Find potential fragments from crystal

CrystalDock: A Novel Approach to Fragment-Based Drug Design

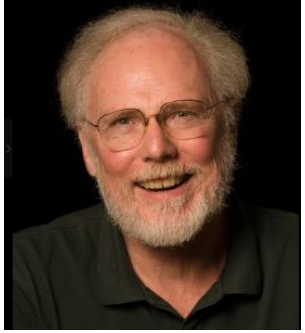
UCSD: Jacob D. Durrant*, J. Andrew McCammon

J. Chem. Inf. Model. **2011**, *51* 2573–2580

2012.3.9

mccammon.ucsd.edu

The authors

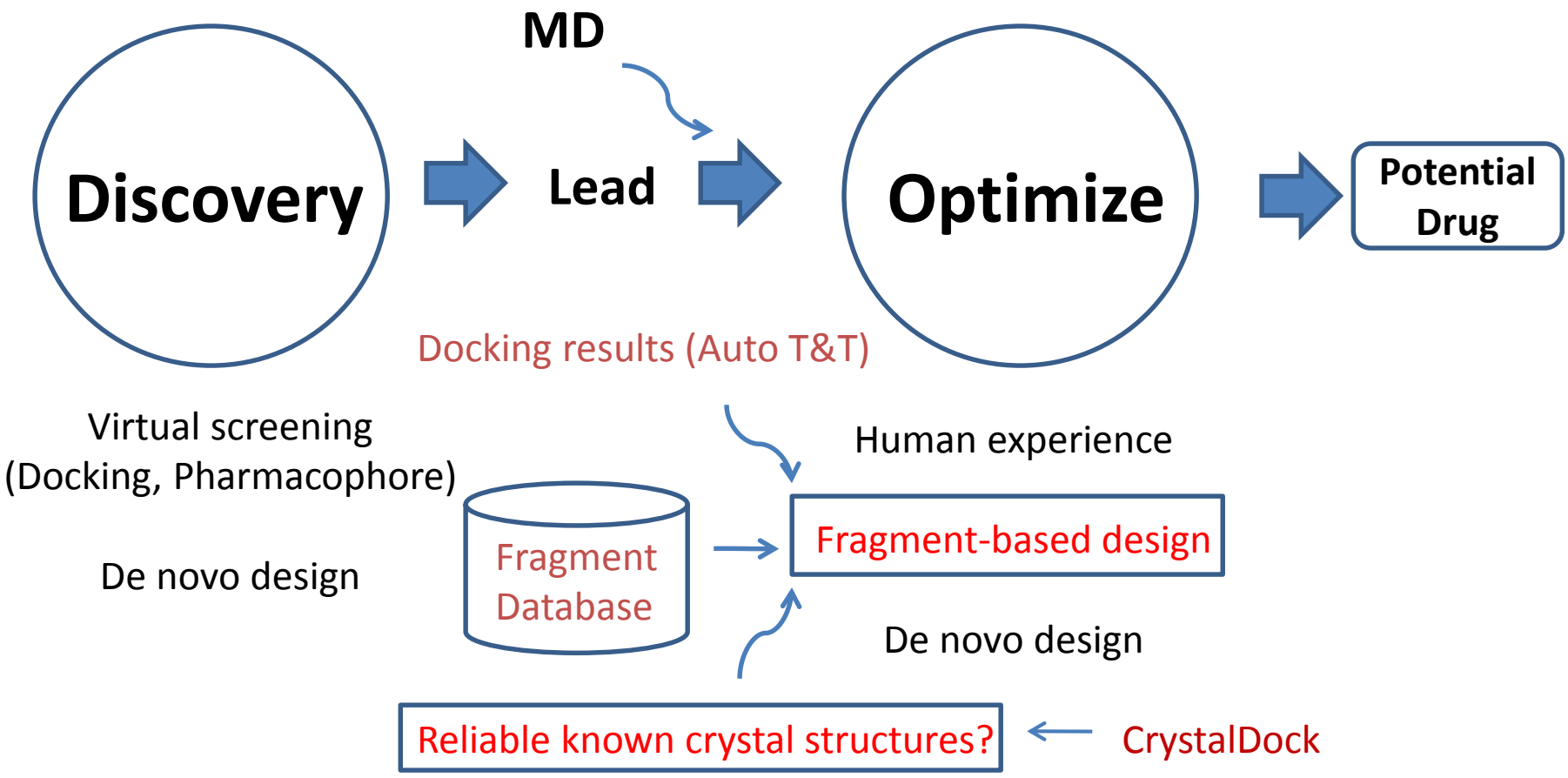


- Jacob D. Durrant is doing CADD research on infectious disease and inflammation in humans. He designs many algorithms and programs on ligand-protein interaction and MD, such as NNScore(Neural-Network-Based), Hbonanza(MD H-bond analysis) and so on(>20 papers during PhD).

- J. Andrew McCammon is a professor in UCSD and was elected to the NAS in 2011. He is the pioneer in MD study and published the first MD simulation of protein with M. Karplus in 1977 and also as one of the author of the book “Dynamics of Proteins and Nucleic Acids”(1987).

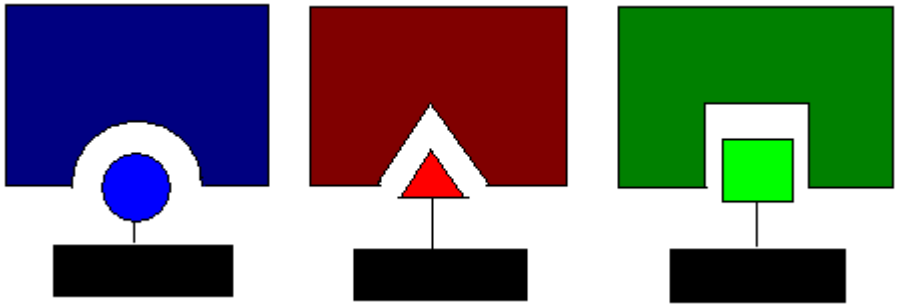
McCammon, J. A.; Gelin, B. R.; Karplus, M. (1977). "Dynamics of folded proteins". *Nature* **267** 585–590

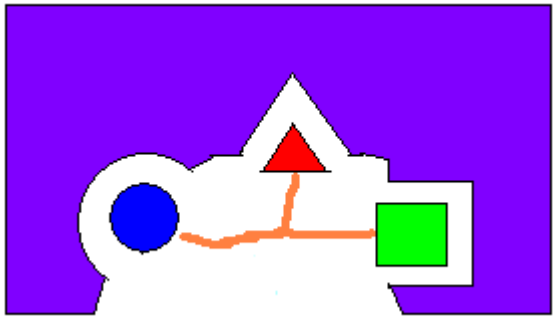
Core in CADD: Discover and Optimize Lead



Aim of CrystalDock

Known crystal structures of Protein-Ligand interaction





Fragment-Based method

Extract the microenvironment and fragments

Microenvironment: the adjacent 3~5 residues near the fragment. (For green fragment, the green box means to microenvironment)

Use of the known crystal structures information of the P-L interaction
To predict the potential binding fragments and optimize the lead.

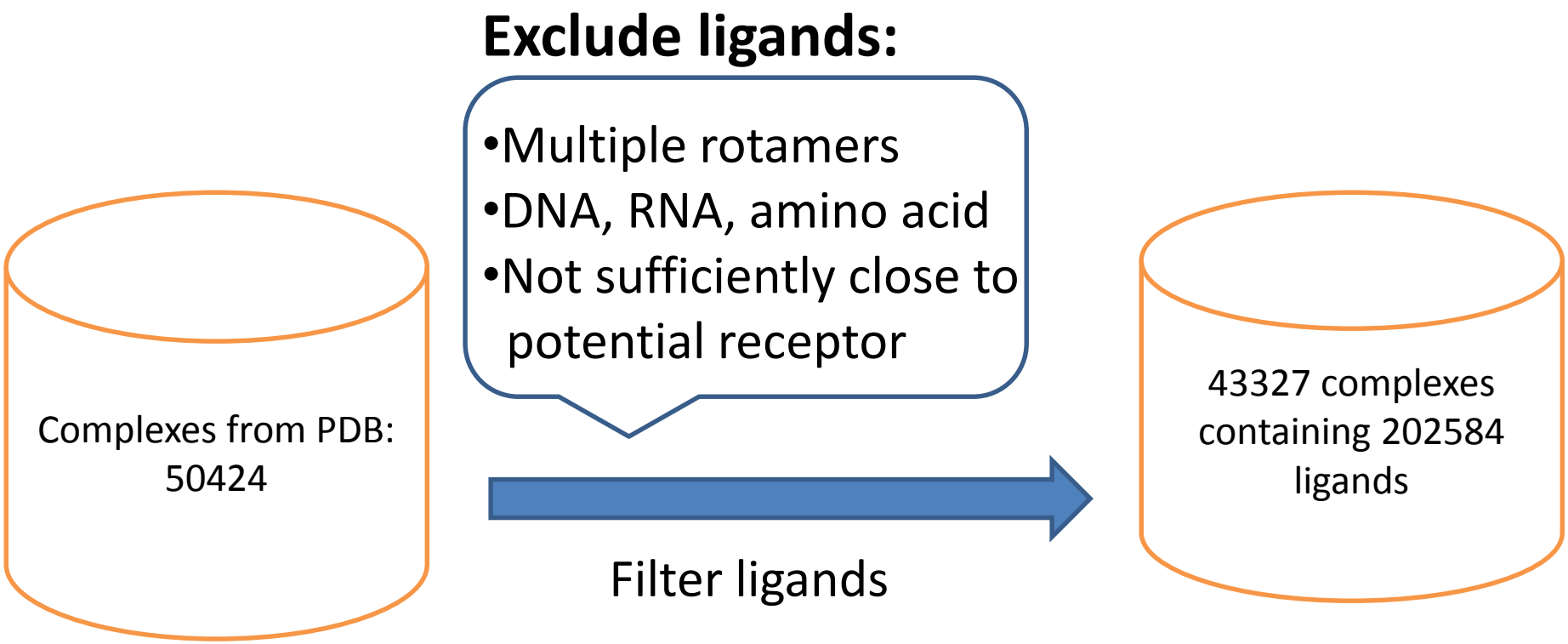
Note: CrystalDock is not a traditional docking program, but a tools for identify potential binding fragments (used for fragment-based drug design).

Contents

- Building the fragment-microenvironment database
- Procedure of Crystal-dock
- Application and Evaluation

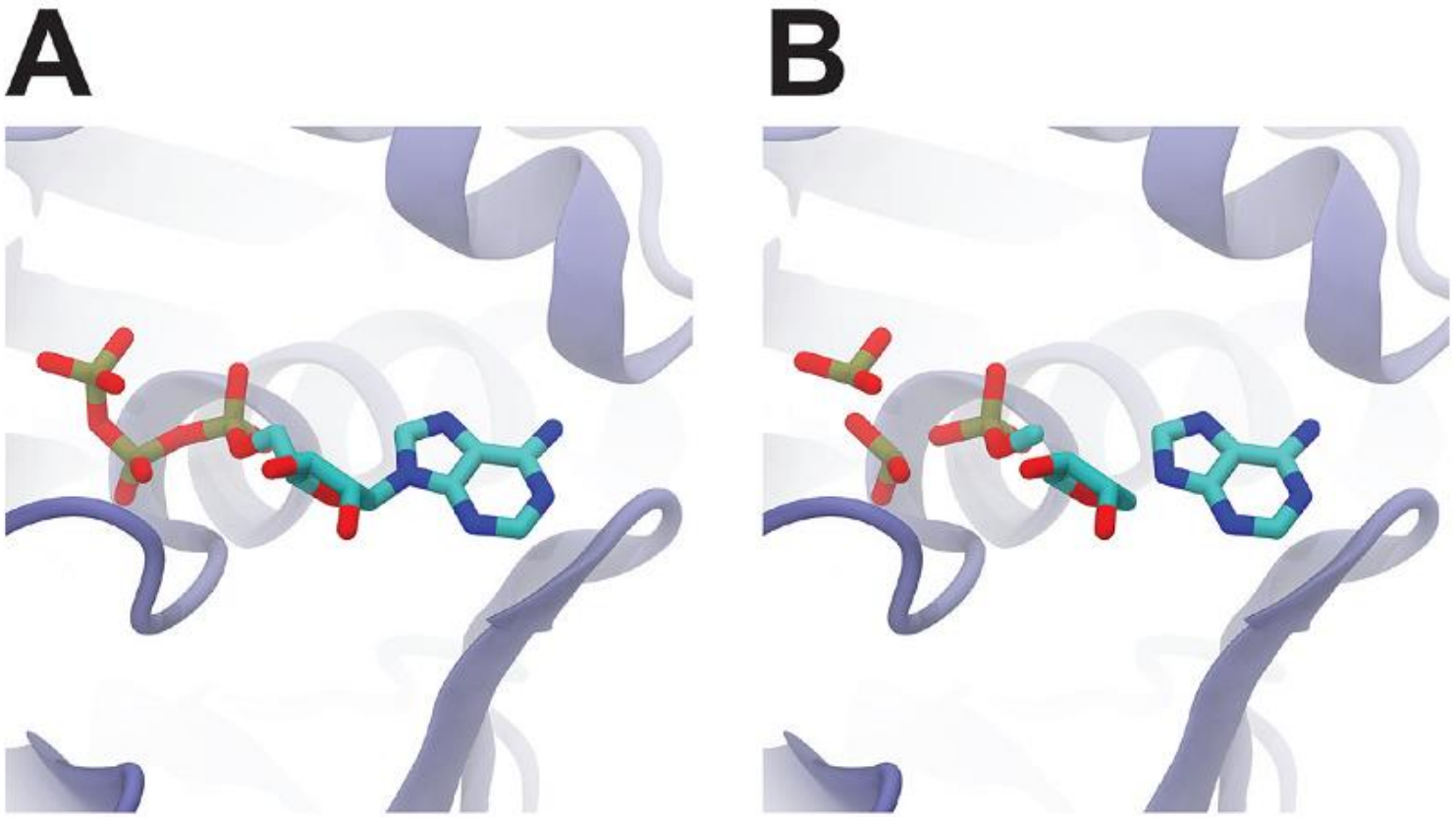
The program is written in python, and could be used in Windows, Linux, Mac.

Complex crystal database building



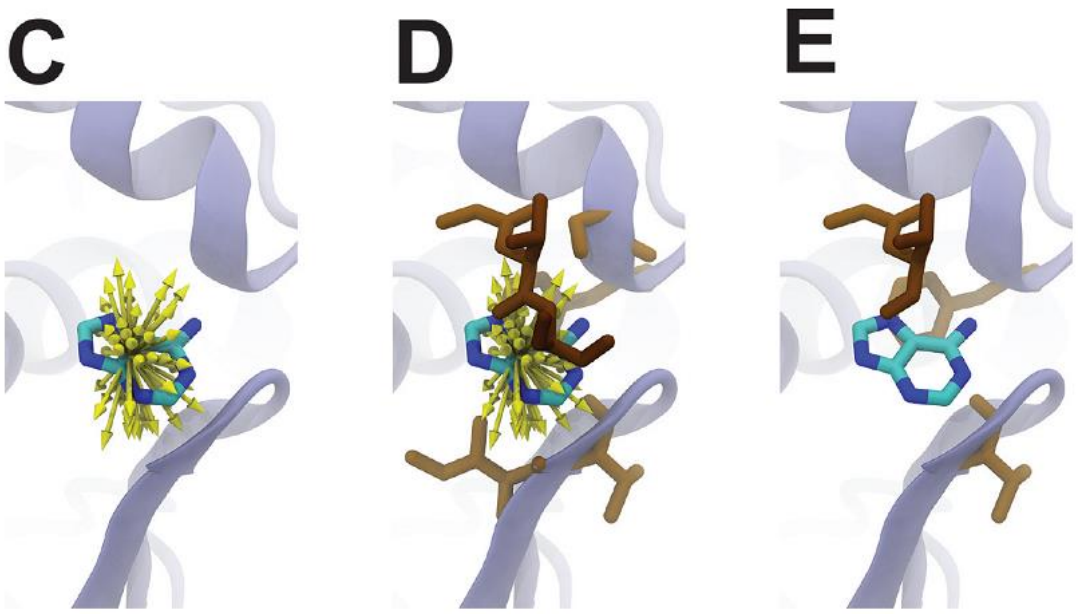
The ligands include the additives in crystal !
Indeed, some ions are also included…

From molecule to fragments



- All bonds between heavy atoms not belonging to the same ring were cut
- Any fragment with fewer than three heavy atoms was merged with the neighboring fragment that had the fewest atoms.

Find residues to build microenvironment



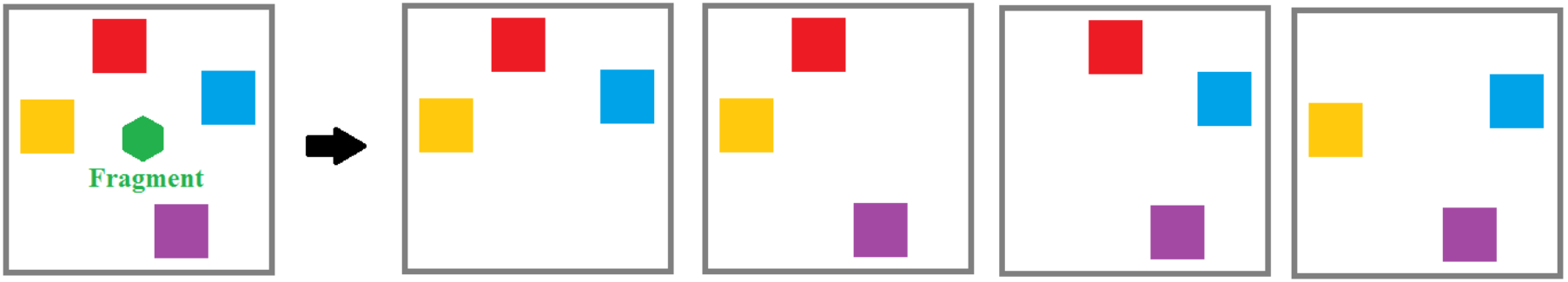
- Geometric rays, separated by 10^{o} in all directions, were extended from each fragment atom out into space(0 to 4 Å, 0.5 Å per step). When the ray find any atoms, record the residue and end the ray.
- The distance cutoff was gradually scaled back from 4 to 0 Å to build multiple fragment-microenvironment for each fragment
- Only those microenvironments with 3, 4, and 5 receptor residues (823,460 in total) were considered

The main procedure of CrystalDock

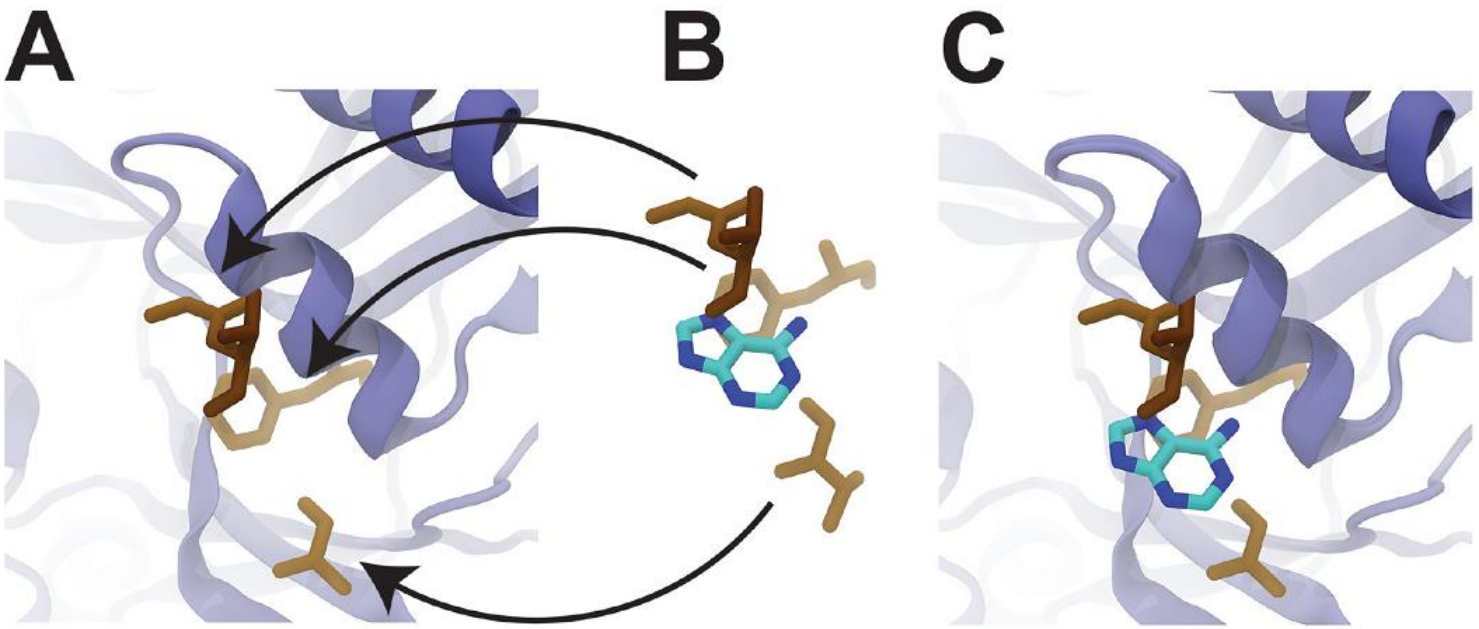
Identify the microenvironment of active-site of assigned protein

- The residues could be manually assigned or be found by offering a point(x,y,z)
- Geometric rays were also used to identify microenvironment receptor residues (default radius 5 Å)
- Get all combinations of active-site residues to build microenvironment containing 3,4,5 residues

(such as: abc,abd,acd,bcd; C_n^3)



Screening the crystal microenvironments

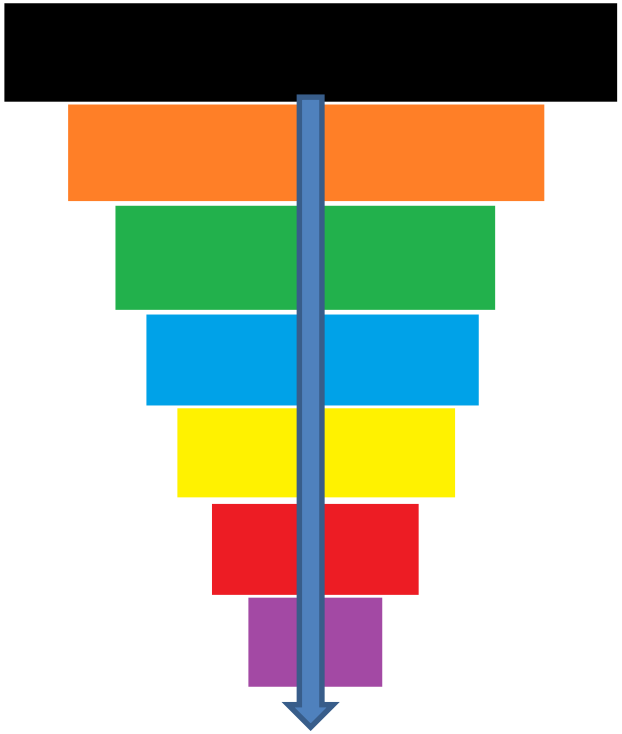


For each active-site microenvironment, the matched microenvironments in the crystal microenvironment library are found by screening. After that, the corresponding fragments Of the matched microenvironment are deposited into the pocket.

Microenvironment screening

Six Filter

1. Residue identification(same or similar)
2. Geometric similar(span and fingerprint)
3. RMSD alignment(Alpha carbon)
4. Side-chain orientation
5. RMSD alignment(heavy atoms)
6. Clash fragment exclusion



These six steps for filtering the microenvironments are most crucial and time-consuming for the program

Residue identification

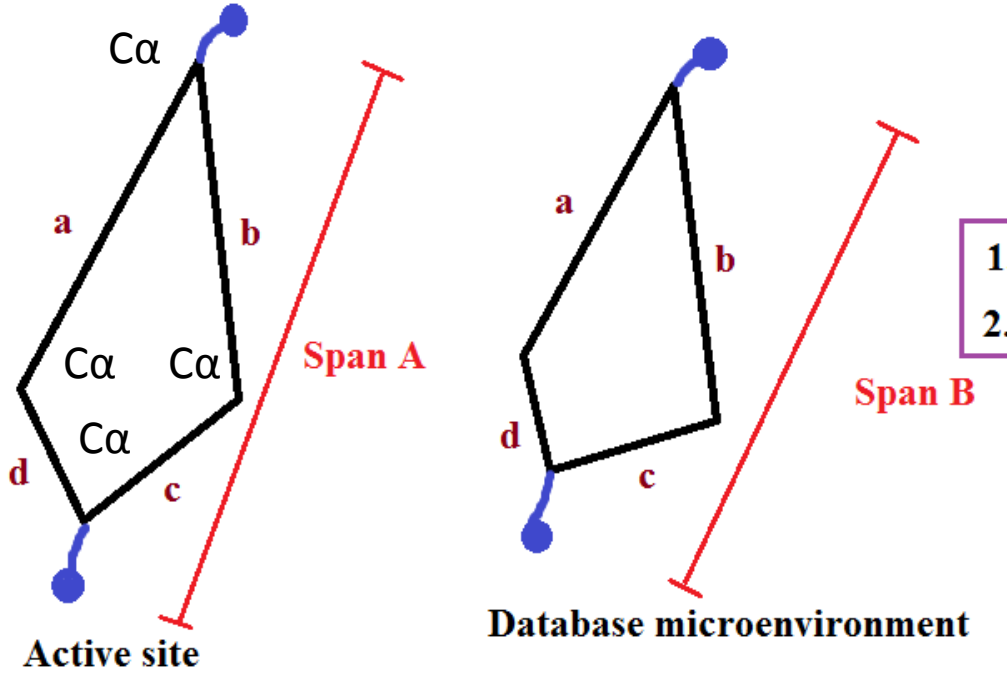
	C	S	T	A	V	Q	N	D	E	H	R	K	M	I	L	F	Y	W
C		X	X															
S	X		X															
T	X	X																
A					X													
V				X									X	X	X			
Q							X											
N						X												
D									X									
E								X										
H												X						
R												X						
K										X	X							
M					X									X	X			
I					X								X		X			
L					X								X	X				
F																	X	X
Y																X		X
W																X	X	

Residue chemical similarity

Only the microenvironments containing the identical or chemically similar residues to the target are used for further study. The similarity matrix above is based on BLOSUM62.

Geometric similar

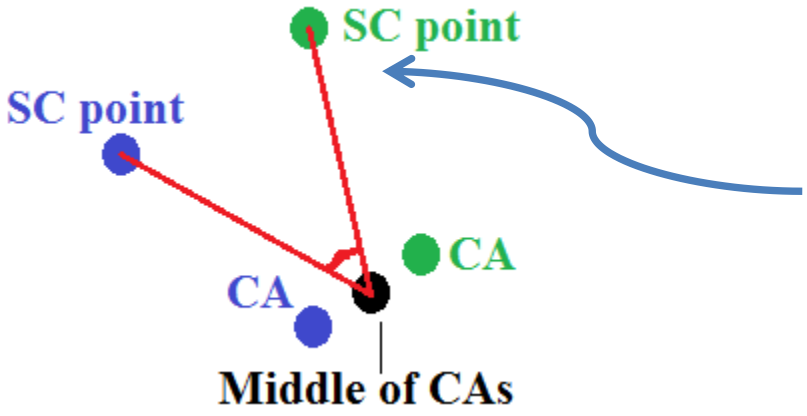
- To find the maximum distance between any two atoms(the span of the microenvironment) in the active-site microenvironments. The difference of maximum distance between target and database structures should less than 2Å
- Difference of the distance in the sorted list of all the pairwise distances among all the alpha carbons between active-site and database microenvironments should less than 2Å



Geometric similar judgment could reduce computational intensive.

Side-chain orientation filter

- After RMSD alignment of Cα, those RMSD larger than 2.5Å were discarded.
- The side chain orientation were compared between two MEs. Angle of (sidechainA -- middle point of Cα -- sidechainB) larger than 100° was discarded.



Side-chain orientation filter

	CB	CZ	CG	SG	CD	NZ	CE	OG
ALA	X							
ARG		X						
ASN			X					
ASP			X					
CYS				X				
GLU					X			
GLN					X			
HIS			X					
ILE	X							
LEU			X					
LYS						X		
MET							X	
PHE			X					
SER								X
THR	X							
TRP			X					
TYR			X					
VAL	X							

Selected sidechain point to determine the orientation of protein side chains

Further steps

- RMSD alignment of heavy atoms(without those different atoms in similar residue)<1.5Å
- Discard the fragments that within the distance <2Å to the active-side to avoid steric clash.
- Identify and keep the non-redundant results(RMSD>=0.5 Å)

- Crude ranking based on three principles:

  1.same residues >similar residues;

  2.Residues of microenvironment:5>4>3;

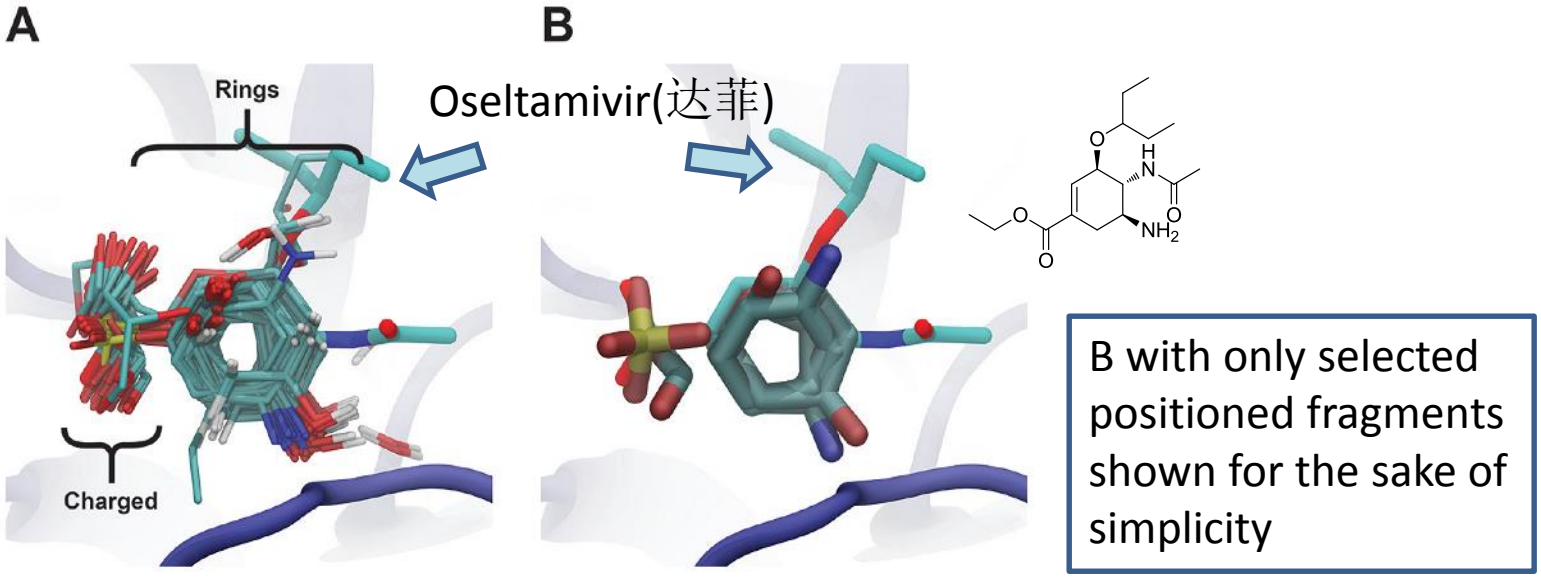
  3.RMSD, the smaller, the better.

Application and Evaluation

- Influenza neuraminidase(神经氨酸酶), a target with abundant ligand-bound structures (222 structures with ~70 unique ligands in PDB)

- Trypanosoma brucei RNA editing ligase 1 (TbREL1),a protein with only one PDB-deposited crystal structure bound to a single ligand (ATP)

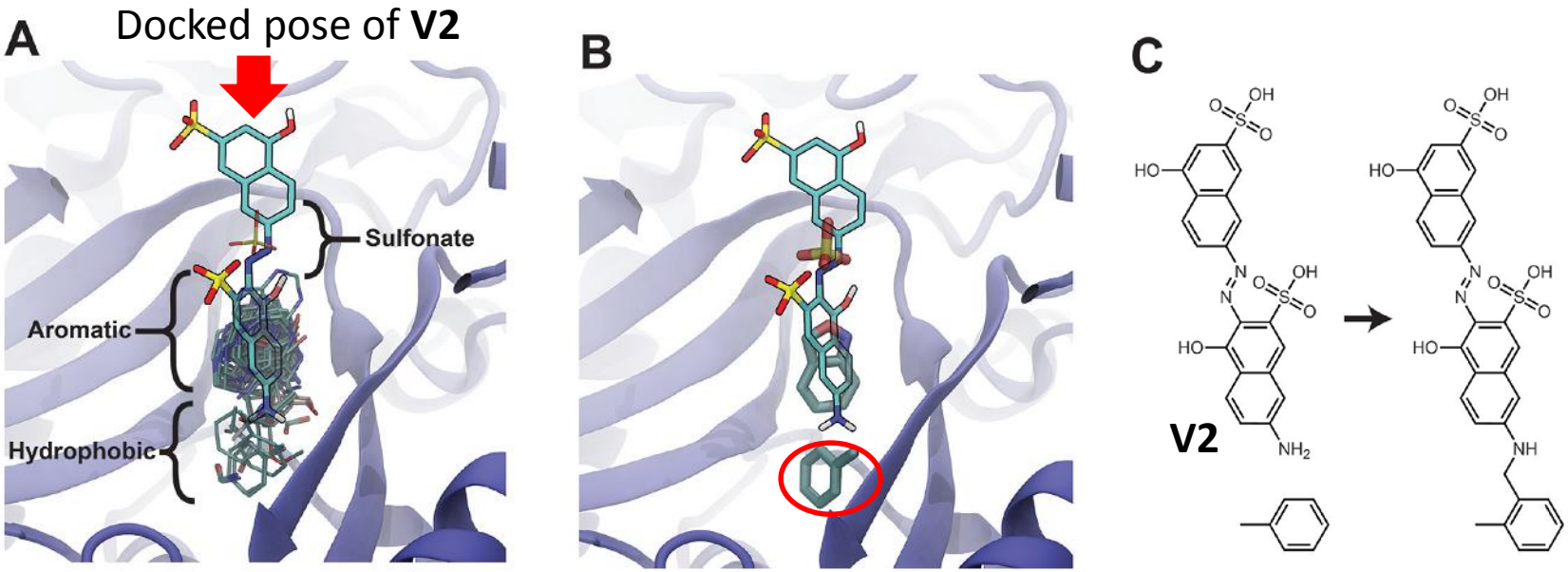
Results of influenza neuraminidase



B with only selected positioned fragments shown for the sake of simplicity

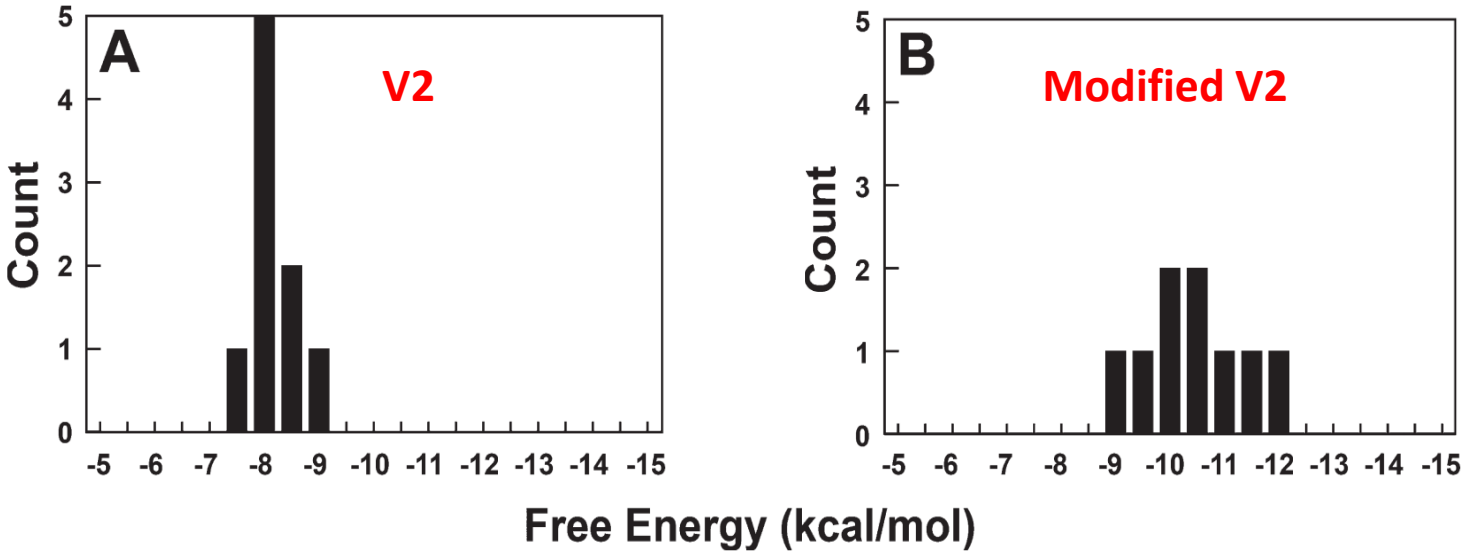
A shows 95 distinct PDB structures representing 39 unique ligands. As expected, most of the identified ringed fragments were derived from known neuraminidase ligands and other experimental inhibitors. A single fragment was also obtained from pentaethylene glycol bound to D-lactate dehydrogenase confirming that CrystalDock is able to identify binding fragments from even distantly related proteins. Interestingly, CrystalDock placed a sulfate ion near the location of the charged oseltamivir carboxylate group (Figure B),implying the potential structural modification method.

Suggested Modifications of a TbREL1 Naphthalene-Based Inhibitor



The CrystalDock-positioned fragments can be generally clustered into three groups. The third group, which is represented by several mostly hydrophobic fragments, does not correspond to any **V2** substructure (lead compound), suggesting a possible route for improving potency. Toluene fragments derived from two unique inhibitors of P38 mitogen-activated protein kinases may be ideal for modification

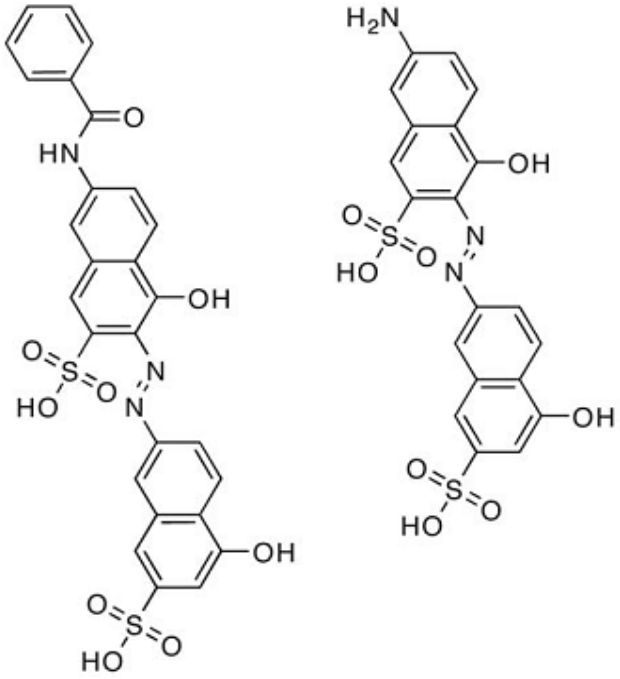
Further evaluation by IT-TI



- The ligand binding energies were predicted by Independent-trajectories thermodynamic integration (IT-TI).
- Six TI runs were executed for each system: three in which the protein-bound ligand was annihilated, and three in which the solvated ligand was annihilated. From these six TI runs, nine binding-energy estimates were calculated. The predicted binding affinity of **V2** was -8.4 $\pm$ 0.5 kcal/mol, which correlates well with the experimentally measured IC50 value of 1.53 µM.
- The predicted binding energy of the new composite compound was -10.7 $\pm$ 0.9 kcal/mol, representing a 2.3 kcal/mol improvement

However…

V1 **V2**



	V1	V2
IC_{50}	2.16±1.20	1.53±1.17
EC_{50}	>100	>100
$AutoDock_{Crystal}$	−11.8	−11.3
$Rank_{Crystal}$	11	20

In the previous study by the author, the compound V1, also containing the hydrophobic fragment and with better score in Autodock, shows worse activity.

PLoS Negl. Trop. Dis. 2010 4(8): e803

Shortcoming

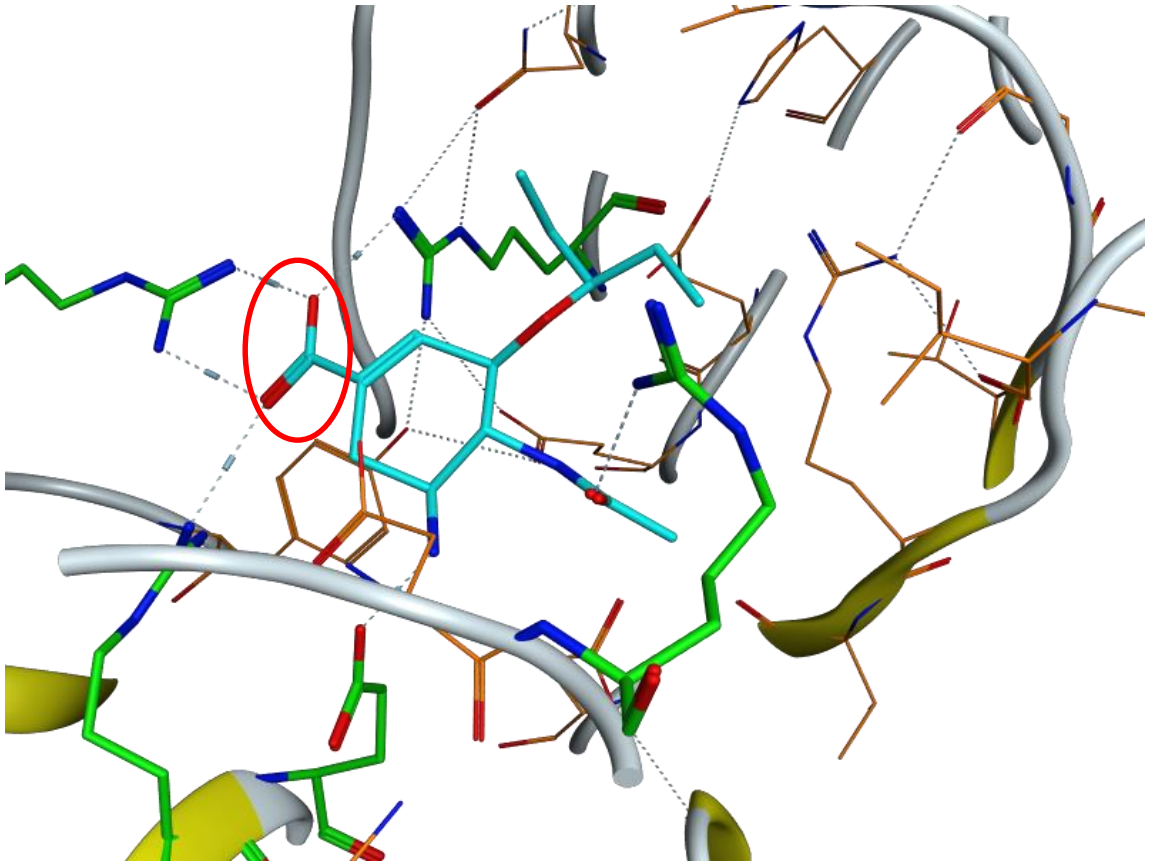
- Neglect the long-distance electrostatic interaction.
- Only offer the possible fragments but not the contact molecules.
- Though called as “dock”, it is Not suitable for high-throughput screening.
- Lack of efficient scoring function.
- Novel fragments which don’t exist in similar microenvironment but are valuable may be neglected.

Conclusion

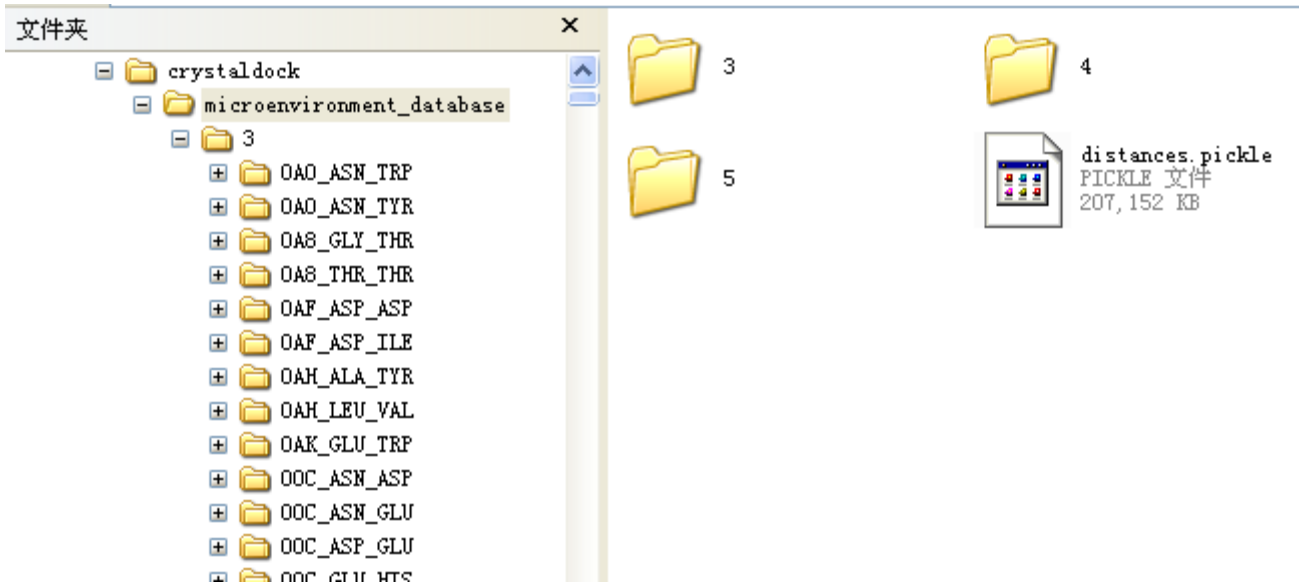
- CrystalDock is a new algorithm for fragment-based drug design based on the fragment-microenvironment information from known crystal complex structures from PDB.
- It could be used for identification of the potential binding fragments, even further optimization of lead compounds.

Thank you!

Carboxylate group in Oseltamivir



Fragment-Microenvironment database



3.5h in p380 when docking 2838 microenvironments for 13 residues in active-site(just consider the 3 residues microenvironments).