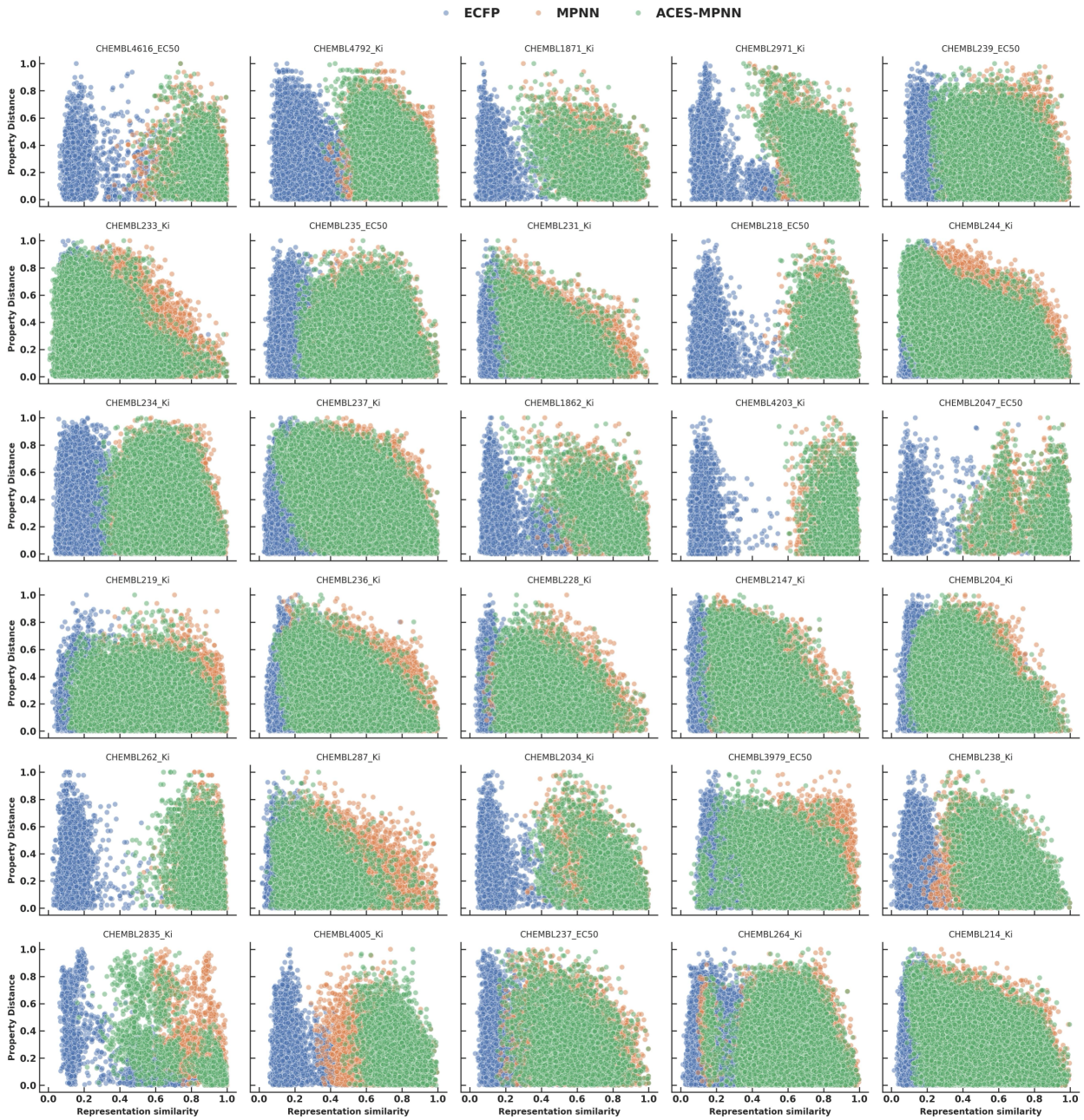

Figure S14. RPSMap of all 30 target datasets using Tanimoto as the representation similarity metric. The blue, orange, green dots represent the distribution of the representations generated from ECFP, MPNN and ACES-MPNN, respectively. All the data points are from the respective test set molecules

References

1. Bergstra, J.; Komer, B.; Eliasmith, C.; Yamins, D.; Cox, D. D., Hyperopt: a python library for model selection and hyperparameter optimization. *Computational Science & Discovery* **2015,** *8* (1), 014008.
2. Sheridan, R. P., Interpretation of QSAR models by coloring atoms according to changes in predicted activity: how robust is it? *Journal of chemical information and modeling* **2019,** *59* (4), 1324-1337.
3. Zhang, Z.; Bian, Y.; Xie, A.; Han, P.; Zhou, S., Can Pretrained Models Really Learn Better Molecular Representations for AI-Aided Drug Discovery? *Journal of Chemical Information and Modeling* **2023**.