Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Material

DLSSAffinity: protein-ligand binding affinity prediction via a deep learning model

Huiwen Wang[1*], Haoquan Liu[2], Shangbo Ning[2], Chengwei Zeng[2], Yunjie Zhao[2*]

[1]School of Physics and Engineering, Henan University of Science and Technology, Luoyang 471023, China

[2]Institute of Biophysics and Department of Physics, Central China Normal University, Wuhan 430079, China

* Corresponding author: (huiwenwang@haust.edu.cn) H.W.

(yjzhaowh@mail.ccnu.edu.cn) Y.Z.

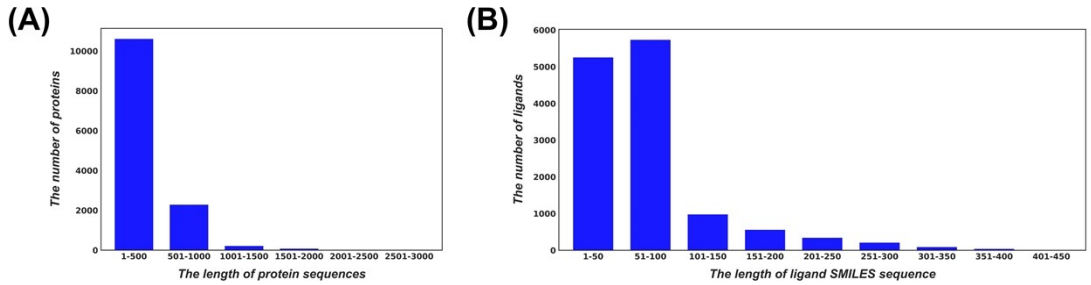

Fig. S1 (A) The distribution of the protein sequence lengths. (B) The distribution of the ligand SMILES sequence lengths.

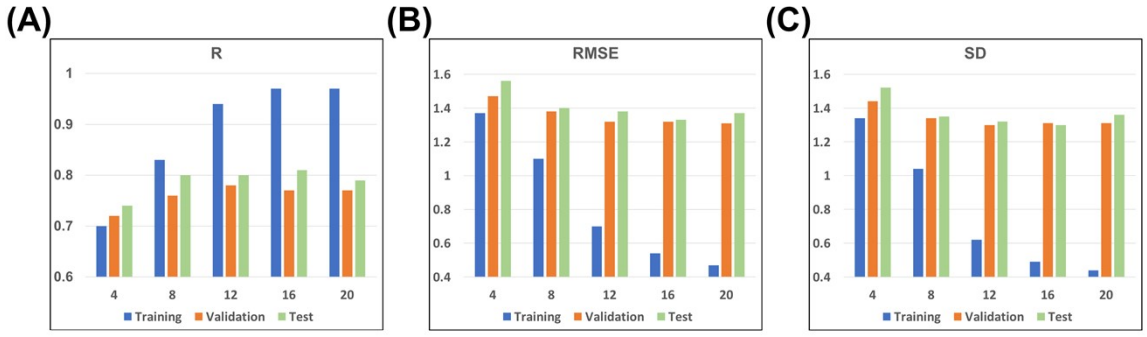

Fig. S2 The prediction accuracy of DLSSAffinity in training, validation, and test sets. (A) The R values are 0.70, 0.83, 0.94, 0.97, and 0.97 in training set while they are 0.74, 0.80, 0.80, 0.81, and 0.79 in the test set when the number of epochs = 4, 8, 12, 16, and 20, respectively. (B) The RMSE values are 1.37, 1.10, 0.7, 0.54, and 0.47 in the training set, while they are 1.56, 1.40, 1.38, 1.33, and 1.37 in the test set when the number of epochs = 4, 8, 12, 16, and 20, respectively. (C) The SD values are 1.34, 1.04, 0.62, 0.49, and 0.44 in the training set, while they are 1.52, 1.35, 1.32, 1.30, and 1.36 in the test set when the number of epochs = 4, 8, 12, 16, and 20, respectively.

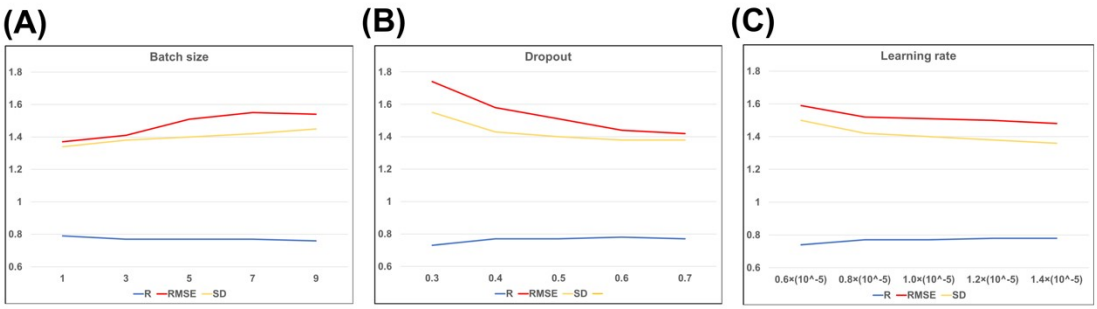

Fig. S3 Detecting the optimal parameters for DLSSAffinity model. (A) Choosing batch size = 1 for DLSSAffinity model. (B) Choosing dropout = 0.6 for DLSSAffinity model. (C) Choosing learning rate = 1.4×10^{-5} for DLSSAffinity model.

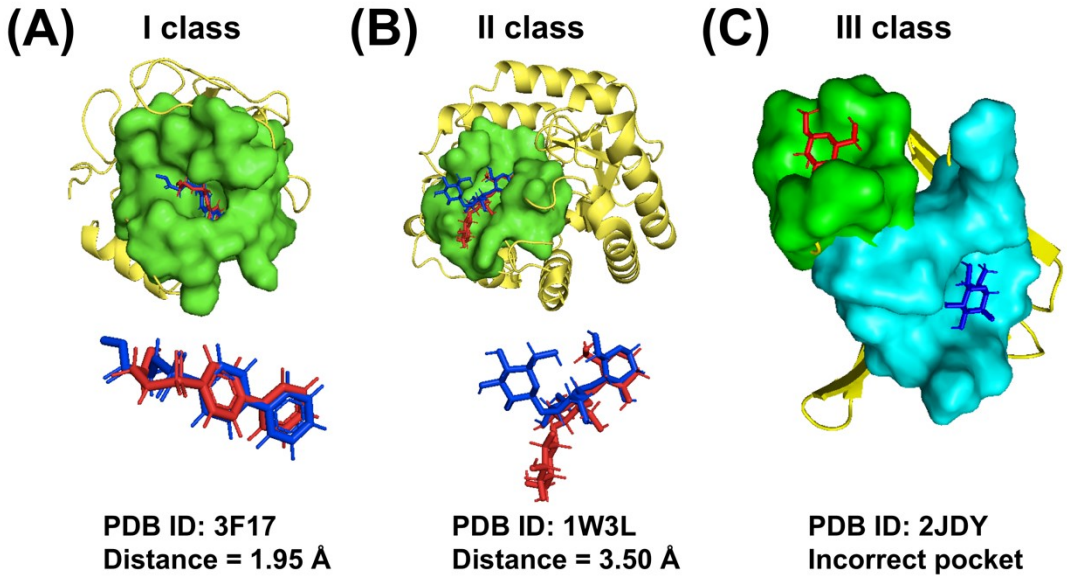

Fig. S4 Three types of predicted protein-ligand complex structures. (A) the ligand targeted in a correct pocket with the correct pose, that is, the average distance error of the ligand atoms is less than 2 Å; (B) the ligand targeted in the correct pocket with an incorrect ligand pose, that is, the average distance error of the ligand atoms is greater than 2 Å; (C) the ligand targeted in an incorrect pocket.

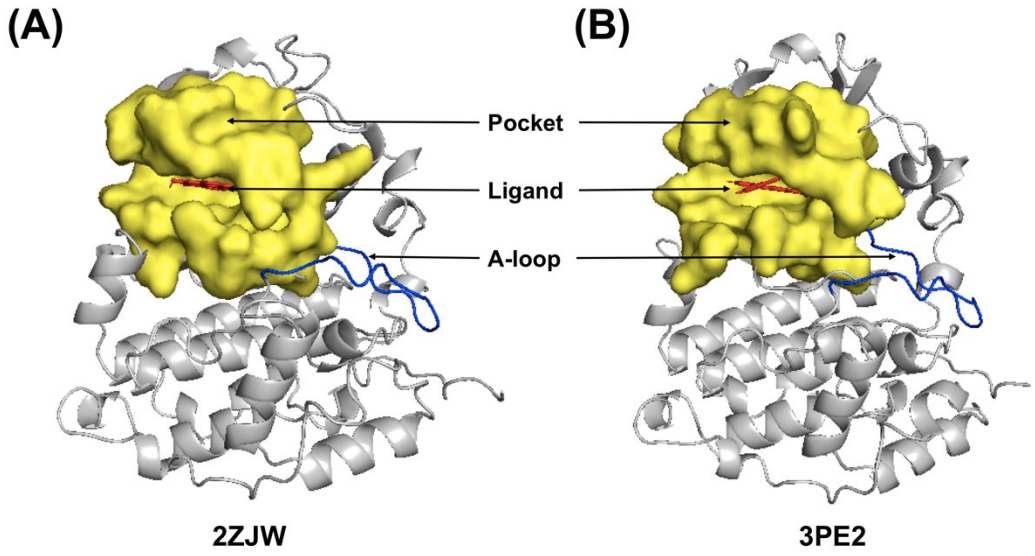

Fig. S5 The tertiary structures of two kinase-ligand complexes in CASF-2013 benchmark. The blue loop is the activation loop (A-loop) that can affect the interaction of ligand (red) targeting ATP pocket (yellow) through long-range indirect interactions.

Table S1. The prediction accuracy of DLSSAffinity in training, validation and test sets with the number of epochs = 4, 8, 12, 16, and 20, respectively.

Epochs	R			RMSE			SD		
	Training	Validation	Test	Training	Validation	Test	Training	Validation	Test
4	0.70	0.72	0.74	1.37	1.47	1.56	1.34	1.44	1.52
8	0.83	0.76	0.79	1.10	1.38	1.40	1.04	1.34	1.35
12	0.94	0.78	0.80	0.70	1.32	1.38	0.62	1.30	1.32
16	0.97	0.77	0.81	0.54	1.32	1.33	0.49	1.31	1.30
20	0.97	0.77	0.79	0.47	1.31	1.37	0.44	1.31	1.36