

Prediction of Protein-ligand binding affinity

汇报人：李哲



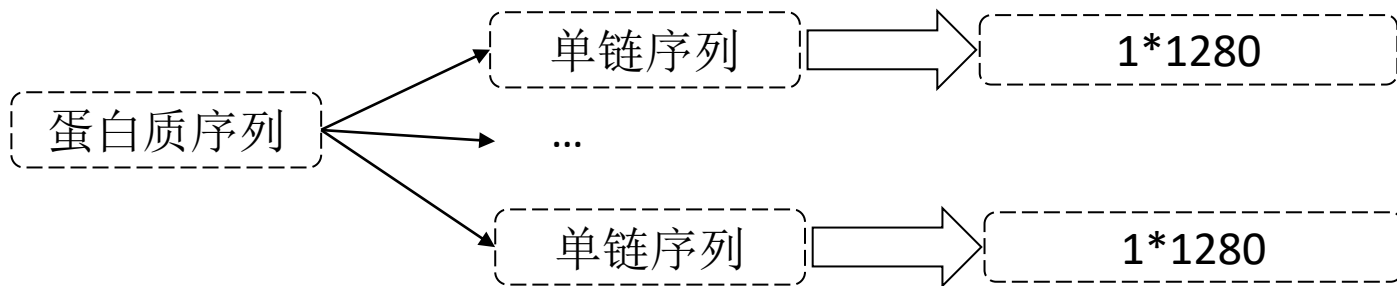
Dataset

dataset: PDBbind V2016/V2020	
training set	11906
validation set	1000
test set	290
test105 set	105
test71 set	71

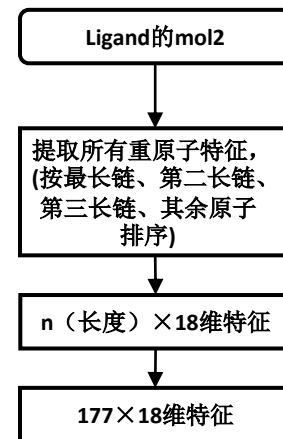
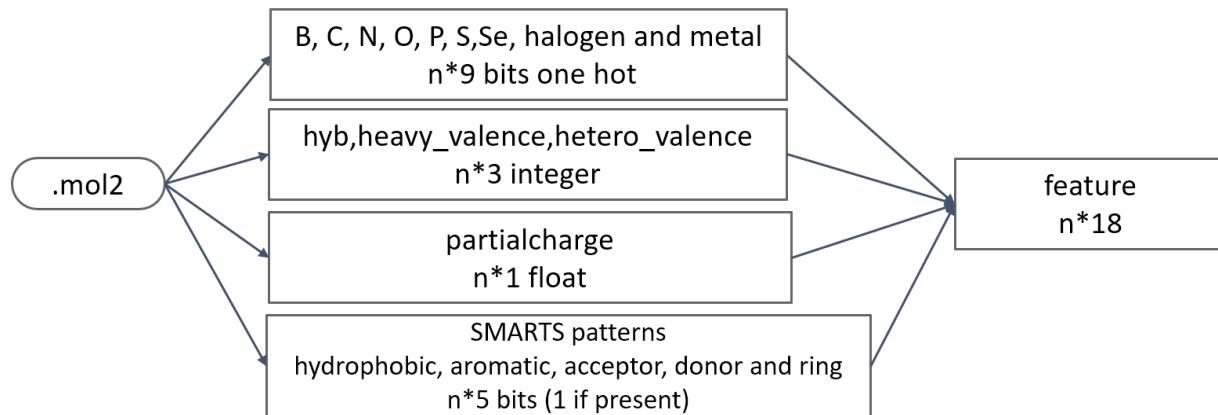
Data pre-process (protein)

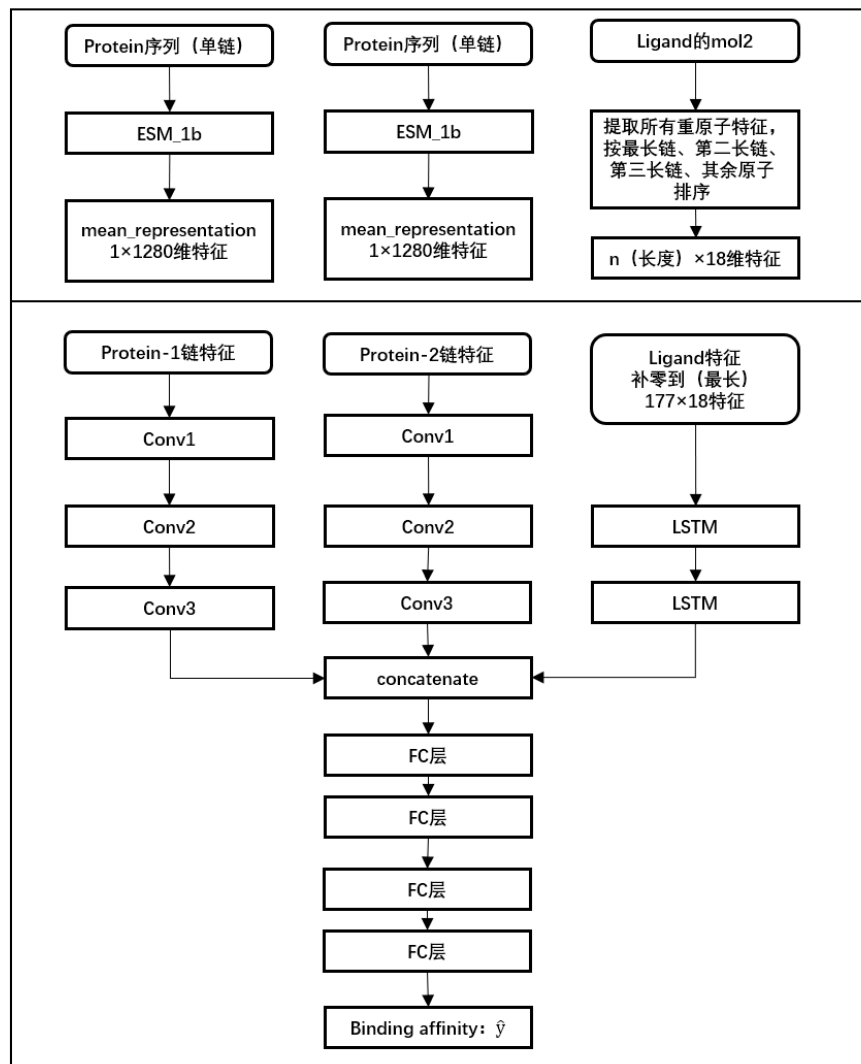


1a4k蛋白质及其配体



Data pre-process (ligand)







Result

	Datasets	RMSE	MAE	R	SD	CI
DeepDTAF 原	Training	0.952	0.731	0.864	0.942	0.839
	Validation	1.367	1.054	0.747	1.365	0.779
	Test	1.355	1.073	0.789	1.337	0.799
	Test105	1.247	0.966	0.766	1.149	0.801
	Test71	1.273	0.998	0.480	1.194	0.656

	Datasets	RMSE	MAE	R	SD	CI
ZCNNLSTM	Training	0.645	0.421	0.953	0.642	0.911
	Validation	0.997	0.697	0.873	0.990	0.855
	Test	1.200	0.887	0.814	1.188	0.826
	Test105	1.336	0.990	0.694	1.268	0.773
	Test71	1.372	1.150	0.457	1.205	0.644

	RMSE	MAE	R	SD	CI
ZCNNLSTM	1.200	0.887	0.814	1.188	0.826
DeepDTAF	1.355	1.073	0.789	1.337	0.799
DeepDTA	1.443	1.148	0.749	1.445	0.771
Pafnucy	1.418	1.129	0.775	1.375	0.789
TopologyNet	3.713	3.151	0.173	2.142	0.555

最优	RMSE	MAE	R	SD	CI
ZCNNLSTM	1.158	0.792	0.831	1.138	0.838



Result-Test105 set

	RMSE	MAE	R	SD	CI
ZCNNLSTM	1.336	0.990	0.694	1.268	0.773
DeepDTAF	1.247	0.966	0.766	1.149	0.801
DeepDTA	1.425	1.134	0.652	1.432	0.738
Pafnucy	1.392	1.169	0.750	1.176	0.782
TopologyNet	4.143	3.841	0.444	1.530	0.646

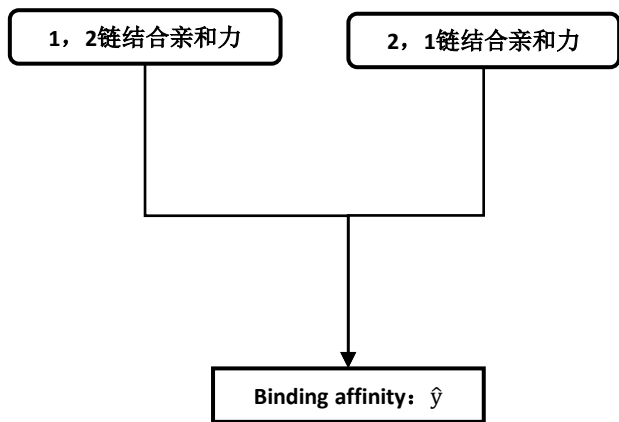
最优	RMSE	MAE	R	SD	CI
ZCNNLSTM	1.270	0.902	0.698	1.262	0.767



Result-Test71 set

	RMSE	MAE	R	SD	CI
ZCNNLSTM	1.372	1.150	0.457	1.205	0.644
DeepDTAF	1.273	0.998	0.480	1.194	0.656
DeepDTA	1.517	1.144	0.417	1.527	0.641
Pafnucy	1.442	1.210	0.427	1.230	0.628
TopologyNet	4.157	3.913	0.192	1.308	0.559

最优	RMSE	MAE	R	SD	CI
ZCNNLSTM	1.214	0.965	0.562	1.120	0.693



$$y = \frac{1}{n} \sum_1^n y$$

Result-单个蛋白质求平均

	Datasets	RMSE	MAE	R	SD	CI
ZCNNLSTM 平均	Training	0.958	0.729	0.860	0.954	0.834
	Validation	1.317	1.026	0.772	1.305	0.788
	Test	1.405	1.134	0.773	1.382	0.779
	Test105	1.533	1.159	0.576	1.460	0.717
	Test71	1.450	1.161	0.392	1.252	0.620

	Datasets	RMSE	MAE	R	SD	CI
ZCNNLSTM 原	Training	0.645	0.421	0.953	0.642	0.911
	Validation	0.997	0.697	0.873	0.990	0.855
	Test	1.200	0.887	0.814	1.188	0.826
	Test105	1.336	0.990	0.694	1.268	0.773
	Test71	1.372	1.150	0.457	1.205	0.644



Feature Plan

- 1.配体用Graph
- 2.蛋白质不使用经esm-1b输出的平均特征，使用全部特征进行训练
- 3.改进模型直接输入一批蛋白质和配体进行训练
- 4.链组合结合亲和力可以加权重求平均
- 5.再想一些策略