1/18 면담 내용
1. CAPRI가 무슨 뜻인지 알아보기
[CAPRI ROUND 28 (ebi.ac.uk)]

카프리 기준

CAPRI is a blind test of the ability of protein-protein docking algorithms to predict the mode of association of two proteins based on their three-dimensional structure.

Short description of the peptide assessment procedure

With respect to the regular assessment of protein-protein interaction in CAPRI, the assessment criteria have been slightly tightened for theassessment of protein-peptide interaction.

The following distance cut-offs are employed:

- a distance cut-off of 8 A between any two CB atoms (or CA for Gly) to define interface residues
- a distance cut-off of 4 A between any two atoms to define native contacts
- a clash distance of 3 A

Also the classification criteria have been tightened:

- f(nat) thresholds of 0.2, 0.5 and 0.8 now distinguish between high (f(nat)>0.8), medium (f(nat)>0.5), acceptable (f(nat)>0.2) and incorrect models (f(nat)<0.2), or:
  * fnat: incorrect < 0.2 < acceptable < 0.5 < medium < 0.8 < high
- and similarly L-rms and I-rms:
  * lrms: incorrect > 4.0 > acceptable > 2.0 > medium > 1.0 > high
  * irms: incorrect > 2.0 > acceptable > 1.0 > medium > 0.5 > high

in other words: a high-quality protein-peptide model has f(nat) above 0.8 and either L-rms below 1.0 or I-rms below 0.5.

카프리는 정기적으로 열리는 프로틴 도킹 알고리즘 대회…? 커뮤니티 거기의 기준을 사용한다.

2. Dockground에서 Y가 1인(correct)한 decoy가 무슨 뜻인지

1a2k_u1, 1a2k_u2 is the unbounded form of receptor and ligand, which is the input for those docking methods. They are trying to check how to bind them together with correct conformations.

what is y? This is the correctness of generated structures, which is based on CAPRI criteria to compare against the bounded form 1a2k_p1 and 1a2k_p2. To get y, you need to calculate iRMSD, lRMSD and fnat by comparing decoys with 1a2k_p1 and 1a2k_p2. However, if you are working on DOCKGROUND dataset, it should be simpler. All the correct decoys typically with large numbers like 365130, and those decoys named with r-l-1.pdb to r-l-100.pdb are typically incorrect structures.

> unbounded form을 input으로 docking 한 decoy들을 bounded form (native, correct) 와 비교해서 iRMSD, lRMSD and fnat을 계산해서 Y를 얻는데, Dockground만 사용하면 correct를 1, incorrect를 0으로 사용하면 된다고 합니다.

3. DOVE에 따로 hit rate를 설명한게 있는지 확인하기

> the fraction of target complexes in the ZDOCK dataset for which a method produced at least one correct (i.e. CAPRI acceptable) model within top *k* rank.
이런식으로 GNN DOVE에서와 비슷한 설명을 하였습니다.

4. hit rate를 물어보기

In the paper, a hit rate of a method is the fraction of target complexes where the method ranked at least one acceptable model based on the CAPRI criteria within each top rank. Figure2 (A) The panel shows the fraction of target complexes among the 58 complexes in the benchmark set for which a method selected at least one acceptable model (within top *x* scored models).

some method	**rank 1**	**2**	**3**	**4**	**…**
complex1	hit				
complex2		hit			
complex3	hit				
complex4				hit	

For example, there are only four complexes. And the complexes are re-ranked by some method.
If top 1 rank is considered, hit rate is 50%.
Top 2 rank considered, hit rate is 75%.
Top 3 rank considered, hit rate is 75%.
Top 4 rank considered, hit rate is 100%.

Is Hit rate scored like this?
> Yes. Your understanding is correct. That's how we define hit rate in our paper. 이라고 답장이 왔습니다.

6. hit rate 그려보기

Considering the settings of 4-fold, fold 1 model in our paper means model trained and validated with data from fold 2,3,4, and the model is finally tested on fold 1.

4-fold로 train하고 test를 진행했다고 합니다. 그래서fold 별로 다 진행해서 해봐야하는데….

2. l-rmsd, r-rmsd, fnat 무슨 뜻인지 알아보기

Define an interface region between both chains in the native structure. Basically, check for all atoms of chain A which atoms of chain B are within 10Å.

Rank : the rank by Gramm-x in scanning stage.
R_rmsd : the RMSD of backbone atoms (N, Ca, C, O) of receptor residues calculated after finding the best superposition of bound and unbound structure.
L_rmsd : the RMSD of the backbone atoms of the ligand after receptor was optimally superimposed.
I_rmsd : the RMSD of the backbone atoms of the interface residues after they have been optimally superimposed.
fnat : the number of native (correct) residue-residue contacts in the predicted complex divided by the number of contacts in the native complex.
fnon-nat: the number of non-native (incorrect) residue-residue contacts in the predicted complex divided by the total number of contacts in that complex.

rmsd - native와의 거리

R- rmsd - backbone atoms (N, Ca, C, O) of receptor residues 거리
L-rmsd - the RMSD of the backbone atoms of the interface residues

fnat : native residue-residue 연결 비율 fraction of native contact

7. fnat 계산해보기 DockQ

rmsd.list라는 파일이 있어서 각 파일의 R-rmsd, L-rmsd, I-rmsd, fnat, fnon-nat이 있습니다

Rank	R_rmsd	L_rmsd	I_rmsd	fnat	fnon-nat
1	0.69	53.23	18.31	0.00	1.00
2	0.69	55.35	17.46	0.00	1.00
3	0.69	55.36	18.47	0.02	0.99
4	0.69	54.80	17.72	0.00	1.00
5	0.69	54.20	16.95	0.00	1.00
6	0.69	56.13	16.26	0.00	1.00
7	0.69	54.06	18.63	0.00	1.00
8	0.69	54.08	16.91	0.00	1.00
9	0.69	52.22	18.20	0.00	1.00
10	0.69	46.62	13.92	0.00	1.00

list에 있는 것

Rank	R_rmsd	L_rmsd	I_rmsd	fnat	fnon-nat
365130	1.32	4.78	2.89	0.62	0.51

DockQ를 돌린 것
Model : decoys/r-l_365130.pdb
Native : decoys/native.pdb
Fnat 0.075 4 correct of 53 native contacts
Fnonnat 0.940 63 non-native of 67 model contacts

iRMS 7.188
LRMS 9.522
DockQ 0.187

왜 fnat이 다르게 나오는지 모르겠습니다

native파일을 잘못설정한건가 싶어서 인터넷에서 새로 다운받아서 해봐도 같은 결과가 나왔습니다.