Supporting Information

Local structure–activity landscape roughness detects the activity-cliff failure mode that standard QSAR reliability methods miss

Krishna Harish
Elkins High School, Missouri City, Texas, United States
krishnaharish2009@gmail.com

This document contains: per-target Spearman correlations between each roughness descriptor and per-compound random-forest error, with the applicability-domain-controlled partial correlation and the activity-cliff fraction, for all 30 benchmark targets (Table S1); the full robustness sweep across neighborhood size k and molecular distance metric (Table S2); and extended graph-network architecture and training details with per-target errors (Table S3). All values are regenerated from the analysis pipeline by `src/make_si_tables.py`.

S1. Per-target correlations

For each target, the Spearman rank correlation ρ between each descriptor and the per-compound random-forest error is reported, computed on that target's test set. "Cliff %" is the fraction of test compounds that participate in at least one benchmark-labeled activity cliff. "Nbr. disp. | AD" is the partial Spearman correlation of the activity-free neighborhood dispersion with error after controlling (by rank residualization) for nearest-neighbor similarity and local density. Dirichlet and Lipschitz use the query's own activity; neighborhood dispersion and SALI density do not.

Table S1: Per-target Spearman correlations with per-compound random-forest error, and activity-cliff fractions, for all 30 MoleculeACE targets ($k = 10$, ECFP4/Tanimoto).

Target	n	Cliff %	Dirichlet	Lipschitz	Nbr. disp.	SALI mean	Nbr. disp. | AD
CHEMBL1862_Ki	161	41.6	0.62	0.38	0.33	0.23	0.25
CHEMBL1871_Ki	134	23.9	0.76	0.51	0.32	0.21	0.25
CHEMBL2034_Ki	152	34.2	0.77	0.65	0.45	0.34	0.35
CHEMBL2047_EC50	128	39.1	0.69	0.51	0.34	0.28	0.29
CHEMBL204_Ki	553	39.8	0.64	0.53	0.40	0.30	0.41
CHEMBL2147_Ki	294	39.5	0.50	0.40	0.19	0.15	0.16
CHEMBL214_Ki	666	36.8	0.65	0.43	0.27	0.16	0.22
CHEMBL218_EC50	208	36.5	0.69	0.44	0.30	0.16	0.26
CHEMBL219_Ki	374	39.6	0.70	0.50	0.35	0.15	0.29
CHEMBL228_Ki	342	37.1	0.65	0.45	0.25	0.12	0.19
CHEMBL231_Ki	197	24.4	0.67	0.45	0.32	0.22	0.21
CHEMBL233_Ki	630	41.1	0.69	0.47	0.33	0.22	0.29
CHEMBL234_Ki	733	43.7	0.63	0.43	0.28	0.13	0.18

continued on next page

Table S1 (continued)

Target	n	Cliff %	Dirichlet	Lipschitz	Nbr. disp.	SALI mean	Nbr. disp. | AD
CHEMBL235_EC50	472	37.7	0.66	0.43	0.26	0.16	0.24
CHEMBL236_Ki	521	39.2	0.64	0.40	0.28	0.21	0.27
CHEMBL237_EC50	193	47.2	0.63	0.31	0.22	0.16	0.20
CHEMBL237_Ki	522	42.7	0.67	0.39	0.27	0.14	0.26
CHEMBL238_Ki	213	25.8	0.63	0.45	0.21	0.21	0.19
CHEMBL239_EC50	347	40.6	0.65	0.42	0.30	0.16	0.27
CHEMBL244_Ki	621	47.7	0.63	0.51	0.34	0.25	0.31
CHEMBL262_Ki	173	18.5	0.63	0.39	0.12	0.11	0.13
CHEMBL264_Ki	574	41.6	0.62	0.37	0.27	0.17	0.20
CHEMBL2835_Ki	126	10.3	0.74	0.59	0.49	0.36	0.20
CHEMBL287_Ki	267	38.6	0.61	0.38	0.19	0.01	0.15
CHEMBL2971_Ki	197	17.3	0.73	0.52	0.54	0.32	0.24
CHEMBL3979_EC50	226	41.6	0.60	0.41	0.10	0.02	0.05
CHEMBL4005_Ki	193	42.0	0.71	0.63	0.13	0.14	0.06
CHEMBL4203_Ki	149	8.7	0.62	0.61	0.19	0.11	0.16
CHEMBL4616_EC50	139	52.5	0.73	0.48	0.15	0.02	0.13
CHEMBL4792_Ki	297	53.9	0.72	0.47	0.23	0.16	0.24

S2. Robustness across neighborhood size and distance metric

The central correlations were recomputed for five neighborhood sizes k under three distance metrics: Tanimoto similarity over ECFP4 (radius 2) and ECFP6 (radius 3) fingerprints, and Euclidean distance in a standardized 12-descriptor physicochemical space. *local_var* is a uniform-weight landscape roughness (mean squared activity gap to the k neighbors, which uses the query's activity); *nbr_disp* is the activity-free neighborhood dispersion; *nbr_disp | AD* is its partial correlation after controlling for nearest-neighbor distance and local density. Each cell is the median across the 30 targets. All fifteen k–metric combinations give a positive partial correlation.

S3. Graph-network details and per-target errors

Architecture. The graph model is a Graph Isomorphism Network (GIN-0). Each atom carries a 16-dimensional feature vector: a one-hot atomic-number indicator over {C, N, O, F, P, S, Cl, Br, I, other}, degree (scaled by 4), formal charge, an aromaticity flag, total hydrogen count (scaled by 4), and a ring-membership flag. Three message-passing layers of hidden width 64 update node states as $\mathbf{H}^{(l+1)} = \mathrm{MLP}^{(l)}(\mathbf{H}^{(l)} + \mathbf{A}\,\mathbf{H}^{(l)})$, where $\mathbf{A}$ is the bond adjacency matrix and each $\mathrm{MLP}^{(l)}$ is a two-layer perceptron with ReLU activation. The graph embedding is the sum over layers of the masked mean of node embeddings, read out by a two-layer perceptron head.

Training. Models are trained with Adam (learning rate 10^{-3}) to minimize mean squared error on training-standardized targets. The *fixed* regimen runs a flat schedule of 60 epochs. The *tuned* regimen holds out 10% of the training set as a validation split, trains for up to 150 epochs with learning-rate reduction on plateau and best-checkpoint early stopping (patience 20), and uses size-bucketed mini-batches of 128 to minimize padding. Per-compound test error is $|\hat{y}_i - y_i|$ in log-potency units. The tuned regimen lowers mean MAE on 14 of 30 targets and does not change the

Table S2: Median Spearman correlation (across 30 targets) with per-compound random-forest error, across neighborhood size and distance metric.

Metric	k	local_var	nbr_disp	nbr_disp | AD
ECFP4 / Tanimoto	5	0.641	0.273	0.226
ECFP4 / Tanimoto	10	0.638	0.275	0.230
ECFP4 / Tanimoto	15	0.628	0.265	0.207
ECFP4 / Tanimoto	20	0.615	0.240	0.188
ECFP4 / Tanimoto	30	0.587	0.218	0.186
ECFP6 / Tanimoto	5	0.633	0.264	0.211
ECFP6 / Tanimoto	10	0.638	0.287	0.242
ECFP6 / Tanimoto	15	0.623	0.258	0.210
ECFP6 / Tanimoto	20	0.604	0.250	0.197
ECFP6 / Tanimoto	30	0.583	0.219	0.165
RDKit desc / Euclidean	5	0.464	0.149	0.135
RDKit desc / Euclidean	10	0.482	0.163	0.143
RDKit desc / Euclidean	15	0.457	0.134	0.111
RDKit desc / Euclidean	20	0.460	0.138	0.115
RDKit desc / Euclidean	30	0.461	0.131	0.104

roughness ordering reported in the main text.

Table S3: Per-target mean absolute test error (log units) for the random forest and the two graph-network regimens, with the per-target ROGI global-roughness index.

Target	n	RF MAE	GNN (fixed-60)	GNN (tuned)	ROGI
CHEMBL1862_Ki	161	0.59	0.78	0.89	0.049
CHEMBL1871_Ki	134	0.53	0.61	0.62	0.056
CHEMBL2034_Ki	152	0.50	0.59	0.83	0.074
CHEMBL2047_EC50	128	0.42	0.64	0.57	0.078
CHEMBL204_Ki	553	0.55	0.80	0.76	0.058
CHEMBL2147_Ki	294	0.47	1.14	0.79	0.053
CHEMBL214_Ki	666	0.51	0.71	0.67	0.063
CHEMBL218_EC50	208	0.53	0.69	0.69	0.064
CHEMBL219_Ki	374	0.55	0.77	0.74	0.073
CHEMBL228_Ki	342	0.52	0.67	0.70	0.062
CHEMBL231_Ki	197	0.56	0.70	0.64	0.051
CHEMBL233_Ki	630	0.60	0.76	0.75	0.065
CHEMBL234_Ki	733	0.49	0.66	0.73	0.065
CHEMBL235_EC50	472	0.48	0.61	0.70	0.072
CHEMBL236_Ki	521	0.52	0.81	0.70	0.066
CHEMBL237_EC50	193	0.58	0.83	0.99	0.089
CHEMBL237_Ki	522	0.55	0.70	0.63	0.068
CHEMBL238_Ki	213	0.49	0.63	0.69	0.061

continued on next page

Table S3 (continued)

Target	n	RF MAE	GNN (fixed-60)	GNN (tuned)	ROGI
CHEMBL239_EC50	347	0.51	0.68	0.69	0.085
CHEMBL244_Ki	621	0.56	0.80	0.78	0.069
CHEMBL262_Ki	173	0.57	0.75	0.78	0.063
CHEMBL264_Ki	574	0.48	0.62	0.68	0.066
CHEMBL2835_Ki	126	0.27	0.33	0.47	0.063
CHEMBL287_Ki	267	0.56	0.66	0.75	0.071
CHEMBL2971_Ki	197	0.43	0.58	0.66	0.043
CHEMBL3979_EC50	226	0.49	0.66	0.60	0.076
CHEMBL4005_Ki	193	0.50	0.66	0.62	0.072
CHEMBL4203_Ki	149	0.67	0.79	0.85	0.052
CHEMBL4616_EC50	139	0.51	0.65	0.64	0.110
CHEMBL4792_Ki	297	0.54	0.68	0.79	0.096

S4. Conformal prediction-interval coverage

90% inductive-conformal prediction intervals were built from the held-out per-compound residuals (main-text Methods). For each target the test compounds were split at random into equal calibration and evaluation halves. Coverage was made conditional on a per-compound variable in two standard ways: a locally-weighted (normalized) nonconformity score $|\hat{y}_i - y_i|/\sigma(x_i)$, and a Mondrian scheme that stratifies calibration into quintiles of the variable and calibrates within each. The conditioning variable was the random-forest tree variance, the activity-free roughness (pairwise-SALI density), the applicability domain ($1 - s_{\mathrm{NN}}$), or the combined variance-plus-roughness score. Every variant keeps the marginal coverage guarantee; they differ in where the coverage lands (Table S4, median across the 30 targets over 50 random splits). Conditioning on tree variance or the applicability domain leaves activity-cliff coverage at or below the unconditional level; only roughness restores it toward nominal, and the combined score recovers most of it at no width cost. The same pattern holds across the full roughness range, not only on the labeled-cliff subset: unconditional coverage falls from 0.95 in the smoothest roughness quintile to 0.87 in the roughest, and variance- and applicability-domain-conditioning fall similarly, whereas roughness (or the combined score) holds it within two points of nominal across all quintiles (main-text Figure 5A; `results/conformal_curve.csv`).

Table S4: Median 90% conformal prediction-interval coverage and relative width across the 30 targets, by construction and conditioning variable.

Construction (conditioning variable)	Marginal	Cliff	Non-cliff	Median width (rel.)
Standard (unconditional)	0.904	0.872	0.920	1.00
Mondrian (tree variance)	0.910	0.860	0.946	0.97
Mondrian (applicability domain)	0.910	0.861	0.945	1.01
Mondrian (roughness, SALI)	0.911	0.895	0.928	1.06
Mondrian (variance + roughness)	0.913	0.876	0.941	0.99
Locally-weighted (roughness, SALI)	0.904	0.912	0.898	1.16
Locally-weighted (applicability domain)	0.904	0.820	0.954	1.14