Support material

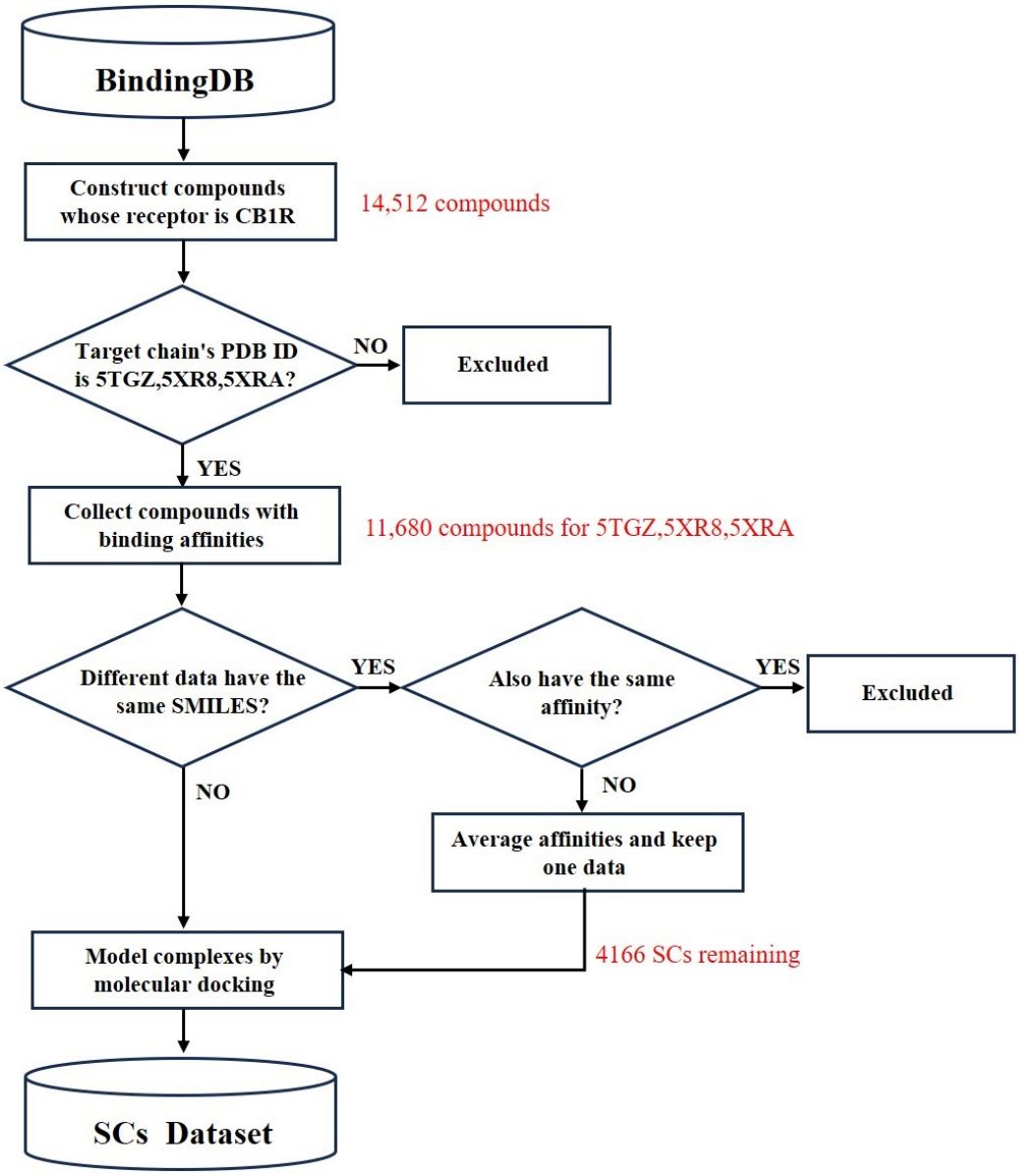


Figure S1. Schematic diagram of the process of building SCsDB.

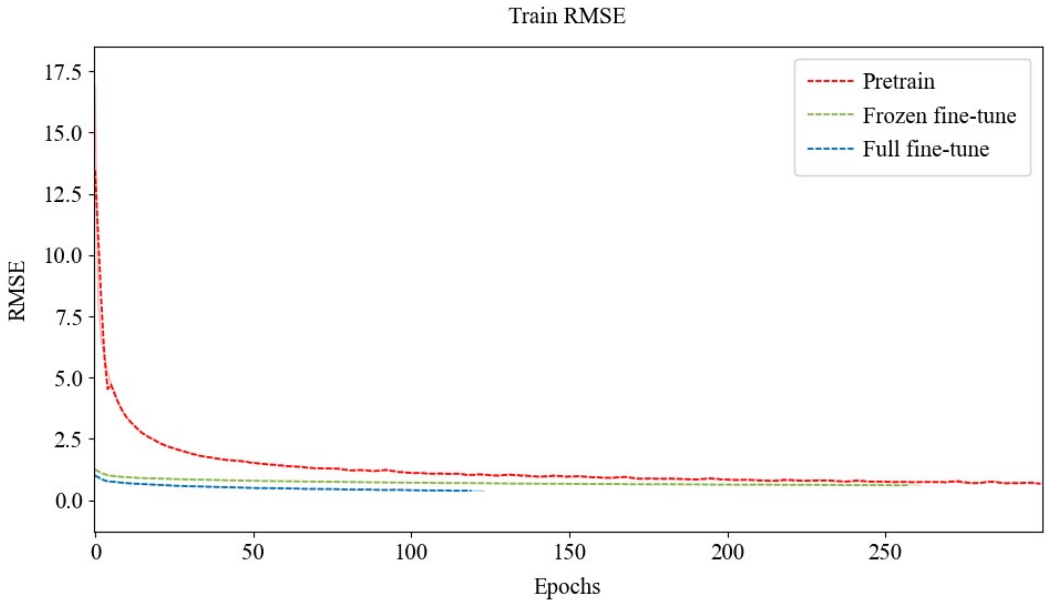


Figure S2. Plot of model train RMSE changes under three different training strategies.

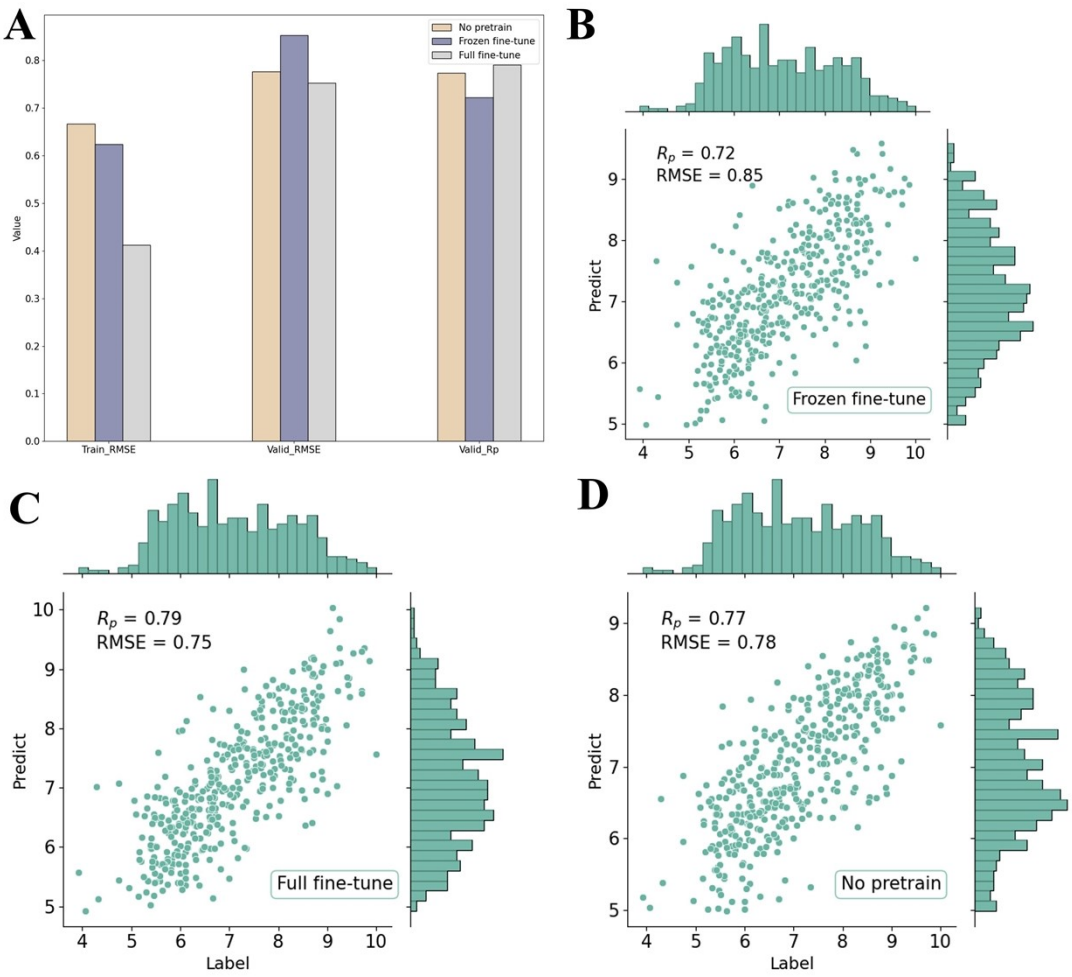


Figure S3. Full fine-tuning model achieved the best metrics among all the training strategies. (A) Comparison of metrics of the three training strategies on the training and validation sets. (B) Frozen fine-tuning model. (C) Full fine-tuning model. (D) No pre-training model.

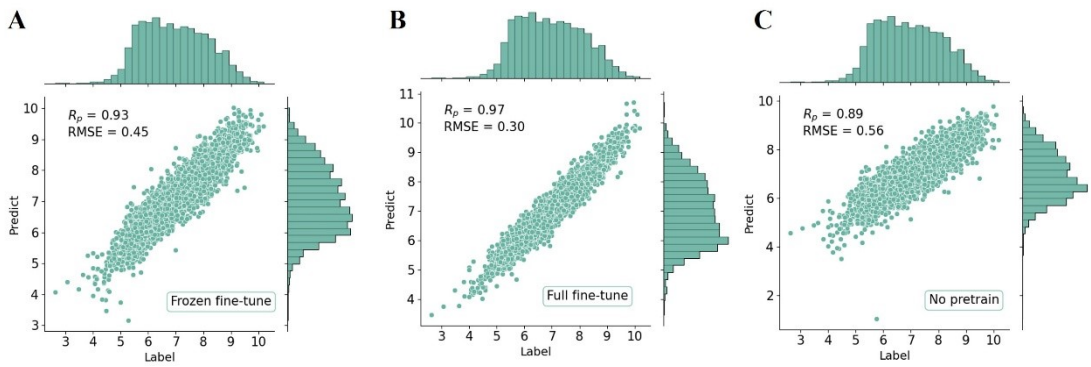


Figure S4. Comparison of metrics of the three training strategies on the training sets.

(A) Frozen fine-tuning model. (B) Full fine-tuning model. (C) No pre-training model.

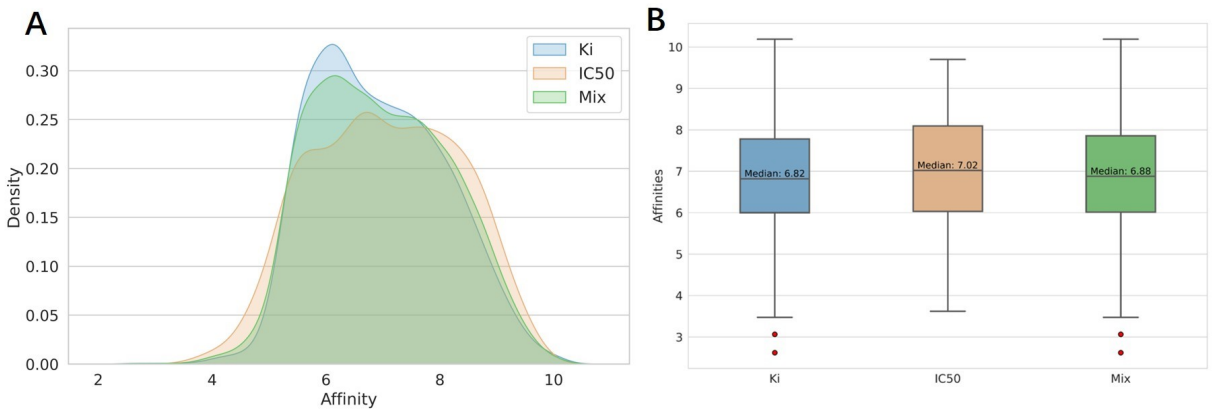


Figure S5. Comparison of distributional differences across labeled datasets.

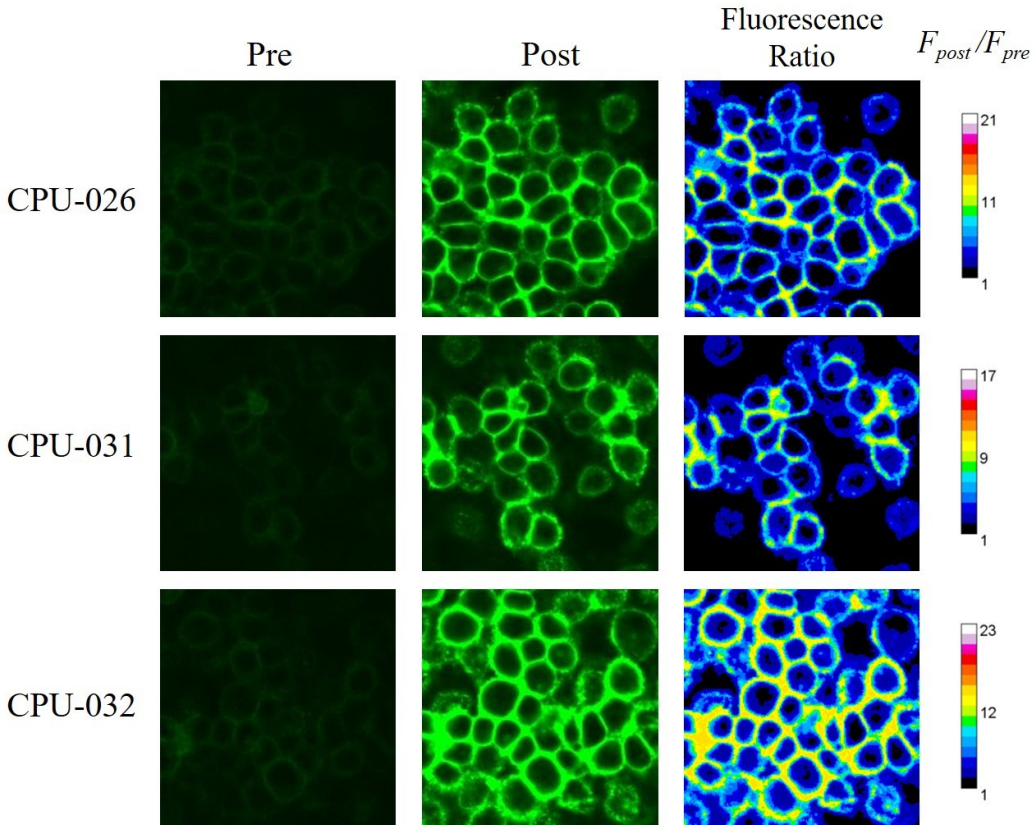


Figure S6. Plot of changes in cellular fluorescence expression.

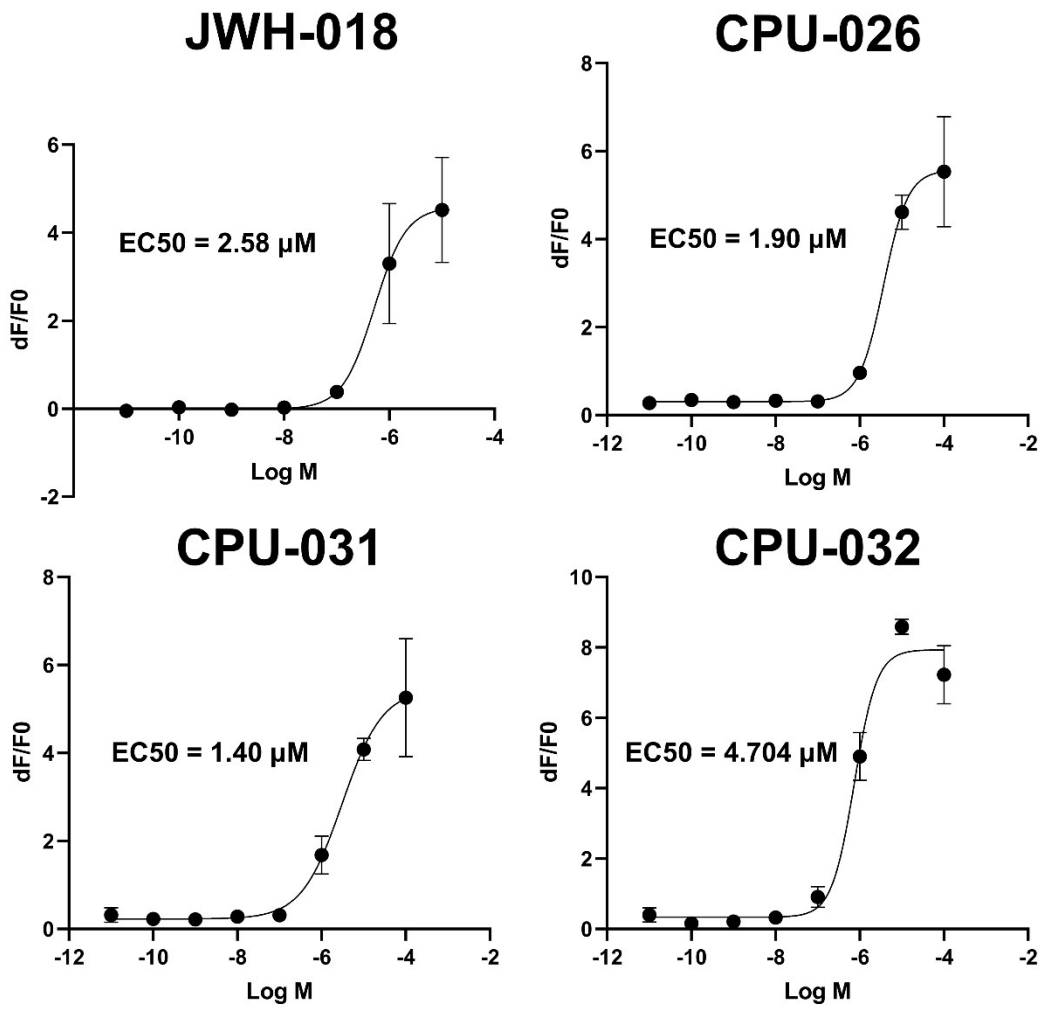


Figure S7. EC50 values of four SCs corresponding to changes in cellular fluorescence expression.

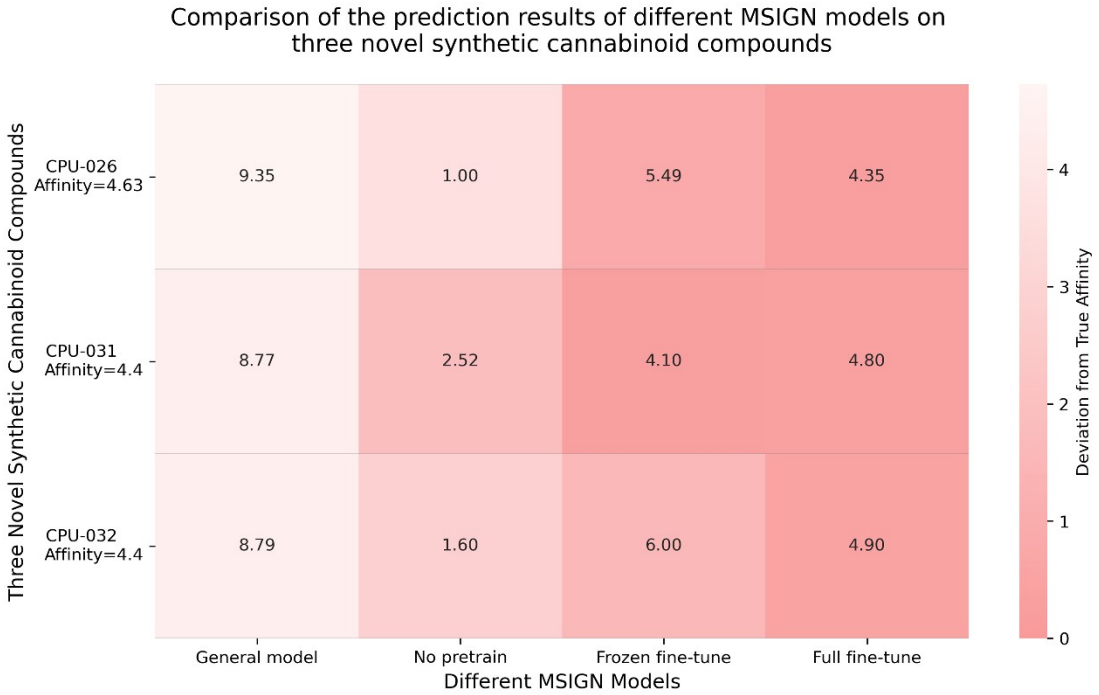


Figure S8. Comparison of the prediction results of different MSIGN models on three novel synthetic cannabinoid compounds.

Table S1. Meanings and vector dimensions of initial features of complex atom interaction diagrams.

Type	Feature	Description	Dimension
Atom Node Features	Atomic symbol	Atom type (C, N, O, S, etc.)	10
	Degree of atom	Number of non-hydrogen atoms directly bonded to the atom	7
	Implicit valence	Undisplayed valence state of the atom	7
	Hybridization type	Hybridization state of atomic orbitals	5
	Aromaticity	Whether the atom is part of an aromatic ring	1
	Number of hydrogens	Number of hydrogen atoms bonded to the atom	5
Bond Features	Bond type	Single, double, triple and aromatic bond	4
	Conjugation	Whether the bond is conjugated	1
	In a ring	Whether the bond is part of a ring	1
Geometric Features	Maximum angle	Maximum angle formed by adjacent atoms	1
	Sum of angles	Sum of all angles	1
	Average angle	Average value of all angles	1
	Maximum triangular area	Maximum triangle area formed by adjacent atoms	1
	Sum of areas	Sum of all triangular areas	1
	Average area	Average value of all triangular areas	1
	Maximum distance	Maximum distance between non-bonded atoms	1
	Sum of distances	Sum of all non-bonded distances	1

	Average distance	Average value of non-bonded distances	1
	Manhattan distance (L1)	Sum of absolute differences (L1 norm-based distance)	1
	Euclidean distance (L2)	Root sum of squared differences (L2 norm-based distance)	1

Table S2. Meaning and vector dimensions of initial features of functional group diagrams of ligand molecules.

Type	Feature	Description	Dimension
Functional Group Features	C	Number of carbon atoms in the functional group	11
	O	Number of oxygen atoms in the functional group	6
	N	Number of nitrogen atoms in the functional group	6
	P	Number of phosphorus atoms in the functional group	6
	S	Number of sulfur atoms in the functional group	6
	Other atoms	Presence of other atoms beyond C, O, N, P, S in the functional group	1
	Halogens	Presence of halogens in the functional group	1
Bond features	Single bonds	Number of single bonds in the functional group	11
	Double bonds	Number of double bonds in the functional group	8
	Triple bonds	Number of triple bonds in the functional group	8
	Aromatic bonds	Number of aromatic bonds in the functional group	8
	In a ring	Whether the functional group forms a ring structure	1

Table S3. Meaning and vector dimension of initial features of protein amino acid residue maps.

Type	Feature	Description	Dimension
Amino Acid Residue Features	Amino acid residue	20 types of amino acid residues	20
	Spatial coodinates	C_α atom's 3D coordinates (x,y,z)	3
	Hydrophobicity	Hydrophobicity property of the amino acid	1
	Polaity	Polarity property of the amino acid	1
	DSSP	Secondary structure features extracted using DSSP	14

Table S4. Additional performance metrics across three external benchmarks.

Model	CASF-2013		CASF-2016		2019 core set	
	MAE	R_s	MAE	R_s	MAE	R_s
Random Forest	1.470 [1.277, 1.671]	0.668 [0.521, 0.778]	1.287 [1.166, 1.413]	0.644 [0.554, 0.722]	1.193 [1.165, 1.220]	0.528 [0.505, 0.551]
XGBoost	1.320 [1.149, 1.501]	0.744 [0.617, 0.831]	1.156 [1.049, 1.265]	0.741 [0.668, 0.800]	1.162 [1.136, 1.189]	0.569 [0.547, 0.590]
MSIGN	1.032 [0.898, 1.177]	0.861 [0.794, 0.902]	0.915 [0.839, 0.992]	0.851 [0.811, 0.882]	1.061 [1.036, 1.086]	0.656 [0.637, 0.674]

Table S5. Comparison of performance indicators between two single-label models and a mixed-label model.

Model	Training set size	Validation set size	RMSE	R_p
MSIGN_Ki	2332	584	0.795	0.716
MSIGN_IC50	972	243	0.803	0.788
MSIGN_Mix	3332	834	0.753	0.790

Table S6. Comparison of model predictions with experimental measurements

SCs	Real affinity	General model	No pretrain	Frozen fine-tune	Full fine-tune
CPU-026	4.63	9.35	1.00	5.49	4.35
CPU-031	4.40	8.77	2.52	4.10	4.80
CPU-032	4.40	8.79	1.60	6.00	4.90

Table S7. Summary of different training and fine-tuning strategies.

Strategy	Initial Weights	Layers Frozen	Layers Updated (Trainable)
General	Random	None	All layers (trained on PDBbind)
No-pretrain	Random	None	All layers (trained only on SCsDB)
Frozen	Pre-trained	All GNN layers	Only the final MLP predictor(fine-tuned on SCsDB)
Full fine-tune	Pre-trained	None	All layers (fine-tuned on SCsDB)

Table S8. Comparison of baseline model predictions with experimental measurements

Model	MAE	RMSE
MSIGN	0.39	0.40
RF	1.66	1.68
XGBoost	1.76	1.76
Vina	3.49	3.49
EHIGN	2.46	2.47
SS-GNN	2.47	2.51

Surface plasmon resonance (SPR) experimental procedure

I. Coating Protein

Protein concentration: Stock solution at 300 μg/mL, diluted 10-fold to 30 μg/mL using sodium acetate buffer (pH 4.0).

Buffer: 1× PBS-P+ buffer; CM5 sensor chip.

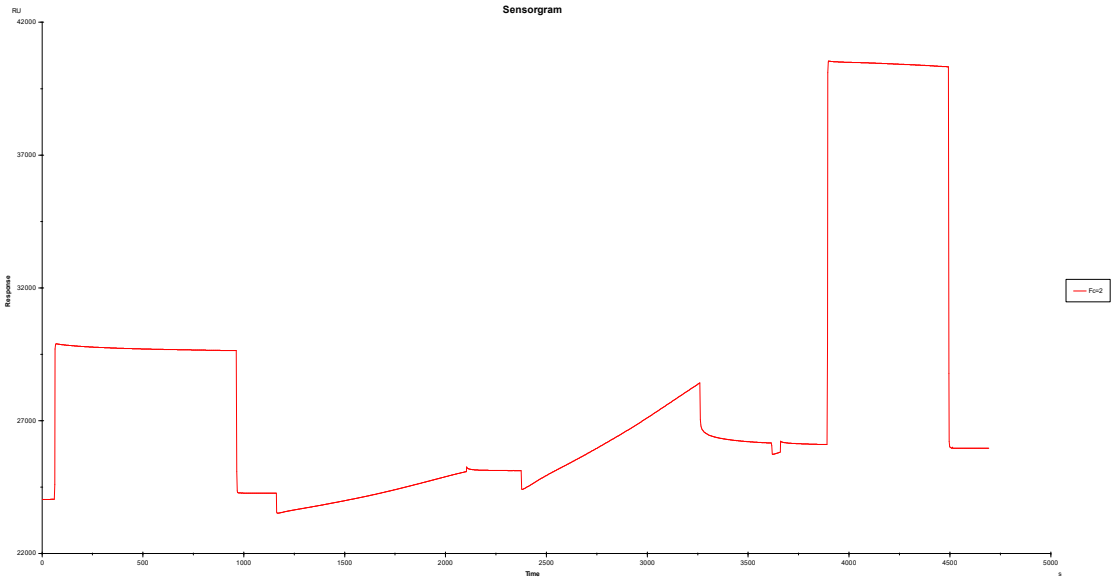
Procedure

Activation: EDC/NHS mixture, flow rate 10 μL/s, activation time 900 seconds.

Coupling: Flow rate 10 μL/s; first injection: 900 seconds, second injection: 900 seconds.

Blocking: Ethanolamine, flow rate 10 μL/s, duration 900 seconds.

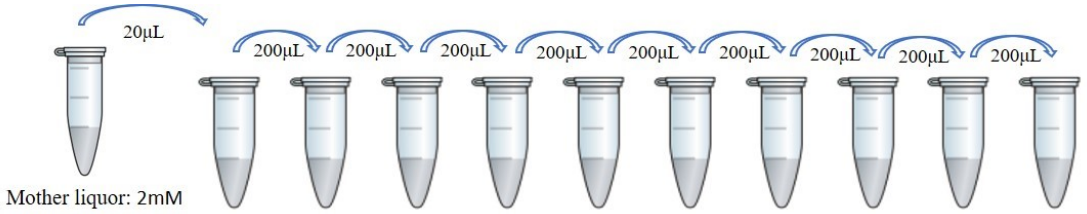
Final coupling amount: Approximately 1900 resonance units (RU).



II. Small Molecule Interaction

Buffer: 1.05× PBS-P+ buffer (containing 5% DMSO).

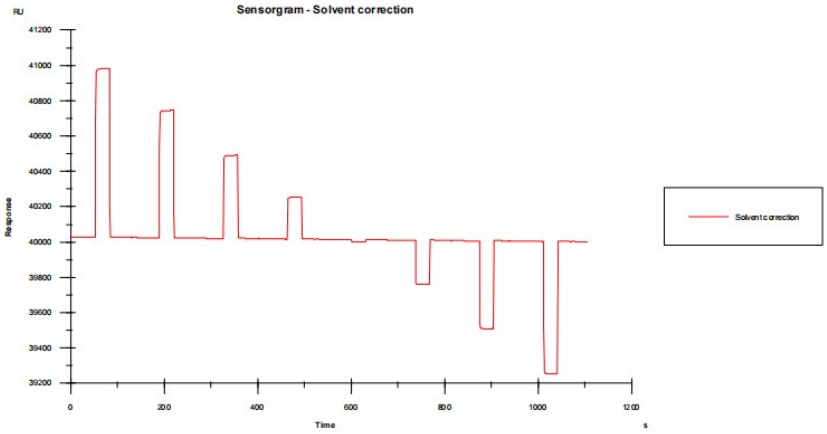
Sample preparation: Samples dissolved in DMSO.



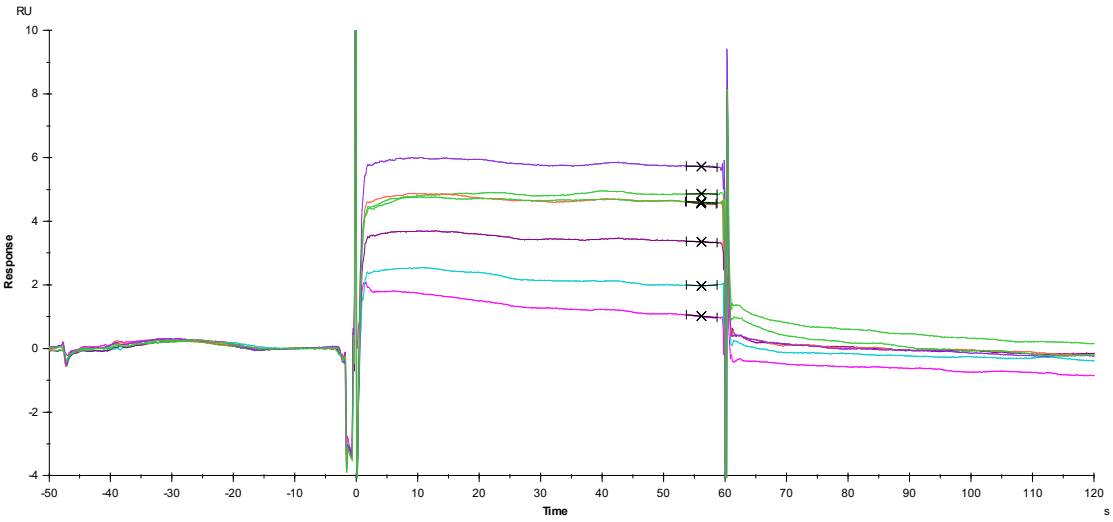
1.05×PBS-P+:	380	-	-	-	-	-	-	-	-	-
Buffer(5%DMSO):	-	200	200	200	200	200	200	200	200	200
Sample concentration(μM)	100	50	25	12.5	6.25	3.125	1.56	0.78	0.39	0.2

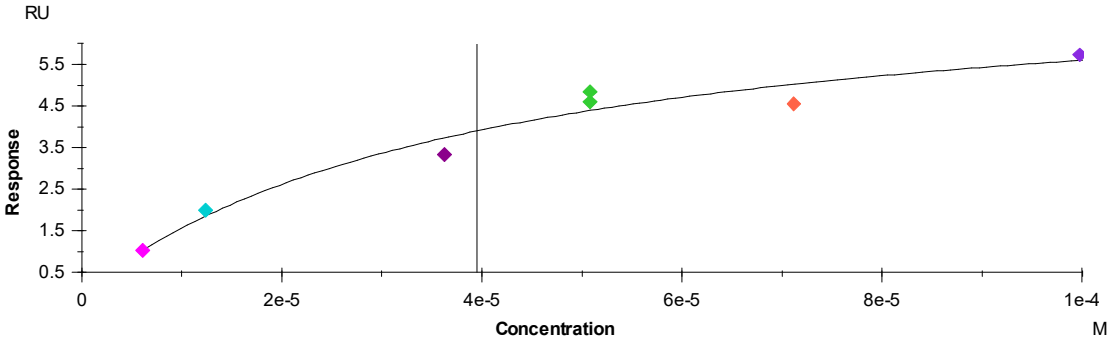
1. Solvent correction curves

Buffer\Vial	1	2	3	4	5	6	7	8
4.5% DMSO		200	400	600	800	1000	1200	1400
5.8% DMSO	1400	1200	1000	800	600	400	200	
Total volume	1400	1400	1400	1400	1400	1400	1400	1400



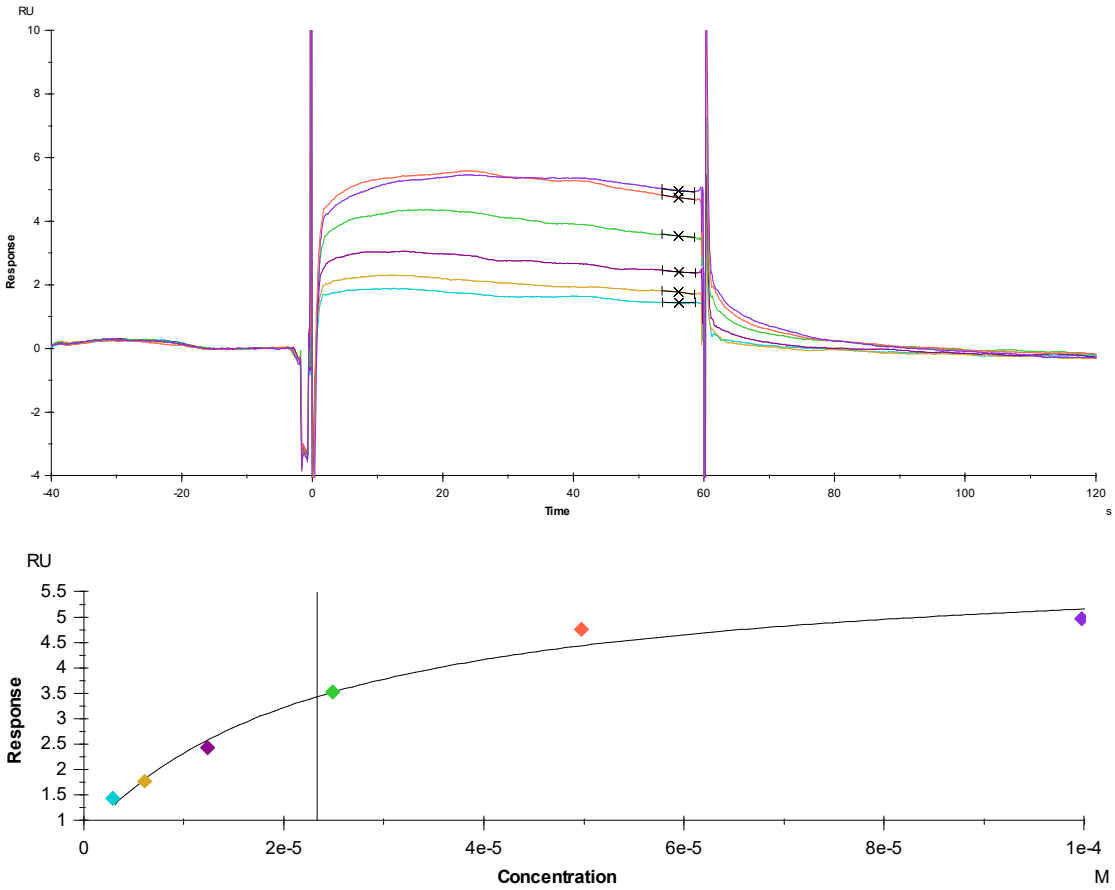
2. Positive control result chart





KD (M)	Rmax (RU)	offset (RU)	Chi² (RU²)
3.949E-5			0.164
	7.831	-0.01490	

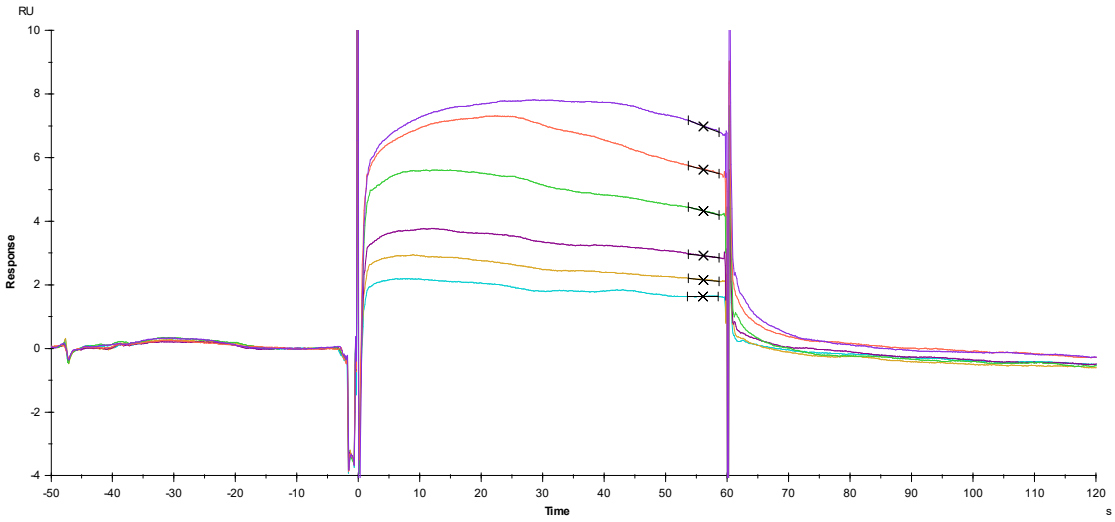
3. CPU-026

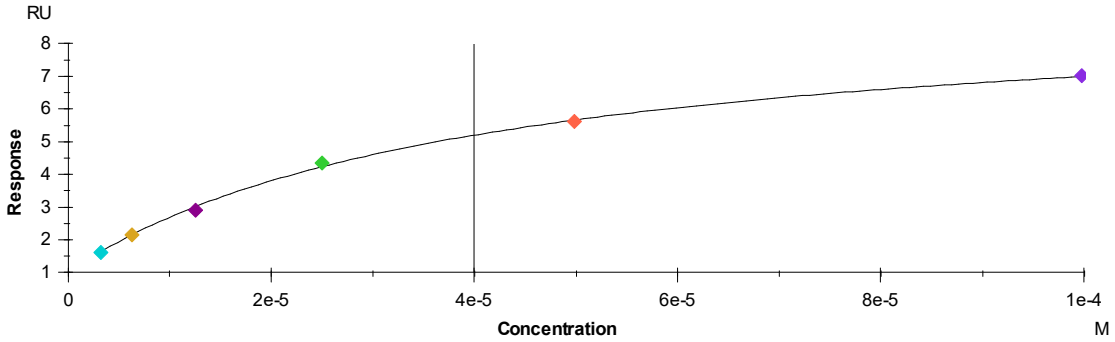


concentrations in that order: 3.125µM, 6.25µM, 12.5µM, 25µM, 50µM, 100µM

KD (M)	Rmax (RU)	offset (RU)	Chi² (RU²)
2.337E-5			0.0602
	5.543	0.6645	

4. CPU-031

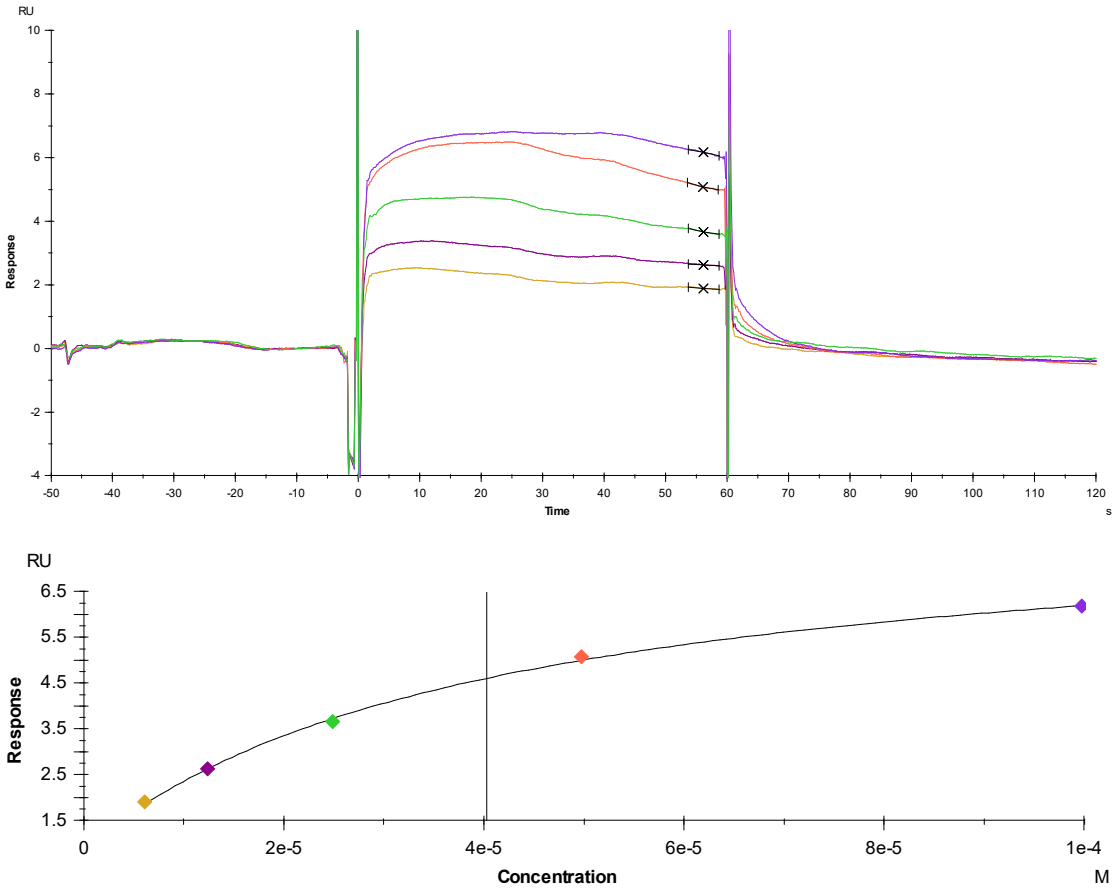




concentrations in that order: 3.125µM, 6.25µM, 12.5µM, 25µM, 50µM, 100µM

KD (M)	Rmax (RU)	offset (RU)	Chi² (RU²)
3.994E-5			0.00639
	8.368	1.016	

5. CPU-032



concentrations in that order: 6.25µM, 12.5µM, 25µM, 50µM, 100µM

KD (M)	Rmax (RU)	offset (RU)	Chi² (RU²)
4.029E-5			0.00553
	7.464	0.8752	