Supporting Information

Delta Machine Learning to Improve Scoring-Ranking-Screening Performances of Protein-Ligand Scoring Functions

Chao Yang[1] and Yingkai Zhang[1,2]*

[1]Department of Chemistry, New York University, New York, NY 10003, United States
[2]NYU-ECNU Center for Computational Chemistry at NYU Shanghai 200062, Shanghai, China
Email: yingkai.zhang@nyu.edu

Table S1. Training and validation sets.

	Subset	Source	Size	Binding affinity	Note
Training set	Binder set	6816 binders from Δ_{Vina}XGB's training set	9117	$pK_d(exp)$	Local optimized poses with and without waters
		1556 weak binders (pK_d<6) and 510 strong binders (pK_d>9) from PDBbind v2018 general set			
		235 strong binders obtained by flexible-redocking (pK_d>9 and RMSD<=1.0)			Docked poses with and without waters
	Decoy set 1	6321 decoys from Δ_{Vina}XGB's training set	7111	$pK_d(est1)^a$	CSAR decoys without water
		790 decoys obtained by flexible docking of very weak binders (pK_d<3) from BindingDB K_d data		$pK_d(est2)^b$	Docked decoys with and without waters
	Decoy set 2	E2E top1 docked poses of train Binder set ($RMSD^{E2E}$ - $RMSD^{opt}$ > 0.5 and $\lvert pK_d - Lin\_F9 \rvert$ < 3)	5715	$pK_d(exp)$	Docked poses with and without waters
Validation set	Binder set	Same as Δ_{Vina}XGB's validation set	946	$pK_d(exp)$	Local optimized poses with and without waters, and crystal poses
	Decoy set	E2E top1 docked poses of validation Binder set	632	$pK_d(exp)$	Docked poses with and without waters

[a] $pK_d(est1)$: the estimated pK_d for CSAR decoys; [b] $pK_d(est1)$: the estimated pK_d for BindingDB docked decoys.

Table S2. Feature set of Δ_{Lin_F9}XGB.

Feature type	Detail	Number
bSASA features	bSASA ligand terms	9
	bSASA protein terms	9
	bSASA complex terms	10
Vina features	polar-polar gauss terms	7
	polar-nonpolar gauss terms	7
	nonpolar-nonpolar gauss terms	7
	hydrogen bond gauss terms	5
	anti-hydrogen bond gauss terms	5
	metal bond gauss term	6
	ligand terms (number of torsions, number of rotors, ligand length, number of heavy atoms, number of hydrophobic atoms, maximum number of possible hydrogen bonds)	6
	repulsion, ad4_solvation, electrostatic	5
Bridge water features	Number of bridge waters Sum of Lin_F9 scores of protein-bridge water Sum of Lin_F9 scores of ligand-bridge water	3
Beta-cluster features	ligand beta score, ligand coverage	2
	ligand efficiency[a]	1
Ligand features	HeavyAtomMolWt, NumValenceElectrons, FpDensityMorgan{1,2,3}, LabuteASA, TPSA, NHOHCount, MolLogP, MolMR	10

[a]Ligand efficiency: pK_d(Lin_F9) divided by number of heavy atoms.

Table S3. The performances of Δ_{Lin_F9}XGB on the train and validation sets.

Dataset	Pearson's R	RMSE (in pK_d)	MAE (in pK_d)
Train set	1.000	0.051	0.035
Validation set	0.789	1.448	1.146

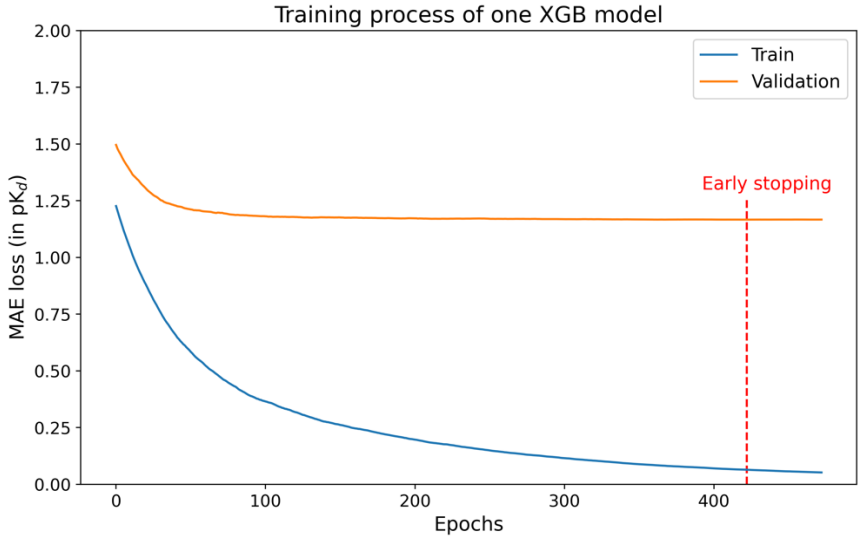

Table S4. CASF-2016 screening power at top1%, 5%, 10% for Vina, Δ_{Vina}XGB, Δ_{Lin_F9}XGB. (EF: enhancement factor, SC: success rate)

Scoring Functions	$EF_{1\%}$	$EF_{5\%}$	$EF_{10\%}$	$SC_{1\%}$	$SC_{5\%}$	$SC_{10\%}$
Vina	7.70	4.01	2.87	29.8%	40.4%	50.9%
Δ_{Vina}XGB	13.14	4.30	2.91	36.8%	52.6%	61.4%
Δ_{Lin_F9}XGB	12.61	4.42	3.02	40.4%	59.6%	68.4%

Table S5. Mean AUC values comparison of Vina, $\Delta_{Vina}RF_{20}$, Lin_F9 and $\Delta_{Lin\_F9}XGB$ on LIT-PCBA dataset.

Target set	Scoring Function				PDB	Number of	Number of
	Vina	$\Delta_{Vina}RF_{20}$	Lin_F9	$\Delta_{Lin\_F9}XGB$	Templates	Actives	Inactives
ADRB2*	0.356	**0.399**	0.312	0.393	4	17	312,433
ALDH1	0.582	0.590	0.586	**0.715**	2	7167	137,822
ESR1-ago*	0.703	0.687	0.675	**0.713**	15	13	5,582
ESR1-ant*	0.660	0.681	0.657	**0.696**	15	102	4,947
FEN1	**0.505**	**0.505**	0.463	0.460	1	369	355,323
GBA	0.620	0.595	**0.641**	0.615	3	166	294,202
IDH1	0.592	0.510	**0.599**	0.546	10	39	361,691
KAT2A	0.430	0.396	0.419	**0.566**	1	194	348,257
MAPK1*	**0.643**	0.605	0.606	0.591	15	308	62,522
MTORC1*	0.515	0.515	**0.536**	0.523	11	97	32,972
OPRK1*	0.525	0.487	**0.874**	0.851	1	24	269,776
PKM2	0.606	**0.641**	0.595	0.603	2	546	245,485
PPARG*	**0.825**	0.815	0.806	0.803	15	27	5,210
TP53*	0.609	0.596	**0.648**	0.612	6	79	4,168
VDR	**0.394**	0.373	0.373	0.364	1	882	355,094
Average	0.571	0.560	0.586	0.603			

PDB templates same as the original benchmark used are highlighted in green color. The 8 targets using cell-based phenotypic assays are marked with *.

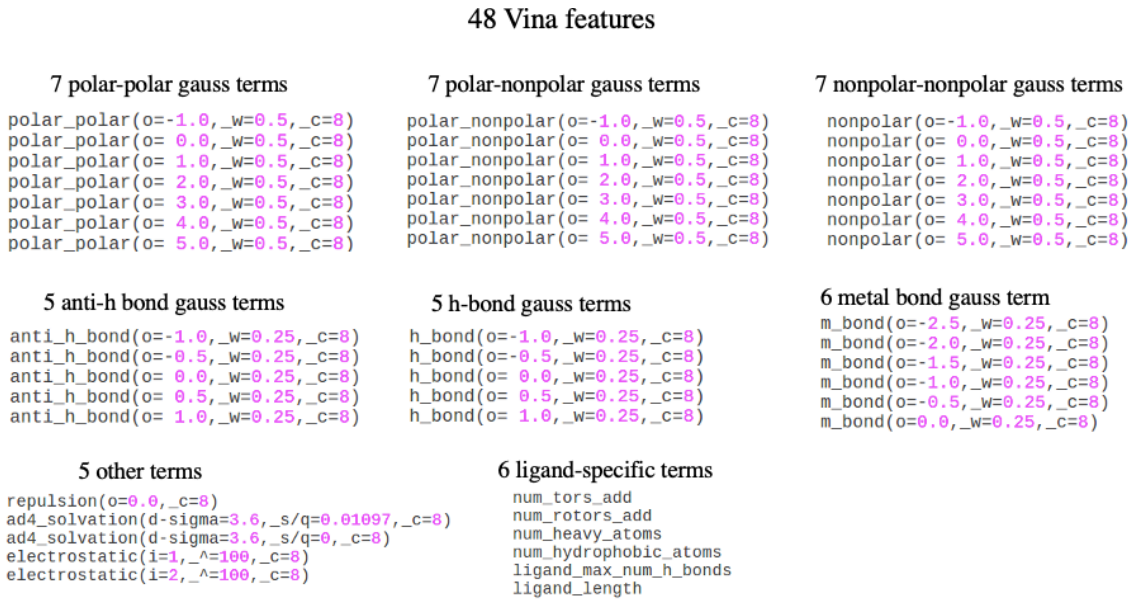

Figure S1. Details of 48 Vina features used in $\Delta_{\text{Lin_F9}}$XGB. The polar and nonpolar atom types are characterized based on XScore atom types. A series of gauss terms are used to describe polar-polar, polar-nonpolar, nonpolar-nonpolar interactions, anti-hbond, hbond and metal bond in different distances. The parameter o and _w represents the center and the width, respectively, of a gauss term. The anti-hbond term describes polar-polar atoms that can't possibly be hydrogen bond. For metal bond term, it describes metal-ligand interactions. For 6 ligand-specific terms and 5 other terms (1 repulsion term and 2 ad4_solvation terms and 2 electrostatic terms), these terms are obtained from original Vina 58 features. The cutoff (_c) distance is 8 Å.

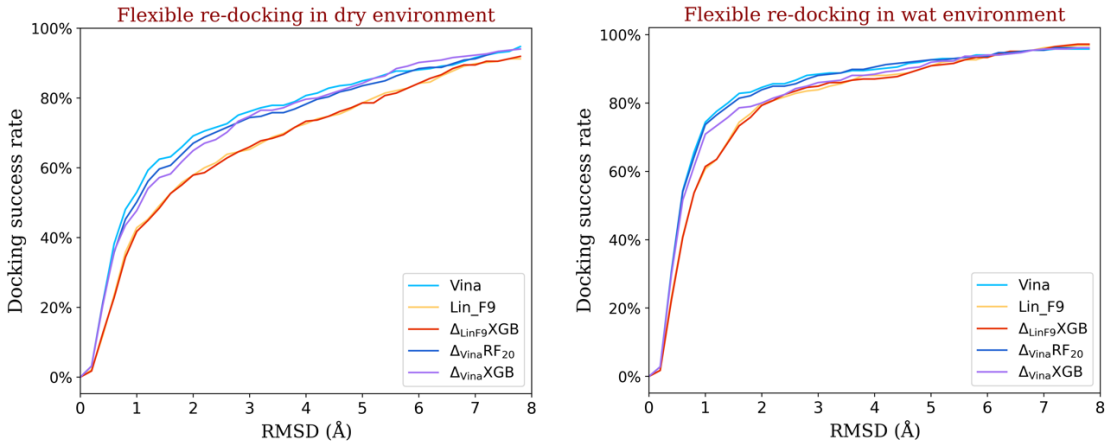

Figure S2. Success rates of flexible re-docking pose in different RMSD values. (A) and (B) show the docking success rates of the best-scored pose in dry environment and in water environment, respectively. Performances of Δ_{Lin_F9}XGB, Lin_F9 and Vina are colored red, orange and cyan, respectively.

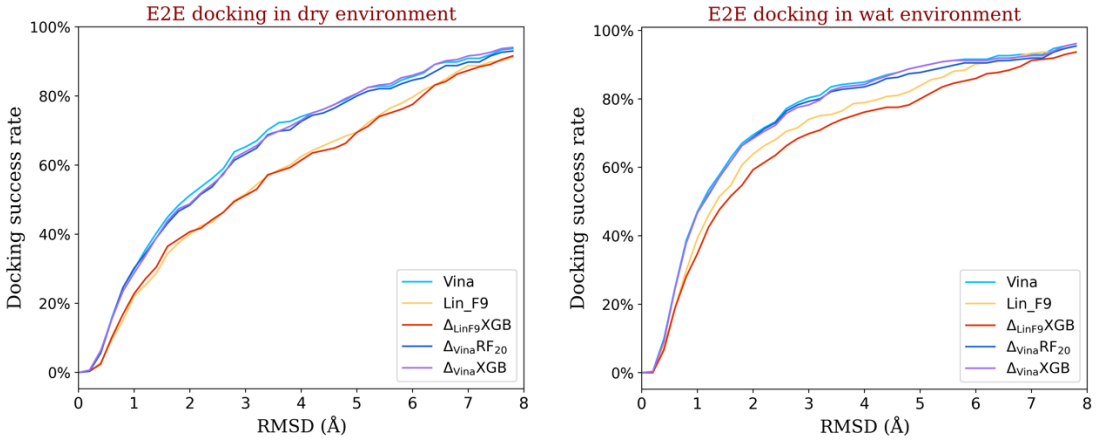

Figure S3. Success rates of E2E docking pose in different RMSD values. (A) and (B) show the docking success rates of the best-scored pose in dry environment and in water environment, respectively. Performances of Δ_{Lin_F9}XGB, Lin_F9 and Vina are colored red, orange and cyan, respectively.

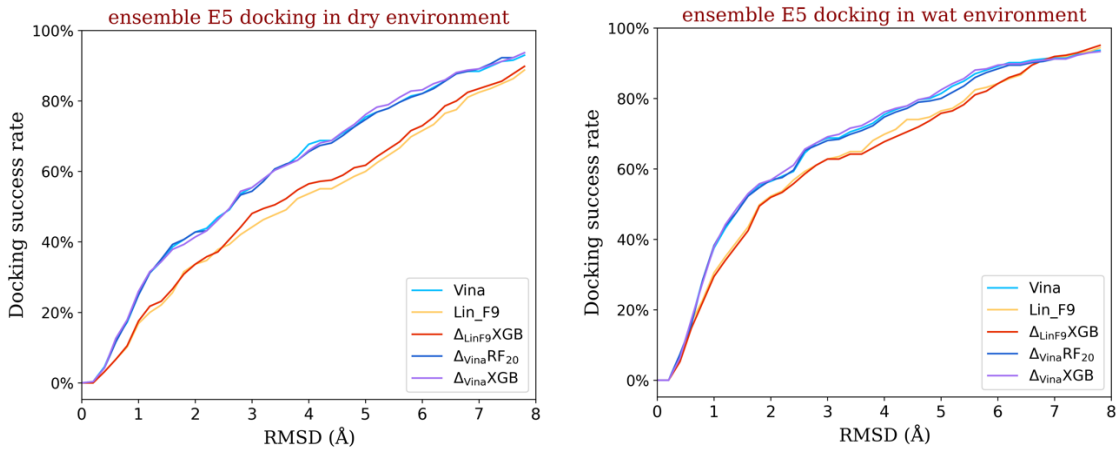

Figure S4. Success rates of ensemble E5 docking pose in different RMSD values. (A) and (B) show the docking success rates of the best-scored pose in dry environment and in water environment, respectively. Performances of Δ_{Lin_F9}XGB, Lin_F9 and Vina are colored red, orange and cyan, respectively.

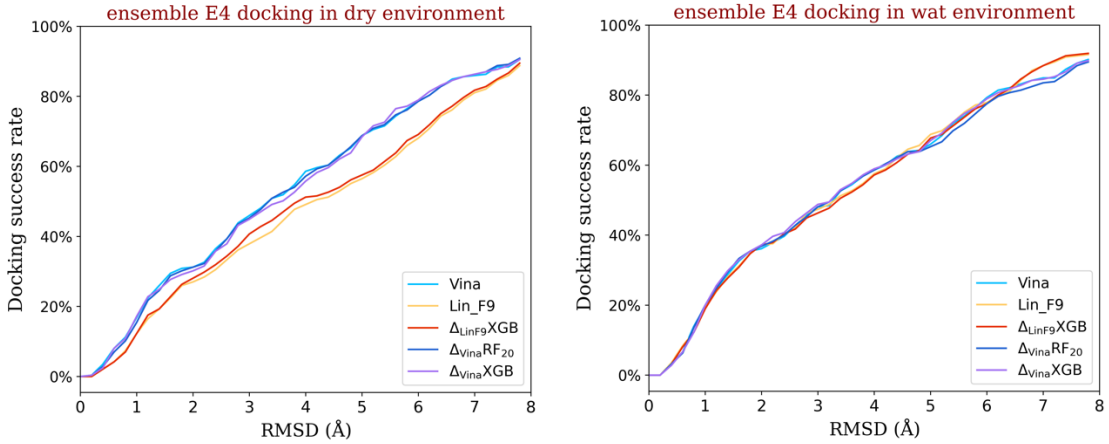

Figure S5. Success rates of ensemble E4 docking pose in different RMSD values. (A) and (B) show the docking success rates of the best-scored pose in dry environment and in water environment, respectively. Performances of Δ_{Lin_F9}XGB, Lin_F9 and Vina are colored red, orange and cyan, respectively.

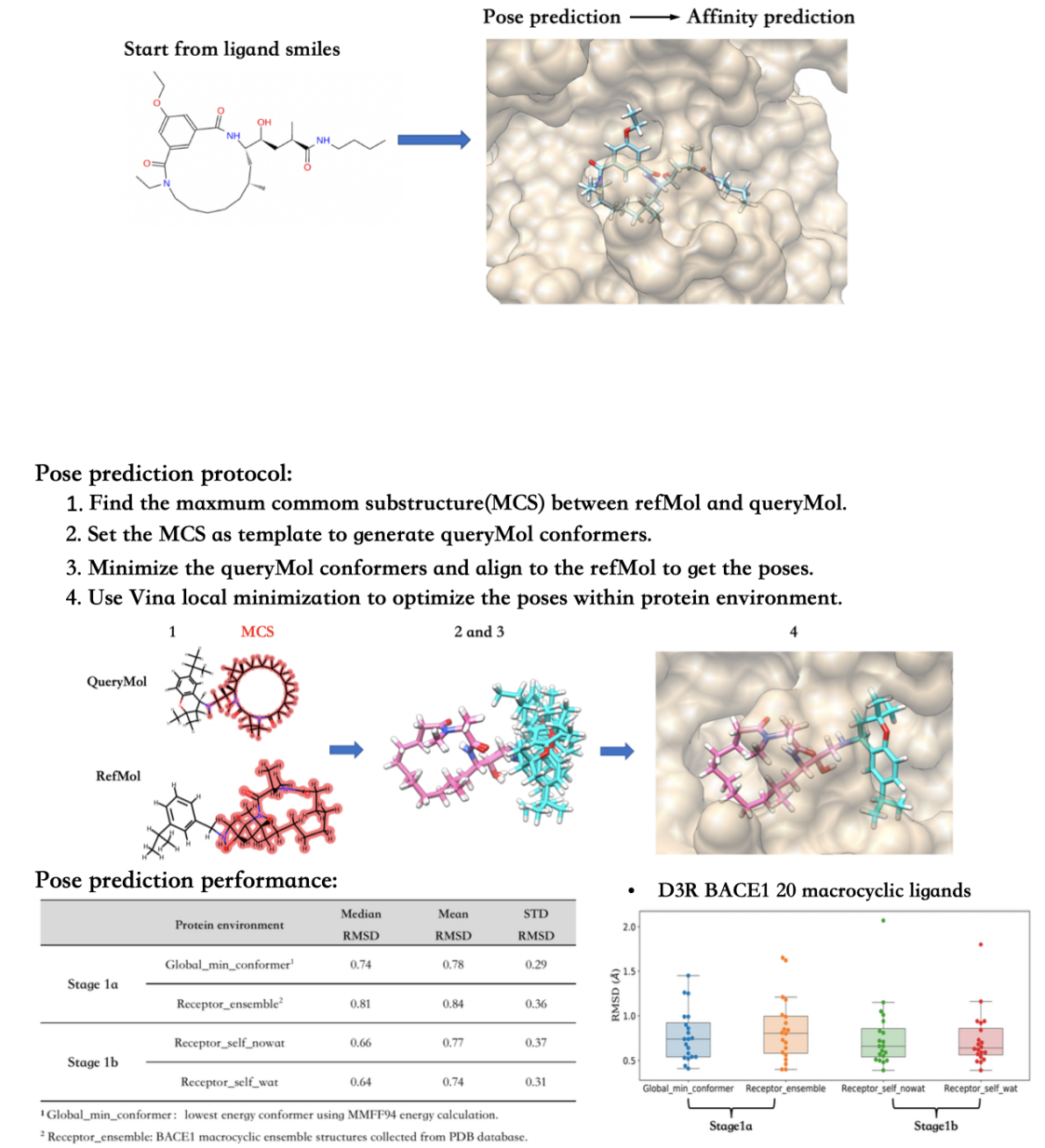

Figure S6. Pose prediction protocol of similarity-based constraint docking method and its performance on BACE1 Stage1a and Stage 1b.

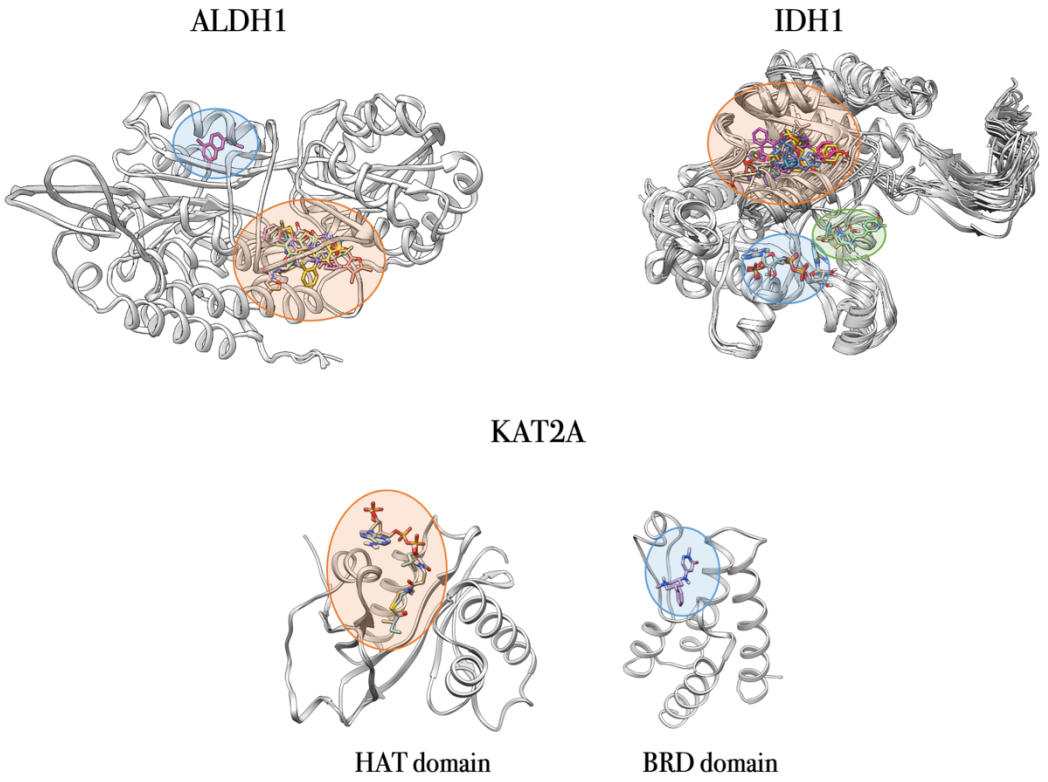

Figure S7. Three target sets (ALDH1, IDH1 and KAT2A) in LIT-PCBA benchmark have more than one ligand binding site in the PDB templates. For each target set, the selected docking site used in our study is highlighted in orange color. For ALDH1, the orthosteric aldehyde binding site was selected as the docking site. For IDH1, the well-known allosteric binding site was selected as the docking site. For KAT2A, the catalytic site at HAT domain was selected as the docking site.

Target	Notes
ALDH1 (bioassay/1030)	Enzymatic assay, most of current available co-crystallized inhibitors bind to aldehyde binding site (color orange), not NAD^+ binding site (color blue). It is very difficult to screen NAD^+ competitive inhibitors.
IDH1 (bioassay/602179)	Enzymatic assay, most of current available co-crystallized inhibitors bind to the well-known allosteric binding site (color orange), not NAD^+ binding site (color blue). In addition, most compounds are selective for R132H, not wild type. The R132H residue locates at this allosteric binding site.
KAT2A (bioassay/504327)	Enzymatic assay, it uses acetyl-CoA as substrate and measures the formation of reduced CoA based on fluorescent signal. Thus, the catalytic site at HAT domain (color orange) was selected as the docking site, not BRD domain (color blue).

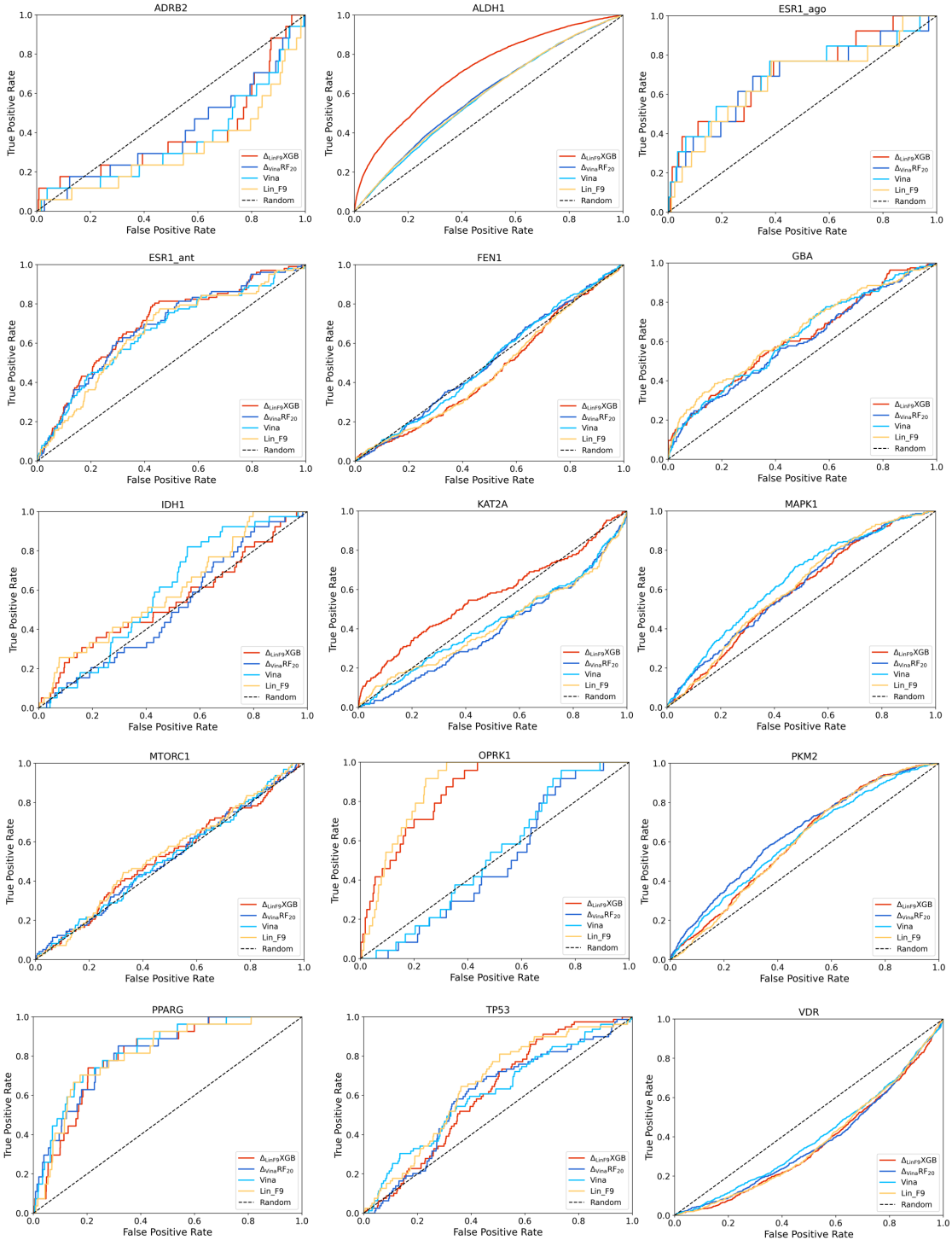

Figure S8. ROC curves for 15 targets in docking-based virtual screening of LIT-PCBA benchmark.

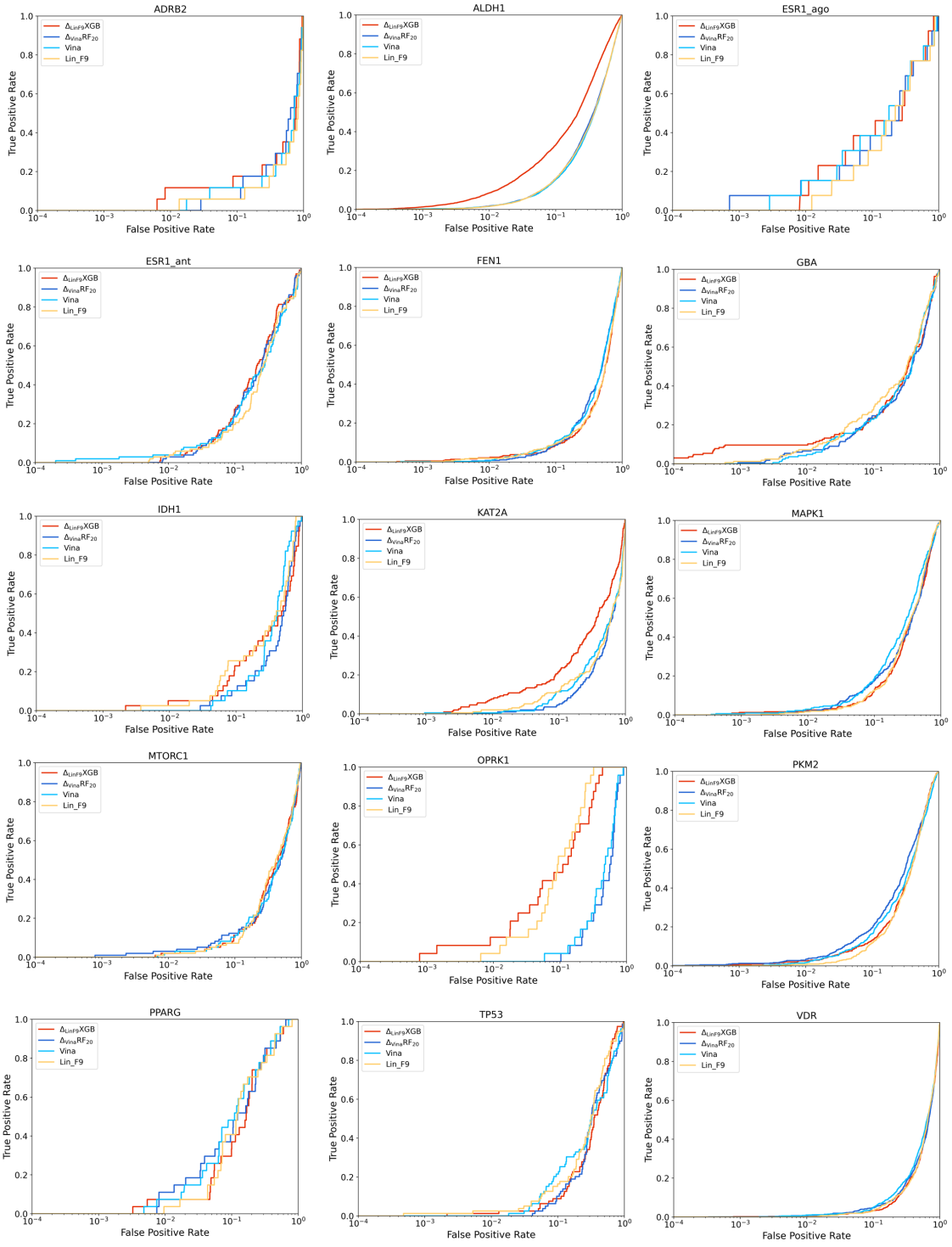

Figure S9. ROC curves for 15 targets in docking-based virtual screening of LIT-PCBA benchmark. Different from Figure S8, here focus is given to early enrichments by scaling false positive rates in logarithmic units.