Protein complex structure prediction using GALAXY in CASP13

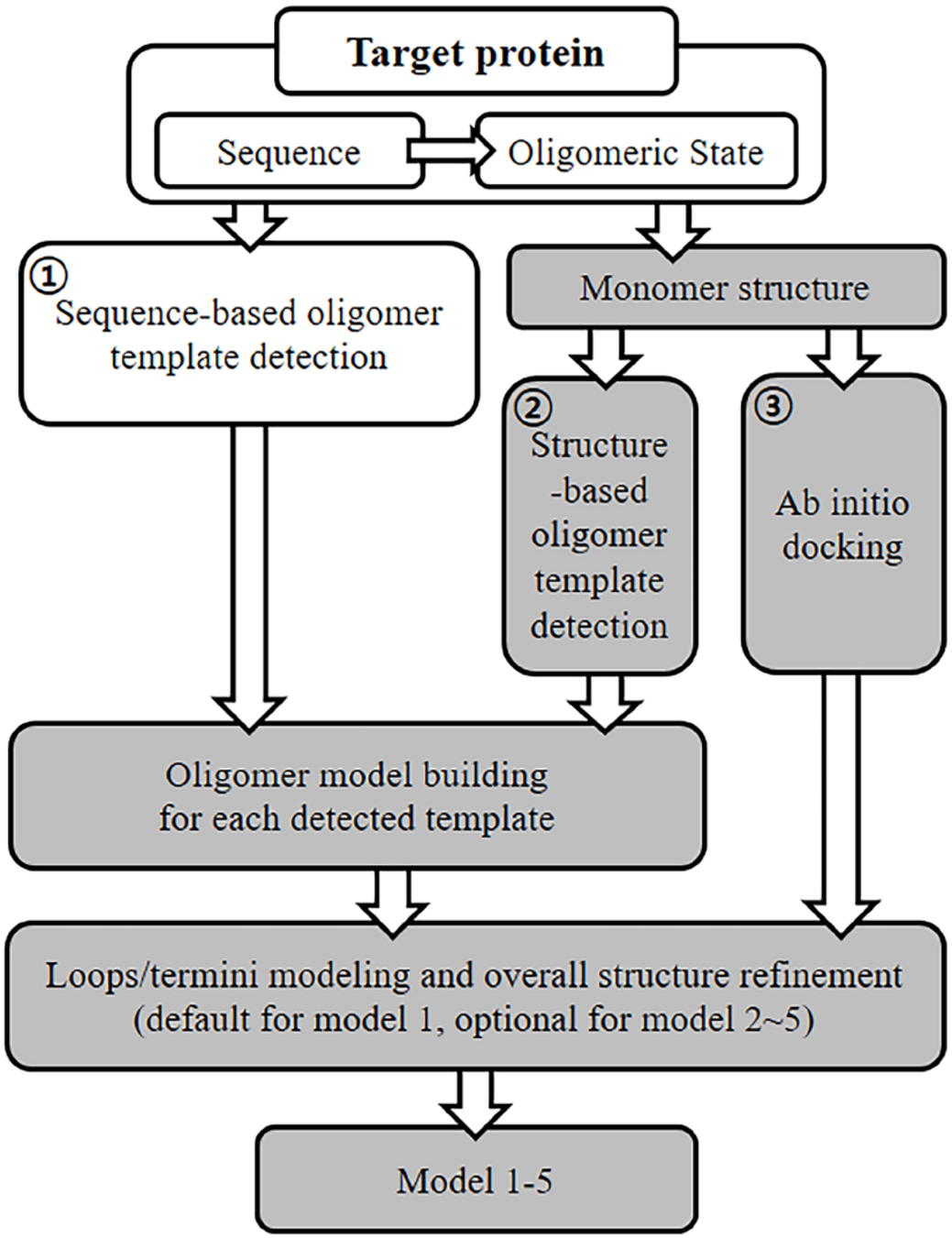
CASP13 Meeting, Dec 2018

Minkyung Baek, Taeyong Park, Hyeonuk Woo, and Chaok Seok
(Seok, Seok-assembly, Seok-naive_assembly)

Lab of Computational Biology and Biomolecular Engineering
Department of Chemistry, Seoul National University

Homo-oligomer structure prediction

GalaxyHomomer (Seok-assembly): Automated homo-oligomer structure prediction



All successful predictions came from ***template-based*** approach

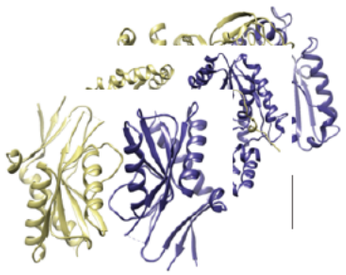
Baek, M., Park, T., Heo, L., Park, C., & Seok, C. (2017). GalaxyHomomer: a web server for protein homo-oligomer structure prediction from a monomer sequence or structure. *Nucleic acids research*, *45*(W1), W320-W324.

GalaxyHomomer: Details on the template-based approach

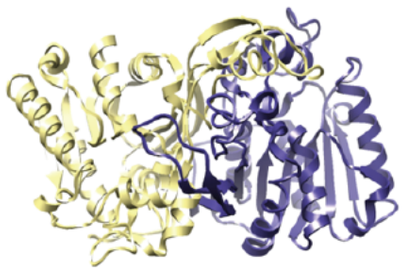
Sequence

Oligomer state

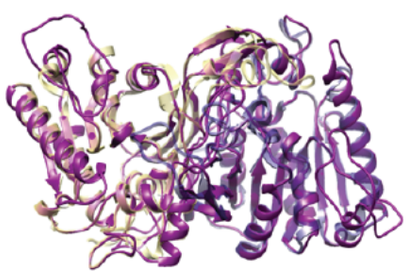
Template-based initial model building



HHsearch on oligomer database

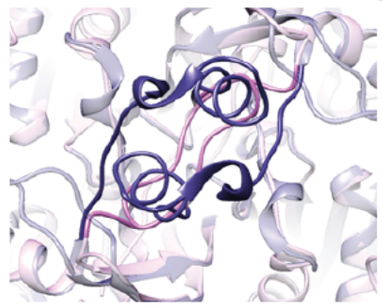


Rescoring & Template selection

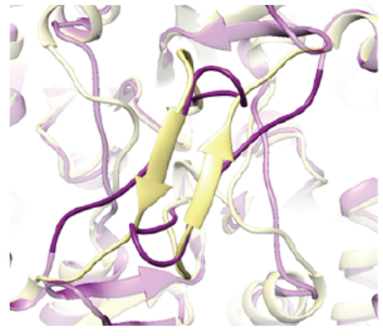


Oligomer model building (GalaxyCassiopeia)

Symmetric ULR modeling

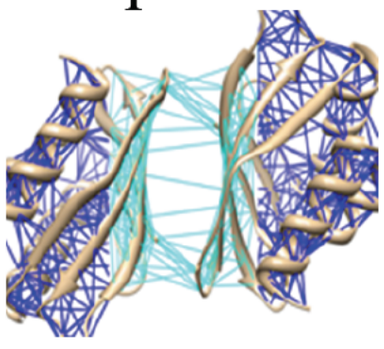


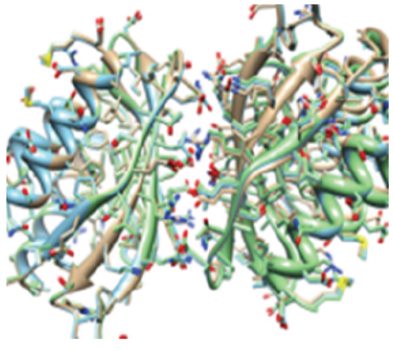
Automatic ULR detection



ULR modeling (GalaxyLoop)

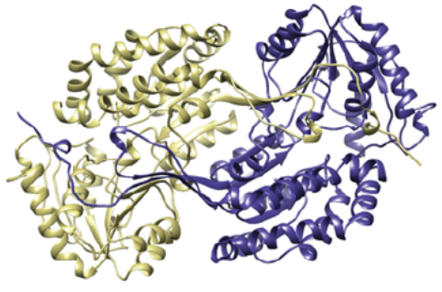
Complex structure refinement





Short MD relaxation (GalaxyRefineComplex)

Final model



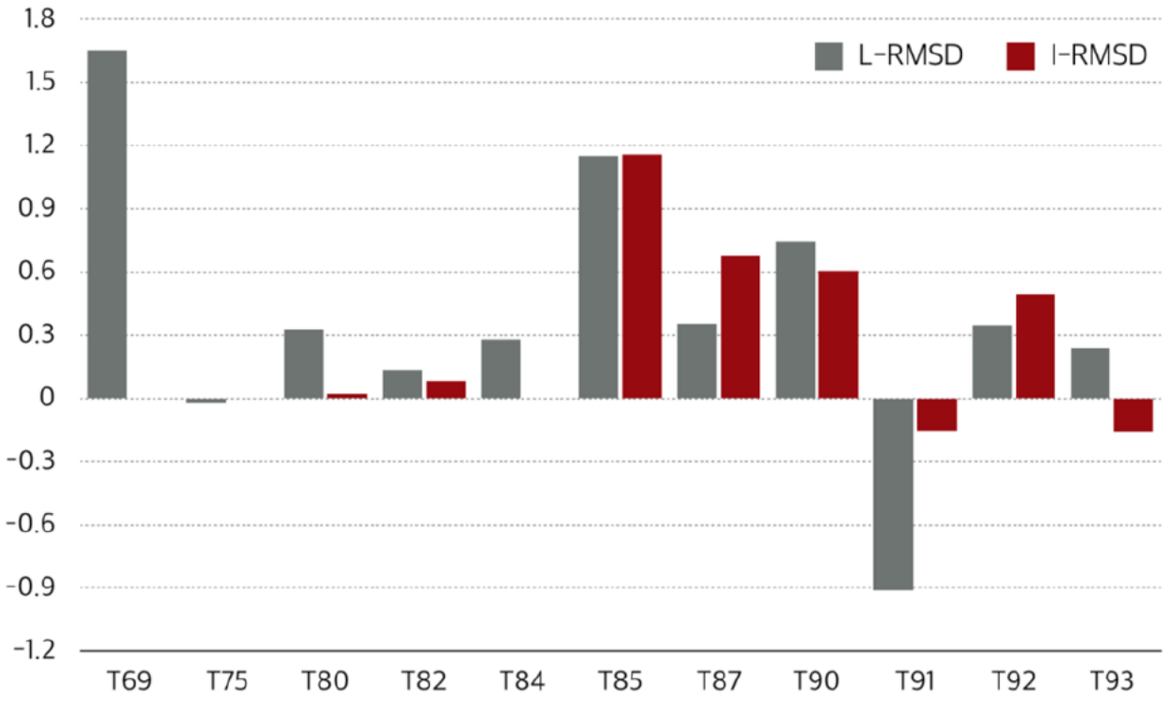
E_{min} model

Baek, M., Park, T., Heo, L., Park, C., & Seok, C. (2017). GalaxyHomomer: a web server for protein homo-oligomer structure prediction from a monomer sequence or structure. *Nucleic acids research*, *45*(W1), W320-W324.

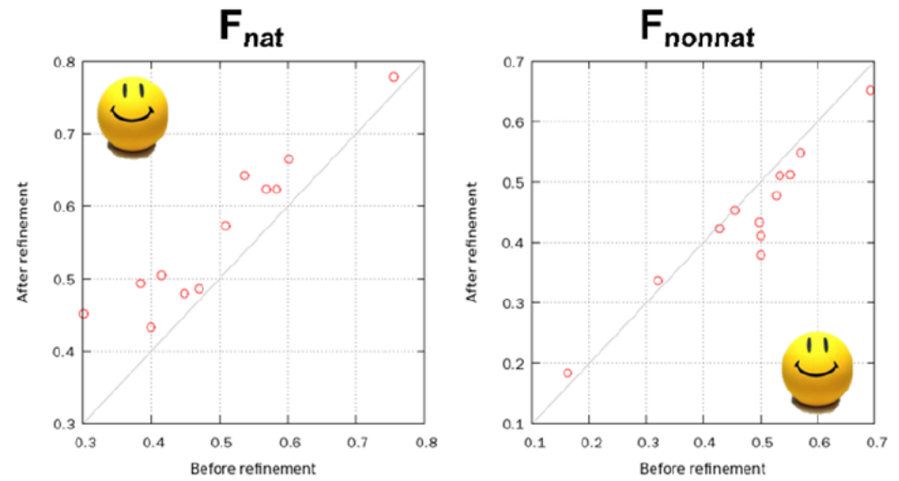
GalaxyHomomer: Impact of model refinement tested on CASP11 targets

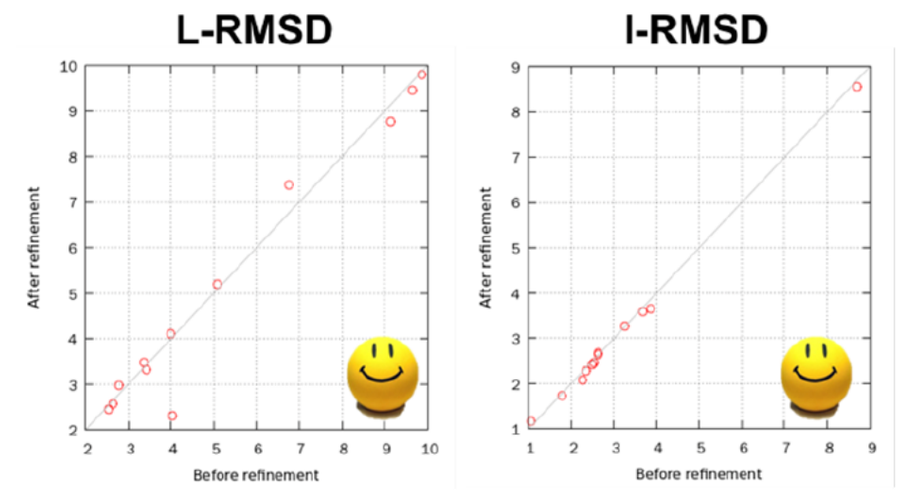
- Tested on the 20 CASP11-CAPRI30 targets

Ligand / Interface RMSD improvement by ULR modeling



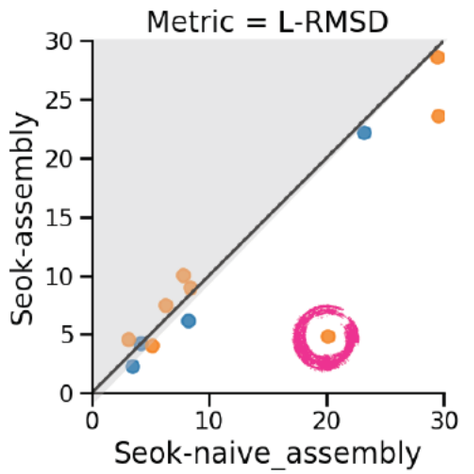
Model quality improvement by overall refinement

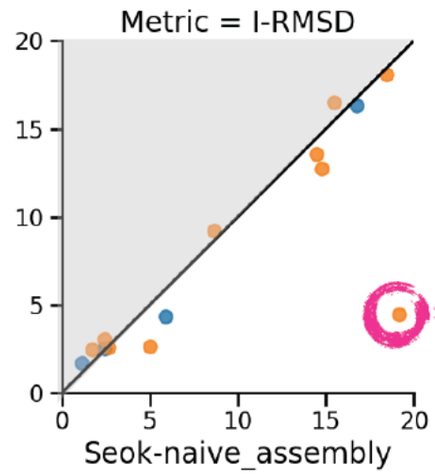


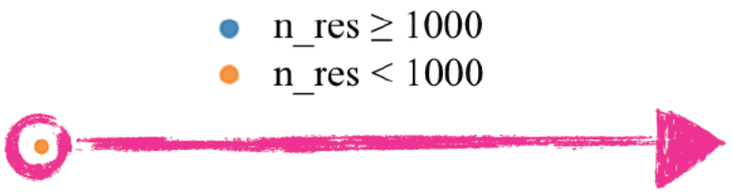


GalaxyHomomer: Performances in CASP13

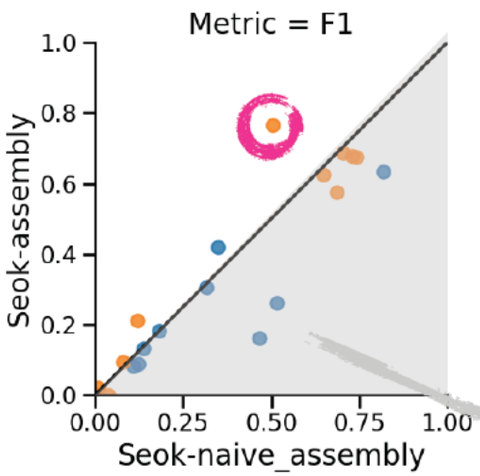
- Compared to Seok-naive_assembly (HHsearch + MODELLER)

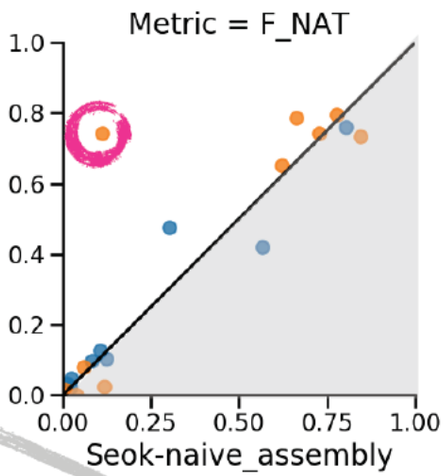


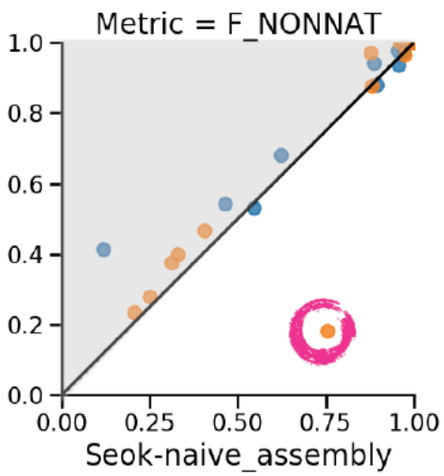




T0979
(GalaxyHomomer detect better template by re-scoring)







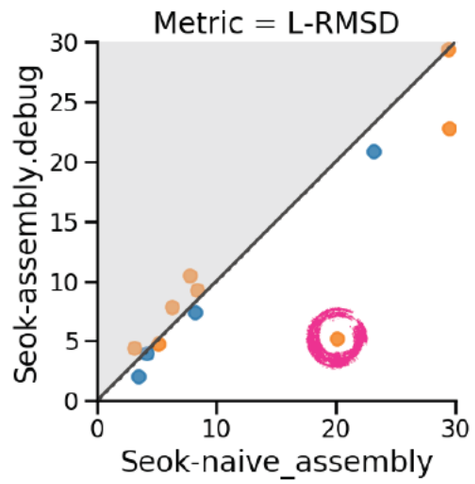
GalaxyHomomer is worse than the naive approach in contact-based measure. Why?

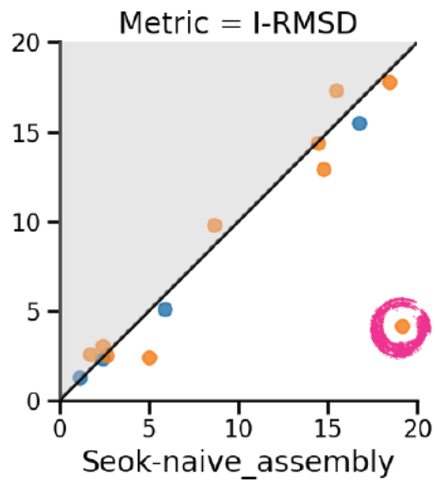
GalaxyHomomer: Performances in CASP13

- Unfortunately, we found critical bugs in our code
  - Some of inter-chain restraints were ignored for large complexes (n_res ≥ 1000)
  - Symmetry constraint was ignored in side-chain sampling during overall refinement

GalaxyHomomer: Performance in CASP13 after debugging

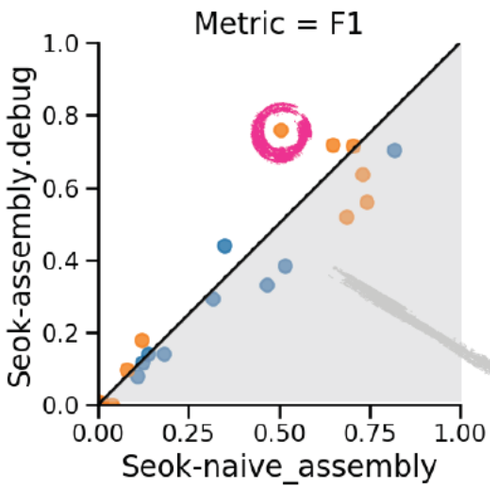
- Compared to Seok-naive_assembly (HHsearch + MODELLER)

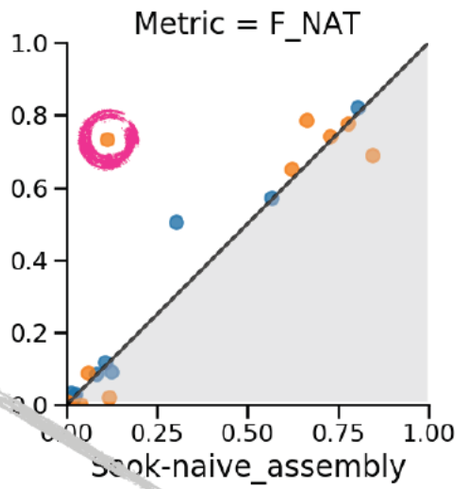


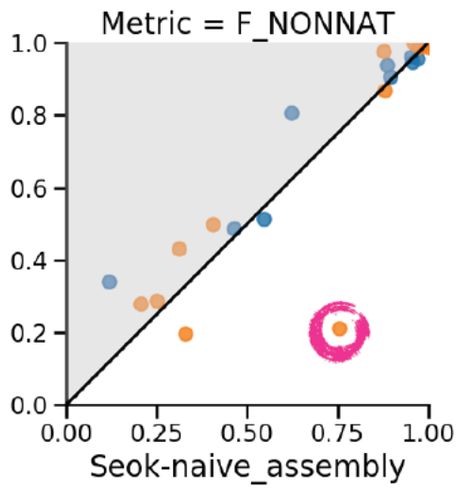


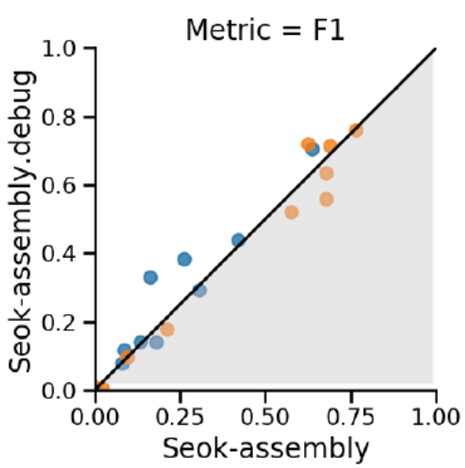
n_res ≥ 1000
n_res < 1000

T0979
(GalaxyHomomer detect better template by re-scoring)





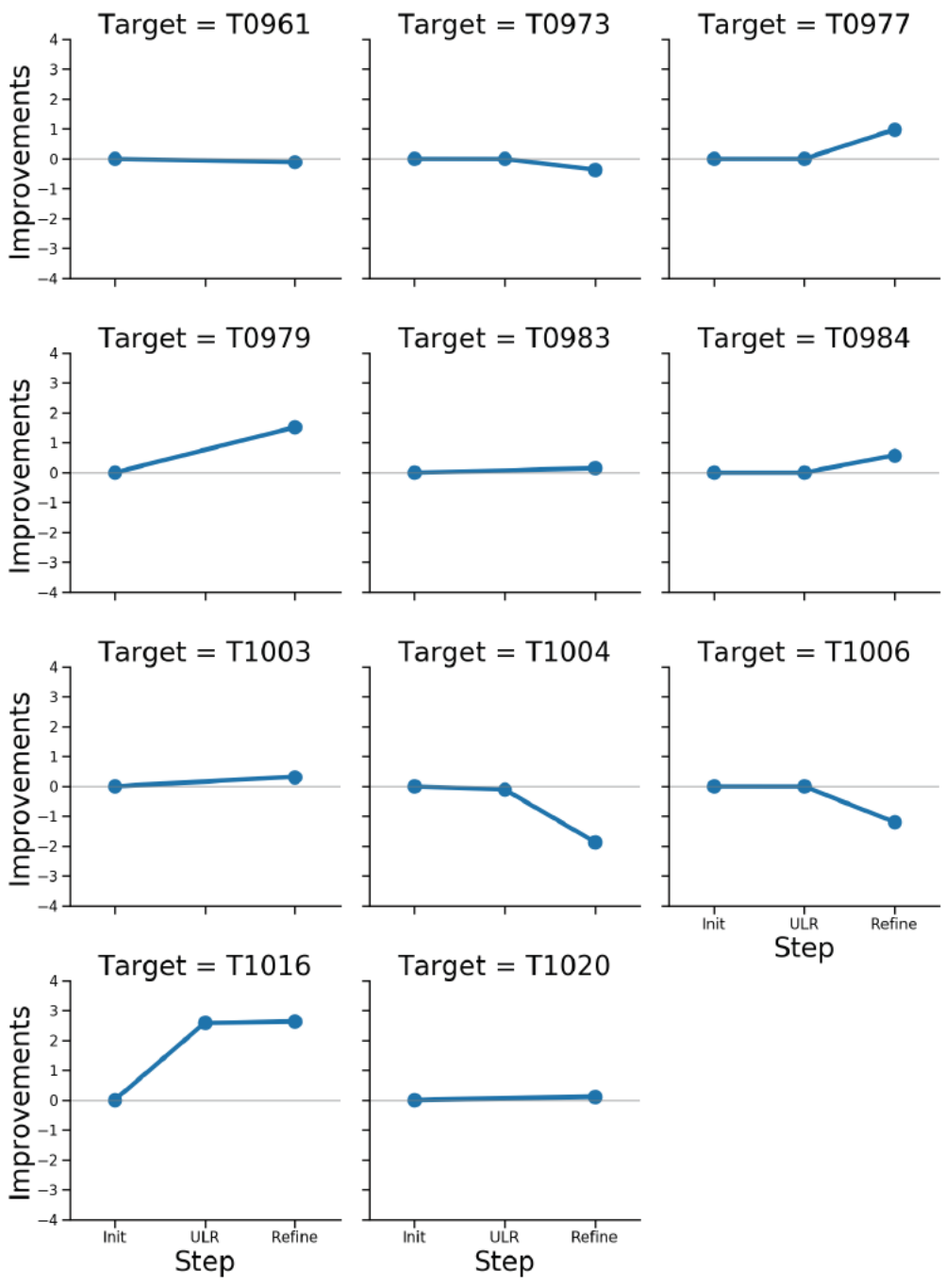




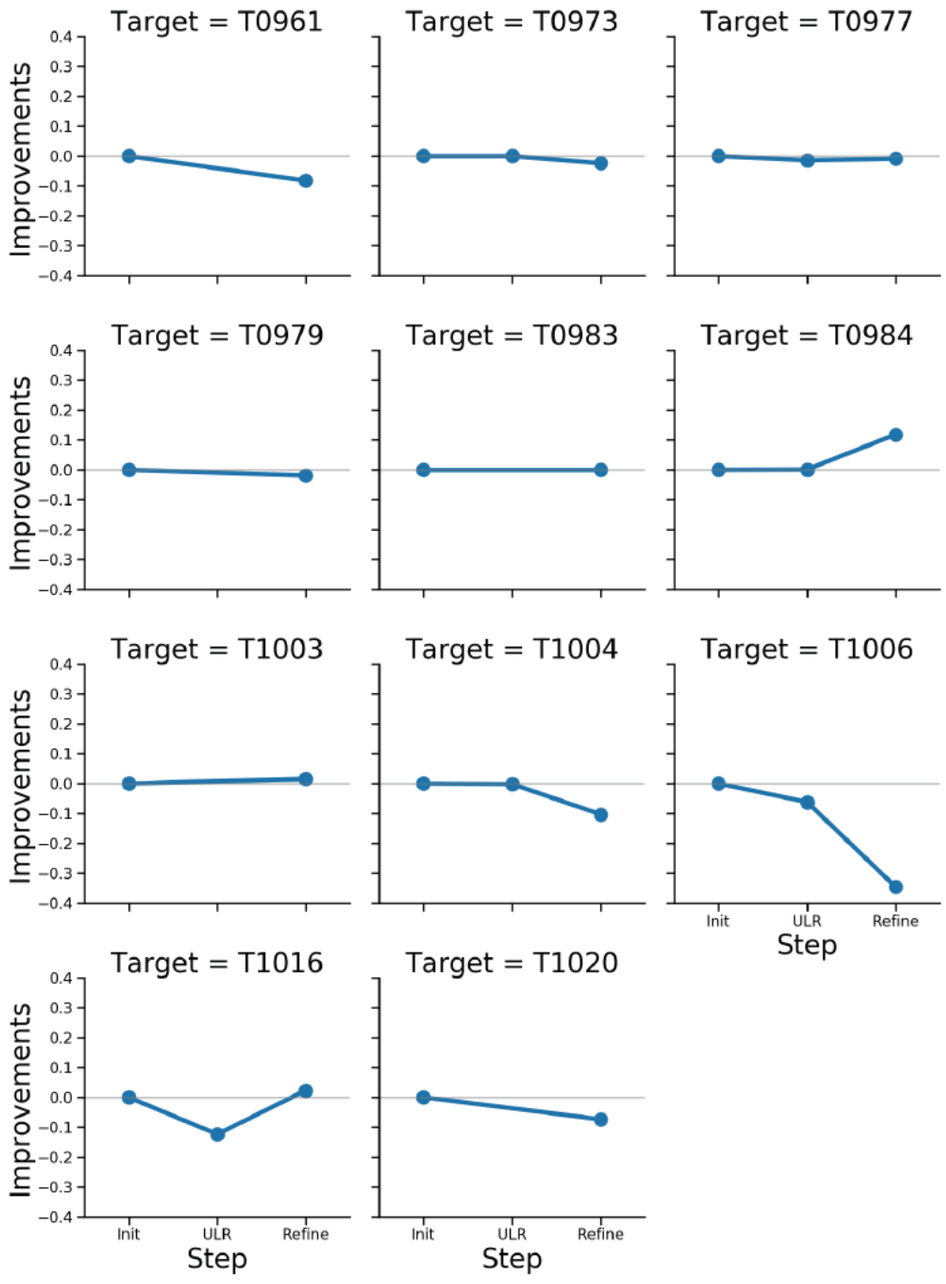
Improved in contact-based measure after bug-fix, but still a bit worse than the naive approach.

GalaxyHomomer: Performance contribution of each refinement step

I-RMSD improvements

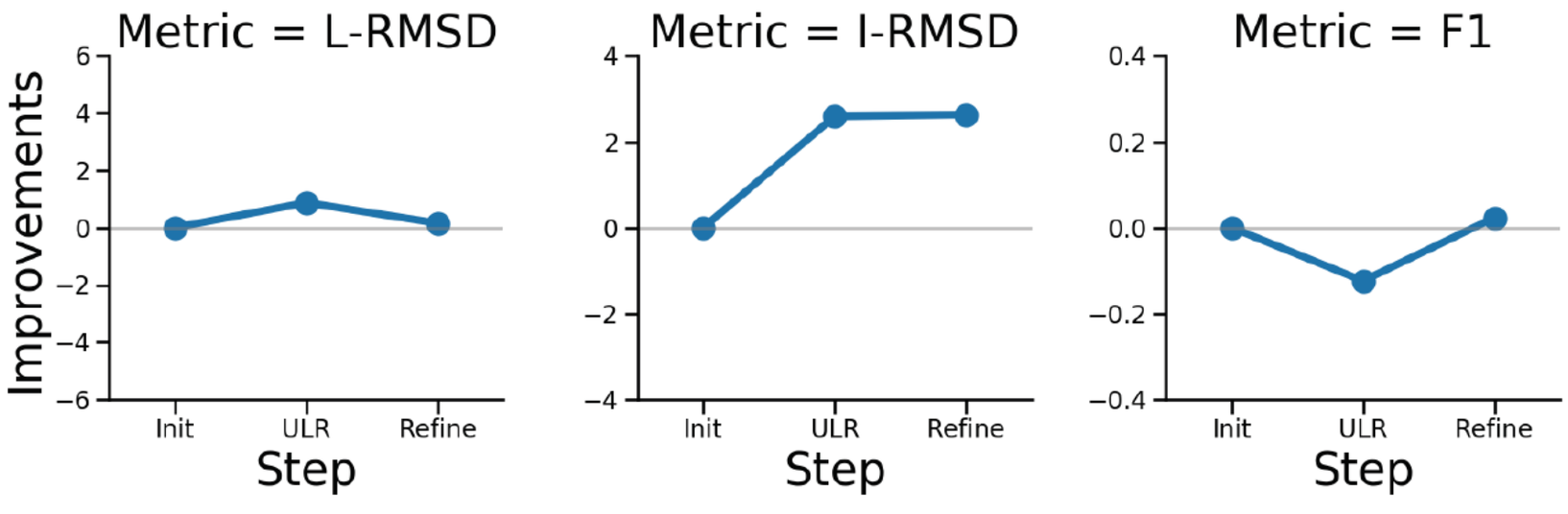


F1 improvements

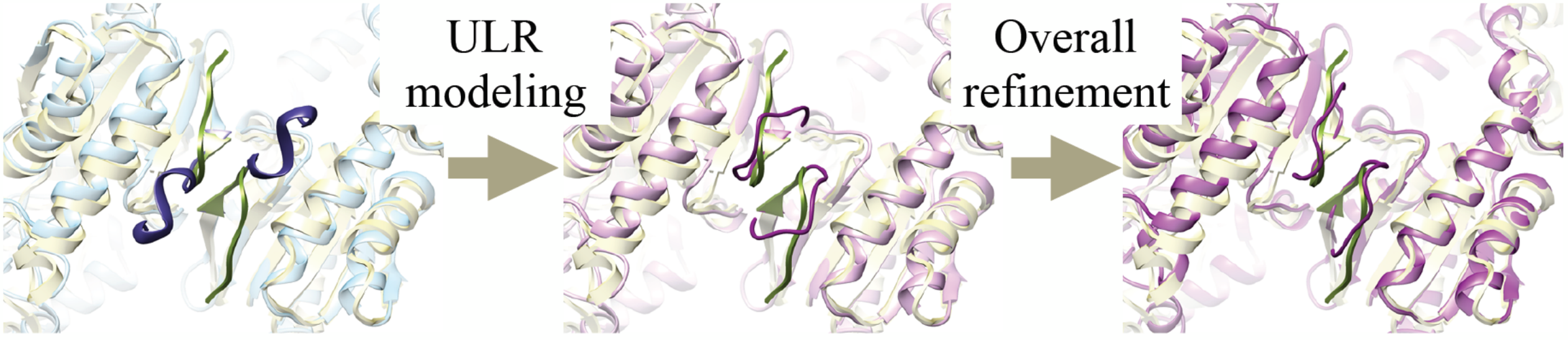


Analysis on the targets having F1 > 0.3

GalaxyHomomer: What went right (T1016)



Crystal structure
Initial structure
Refined structure

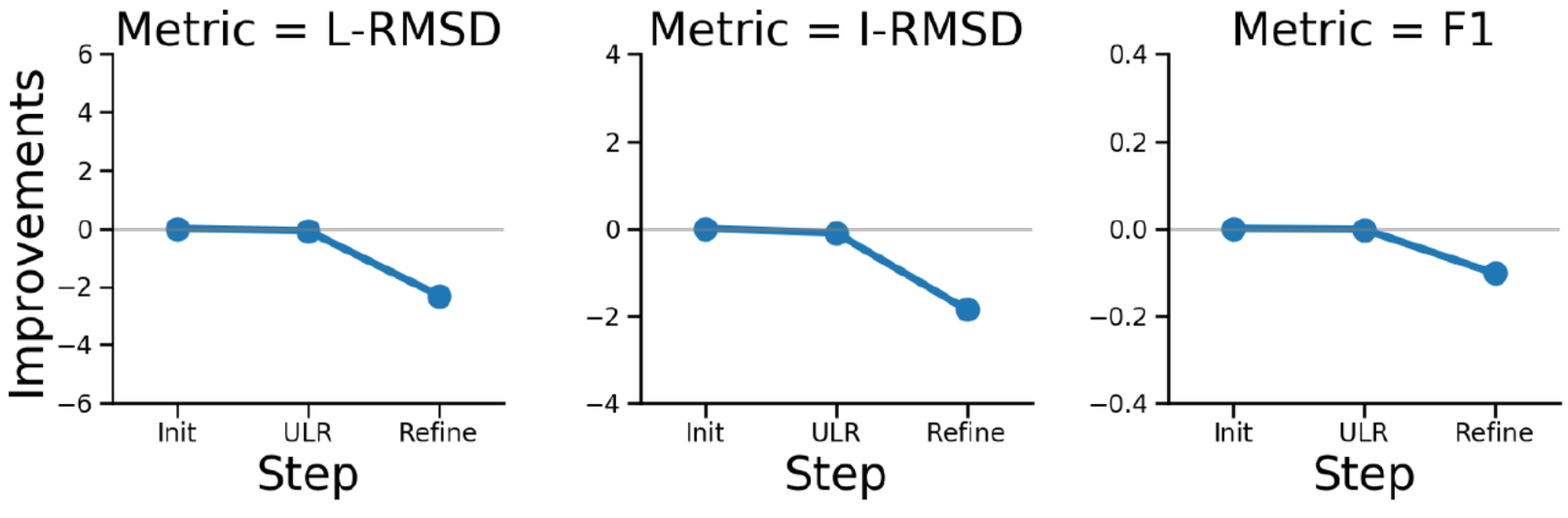


L-RMSD: 4.93 Å
I-RMSD: 5.04 Å
F1: 0.613

L-RMSD: 4.24 Å
I-RMSD: 2.45 Å
F1: 0.489

L-RMSD: 4.78 Å
I-RMSD: 2.41 Å
F1: 0.636

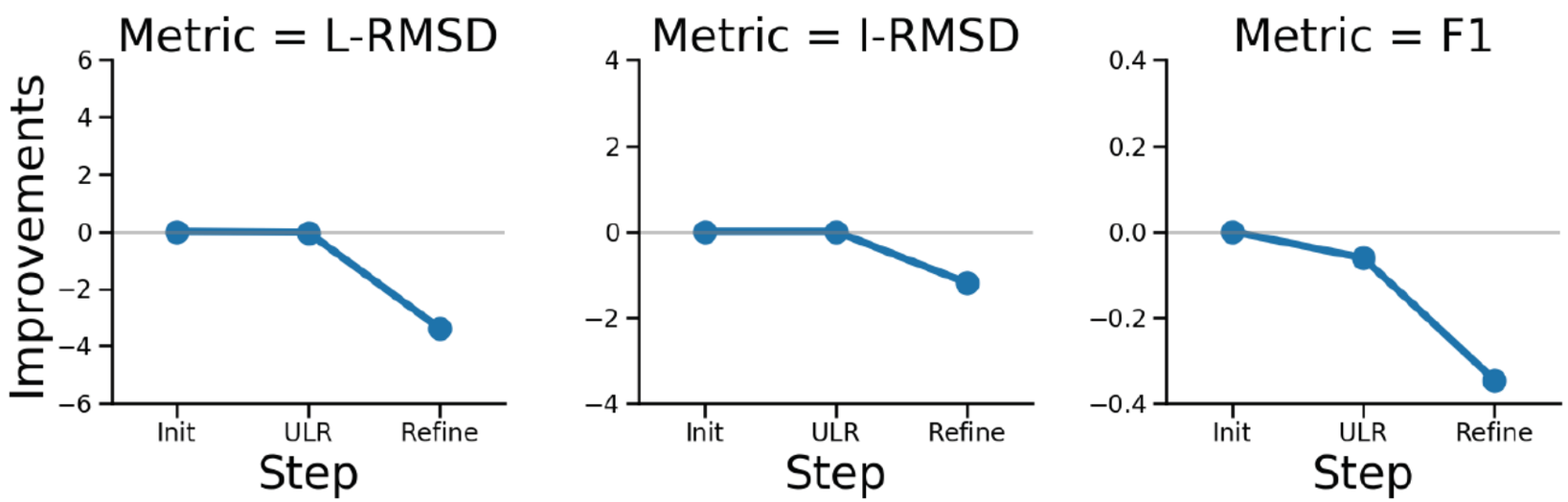
GalaxyHomomer: What went wrong (T1004: multiple modeling units)

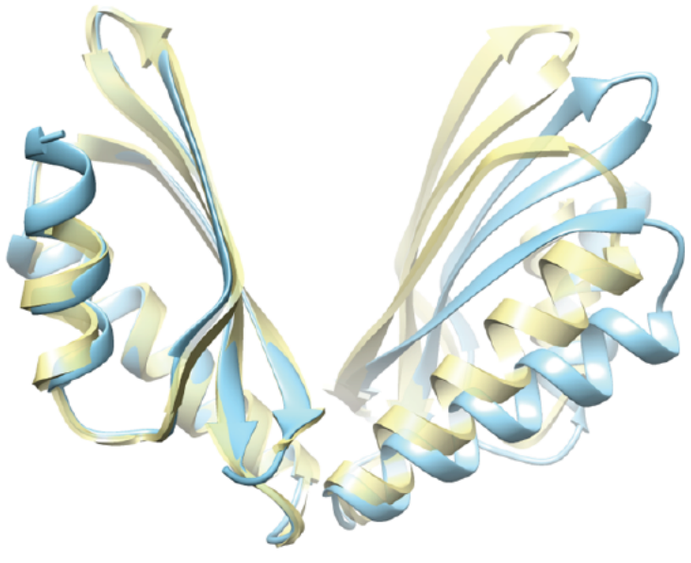


Crystal structure
Initial structure
Refined structure

CASP:
Images redacted

GalaxyHomomer: What went wrong (T1006: small interface)



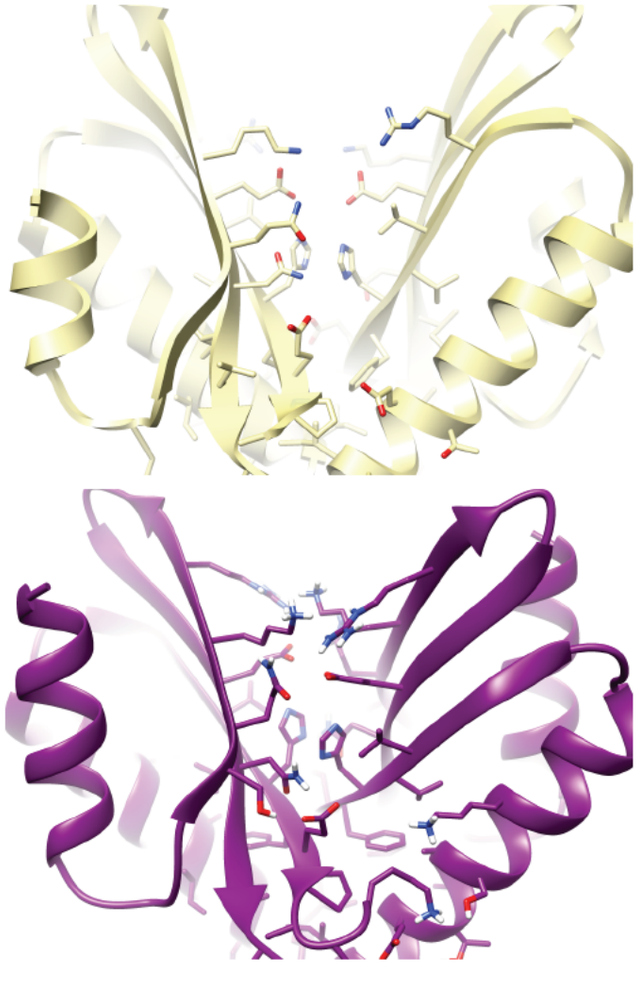


L-RMSD: 4.50 Å
I-RMSD: 1.37 Å
F1: 0.870

Overall refinement

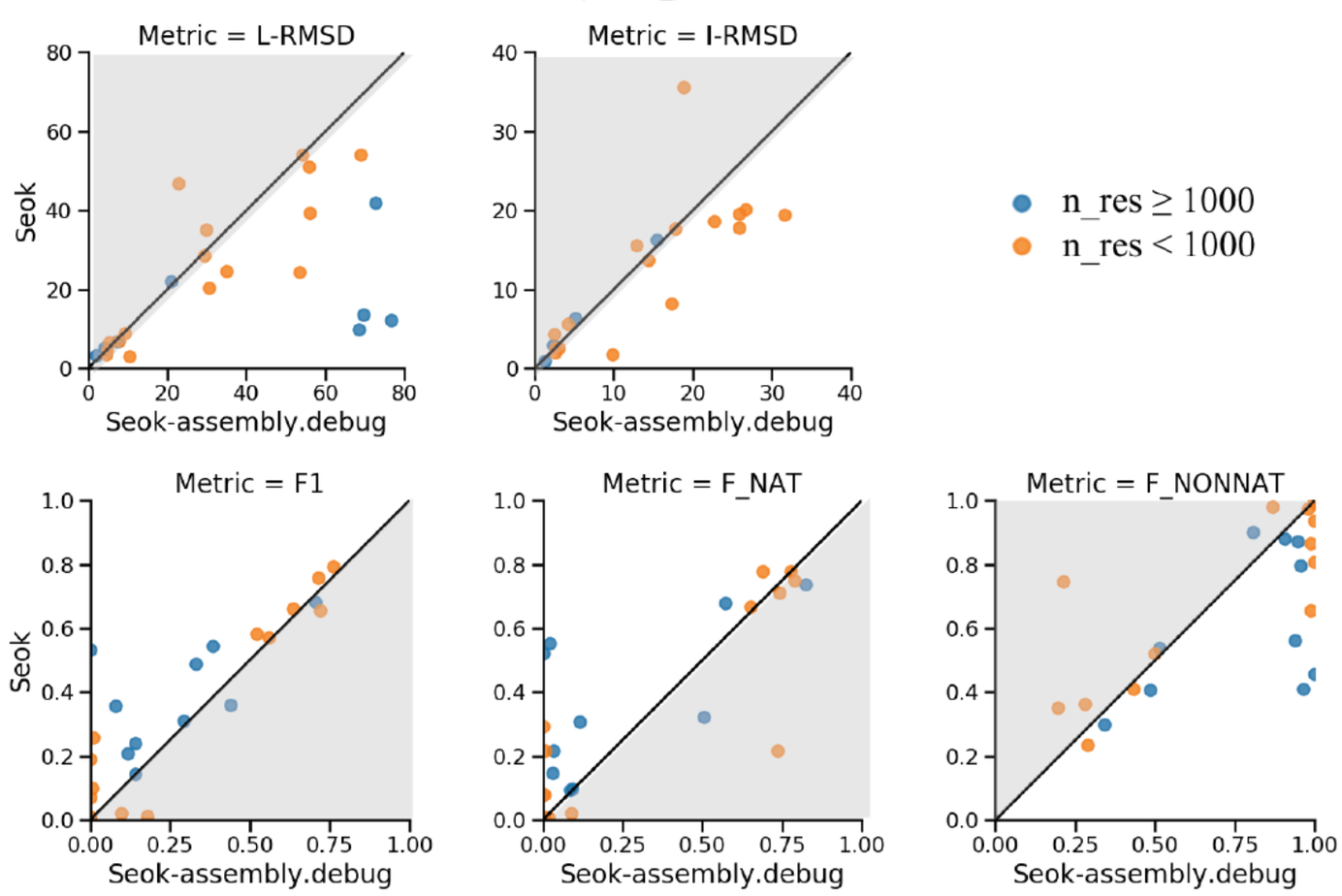
L-RMSD: 7.87 Å
I-RMSD: 2.57 Å
F1: 0.593

Crystal structure
Initial structure
Refined structure

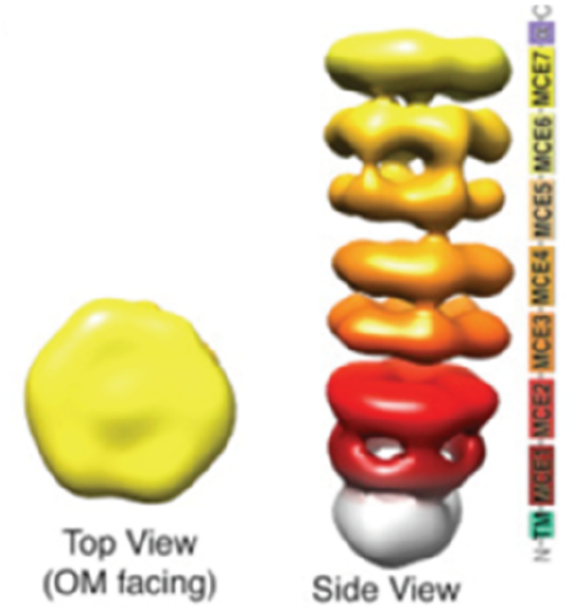


Incorporating information in human prediction

- Made corrections to the server models
  - Multiple modeling units, alignment errors, etc.
- Experiment results from literature search (e.g. low resolution EM map)



What went right, but could have been better
T0996: utilizing low-resolution EM data



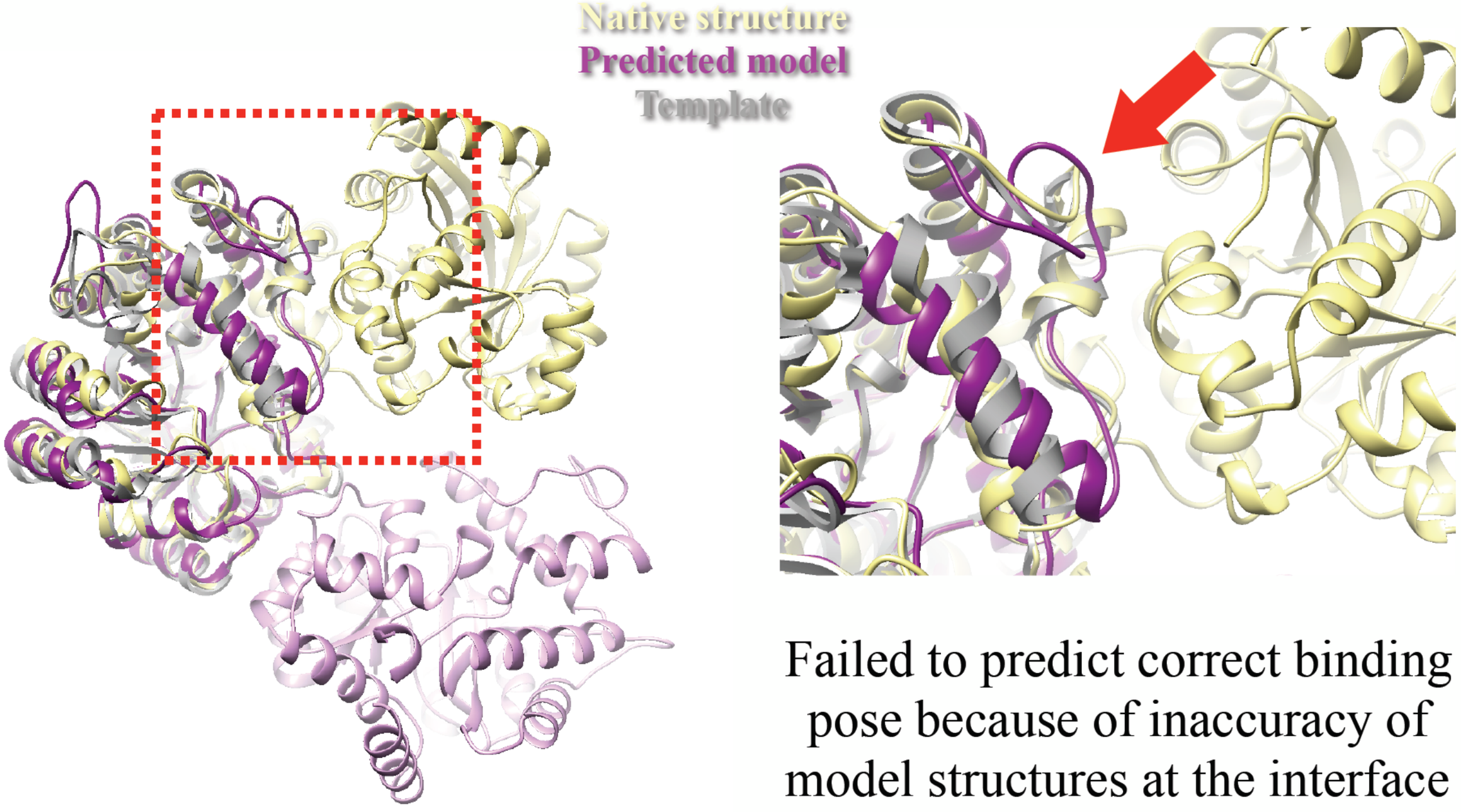
- Very low resolution (~20 Å) EM map
- Each ring was modeled first, and the modeled rings were stacked by fitting into the EM map using UCSF Chimera

CASP:
Images redacted

Native structure
Predicted model
EM map

T0996	Ligand RMSD	Interface RMSD	F1
Model 1	9.81 Å	9.88 Å	0.357

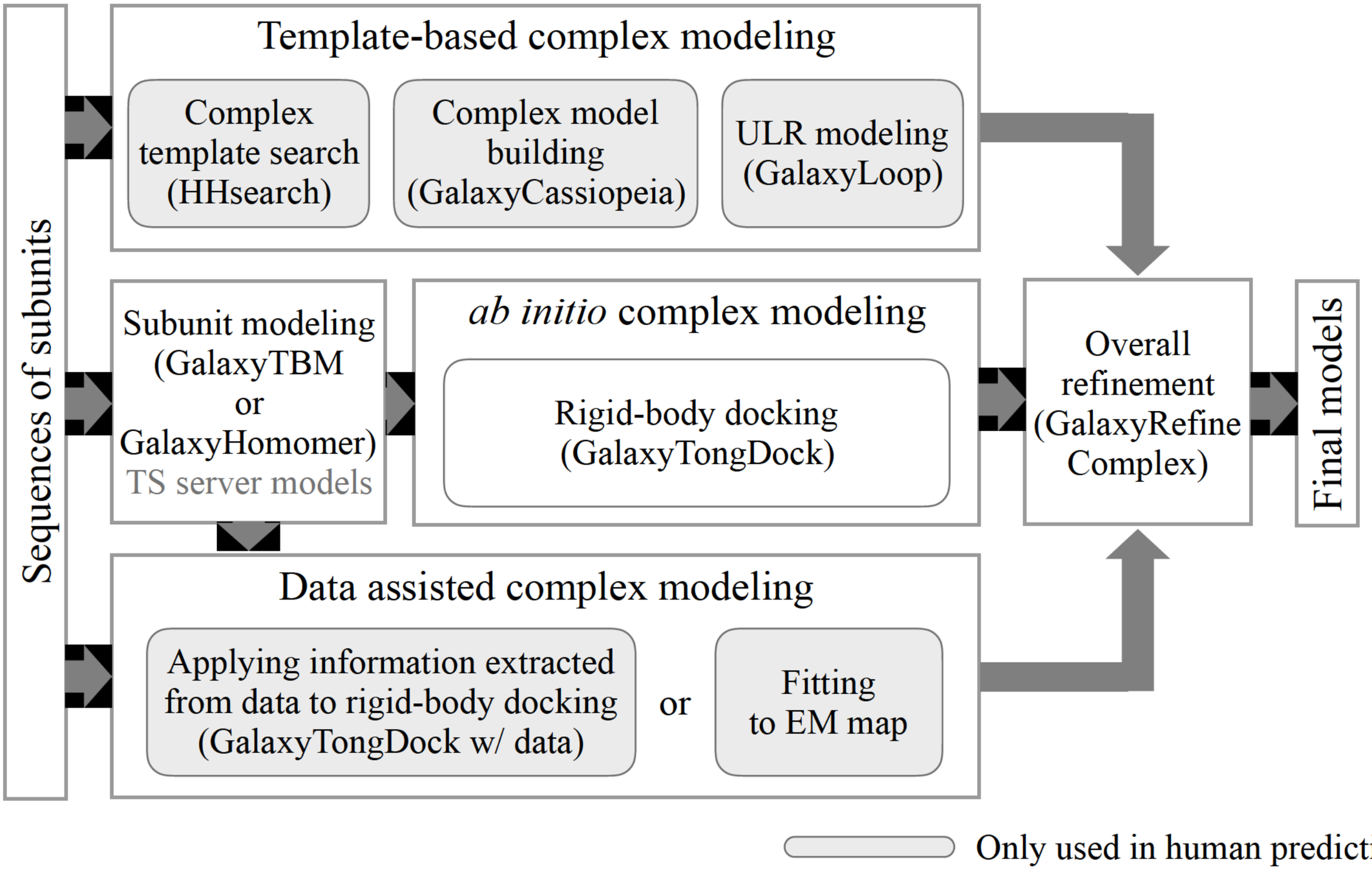
What went wrong
T1018: failed to predict via *ab initio* docking



Failed to predict correct binding pose because of inaccuracy of model structures at the interface

Hetero-oligomer structure prediction

Methods used in hetero-oligomer prediction



Server protocol failed for all hetero-oligomer targets

- It was hard to model subunit structures for the most of targets
- Even when subunits were modeled relatively well, docking failed due to local inaccuracy of the subunit models

CASP:
Images redacted

Information from template or experimental data enabled more accurate complex structure predictions

- All successful predictions came from human experts
  - by building complex structures based on complex templates
  - by utilizing experimental data in docking (e.g. low resolution EM map)

What went right
H0974: template-based complex modeling

- Using homo-dimer structure (1Y7Y) as a template

90

CASP:
Images redacted

Crystal structure
Template structure
Predicted model

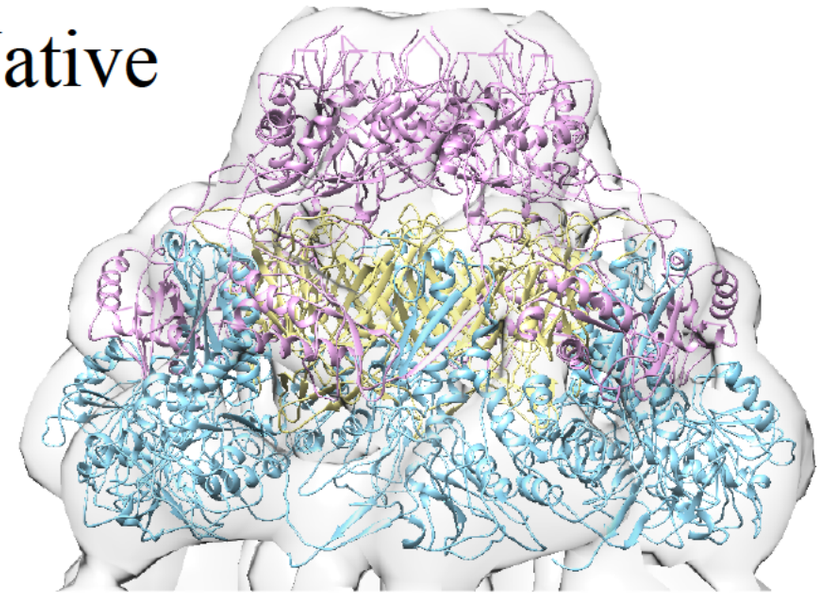
H0974	Ligand RMSD	Interface RMSD	F1
Model 4	4.94 Å	3.09 Å	0.352

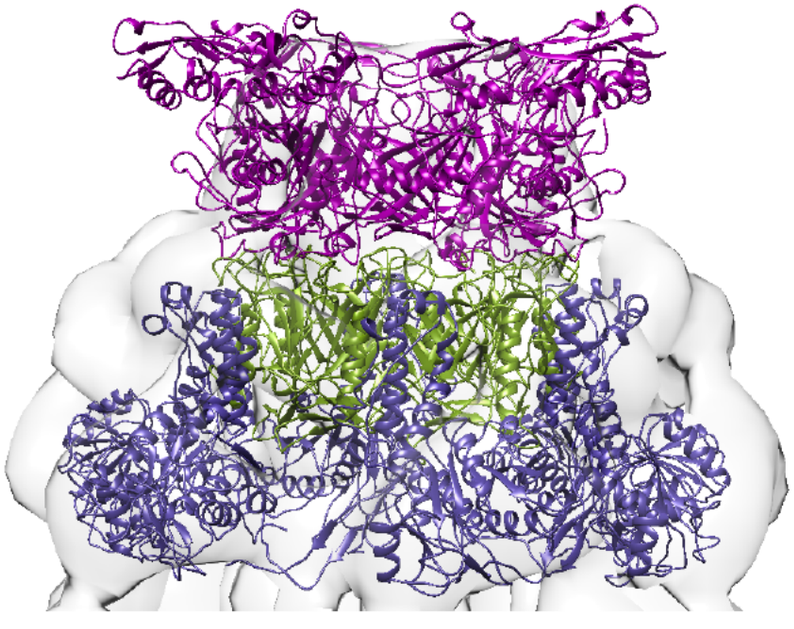
Seok (human)

What went right, but could have been better
H1021: utilizing low-resolution EM data

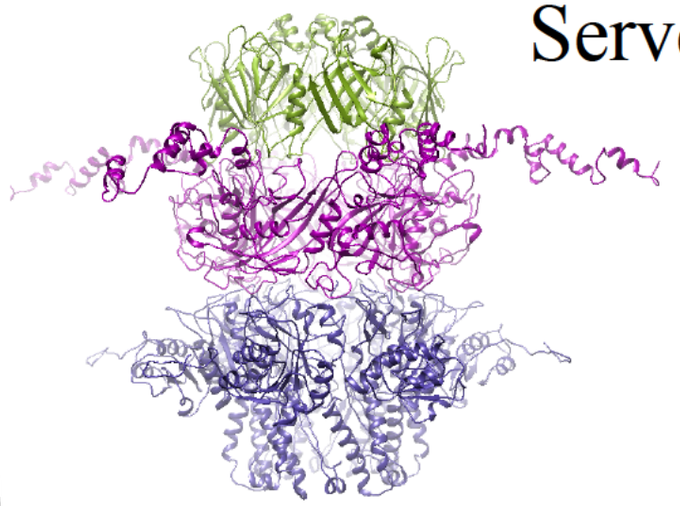
Subunit 1
Subunit 2
Subunit 3

Native





Server



Model 5
H1021
F1: 0.253

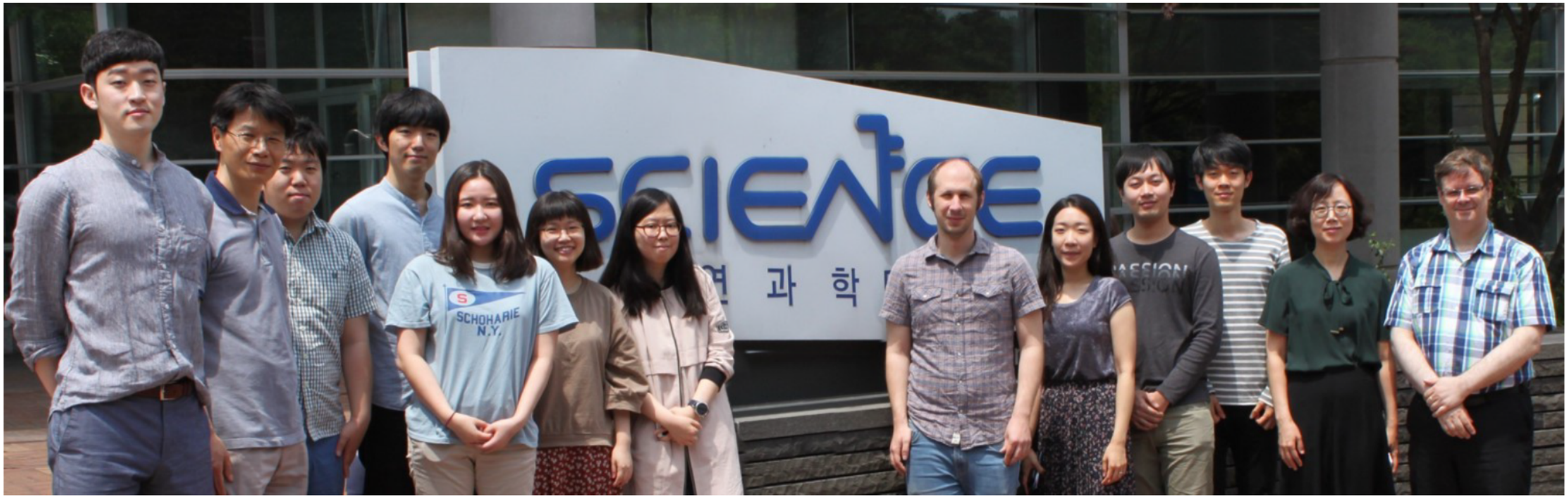
Conclusion

- *Ab initio* docking of model structures is still challenging due to local inaccuracy of models especially at the interface

- Template-based approach is really powerful if proper template can be found
    - Additional model refinement can improve model quality, but should be carefully applied.

- Experimental data like low-resolution EM map can assist oligomer modeling, but requires additional optimization

Acknowledgement



Chaok Seok

Lab members (especially, Taeyong Park and Hyeonuk Woo)