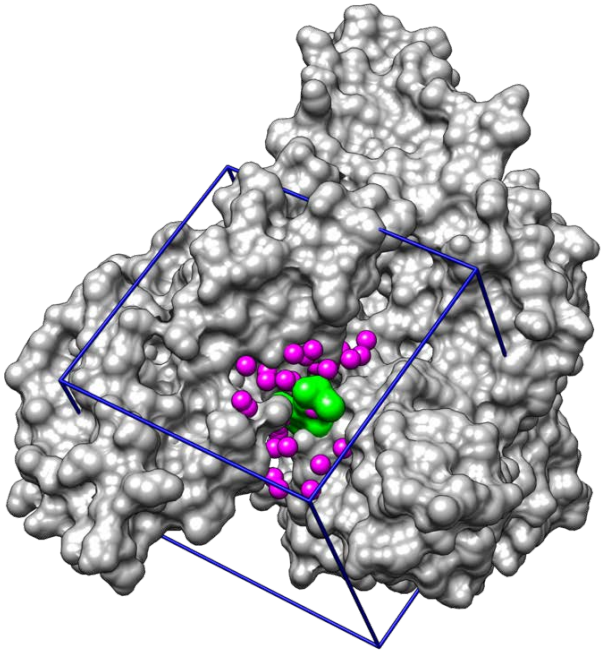
江蘇理工學院
JIANGSU UNIVERSITY OF TECHNOLOGY

普美瑞生物
PRIMARY BIOTECH

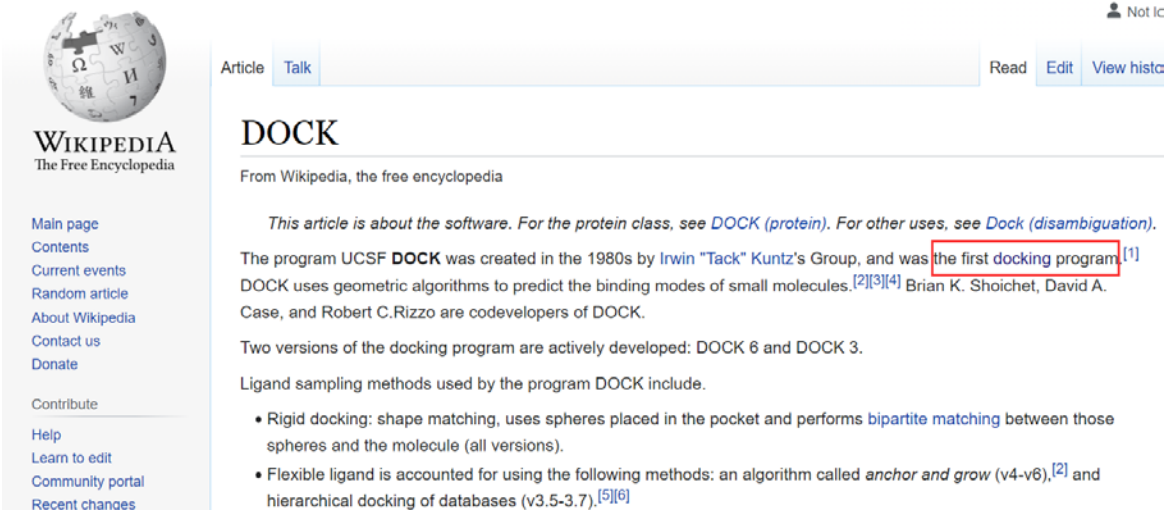
CoDock: Template-based docking and AI-based scoring used in ligand binding prediction

Shan Chang & Ren Kong
Jiangsu University of Technology
Primary Biotechnology Inc.
Dec 12, 2022

UCSF DOCK: the first docking program



2019 DOCK tutorial 3 with PDBID 3JQZ



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Learn to edit
Community portal
Recent changes

Article Talk Read Edit View histo

Not lo

DOCK

From Wikipedia, the free encyclopedia

This article is about the software. For the protein class, see DOCK (protein). For other uses, see Dock (disambiguation).

The program UCSF **DOCK** was created in the 1980s by Irwin "Tack" Kuntz's Group, and was the first docking program.[1] DOCK uses geometric algorithms to predict the binding modes of small molecules.[2][3][4] Brian K. Shoichet, David A. Case, and Robert C.Rizzo are codevelopers of DOCK.

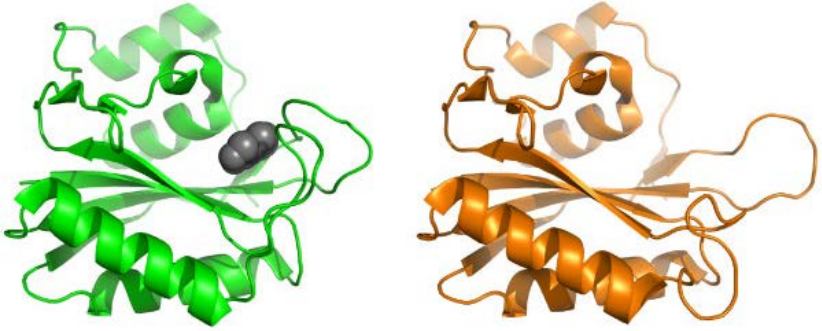
Two versions of the docking program are actively developed: DOCK 6 and DOCK 3.

Ligand sampling methods used by the program DOCK include.

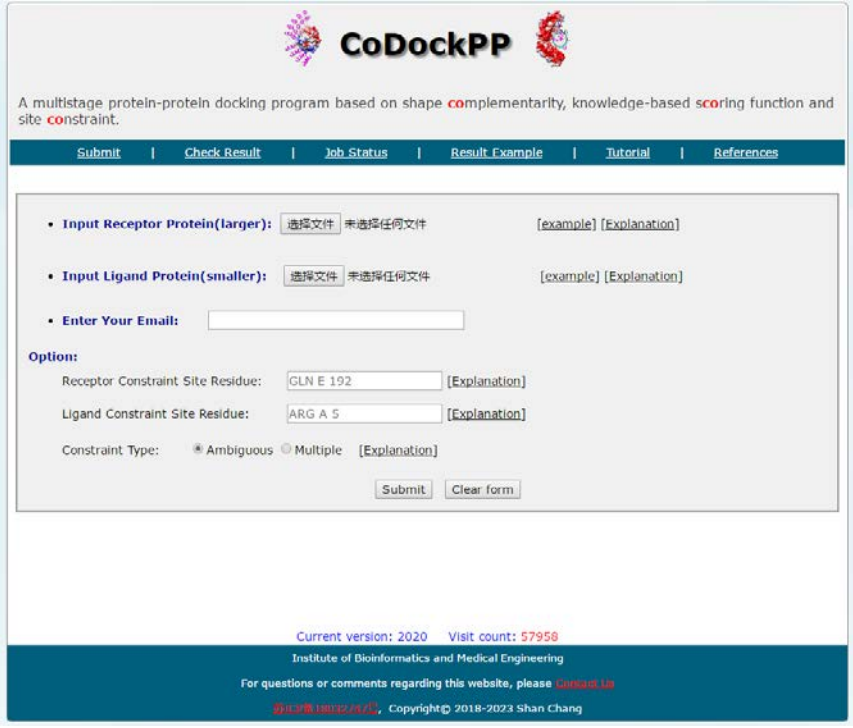
- Rigid docking: shape matching, uses spheres placed in the pocket and performs bipartite matching between those spheres and the molecule (all versions).
- Flexible ligand is accounted for using the following methods: an algorithm called *anchor and grow* (v4-v6),[2] and hierarchical docking of databases (v3.5-3.7).[5][6]

Ref: Kuntz, ID; Blaney, JM; Oatley, SJ; Langridge, R; Ferrin, TE (1982). "A geometric approach to macromolecule-ligand interactions". Journal of Molecular Biology. 161 (2): 269–88.

CASP ligand binding challenge: Unbound Docking



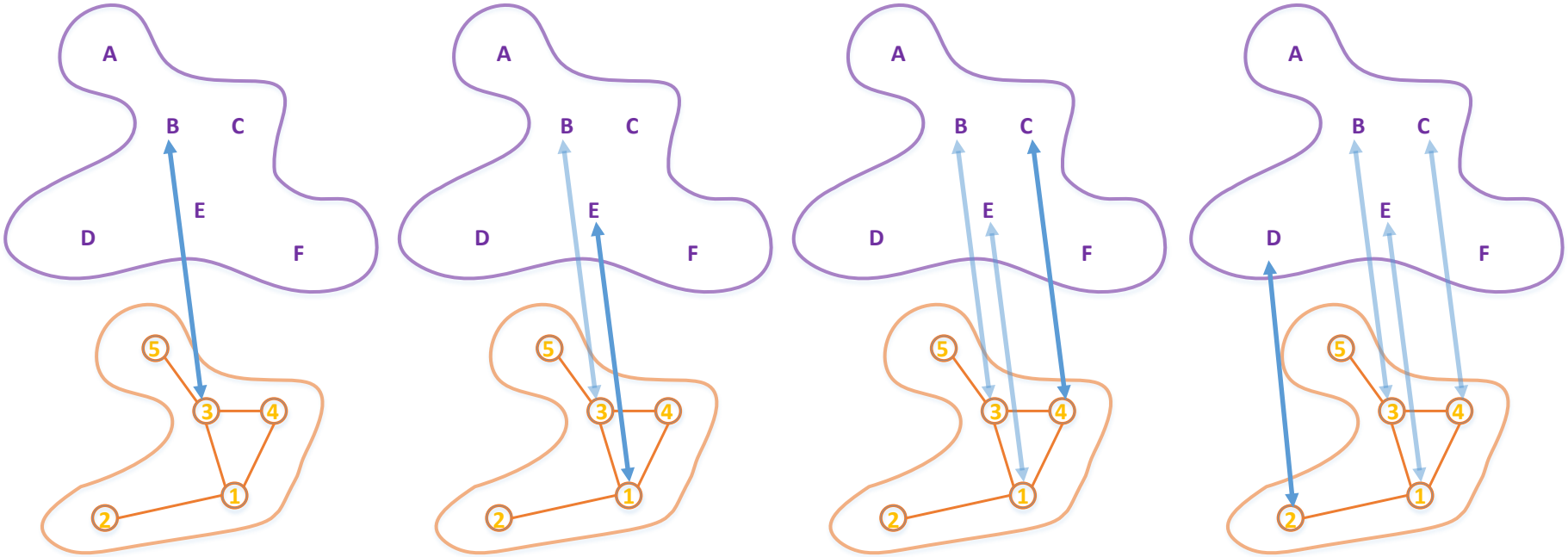
Example of a protein structure pair between ligand-bound and unbound states.



Ref: Morita et al.: BUDDY-system: A web site for constructing a dataset of protein pairs between ligand-bound and unbound states. BMC Research Notes 2011 4:143.

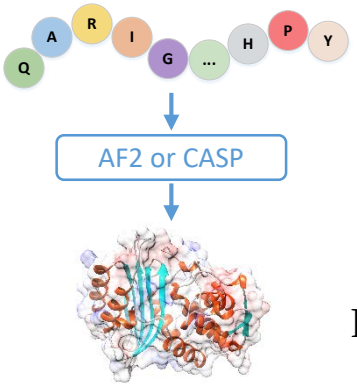
Kong R, Liu R R, Xu X M, Zhang D W, Xu X S, Shi H, Chang S. Template-based modeling and ab-initio docking using CoDock in CAPRI. Proteins. 2020; 88(8): 1100-1109.

Structure-based 3D align algorithm

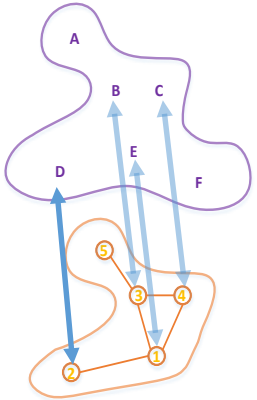


Ref: Ewing T J A, Kuntz I D. Critical evaluation of search algorithms for automated molecular docking and database screening. J Comput Chem. 1997; 18: 1175-1189.

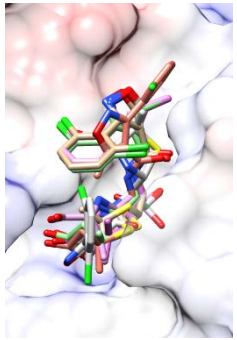
The flow chart of our docking strategy



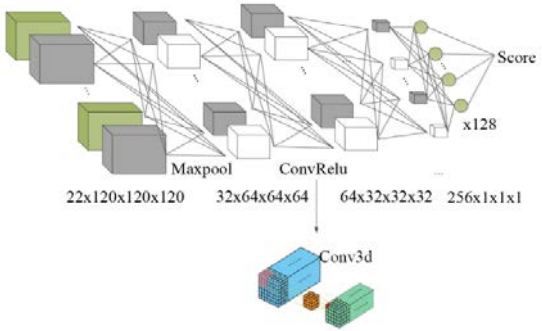
Structure preparation



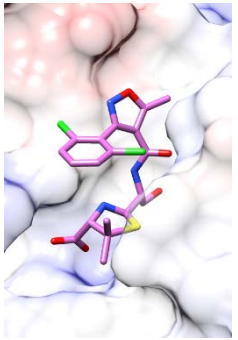
Pocket template searching



Ligand alignment



3D CNN Scoring



Results

Tested on CASF-2016

	Vina 1.2.2 (percent%)		Our docking method (percent%)	
RMSD / Top	<2Å	<1Å	<2Å	<1Å
Top1	62.90	47.70	83.86	76.49
Top3	72.44	53.36	88.07	84.56
Top5	75.62	54.42	89.47	86.67
Top10	81.63	55.48	90.18	87.72

Ref: Su M, Yang Q, Du Y, Feng G, Liu Z, Li Y, Wang R. Comparative Assessment of Scoring Functions: The CASF-2016 Update. J Chem Inf Model. 2019; 59(2): 895-913.

Performance in CASP15 (T1114-1152)

CASP ID	Template	Oligomeric state	Ligand	Lowest RMSD
H1114 [a]	5Y4N, 4UE3, 4KO2, 2FRV	A4B8C8	7/56	0.65(001 3NI) [c] 1.20(009 F3S) 0.56(017 F3S) 0.69(025 F3S) 0.82(033 FCO) 1.28(041 MG) 3.87(049 MQ7)
R1117v2	2L1V, 3FU2	A1	1/1	1.17(001 PRF)
T1118v1 [b]	5JJ5,6Z8A	A1	5/9	4.15(001 FE) 0.40(002 FE) 9.13(003 LIG) 4.35(004 LIG) 4.73(005 LIG)
T1124	7WDW	A2	2/4	1.92(001 SAH) 2.71(003 TYR)
R1126	5OB3, 7L0Z	A1	1/1	27.29(001 K)
T1127v2	3BJ7, 3BJ8, 2B4D, 2B4B, 2B58	A2	2/5	3.17(001 COA) 2.16(003 EPE)
H1135	6R16, 6R15, 6R2I	A9B3	2/12	11.32(001 CL) 0.08(004 K)
R1136	4KZD, 5OB3	A1	3/3	72.05(001 1TU) 24.27(002 J93) 64.57(003 K)
T1146	4Q5K, 4Q68	A1	1/1	0.58(001 NAG)
T1152	4BH9, 4B8V	A2	1/1	0.81(001 NAG)

The results were analyzed based on the structure provided by CASP.

[a] In these targets, our group achieved successful prediction in protein-ligand experiments.

[b] In these targets, our group achieved partial successful prediction in protein-ligand experiments.

[c] Red ligand is RMSD less than 2Å. Purple ligand is RMSD more than 2Å but less than 5Å. Blue ligand is RMSD more than 5Å.

Performance in CASP15 (T1158-1188)

CASP ID	Template	Oligomeric state	Ligand	Lowest RMSD (Å)
T1158v1	5UJA, 6D3R, 6PZ9, 6PZI, 6UY0	A1	1/1	1.85(001 XPG)
T1158v2	5UJA, 6D3R, 6PZ9, 6PZI, 6UY0	A1	1/1	3.90(001 P2E)
T1158v3	5UJA, 6D3R, 6PZ9, 6PZI, 6UY0	A1	1/1	1.66(001 XH0)
T1158v4	5UJA, 6D3R, 6PZ9, 6PZI, 6UY0	A1	4/4	0.79(001 ATP) 1.24(002 ATP) 0.14(003 2MG) 0.65(004 2MG)
T1170	6CHS	A6	3/9	1.20(001 ADP) 2.45(004 AGS) 0.77(007 MG)
H1171	6CHS	A6B1	3/9	1.52(001 ADP) 2.48(004 AGS) 1.57(007 MG)
H1172	6CHS	A6B2	3/9	2.09(001 ADP) 2.38(004 AGS) 1.06(007 MG)
T1181	5W6H, 4OJ6, 4OJ5, 4OJP, 4OJO	A3	6/8	10.63(001 OAA) 16.33(002 OAA) 36.99(005 ZN) 4.45(006 ZN) 5.69(007 ZN) 27.23(008 CA)
T1186	1FCM, 4R1G	A1	1/1	1.05(001 LIG)
T1187	-	A2	1/2	14.36(NAG)
T1188	2YBT, 6BT9	A1	5/5	3.96(001 DW0) 6.02(002 DW0) 11.01(003CD) 2.05(004 CD) 15.42(005 CO)

The results were analyzed based on the structure provided by CASP.

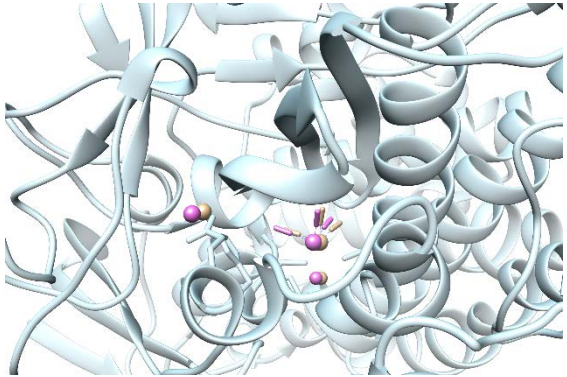
[a] In these targets, our group achieved successful prediction in protein-ligand experiments.

[b] In these targets, our group achieved partial successful prediction in protein-ligand experiments.

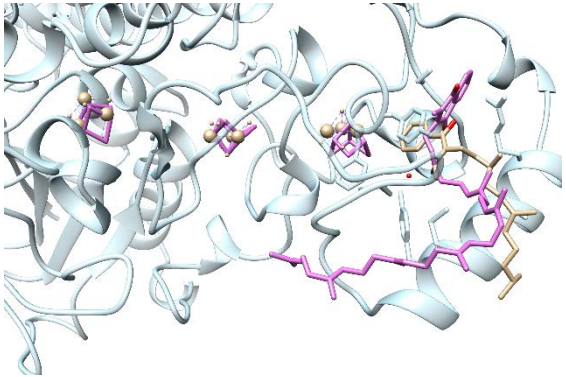
[c] Red ligand is RMSD less than 2Å. Purple ligand is RMSD more than 2Å but less than 5Å. Blue ligand is RMSD more than 5Å.

Successful prediction targets

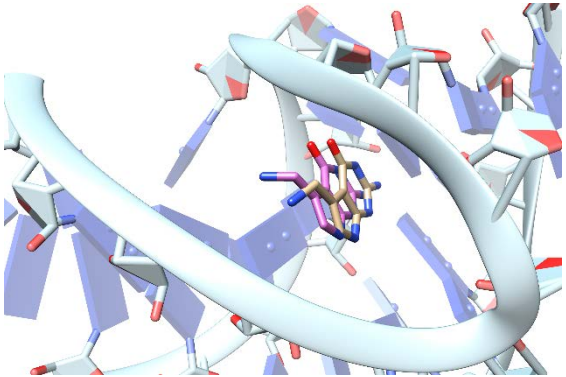
Crystal receptor　Crystal ligand　Prediction ligand



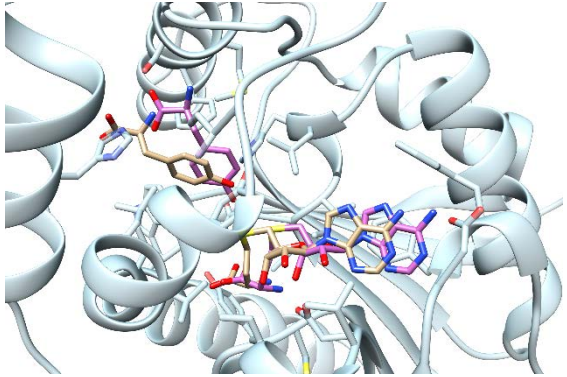
H1114(3NI FCO MG)



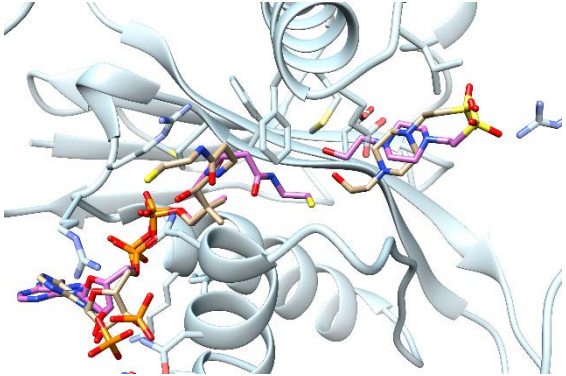
H1114(F3S MQ7)



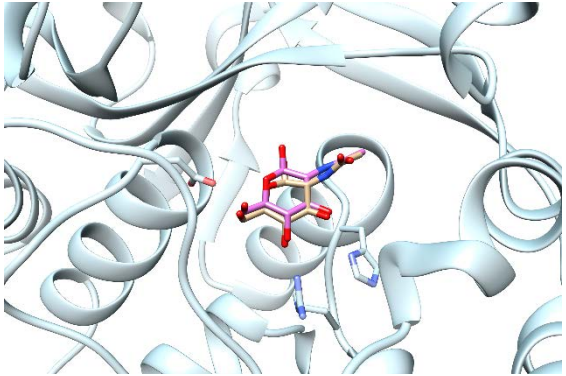
R1117v2



T1124



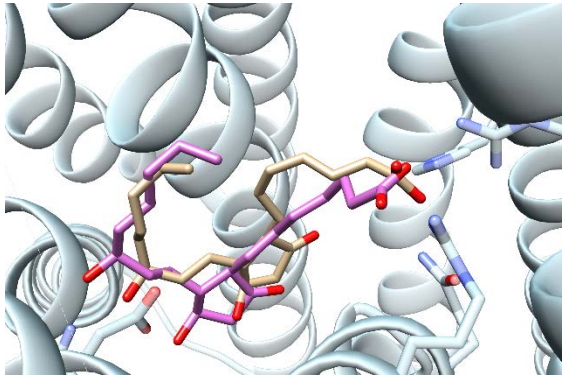
T1127v2



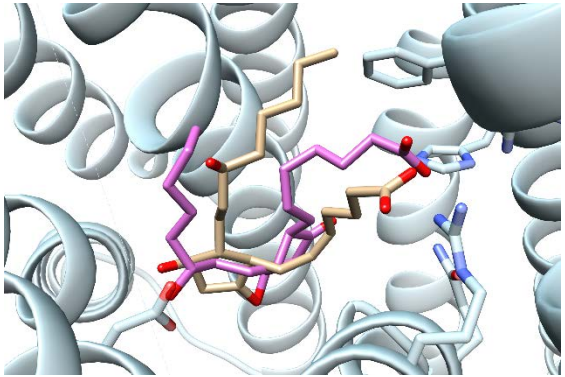
T1146

Successful prediction targets

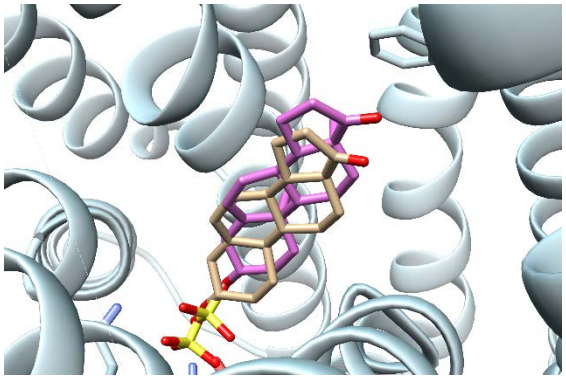
Crystal receptor Crystal ligand Prediction ligand



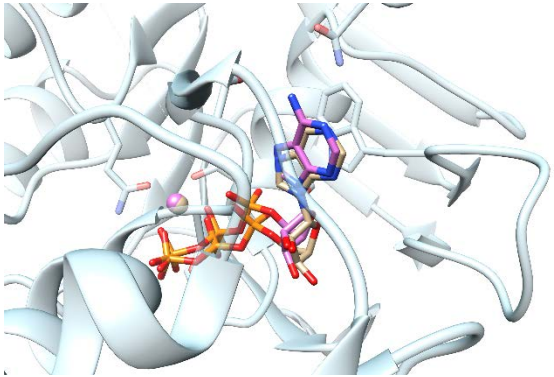
T1158v1



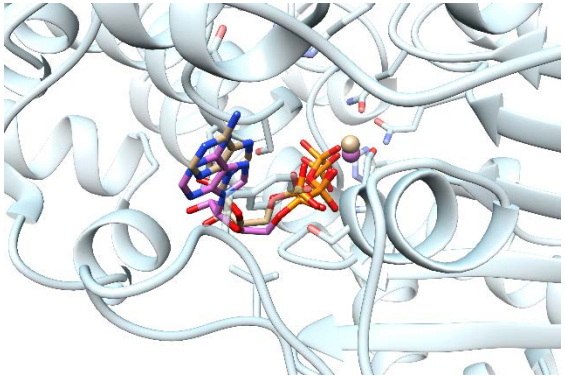
T1158v2



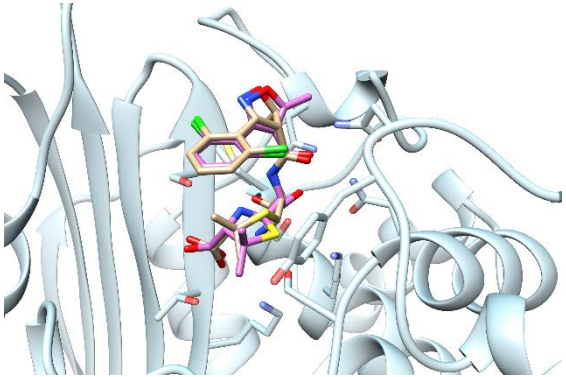
T1158v3



T1158v4(001 ATP&003 2MG)



T1158v4(002 ATP&004 2MG)



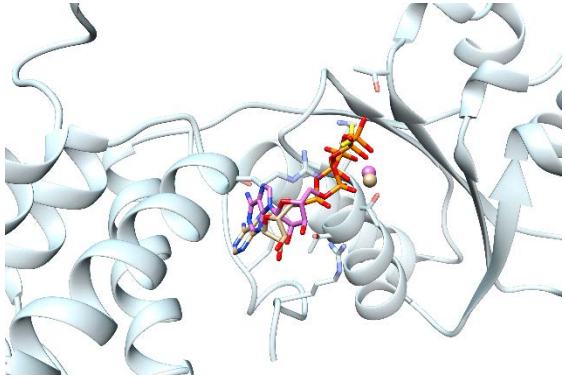
T1186

Successful prediction targets

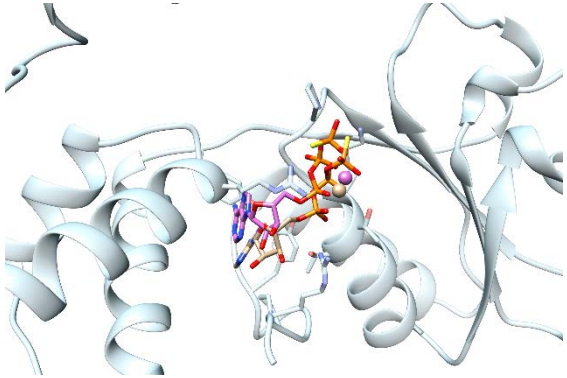
Crystal receptor

Crystal ligand

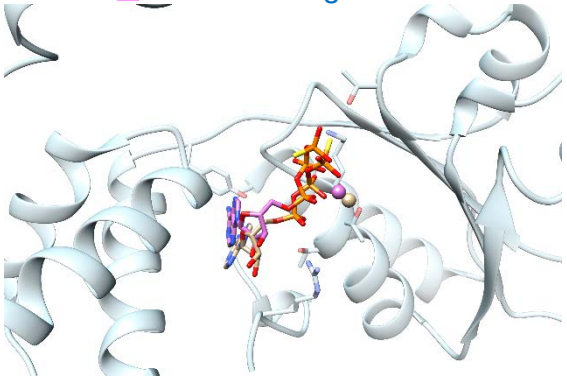
Prediction ligand



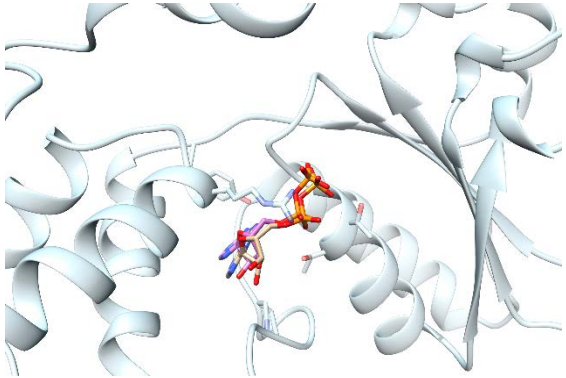
T1170(AGS MG)



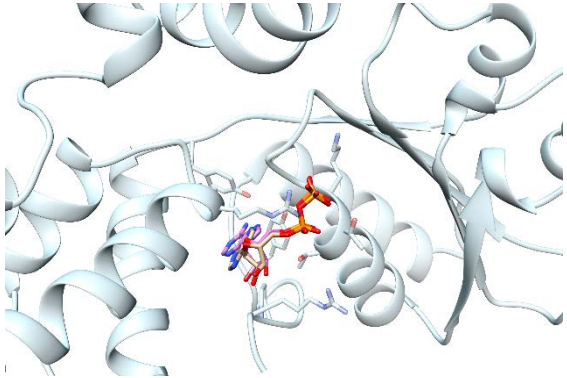
T1171(AGS MG)



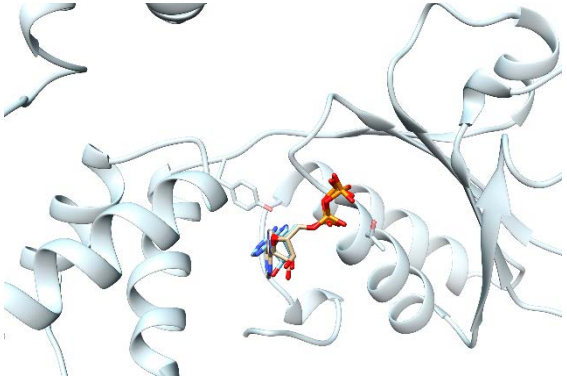
T1172(AGS MG)



T1170(ADP)



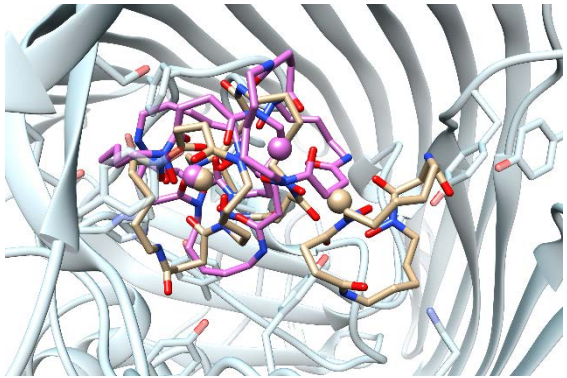
T1171(ADP)



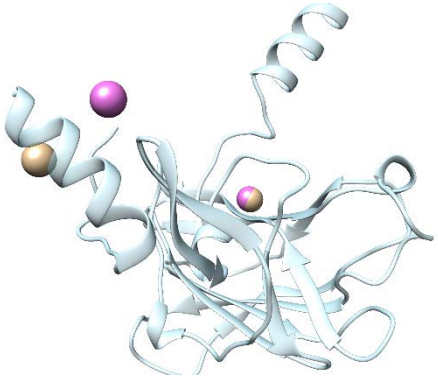
T1172(ADP)

Partial successful prediction targets

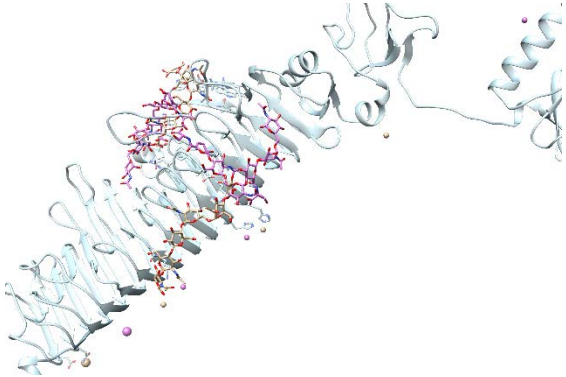
Crystal receptor Crystal ligand Prediction ligand



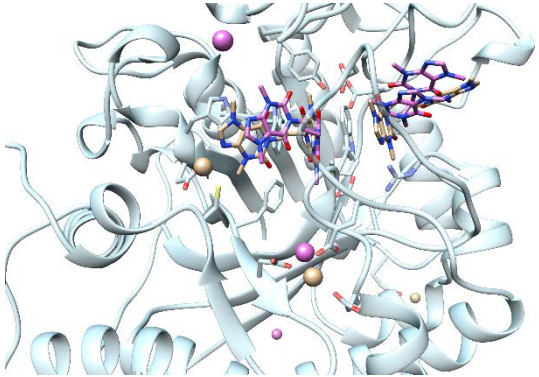
T1118v1



H1135



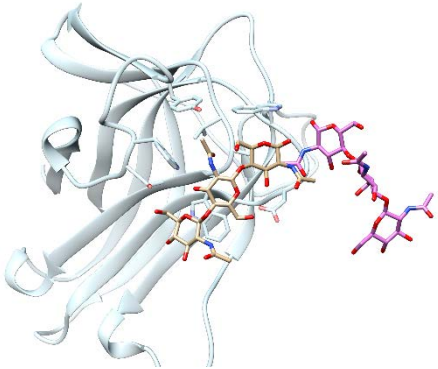
T1181



T1188

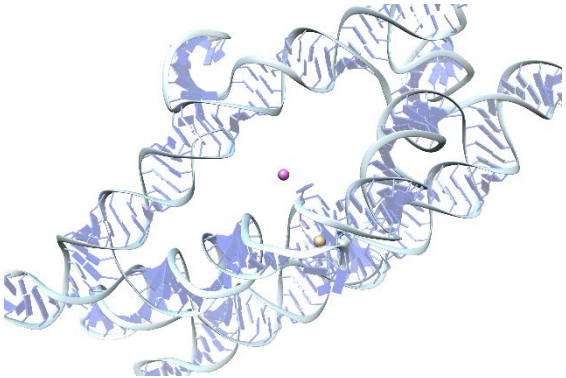
Failed prediction targets

Crystal receptor



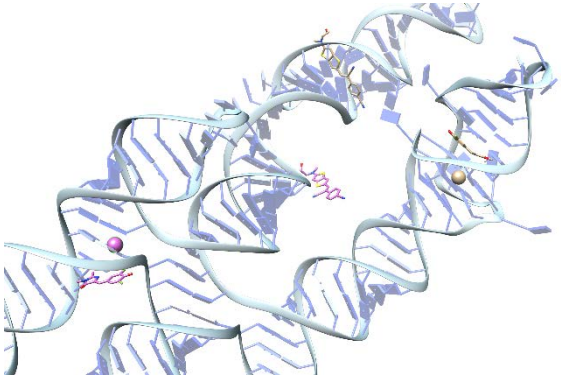
T1187

Crystal ligand



R1126

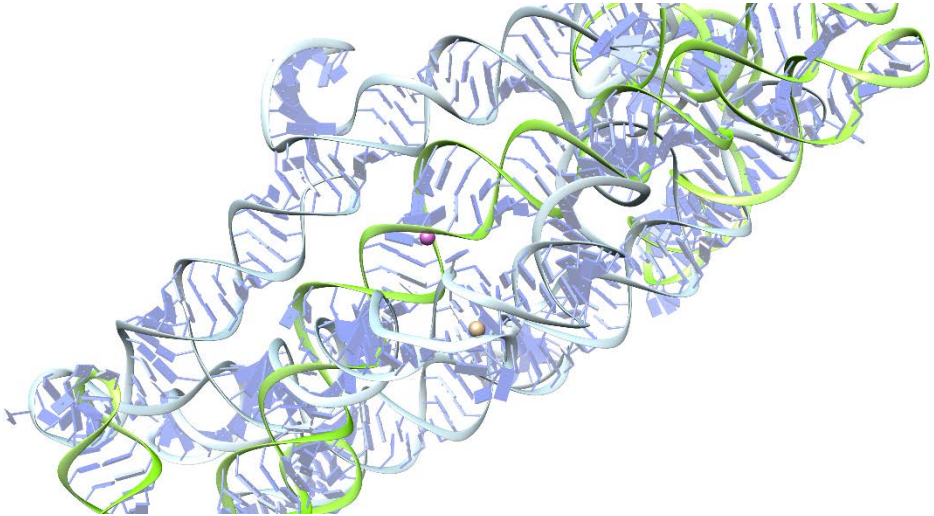
Prediction ligand



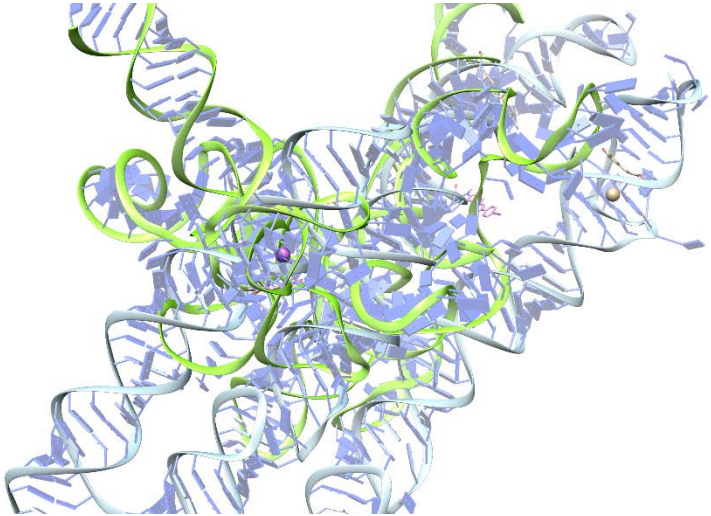
R1136

Challenge 1: incorrect receptor structure

Crystal receptor Prediction receptor Crystal ligand Prediction ligand



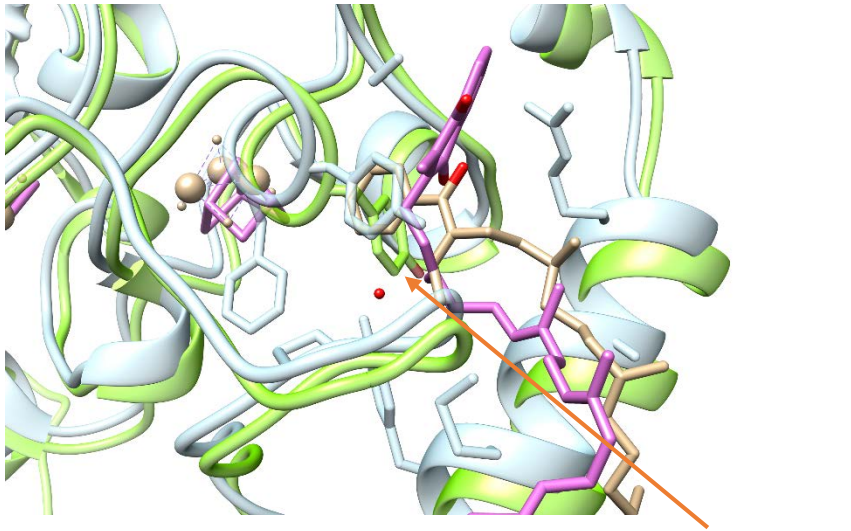
R1126



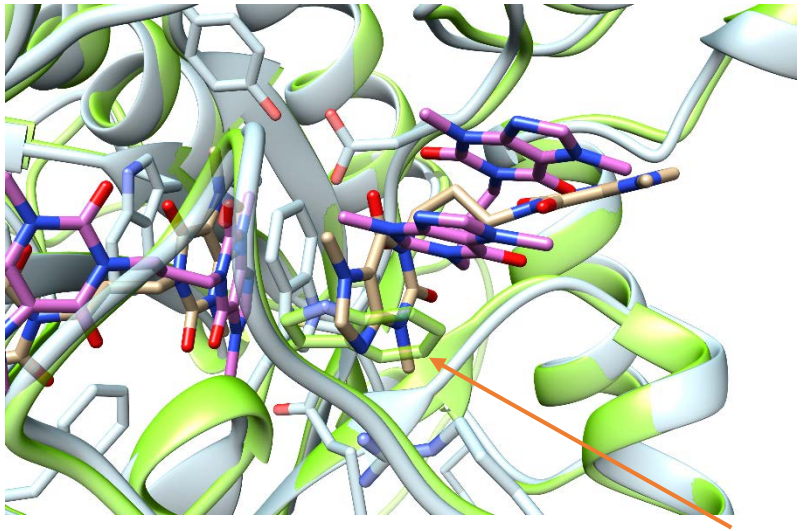
R1136

Challenge 2: incorrect side chain

Crystal receptor Prediction receptor Crystal ligand Prediction ligand



H1114(MQ7) TYR275



T1188(DW0) TRP120

Summary

1、 Our docking protocol combined Template-based and 3D-CNN for protein-ligand structure prediction.

2、 RNA-ligand prediction is not good, which is related to the large RMSD of RNA structure prediction.

3、CASP ligand binding prediction is an unbound prediction, and the side chain of protein has an impact on the ligand structure.

Thank you!

Institute of Bioinformatics and Medical Engineering in JUST is engaged in structure prediction of biological macromolecule, computer-aided drug design and bioassay screening.

Acknowledgement:

Prof. Xiaoqin Zou
Prof. Cunxin Wang
Primary Biotechnology Inc.