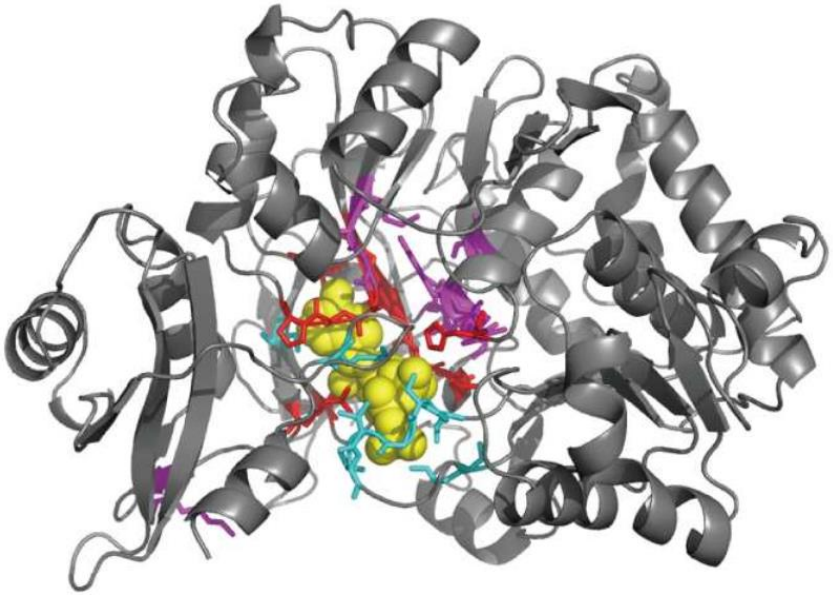
Work Report

Ningxin Jia

1-8-2022

Evaluation



ATP	**TP** true positive	**FN** false positive	**TN** true negative	**FP** false negative

Sensitivity : $Sen = \frac{TP}{TP+FN} \times 100$

Specificity : $Spe = \frac{TN}{TN+FP} \times 100$

Accuracy : $ACC = \frac{TP+TN}{TP+FN+TN+FP} \times 100$

Precision : $Pre = \frac{TP}{TP+FP} \times 100$

Matthews correlation coefficient :

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TN+FN) \times (TP+FN) \times (TN+FP)}}$$

Experiment Results

Statistical composition of the two data sets used in this study.

Data set	N_{pro} [a]	N_{pos} [b]	N_{neg} [c]	$PNratio$ [d]
PATP-388	388	5657	142,086	1: 25.12
PATP-TEST	41	674	14,159	1: 21.01

[a] Number of proteins.

[b] Number of ATP-binding residues.

[c] Number of non-ATP-binding residues.

[d] $PNratio = N_{pos} : N_{neg}$.

Ligand	Method	REC	PRE	MCC	AUC
ATP[e] (ATP-41_Test)	COACH	0.632	0.703	0.652	N/A
	ATPBind	0.631	0.756	0.677	0.915
	TargetS	0.516	0.689	0.580	N/A
	S-SITE	0.570	0.505	0.513	0.801
	DELIA	**0.642**	**0.758**	0.685	**0.947**

	Model	Sen	Spe	Acc	Pre	Mcc
5-fold cross-validation on PATP-388	BiLSTM-based	57.92	98.94	68.70	97.37	0.617
	SVM-based	60.07	98.92	97.43	68.94	0.628
	Integrated	61.36	98.90	97.47	69.02	0.636
on PATP-TEST	BiLSTM-based	62.46	99.13	77.39	97.47	**0.682**
	Integrated	65.13	98.96	74.92	97.43	**0.685**

PPS-align

```
      QEURY_NAME         TEMPL_NAME    PPSscore       Pval   JaccardR
  1A0IA_ATP_BS01     1a6eA_ADP_BS01      0.2109 0.99928909     0.1667

     RMSD_ali    SeqIdent QUERY_SIZE TEMPL_SIZE
       1.3163      0.0000         11         17

=====================================================================
d0 = 1.97302
=====================================================================
Rotation Matrix Information
=====================================================================
i              t[i]         u[i][0]         u[i][1]         u[i][2]
0      4.9306716176   -0.7909825835   -0.6107557955    0.0363855868
1      7.9500699719    0.5629942655   -0.7032649213    0.4341150855
2    -63.4658103601   -0.2395495975    0.3638623486    0.9001223148

Code for rotating Structure Query from (x,y,z) to (X,Y,Z):
for(k=0; k<L; k++)
{
   X[k] = t[0] + u[0][0]*x[k] + u[0][1]*y[k] + u[0][2]*z[k]
   Y[k] = t[1] + u[1][0]*x[k] + u[1][1]*y[k] + u[1][2]*z[k]
   Z[k] = t[2] + u[2][0]*x[k] + u[2][1]*y[k] + u[2][2]*z[k]
}


=====================================================================
Pocket Alignment Information
=====================================================================
Query AA Index:        92  140  211  229
Query AA Type:          E    Y    I    K
Aligned Tag:            :    .    .    :
Aligned Distance:    0.42 1.25 2.06 0.97
Templ AA Type:          P    L    G    T
Templ AA Index:        28  432   79   80


Query pocket number : 1
Template pocket number : 1
Using 0-s to load poc files.
Using 0-s to run PPS-align.
Totally Using 0.000000-s
```

Measure the **structural similarity** between two protein **ligand-binding pockets**.

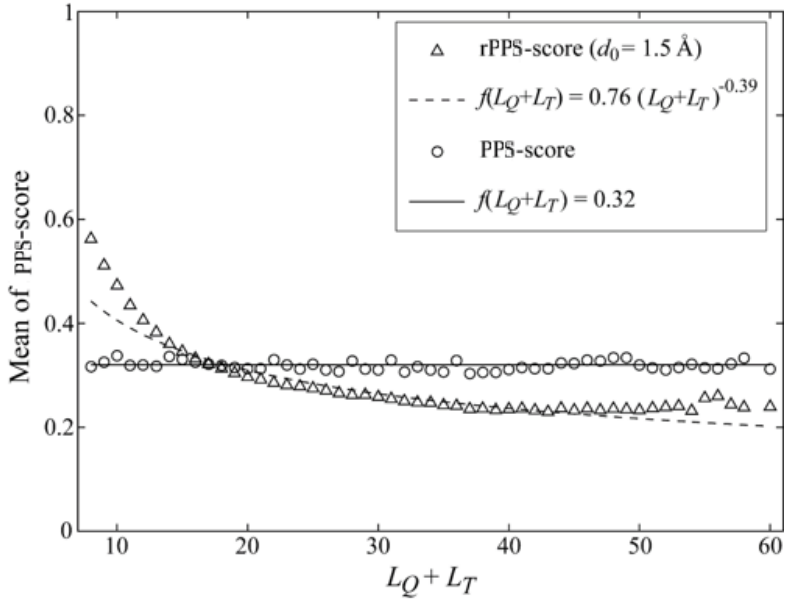
$$PPS-score = Max\left[\frac{2}{L_Q+L_T}\sum_{i=1}^{L_{ali}}\frac{1}{1+\left(\frac{d_i}{d_0}\right)^2}\right]$$

L_Q/L_T : the size of query/template pocket

L_{ali} : the number of aligned residue pairs

d_i : the distance between the C_α atoms of ith aligned residue pair

$d_0 = 0.65 + 0.4\sqrt{\frac{(L_Q+L_T)}{2} - 4}$: a scale to normalize the match difference



* 150,000 pocket pairs from PDB (global sequence identity < 40%; TM-score < 0.5)

Future Works

1\ Post-processing

2\ Write the paper, including drawing pictures and doing experiments.