

Glossary Table

Types of Names	Names	Definition
Datasets	DTC	Drug Target Commons by Tang et al., 2018. A dataset with 14,820,874 bioactivity measurements spanning 1,746,997 compounds and 13,023 targets
	DTC-kinase	~130,000 unique kinase-compound pairs in DTC involving measurement of a binding constant (K _i or K _d)
	KinCo	~130,000 unique kinase-compound pairs, each with an affinity from DTC paired with docked poses. Each kinase-compound pair has > 10,000 docked poses sampling various conformations of the kinase. KinCo is DTC-kinase augmented with docked structures.
	PDBBIND-kinase	~2,000 kinase-ligand pairs derived from PDBBIND2018. Each pair has a crystal structured and a paired binding affinity.
Test sets	easy kinase	Kinases in this test set are also in the training set
	medium kinase	kinases in the this test set share at most 95% sequence identity with those in the training set
	hard kinase	kinases in the this test set share at most 87% sequence identity with those in the training set
	easy compound	compounds in this test share at most 0.8 tanimoto similarity with those in the training set
	medium compound	compounds in this test share at most 0.6 tanimoto similarity with those in the training set
	hard compound	compounds in this test share at most 0.4 tanimoto similarity with those in the training set
Models	Experimental Net	a 3D-CNN model with adapted architecture from Pafnucy, trained on all 16,000 protein-ligand complexes in PDBBIND2018 with crystal structures and paired affinity
	Vina	the scoring function within the docking algorithm, Autodock Vina
	KinCoNet-M₁	a 3D-CNN model with adapted architecture from Pafnucy, trained on PDBBIND-kinase and KinCo. Docked poses from KinCo with the lowest energy predicted by Vina were selected for training
	KinCoNet-M₂	a 3D-CNN model with adapted architecture from Pafnucy, trained on PDBBIND-kinase and KinCo. Docked poses from KinCo with the lowest energy predicted by KinCoNet-M ₁ were selected for training
	DTA	Deep Drug Target Affinity by Öztürk et al., 2018. In this structure-free model architecture, the kinases are represented by its primary sequence and compounds by their SMILES string. The model was trained on binding affinities from DTC-kinase