

Supporting Information

SCOPE: From Shape and Color to Binding Affinity, with
Confidence

Jingyi L. Chen^a, Chris Neale^a, Mireille Krier^a, Someina Khor^b,
Anthony R. Bradley^{c,d}, A. Geoffrey Skillman^a, and Shyamal K.
Nath^{a*}

^aOpenEye, Cadence Molecular Sciences, Santa Fe, NM, USA

^bIsomorphic Labs, London, UK

^cDepartment of Chemistry, University of Liverpool, Liverpool,
UK

^dDalton Tx Ltd, London, UK

1 Methods

1.1 Curation of structure–affinity datasets

Description of how target-specific structure–affinity datasets were curated. Figure S1 illustrates the major steps taken to curate the structure–affinity datasets, followed by details given in text. This workflow is refined and built upon a preliminary Python program developed by Feng Yu, who formerly interned at OpenEye, Cadence Molecular Sciences. This preliminary Python program takes a UniProt accession code and generates a set of ligands with X-ray bound structure and binding affinity data corresponding to a given target.

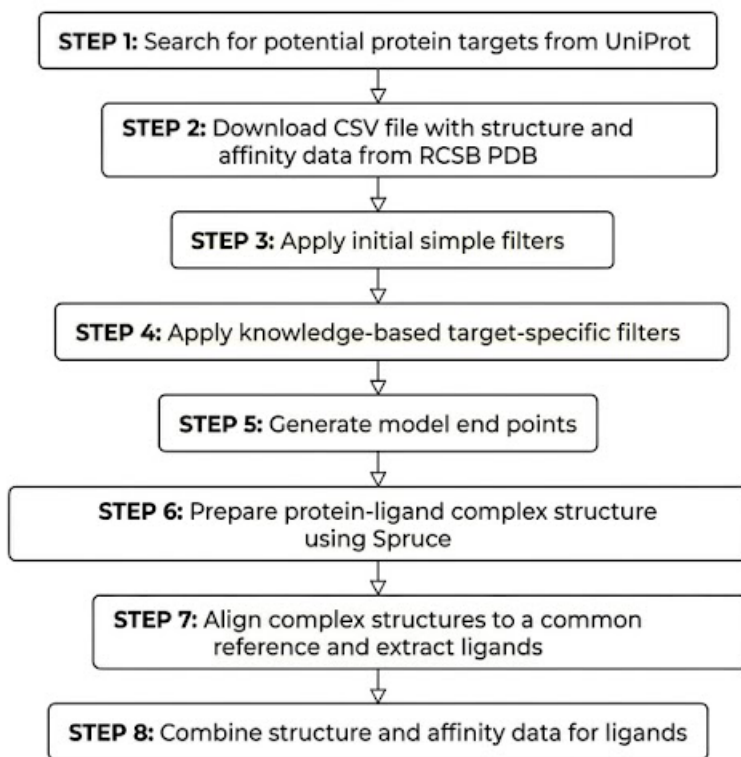


Figure S1: Major workflow to curate target-specific structure–affinity datasets.

1. **Search for potential protein targets.** The protein database UniProt (accessed Jan 26th 2026) [1] was searched for targets with several search criteria ("Reviewed, Taxonomy = 9606 (Homo Sapiens), with 3D structures, Annotation score = 5"). An initial list of 8181 targets with UniProt accession codes was obtained.
2. **Download csv file with structure and affinity data from the RCSB Protein Data Bank (PDB).** For each target, RCSB PDB was searched (accessed Jan 31st 2026) given a UniProt accession code. The csv file fetched includes PDB code for the entry, experimental method and resolution for the structure data, code for the binding ligand, and type, value, and symbol of affinity data. The affinity data is primarily sourced from BindingDB[2].
3. **Apply initial simple filters.** First, only entries with experimental method of X-ray were retained. Then, entries with (1) null ligand or (2) null affinity value or (3) non-null affinity symbol (null implies "="), or (4) energy units (kJ/mol or kcal/mol) were discarded.

Second, because one PDB code can involve multiple biological entities that correspond to different UniProt codes, each unique pair of (PDB code, ligand) was checked to make sure the biological entity that the ligand binds to corresponds to the given target. If not, the entry was filtered out.

Finally, (PDB code, ligand) that were identified to be covalent binders according to reference No.[3] were discarded. For kinase targets, only (PDB code, ligand) that were confirmed to have DFG-in conformers according to the KLIFS database[4] (accessed March 19th 2026) were retained.

Here only targets with at least 50 unique ligands within the same affinity type group were retained for the next step, leading to a total of 21 targets left. A summary of them grouped by target class is given in Table S1.

Table S1: 21 Target proteins grouped by their class after applying initial simple filters.

Target Class	UniProt ID	Protein
Protein kinase	P24941	CDK2
	O14757	CHK1
	Q16539	p38 α
	O60674	JAK2
	P11309	PIM1
	Q13627	DYRK1A
	P43405	SYK
	Q9NWZ3	IRAK4
Serine protease	P00742	Factor Xa
	P00734	Thrombin
	P27487	DPP-IV
Aspartyl protease	P56817	BACE1
	P00797	Renin
Nuclear receptor	P37231	PPAR γ
	P51449	ROR γ
Epigenetic reader	O60885	BRD4
PARP family enzyme	Q9H2K2	TNKS2
Metabolic / redox enzyme	P00918	CAII
	P29475	nNOS
	P34913	sEH
Molecular chaperone	P07900	HSP90 α

4. Apply knowledge-based target-specific filters.

First, proteins known to possess multiple binding domains were identified. Here only the domain with the highest representation of ligand-bound data in PDB was retained for each target. For example, the JH1 domain was selected for JAK2, and the ligand-binding domain (LBD)

for nuclear receptors.

Second, protein conformational flexibility was accounted for by focusing on targets where alternative ligand-bound states are statistically significant and functionally distinct. Selection criteria were data-driven based on structural prevalence within PDB.

- (a) Kinases: Only the DFG-in conformation was retained (performed in previous filtering step to narrow down potential targets).
- (b) BACE1: A "Flap-Closed" filter was applied to exclude fragment-bound "open" structures, which would otherwise introduce inconsistent pocket volumes.
- (c) Thrombin: Only the mature active form was included to ensure catalytic relevance.
- (d) HSP90 α : Only the "lid-open" state was considered.

To account for induced-fit effects in nuclear receptors, ligands were categorized by their functional impact. For PPAR γ , only full agonists (which stabilize Helix 12 in the active position) were included; conversely, for ROR γ , only inverse agonists bound specifically to the orthosteric site were considered.

Finally, the mutational status of the binding sites was evaluated for all entries. Entries with mutations that impact protein–ligand binding were excluded. Following a data-driven approach, we retained the binding site configuration that appeared most frequently in PDB.

5. **Process affinity data and generate model end points.** Following a data-driven approach, the affinity type with the most ligands was selected. All affinity data was converted to log units for model building (e.g., from IC₅₀ to pIC₅₀), reflecting the approximately log-normal distribution of biological activity data[5]. For ligand with multiple end points, the median was used as the final potency data for model building.
6. **Prepare protein–ligand complex structure using Spruce.** Structure data was downloaded for all unique PDB codes. Spruce[6] was used to prepare protein–ligand complex structures in OEDesignUnit (OEDU) format. For ligand with multiple generated OEDUs, the one with the higher Iridium score[7] category was preferred if available.

7. **Align protein–ligand complex structures to a common reference.** A high-quality reference OEDU was manually chosen for each target. Rest of the OEDUs were aligned to the reference based on protein sequence. The ligands were then extracted to form the ligand dataset.

A final check to see whether aligned ligands share spatial proximity was performed by calculating histograms of pairwise distances of center of mass of the ligands. If the histogram shows a visually distinct secondary peak, the outlier ligand(s) was identified and removed from the set.

8. **Combine structure and affinity data.** Affinity data and structure data was combined for each unique ligand to form the final structure–affinity dataset for a protein target.

In the end, we obtained 17 structure–affinity datasets with size above 45, and end points spanning at least 3 log units.

1.2 Exploratory data analysis on curated structure–affinity datasets

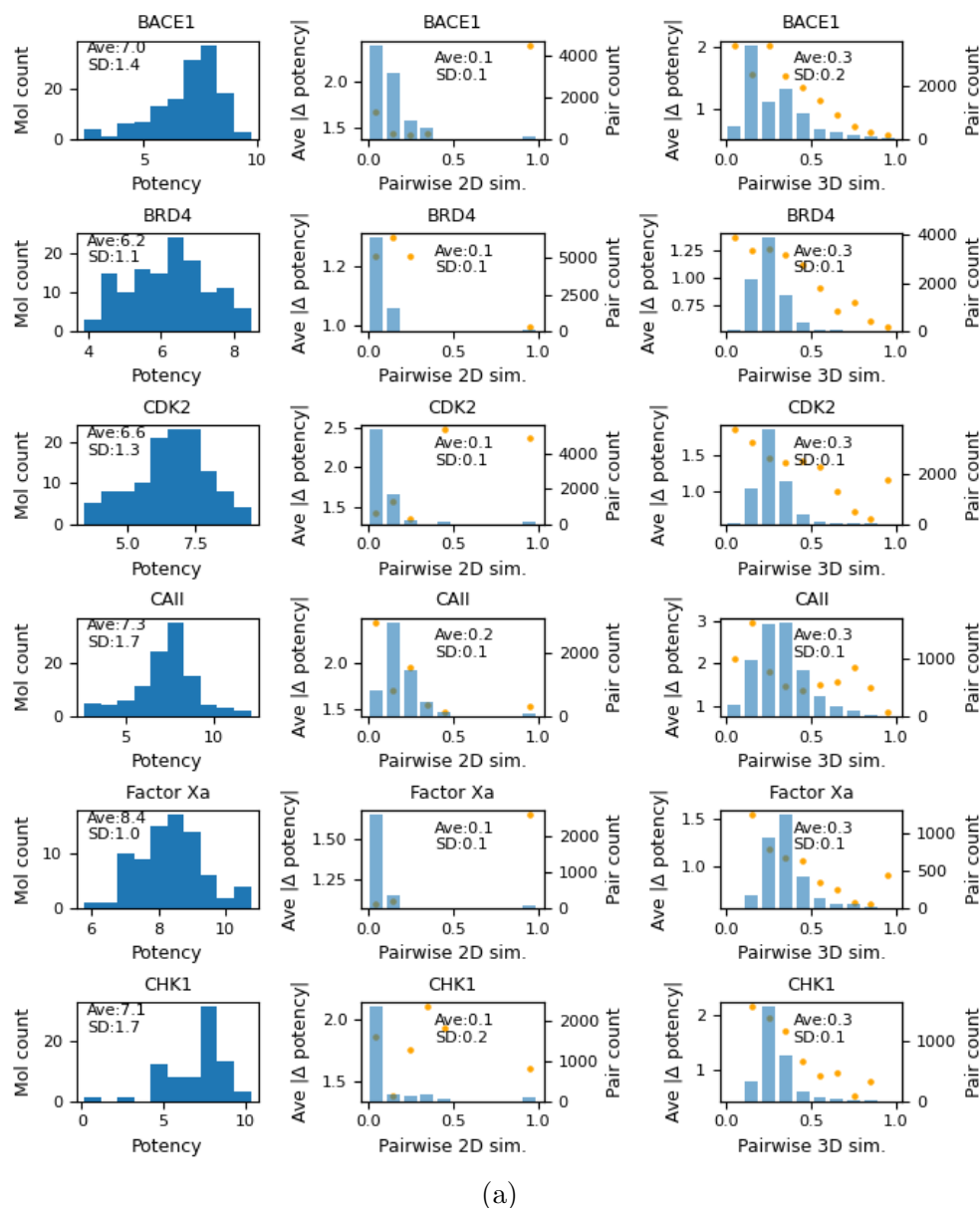
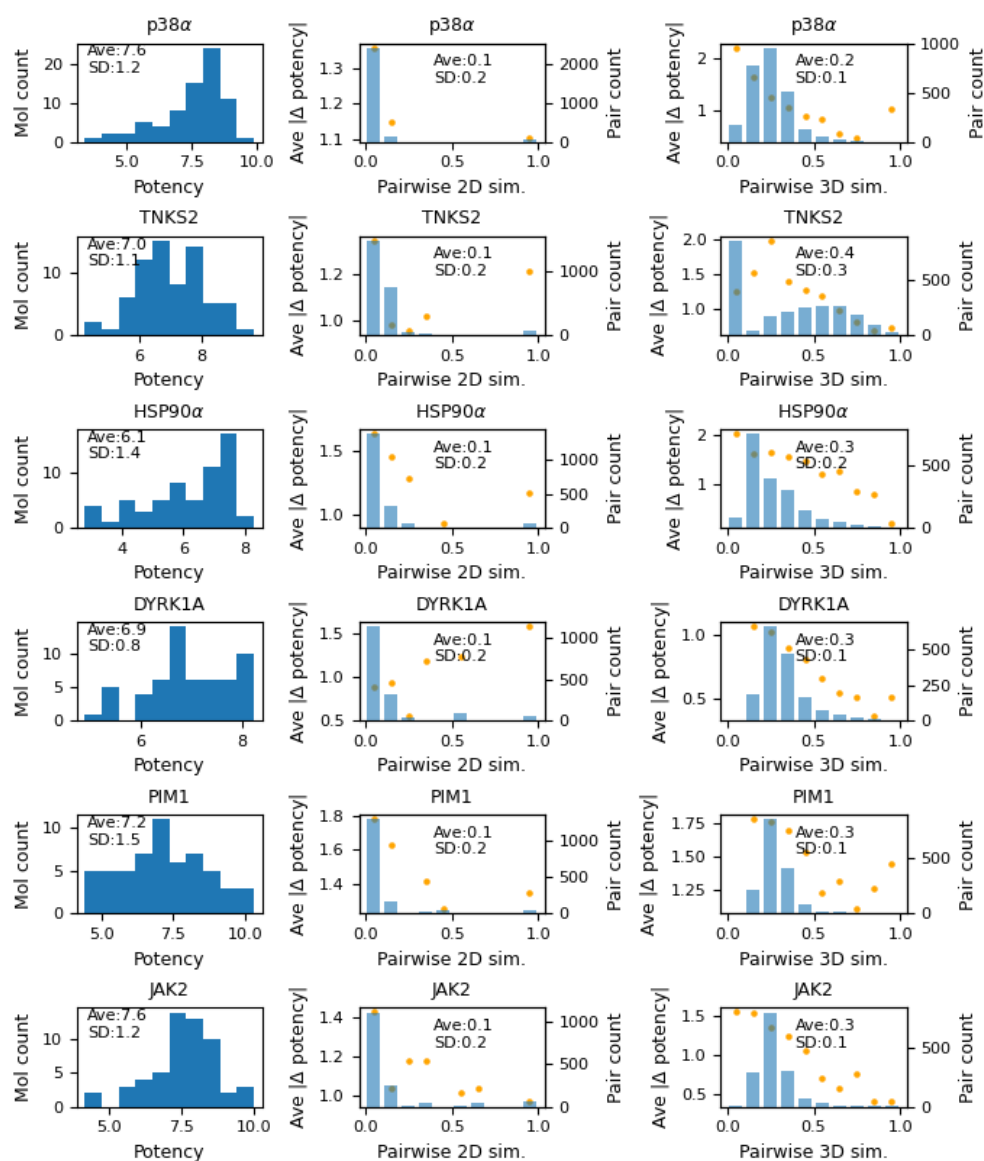
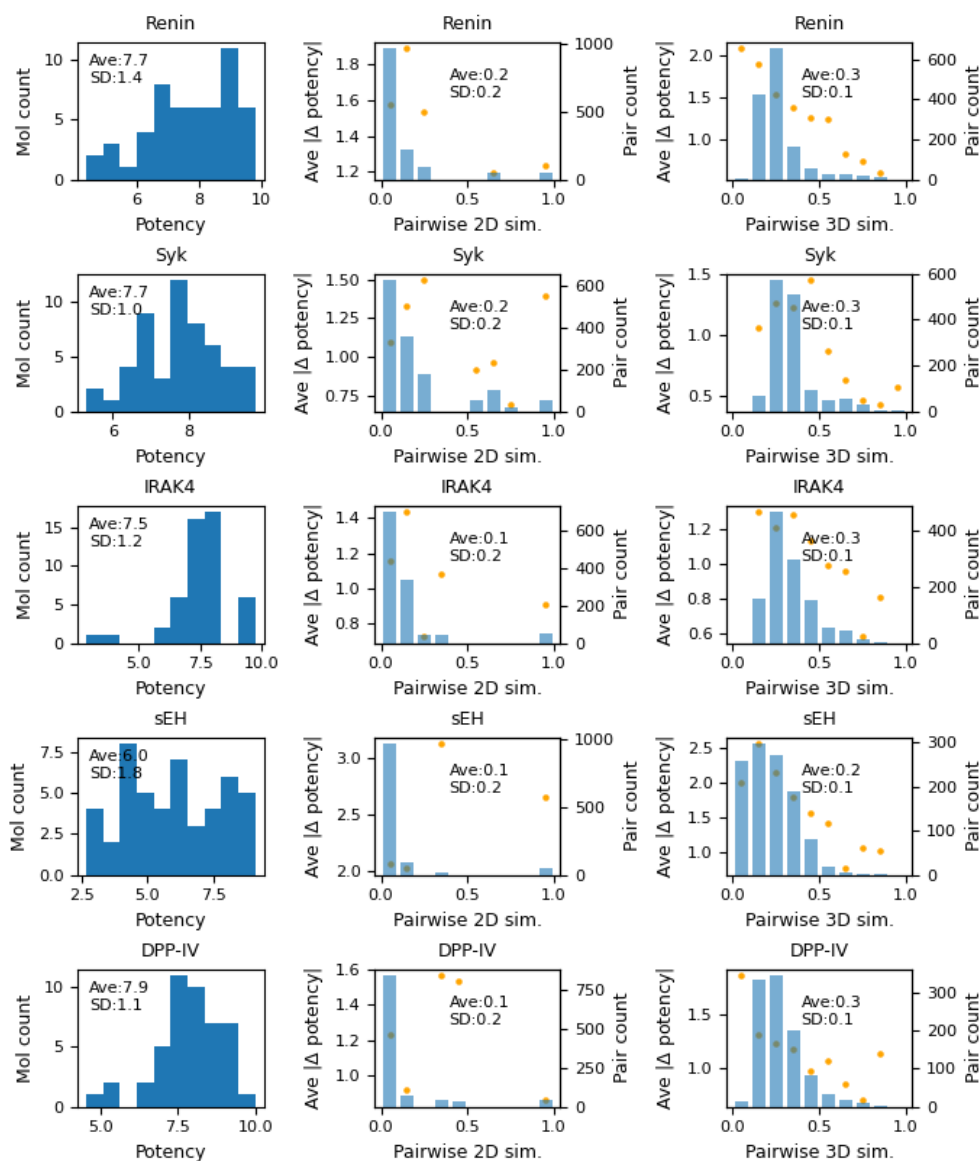


Figure S2: Exploratory data analysis across different subsets of the dataset. (a) datasets 0–6, (b) datasets 6–12, (c) datasets 12–17. For each subset (left to right): potency histogram; 2D similarity histogram (blue) with average absolute pairwise potency difference per bin (orange); normalized 3D ROCS similarity histogram (blue) with average absolute pairwise potency difference per bin (orange). Histograms are annotated with the average value (Ave) and standard deviation (SD).



(b)

Figure S2 (continued)



(c)

Figure S2 (continued)

1.3 Validation of model design

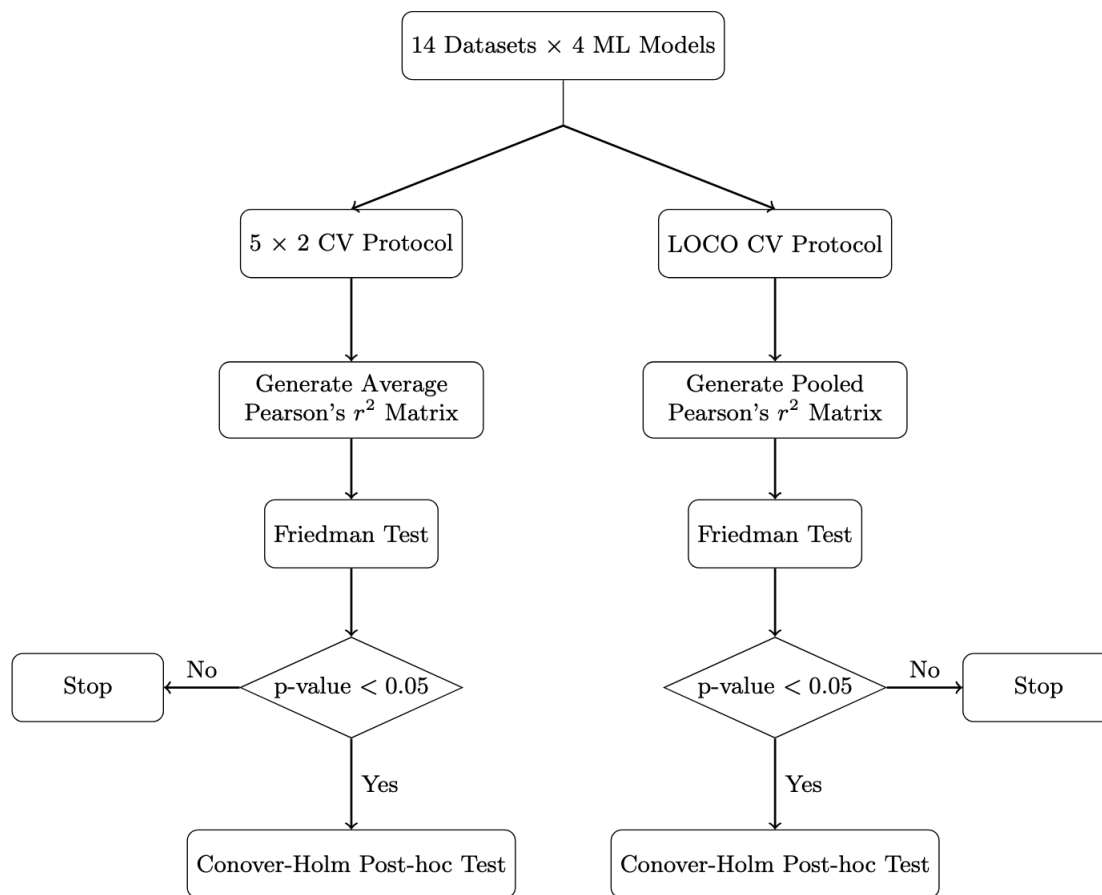


Figure S3: Flow chart for comparing four ML models across fourteen datasets using Pearson's r^2 as an example.

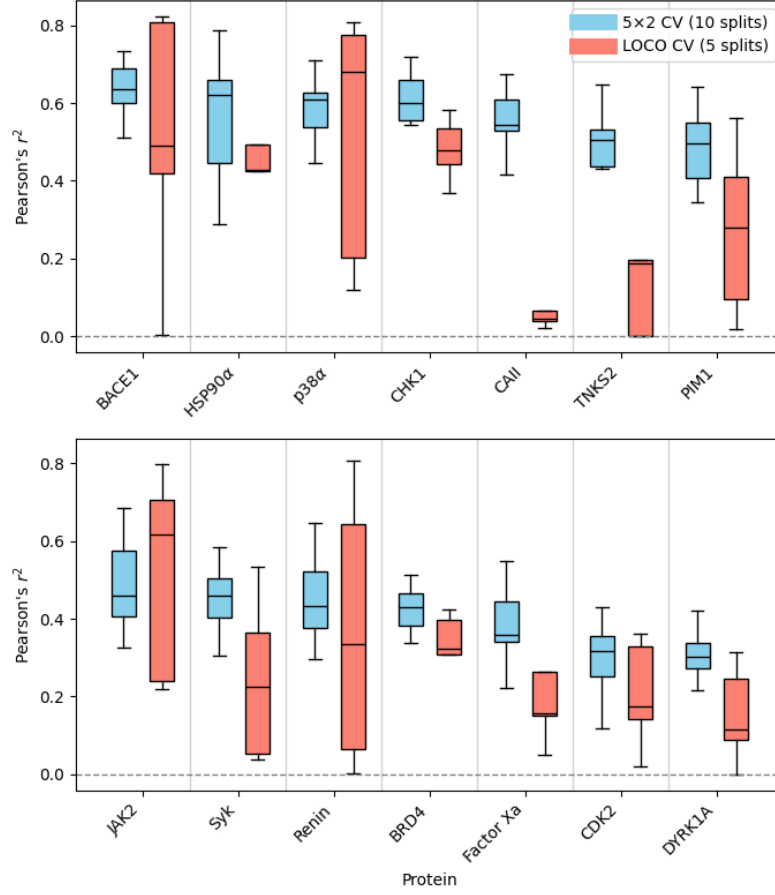


Figure S4: Comparison of SCOPE performance using Pearson's r^2 under 5×2 and LOCO CV across protein targets. Box plots show the distribution of fold-level performance for each target. Targets are ordered by median performance under 5×2 CV and split across two panels for clarity.

Table S2: Performance values (Pearson’s r^2) used as input to the Friedman test under different CV schemes.

CV	Dataset	Ensemble	2D-GPR	ROCS-GPR	ROCS-kPLS
5×2	BACE1	0.64	0.46	0.62	0.54
	BRD4	0.44	0.24	0.38	0.44
	CDK2	0.30	0.13	0.26	0.28
	CAII	0.55	0.59	0.54	0.38
	Factor Xa	0.39	0.26	0.35	0.35
	CHK1	0.61	0.37	0.57	0.61
	p38 α	0.59	0.36	0.51	0.53
	TNKS2	0.49	0.38	0.45	0.44
	HSP90 α	0.56	0.42	0.44	0.58
	DYRK1A	0.32	0.19	0.26	0.30
	PIM1	0.49	0.38	0.47	0.46
	JAK2	0.49	0.25	0.44	0.47
	Renin	0.46	0.30	0.32	0.48
	Syk	0.45	0.37	0.39	0.45
LOCO	BACE1	0.53	0.35	0.45	0.46
	BRD4	0.28	0.13	0.25	0.28
	CDK2	0.15	0.015	0.16	0.15
	CAII	0.20	0.085	0.23	0.17
	Factor Xa	0.23	0.11	0.15	0.22
	CHK1	0.43	0.39	0.43	0.35
	p38 α	0.36	0.24	0.31	0.37
	TNKS2	0.19	0.00090	0.11	0.18
	HSP90 α	0.22	0.11	0.10	0.41
	DYRK1A	0.31	0.067	0.18	0.34
	PIM1	0.13	0.029	0.12	0.16
	JAK2	0.41	0.0012	0.38	0.42
	Renin	0.17	0.20	0.062	0.24
	Syk	0.30	0.31	0.20	0.30

1.4 MD simulation and MM/PBSA calculation

MD simulations in the *NPT* ensemble were performed using the Molecular Dynamics Affinity package (v6.5.8)[8] on the Orion[®] cloud-native platform (OpenEye, Cadence Molecular Sciences)[9]. Specific calculations utilized the **Short Trajectory MD with Analysis Floe**[10], which provides an automated workflow for system preparation, energy minimization, equilibration, and production runs. The default MD engine OpenMM was used[11], with Open Force Field (OpenFF) Sage small-molecule force field (OpenFF 2.2.0)[12, 13] used to parameterize the ligands and the Amber ff14SB force field[14] applied to the protein. Production simulations were run for 2 ns, and trajectories were saved every 0.025 ns.

Computation of the PBSA part for the saved trajectories was performed using the OpenEye ZAP Toolkit[15, 16]. Electrostatic potentials were calculated by solving the Poisson–Boltzmann equation on a regular 3D lattice. The dielectric environment was modeled using an inner dielectric constant of 1.0 and an outer solvent dielectric of 80.0. The nonpolar solvation contribution was estimated from the solvent-accessible surface area, using a surface tension coefficient of 0.006 kcal/Å².

Additional details of of the MD simulation Protein and ligand conformations are combined, and crystallographic water molecules are removed if they are within 1.5 Å of the posed ligand. A cubic volume is constructed with a minimum distance of 6 Å between the solute and the edge of the unit cell. This system is solvated and neutralized by water and 150 mM NaCl by PACKMOL version 20.010[17] using a target pure-water density of 1.03 g/cm³ and a minimal intermolecular distance of 2 Å. Protein and ligand are modeled by AMBER ff14SB[14] and OpenFF 2.2.0 [12, 13] force fields, respectively. Histidine protonation states are set by SPRUCE[6, 9]. The water model is TIP3P[18]. Water molecules are rigidified with SETTLE[19], and other covalent bond lengths involving hydrogen are constrained with SHAKE[20] (tolerance=10⁻⁵). Lennard-Jones (LJ) interactions are truncated at 10 Å. Coulomb interactions are calculated using the smooth particle-mesh Ewald method[21] with error tolerance of 5×10^{-4} and fifth order interpolation. Harmonic position restraints are applied to all protein and ligand heavy atoms with a force constant of 5 kcal/mol/Å² and the system is energy minimized with L-BFGS[22] using mixed-precision OpenMM 8.2.0[11] until reaching a force tolerance of 0.24 kcal/mol/Å. Restraint force constants are reduced to 2 kcal/mol/Å and the system is simulated for 10 ps at constant

volume using velocity Langevin dynamics[23, 24] at 300 K with a friction coefficient of 1 ps^{-1} and an integration time step of 4 fs using hydrogen mass repartitioning[25]. Restraint strength is successively lowered in stages (10 ps at $1 \text{ kcal/mol/\AA}$, 20 ps at $0.5 \text{ kcal/mol/\AA}$, 0.1 ns at $0.2 \text{ kcal/mol/\AA}$, and 0.1 ns at $0.1 \text{ kcal/mol/\AA}$, with position restraints acting only on ligand heavy atoms and protein C_α atoms in the final three stages). Production simulations of 2 ns are conducted in the absence of restraints.

2 Results

2.1 Two complementary CV protocols

Table S3: Cluster size distribution using K-medoids clustering with K=5, and nearest-neighbor test–train similarity under LOCO CV. Values are reported as median (Q1–Q3) of per-test nearest-neighbor similarity for each split.

Protein	Cluster sizes	Split-wise max sim
BACE1	58, 29, 27, 14, 8	0.14 (0.14–0.19), 0.17 (0.15–0.26), 0.13 (0.12–0.14), 0.20 (0.13–0.41), 0.16 (0.15–0.21)
BRD4	34, 33, 25, 21, 15	0.14 (0.12–0.19), 0.14 (0.13–0.21), 0.15 (0.13–0.25), 0.14 (0.11–0.26), 0.13 (0.12–0.23)
CDK2	30, 29, 28, 22, 14	0.18 (0.14–0.23), 0.15 (0.13–0.18), 0.14 (0.12–0.20), 0.15 (0.13–0.19), 0.14 (0.13–0.20)
CAII	59, 17, 15, 11, 7	0.22 (0.17–0.26), 0.28 (0.26–0.31), 0.16 (0.14–0.21), 0.21 (0.21–0.27), 0.18 (0.15–0.18)
Factor Xa	23, 17, 16, 13, 10	0.28 (0.17–0.29), 0.22 (0.19–0.28), 0.29 (0.18–0.30), 0.35 (0.29–0.37), 0.14 (0.12–0.21)
CHK1	20, 17, 15, 13, 12	0.14 (0.12–0.21), 0.13 (0.12–0.21), 0.13 (0.12–0.21), 0.12 (0.11–0.14), 0.12 (0.11–0.23)
p38 α	24, 17, 14, 10, 8	0.13 (0.12–0.15), 0.23 (0.18–0.25), 0.25 (0.17–0.28), 0.18 (0.16–0.25), 0.14 (0.13–0.16)
TNKS2	26, 14, 10, 10, 9	0.13 (0.12–0.18), 0.12 (0.12–0.23), 0.15 (0.13–0.16), 0.12 (0.11–0.15), 0.13 (0.12–0.13)
HSP90 α	16, 16, 13, 8, 8	0.16 (0.13–0.17), 0.10 (0.10–0.13), 0.10 (0.09–0.11), 0.16 (0.14–0.18), 0.16 (0.14–0.18)
DYRK1A	17, 13, 11, 9, 8	0.14 (0.11–0.17), 0.13 (0.10–0.14), 0.16 (0.13–0.17), 0.16 (0.16–0.17), 0.12 (0.11–0.15)
PIM1	17, 16, 11, 8, 5	0.11 (0.10–0.12), 0.12 (0.11–0.14), 0.12 (0.10–0.12), 0.13 (0.11–0.15), 0.12 (0.12–0.12)
JAK2	17, 12, 11, 10, 6	0.19 (0.13–0.26), 0.16 (0.15–0.24), 0.13 (0.11–0.18), 0.12 (0.10–0.13), 0.11 (0.10–0.15)
Renin	16, 12, 10, 9, 6	0.19 (0.15–0.22), 0.22 (0.19–0.30), 0.18 (0.14–0.21), 0.12 (0.12–0.13), 0.12 (0.11–0.13)
Syk	16, 14, 9, 7, 7	0.14 (0.13–0.15), 0.13 (0.12–0.14), 0.23 (0.14–0.31), 0.15 (0.14–0.15), 0.28 (0.23–0.29)

Table S4: Nearest-neighbor test–train similarity under 5×2 CV. Values are reported as median (Q1–Q3) of per-test nearest-neighbor similarity for each split.

Protein	Split-wise nearest-neighbor sim
BACE1	0.39 (0.24–0.51), 0.37 (0.26–0.48), 0.42 (0.29–0.53), 0.37 (0.25–0.52), 0.38 (0.26–0.51), 0.30 (0.22–0.52), 0.33 (0.19–0.50), 0.41 (0.30–0.54), 0.39 (0.27–0.47), 0.41 (0.26–0.52)
BRD4	0.25 (0.17–0.38), 0.23 (0.14–0.39), 0.27 (0.15–0.45), 0.22 (0.13–0.47), 0.26 (0.15–0.48), 0.29 (0.17–0.50), 0.30 (0.19–0.49), 0.23 (0.16–0.44), 0.27 (0.17–0.42), 0.25 (0.15–0.41)
CDK2	0.27 (0.17–0.46), 0.36 (0.20–0.48), 0.25 (0.17–0.41), 0.28 (0.16–0.45), 0.24 (0.17–0.41), 0.24 (0.16–0.43), 0.25 (0.16–0.43), 0.28 (0.17–0.45), 0.27 (0.16–0.43), 0.28 (0.17–0.41)
CAII	0.34 (0.27–0.44), 0.35 (0.28–0.45), 0.30 (0.21–0.40), 0.34 (0.25–0.44), 0.33 (0.29–0.44), 0.36 (0.28–0.45), 0.37 (0.27–0.47), 0.35 (0.26–0.47), 0.32 (0.23–0.45), 0.37 (0.27–0.45)
Factor Xa	0.46 (0.29–0.63), 0.30 (0.20–0.59), 0.44 (0.36–0.53), 0.36 (0.20–0.53), 0.42 (0.21–0.56), 0.42 (0.19–0.60), 0.41 (0.25–0.57), 0.41 (0.21–0.53), 0.45 (0.29–0.63), 0.42 (0.18–0.51)
CHK1	0.30 (0.21–0.39), 0.34 (0.22–0.42), 0.33 (0.21–0.39), 0.35 (0.23–0.44), 0.30 (0.18–0.38), 0.34 (0.19–0.41), 0.27 (0.14–0.39), 0.30 (0.21–0.38), 0.23 (0.18–0.38), 0.32 (0.21–0.39)
p38 α	0.25 (0.14–0.43), 0.25 (0.18–0.51), 0.26 (0.19–0.50), 0.35 (0.18–0.50), 0.31 (0.19–0.50), 0.29 (0.21–0.50), 0.31 (0.22–0.47), 0.38 (0.20–0.50), 0.26 (0.17–0.40), 0.30 (0.17–0.56)
TNKS2	0.44 (0.24–0.61), 0.45 (0.28–0.60), 0.45 (0.19–0.56), 0.50 (0.23–0.61), 0.53 (0.23–0.64), 0.50 (0.28–0.63), 0.44 (0.30–0.60), 0.42 (0.21–0.62), 0.49 (0.24–0.63), 0.45 (0.20–0.63)
JAK2	0.19 (0.13–0.50), 0.24 (0.16–0.48), 0.27 (0.15–0.54), 0.15 (0.12–0.36), 0.27 (0.13–0.48), 0.27 (0.15–0.49), 0.28 (0.14–0.40), 0.31 (0.15–0.53), 0.28 (0.14–0.39), 0.23 (0.13–0.34)
PIM1	0.22 (0.15–0.36), 0.20 (0.14–0.34), 0.19 (0.15–0.31), 0.23 (0.16–0.34), 0.14 (0.11–0.30), 0.17 (0.13–0.29), 0.17 (0.13–0.27), 0.17 (0.13–0.31), 0.26 (0.16–0.34), 0.25 (0.16–0.34)

Continued on next page

Protein	Split-wise nearest-neighbor sim
Renin	0.24 (0.16–0.32), 0.27 (0.20–0.38), 0.40 (0.25–0.52), 0.38 (0.25–0.54), 0.37 (0.24–0.45), 0.30 (0.24–0.42), 0.28 (0.19–0.53), 0.34 (0.22–0.55), 0.29 (0.20–0.41), 0.29 (0.24–0.44)
Syk	0.27 (0.16–0.52), 0.25 (0.15–0.54), 0.27 (0.16–0.57), 0.32 (0.19–0.53), 0.33 (0.16–0.52), 0.22 (0.13–0.45), 0.26 (0.15–0.38), 0.27 (0.23–0.48), 0.27 (0.17–0.43), 0.26 (0.15–0.46)

2.2 Target-agnostic applicability domain

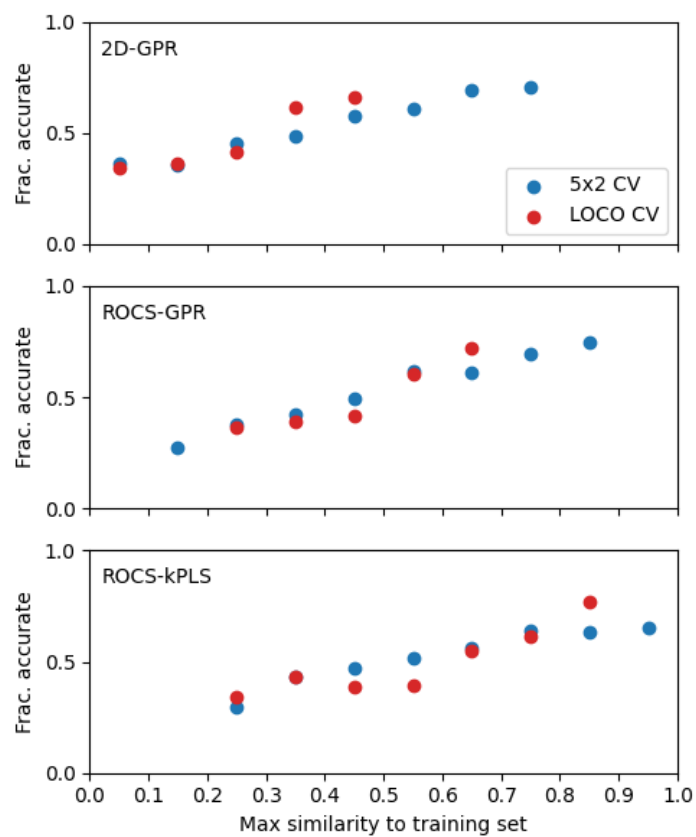


Figure S5: Fraction accurate as a function of maximum similarity to the training set using 5×2 and LOCO CV. Data points with lower than 30 samples in the similarity bin are discarded.

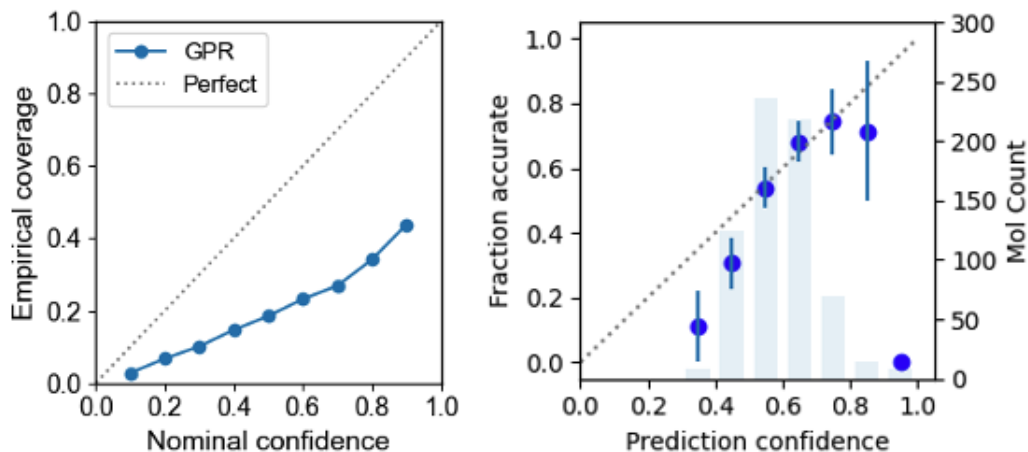


Figure S6: Left: Calibration curve for ROCS-GPR predictive intervals showing empirical coverage versus nominal confidence based on 5×2 CV predictions for BACE1. Empirical coverage at confidence level α is computed as the fraction of samples for which the true value y lies within the two-sided predictive interval $[\hat{y} - z_\alpha \sigma, \hat{y} + z_\alpha \sigma]$, where σ is the returned standard deviation and z_α is the standard normal quantile associated with α ; Right: Observed fraction of accurate predictions as a function of AD-derived prediction confidence for ROCS-GPR using the same prediction set. Uncertainties are estimated using 95% confidence interval through bootstrapping. Blue bars correspond to the number of molecules within each confidence bin.

2.3 Validation of model design

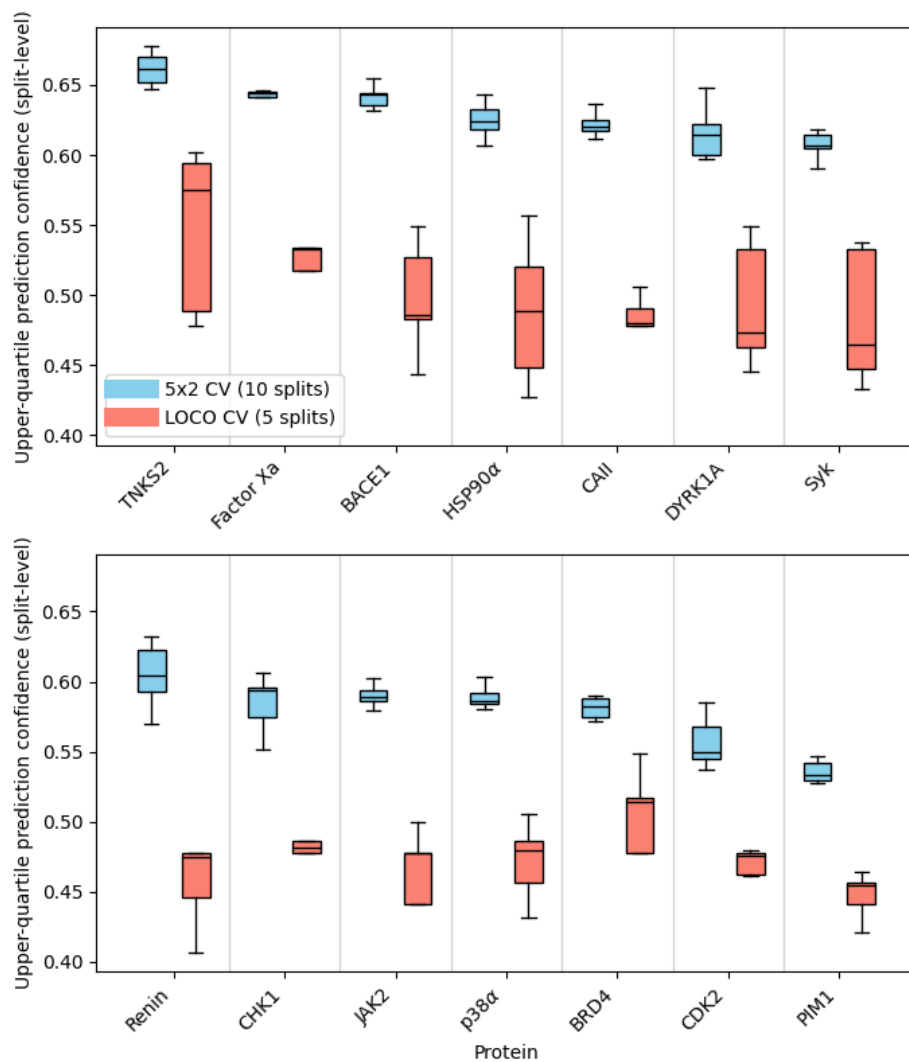


Figure S7: Comparison of SCOPE prediction confidence under 5×2 and LOCO CV across protein targets. Box plots show the distribution of fold-level upper-quartile prediction confidence for each target. Targets are ordered by median value under 5×2 CV and split across two panels for clarity. This shows that confidence is consistently higher under 5×2 CV than LOCO CV, with random splits yielding higher values and narrower variability.

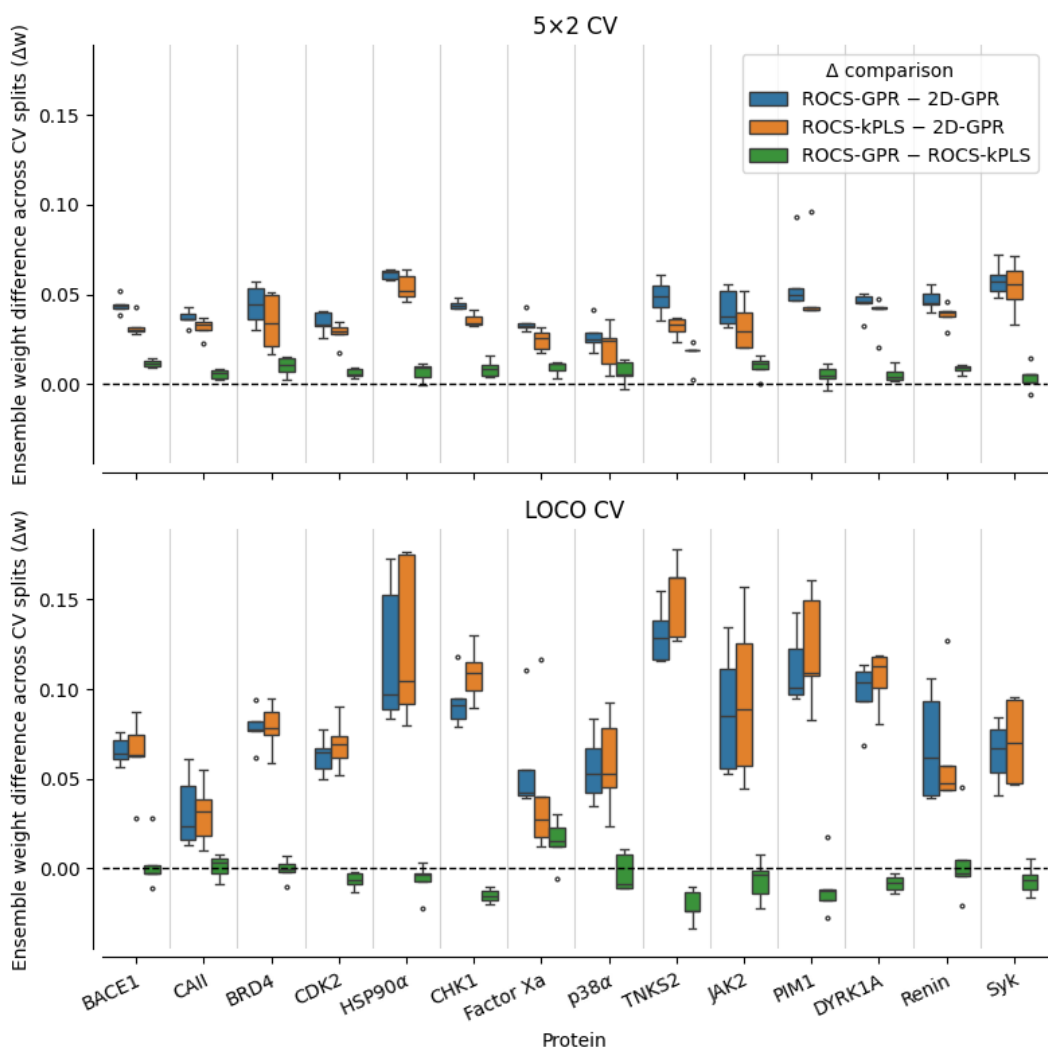


Figure S8: Distribution of fold-level weight differences between individual models in the final ensemble prediction under 5×2 (top) and LOCO CV (bottom) across protein targets. This shows that LOCO CV reflects increased separation under extrapolative conditions and a stronger role for 3D models.

Table S5: P-values from Friedman tests under 5×2 and LOCO CV scheme based on different performance metrics.

CV Scheme	Pearson's r^2	Kendall's τ	MedAE
5×2	2E-5	7E-4	1E-3
LOCO	1E-3	1E-3	3E-3

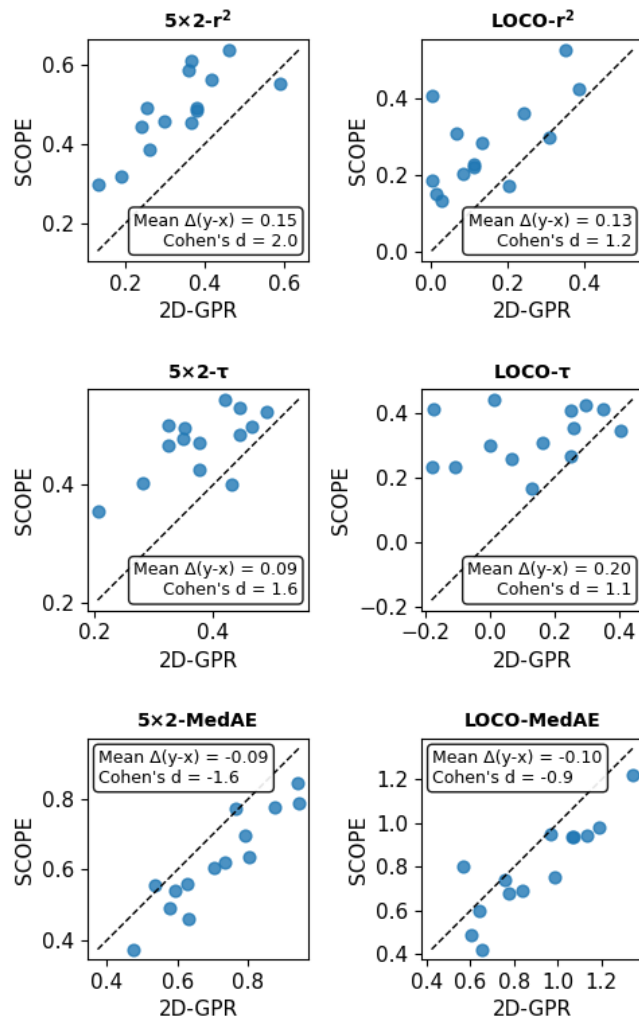


Figure S9: Paired scatter plots across fourteen structure–affinity datasets with effect size analysis between SCOPE and 2D-GPR baseline from 5×2 and LOCO CV using Pearson's r^2 , Kendall's τ , and MedAE as performance metrics.

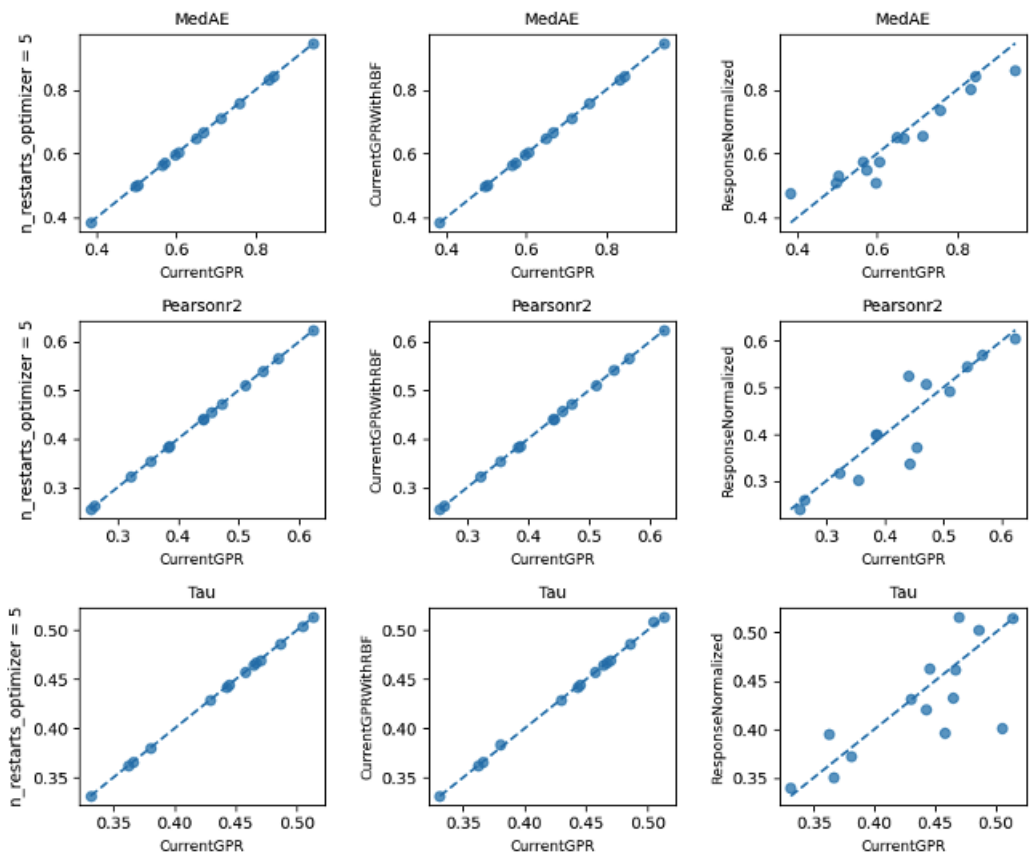


Figure S10: Comparison of alternative GPR configurations evaluated using the 5×2 CV protocol on fourteen structure–affinity datasets across all three performance metrics against current 2D-GPR. The current model is compared against variants with increased optimizer restarts (left), inclusion of an additional RBF kernel (middle), and explicit response normalization (right). Increasing optimizer restarts or adding a nonlinear kernel yielded nearly identical performance. The response-normalized variant showed few pronounced negative deviations in τ .

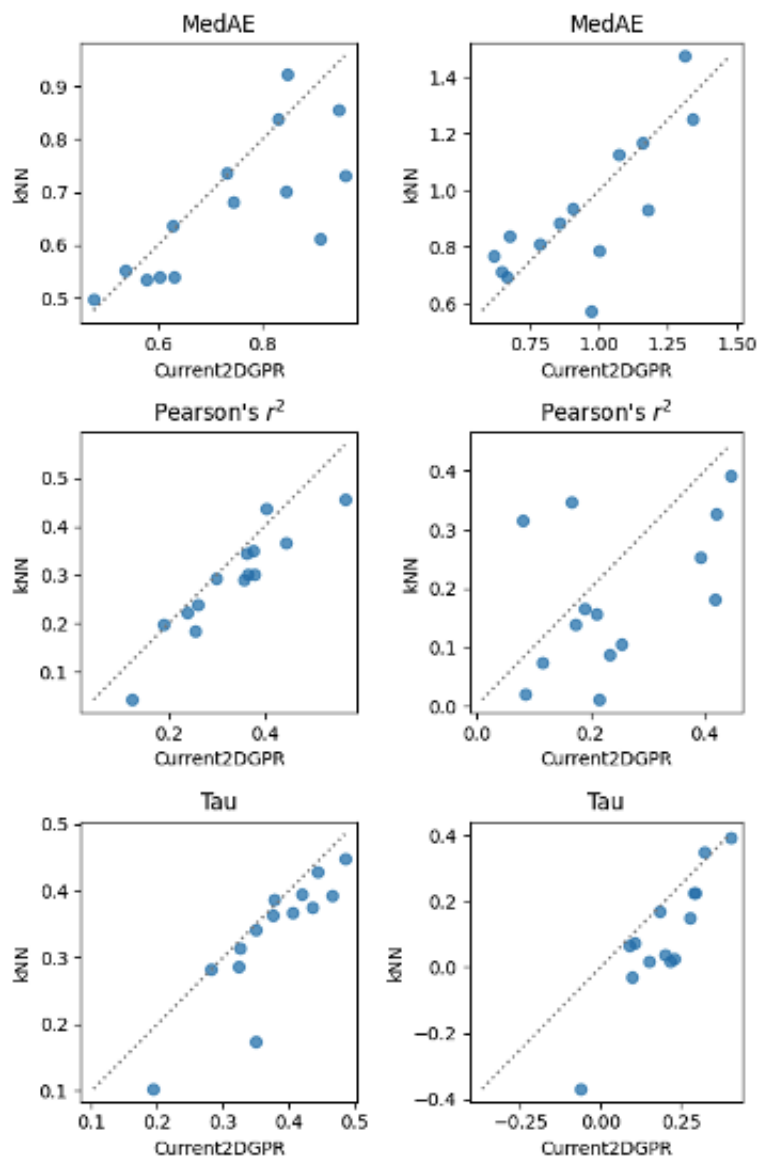


Figure S11: Comparison of kNN (k=3) against current 2D-GPR using the 5×2 (left) and LOCO (right) CV protocols on fourteen structure–affinity datasets across all three performance metrics.

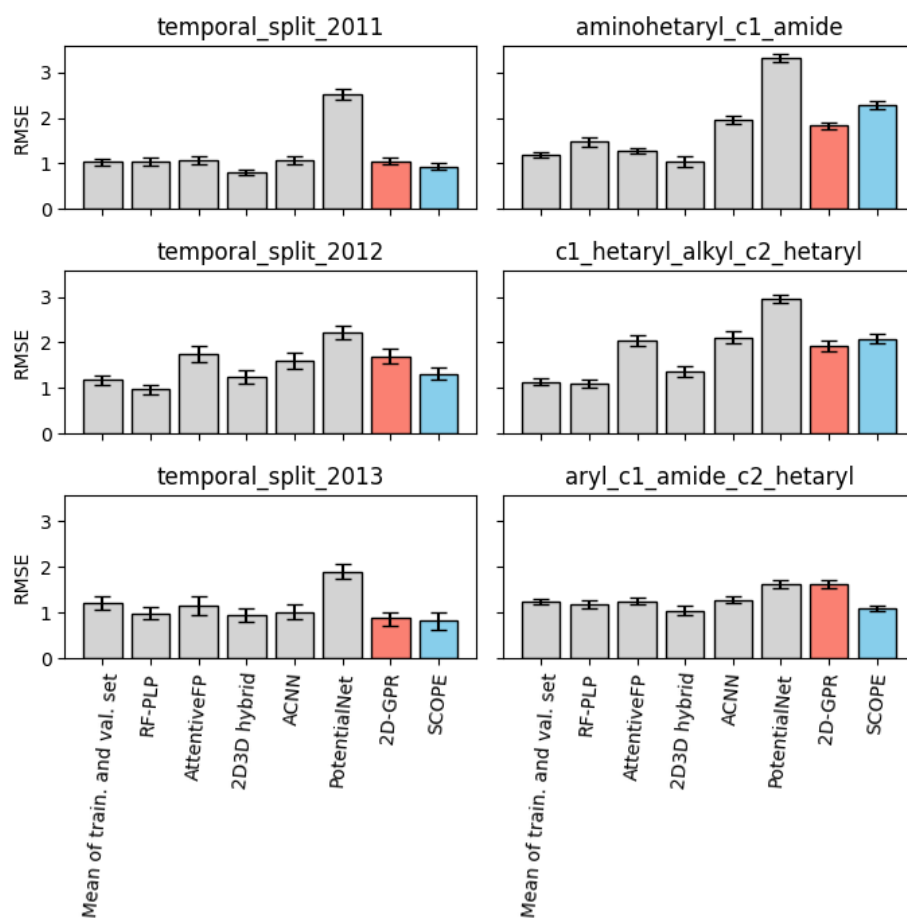


Figure S12: Performance comparison between SCOPE (blue) and 2D-GPR (red) against third-party methods for temporal splits (left) and binding mode splits (right) using RMSE. Prediction error is estimated using 95% confidence interval.

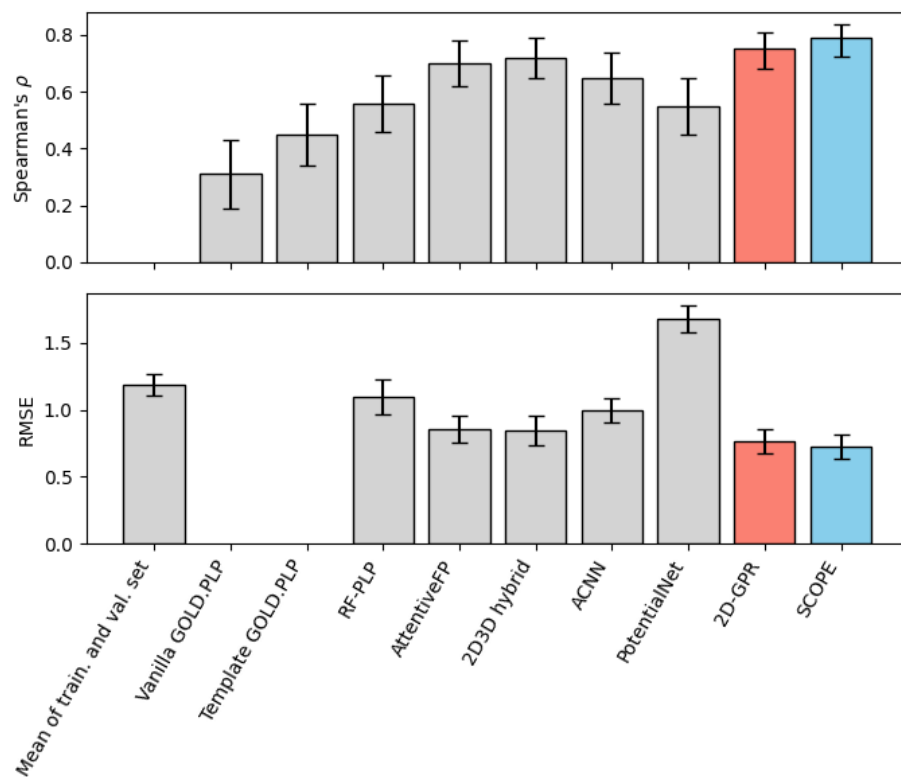


Figure S13: Performance comparison on the PDE10A dataset for random split between SCOPE (blue) and 2D-GPR (red) against third-party methods using Spearman's ρ and RMSE. Prediction error is estimated using 95% confidence interval.

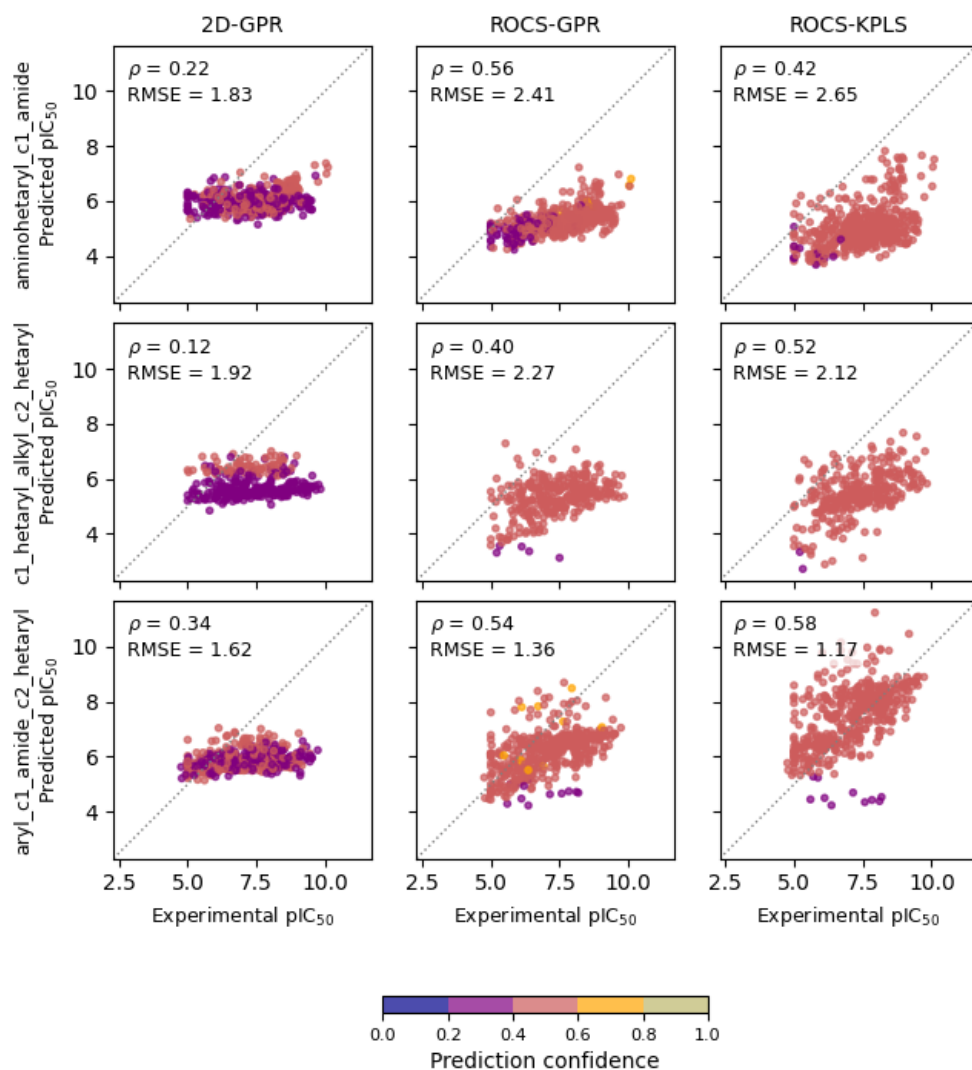


Figure S14: Predicted versus experimental pIC_{50} obtained using 2D-GPR (left), ROCS-GPR (middle), and ROCS-kPLS (right) for the three binding mode splits of the PDE10A dataset. Data points are colored by their corresponding prediction confidence. For each panel, Spearman's ρ and RMSE are reported.

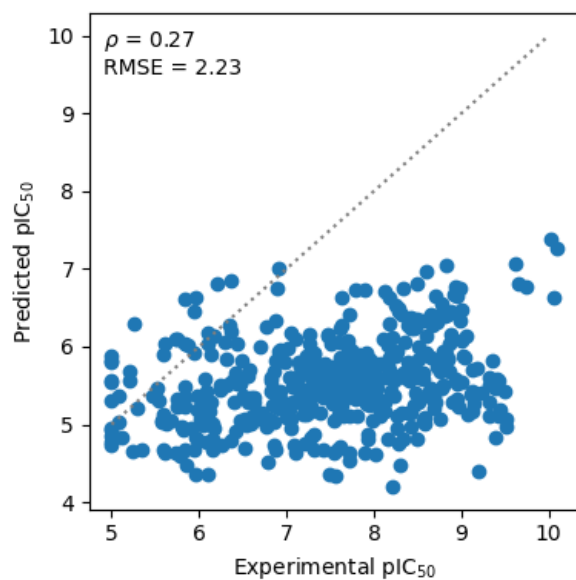


Figure S15: Predicted versus experimental pIC_{50} obtained using an alternative 2D GPR model with Tanimoto kernel. This model exhibits similarly poor performance with flattened predictions, indicating that the observed prediction flattening is not attributable to the choice of similarity representation.

Table S6: Details of the eight Sutherland datasets.

Dataset	Protein name	Activity type	Training size	Test size	Total size
ACE	Angiotensin Converting Enzyme	pIC ₅₀	76	38	114
ACHE	Acetylcholinesterase	pIC ₅₀	74	37	111
BZR	Benzodiazepine Receptor	pIC ₅₀	98	49	147
COX2	Cyclooxygenase-2	pIC ₅₀	188	94	282
DHFR	Dihydrofolate Reductase	pIC ₅₀	237	124	361
GPB	Glycogen Phosphorylase b	pK _i	44	22	66
THERM	Thermolysin	pK _i	51	25	76
THR	Thrombin	pK _i	59	29	88

Table S7: Coefficient of determination (COD) across the eight Sutherland datasets for all evaluated models.

	ACE	ACHE	BZR	COX2	DHFR	GPB	THERM	THR
CoMFA	0.49	0.47	0.00	0.29	0.59	0.42	0.54	0.63
CoMSIA basic	0.52	0.44	0.08	0.03	0.52	0.46	0.36	0.55
CoMSIA extra	0.49	0.44	0.12	0.37	0.53	0.59	0.53	0.63
Open3DQSAR	0.69	0.67	0.17	0.32	0.60	0.50	0.51	0.67
Open3DALIGN/ Open3DQSAR	0.54	0.65	0.24	0.28	0.53	0.41	-0.18	0.30
COSMOsar3D	0.62	0.61	0.13	0.43	0.58	0.63	0.59	0.66
QMFA-water	0.71	0.61	0.21	0.38	0.63	0.58	0.52	0.64
QMFA- formamide	0.60	0.57	0.16	0.38	0.63	0.47	0.52	0.64
QMOD	0.32	0.56	0.27	0.22	0.46	0.46	0.39	0.42
SCOPE	0.63	0.72	0.30	0.29	0.68	0.54	0.43	0.50

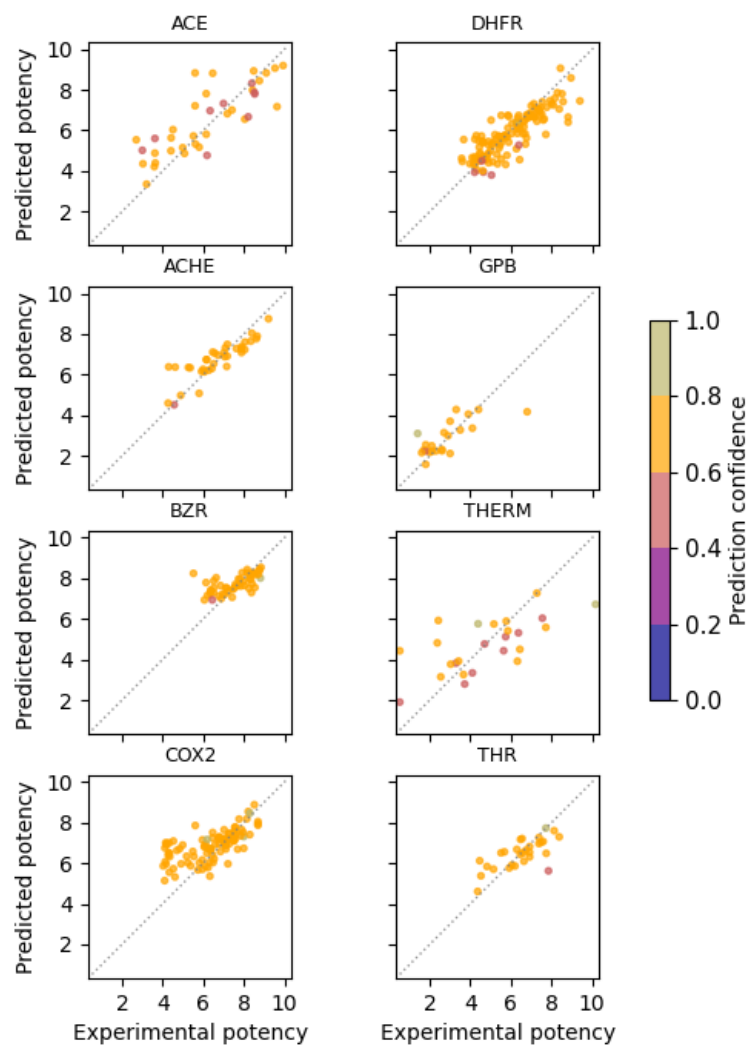


Figure S16: Predicted versus experimental potency for all eight Sutherland datasets. Data points are colored by their corresponding prediction confidence.

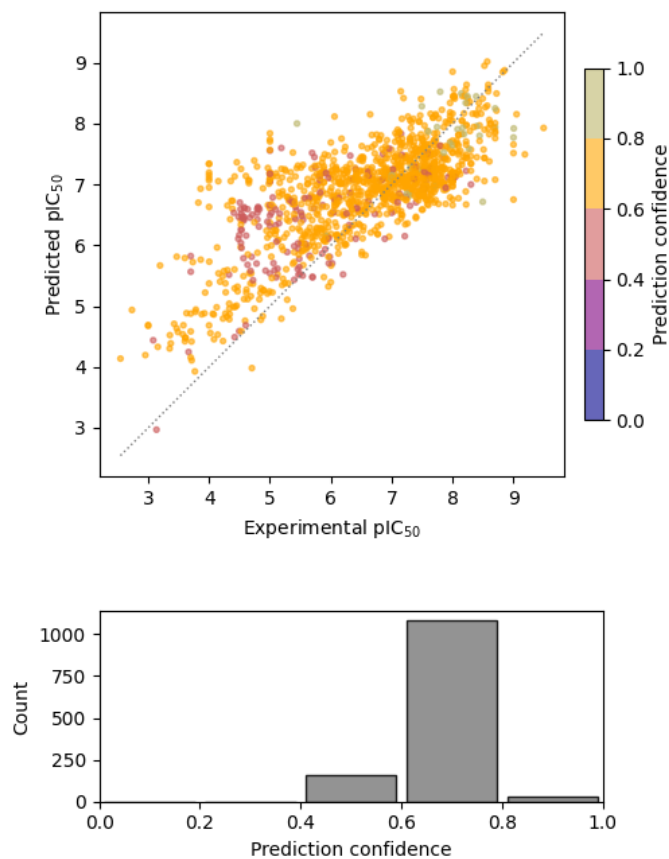


Figure S17: Predicted versus experimental pIC_{50} for BACE1 validation set. Data points are colored by their corresponding prediction confidence. Histogram of prediction confidence is shown on the bottom for reference.

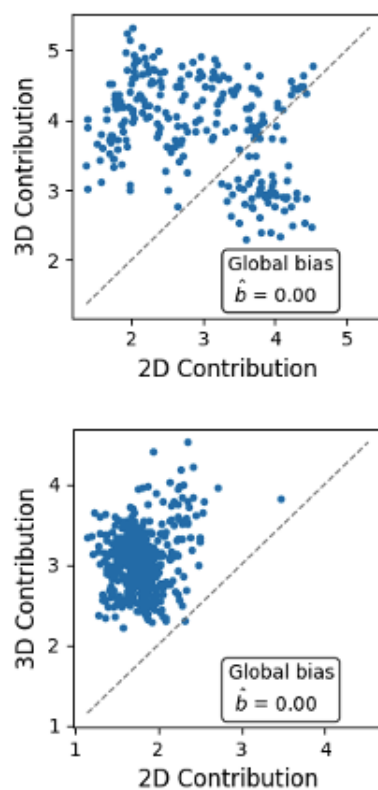


Figure S18: 3D Versus 2D contributions to predicted activities of test set for a representative temporal split (top) and binding mode split (bottom) using the PDE10A dataset. The fitted global bias term is near 0 in both cases.

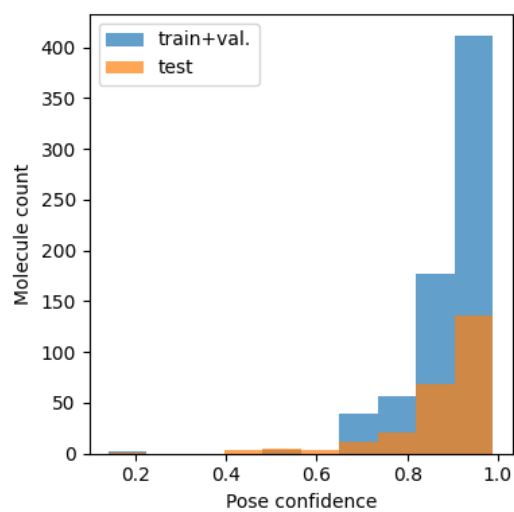


Figure S19: Histogram of POSIT pose confidence for training+validation and test sets under 2011 temporal split for the PDE10A dataset.

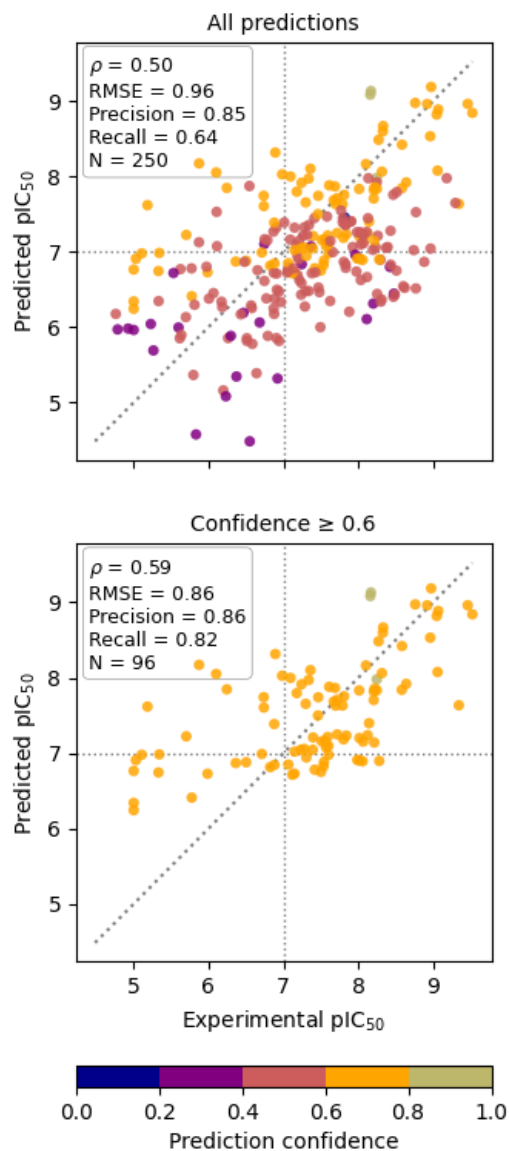


Figure S20: Predicted versus experimental pIC₅₀ values for the PDE10A 2011 temporal split test set. Results are shown for all test-set predictions (top) and after applying a confidence threshold of 0.6 (bottom). Points are colored by prediction confidence.

References

- [1] UniProt Consortium, *Nucleic Acids Res.* **2025**, *53*, D609.
- [2] T. Liu, L. Hwang, S. K. Burley, C. I. Nitsche, C. Southan, W. P. Walters, M. K. Gilson, *Nucleic Acids Res.* **2025**, *53*, D1633.
- [3] H. Du, X. Zhang, Z. Wu, O. Zhang, S. Gu, M. Wang, F. Zhu, D. Li, T. Hou, P. Pan, *Nucleic Acids Res.* **2025**, *53*, D1322.
- [4] G. K. Kanev, C. de Graaf, B. A. Westerman, I. J. P. de Esch, A. J. Kooistra, *Nucleic Acids Res.* **2021**, *49*, D562.
- [5] B. Srinivasan, M. D. Lloyd, *J. Med. Chem.* **2025**, *68*, 2052.
- [6] OpenEye, Cadence Molecular Sciences, Santa Fe, NM, *Spruce* **2026**.
- [7] G. L. Warren, T. D. Do, B. P. Kelley, A. Nicholls, S. D. Warren, *Drug Discov. Today* **2012**, *17*, 1270.
- [8] OpenEye, Cadence Molecular Sciences, Santa Fe, NM, *OpenEye Molecular Dynamics Affinity v6.5.8* **2026**.
- [9] J. Sørensen, C. C. Bannan, G. Calabrø, V. Jain, G. Ovanesyan, A. Smith, S. Zhang, C. I. Bayly, T. A. Darden, M. T. Geballe, D. N. LeBard, M. McGann, J. B. Moon, H. S. Muddana, A. Shewmaker, J. LaFon, R. W. Tolbert, A. G. Skillman, A. Nicholls, Orion® A Cloud-Native Molecular Design Platform, in V. Poongavanam, V. Ramaswamy (Editors), *Computational Drug Discovery*, chapter 24, Wiley-VCH **2024**.
- [10] OpenEye, Cadence Molecular Sciences, Santa Fe, NM, *Short Trajectory MD with Analysis Floe [STMD]*, *OpenEye Molecular Dynamics Affinity v6.5.8* **2026**.
- [11] P. Eastman, R. Galvelis, R. P. Peláez, C. R. A. Abreu, S. E. Farr, E. Gallicchio, A. Gorenko, M. M. Henry, F. Hu, J. Huang, A. Krämer, J. Michel, J. A. Mitchell, V. S. Pande, J. P. Rodrigues, J. Rodriguez-Guerra, A. C. Simmonett, S. Singh, J. Swails, P. Turner, Y. Wang, I. Zhang, J. D. Chodera, G. De Fabritiis, T. E. Markland, *J. Phys. Chem. B* **2023**, *128*, 109.

- [12] S. Boothroyd, P. K. Behara, O. C. Madin, D. F. Hahn, H. Jang, V. Gapsys, J. R. Wagner, J. T. Horton, D. L. Dotson, M. W. Thompson, J. Maat, T. Gokey, L.-P. Wang, D. J. Cole, M. K. Gilson, J. D. Chodera, C. I. Bayly, M. R. Shirts, D. L. Mobley, *J. Chem. Theory Comput.* **2023**, *19*, 3251.
- [13] A. R. McIsaac, P. K. Behara, T. Gokey, C. Cavender, J. Horton, L. Wang, B. R. Westbrook, M. W. Thompson, M. Osato, H. M. Baumann, H. Jang, J. R. Wagner, D. J. Cole, C. I. Bayly, D. L. Mobley, *OpenFF Force Fields: Sage 2.2.0* **2024**.
- [14] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*, 3696.
- [15] J. A. Grant, B. T. Pickup, A. Nicholls, *J. Comput. Chem.* **2001**, *22*, 608.
- [16] OpenEye, Cadence Molecular Sciences, Santa Fe, NM, *ZAP TK* **2026**.
- [17] L. Martínez, R. Andrade, E. G. Birgin, J. M. Martínez, *J. Comput. Chem.* **2009**, *30*, 2157.
- [18] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926.
- [19] S. Miyamoto, P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 952.
- [20] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327.
- [21] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577.
- [22] D. C. Liu, J. Nocedal, *Mathematical programming* **1989**, *45*, 503.
- [23] A. Brünger, C. L. Brooks, M. Karplus, *Chem. Phys. Lett.* **1984**, *105*, 495.
- [24] N. Goga, A. J. Rzepiela, A. H. de Vries, S. J. Marrink, H. J. C. Berendsen, *J. Chem. Theory Comput.* **2012**, *8*, 3637.
- [25] C. W. Hopkins, S. Le Grand, R. C. Walker, A. E. Roitberg, *J. Chem. Theory Comput.* **2015**, *11*, 1864.