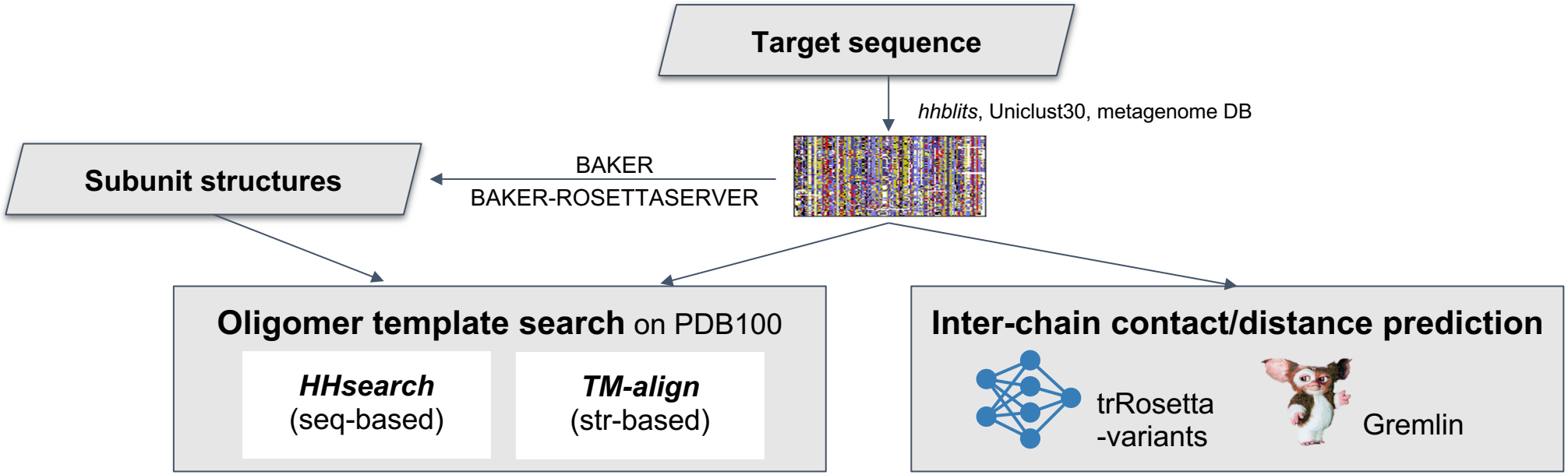
BAKER-experimental (Assembly):
Protein oligomer structure prediction guided by predicted inter-chain contacts

Minkyung Baek, Ivan Anishchenko, Hahnbeom Park, Ian Humphrey, and David Baker

CASP14 meeting, Dec 2020

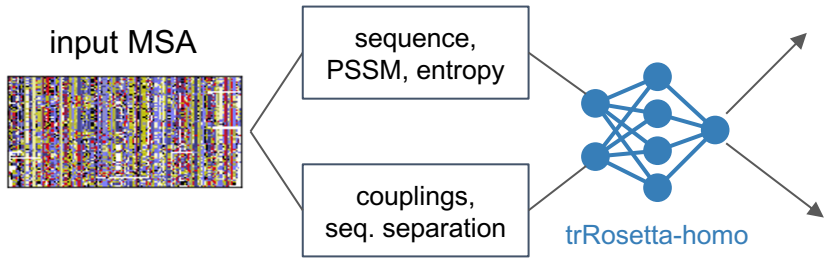
INSTITUTE FOR
Protein Design
UNIVERSITY of WASHINGTON

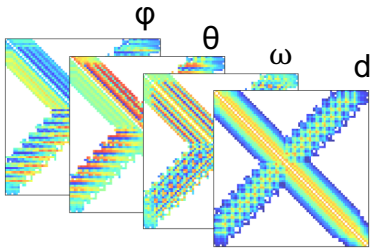
Protocols used in assembly prediction



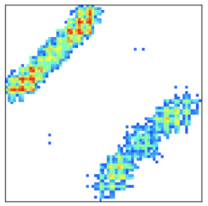
DL-based inter-chain contact & distance prediction

For homo-oligomers



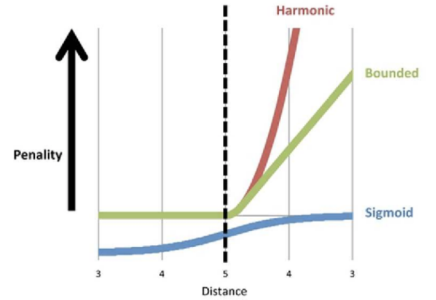


Intra-chain predictions



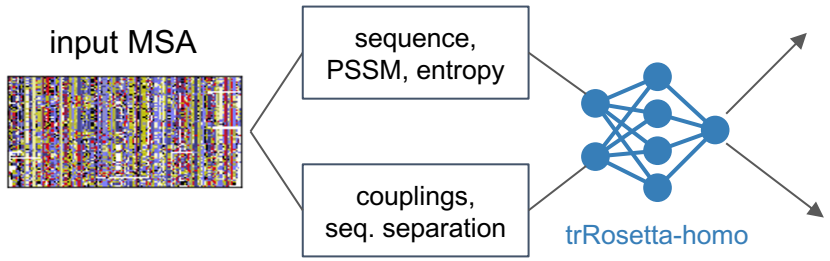
Inter-chain contacts

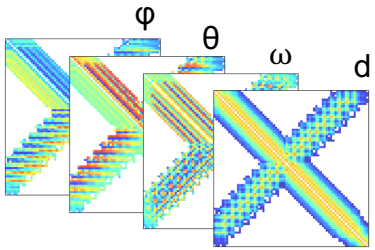
Bounded or sigmoid contact restraints
(P > 0.9; bounded / P < 0.9; sigmoid)



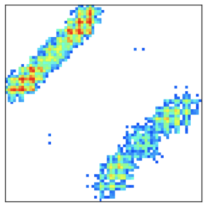
DL-based inter-chain contact & distance prediction

For homo-oligomers



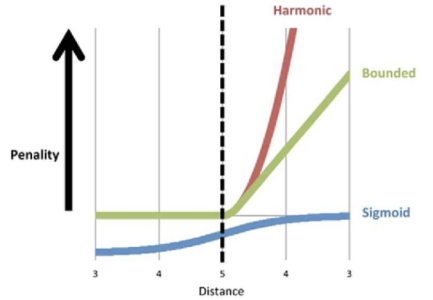


Intra-chain predictions

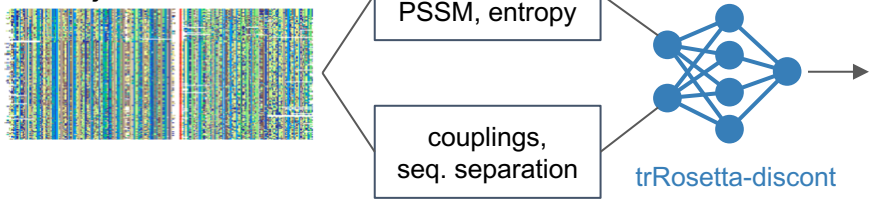


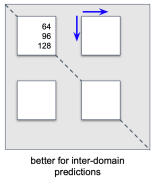
Inter-chain contacts

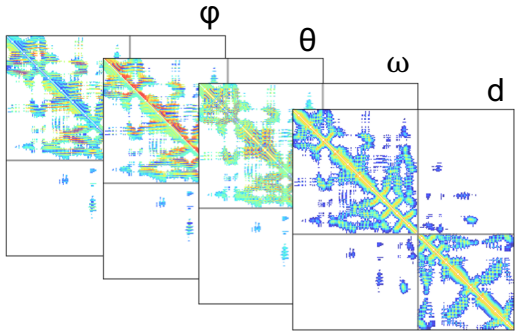
Bounded or sigmoid contact restraints
(P > 0.9; bounded / P < 0.9; sigmoid)



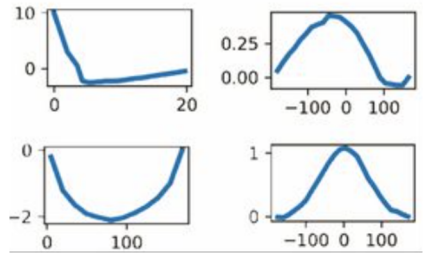
For hetero-oligomers







binned predictions @ d<20Å for entire complex



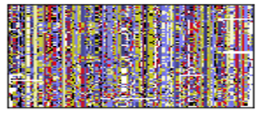
Smoothened distance/angle restraints

+ Gremlin prediction based on given MSA also used as additional ambiguous contact restraints

Protocols used in assembly prediction

Target sequence

hhblits, Uniclust30, metagenome DB



BAKER
BAKER-ROSETTASERVER

Subunit structures

Oligomer template search on PDB100

- ***HHsearch*** (seq-based)
- ***TM-align*** (str-based)

Inter-chain contact/distance prediction

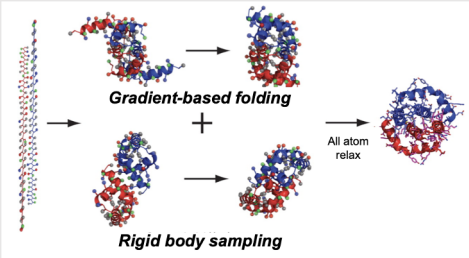
trRosetta
-variants

Gremlin

Template-based approach

- ***trRefineSymm*** (homo-)
- ***Superpose & relax*** (hetero-)

Fold-and-dock approach

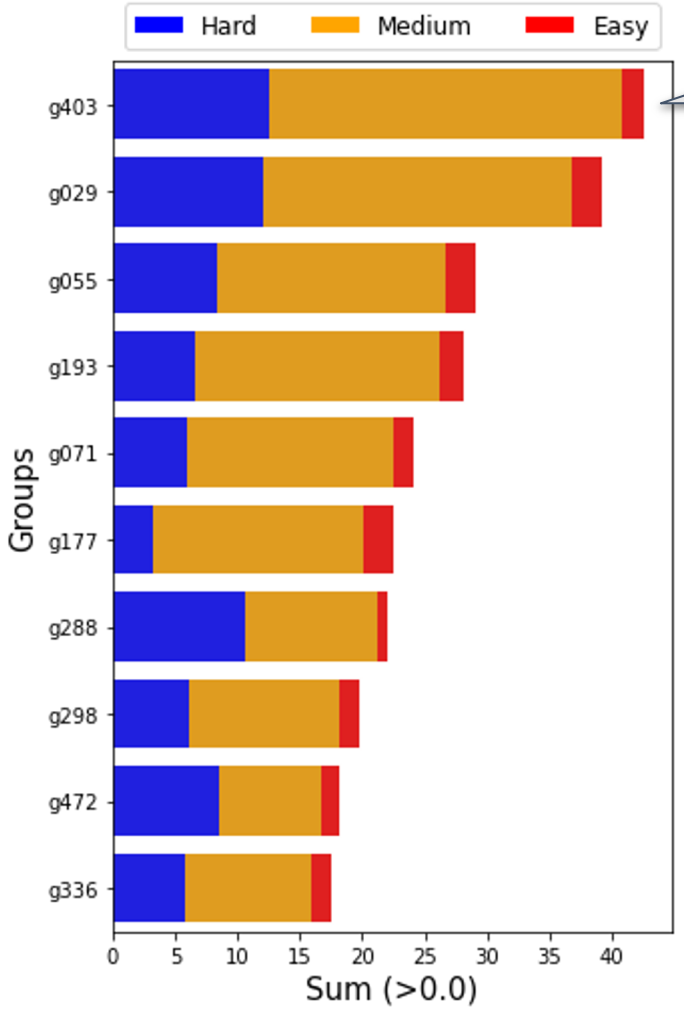


***Ab initio* docking**

- ***SymmDock*** (homo-)
- ***ZDock + RosettaDock*** (hetero-)
- ***Symmetry axes constrained docking***

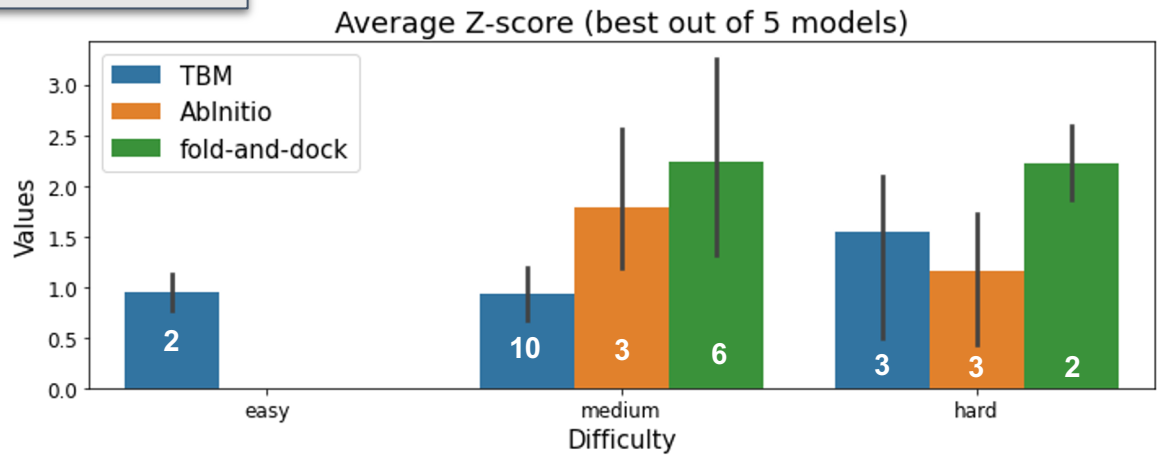
Performance of our group in CASP14

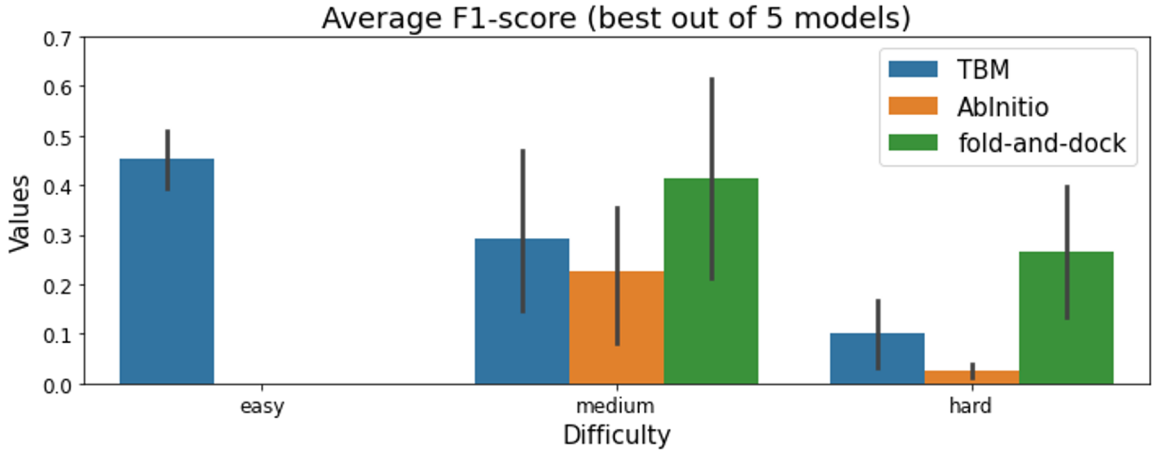
Z-score based performance



Z_Jacc + Z_F1 + Z_LDDT + Z_TM

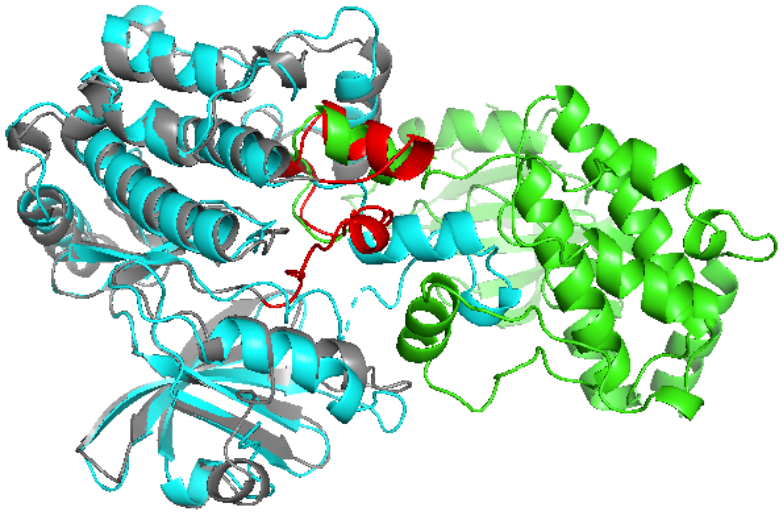
BAKER-experimental
Template-based
Ab initio docking
fold-and-dock

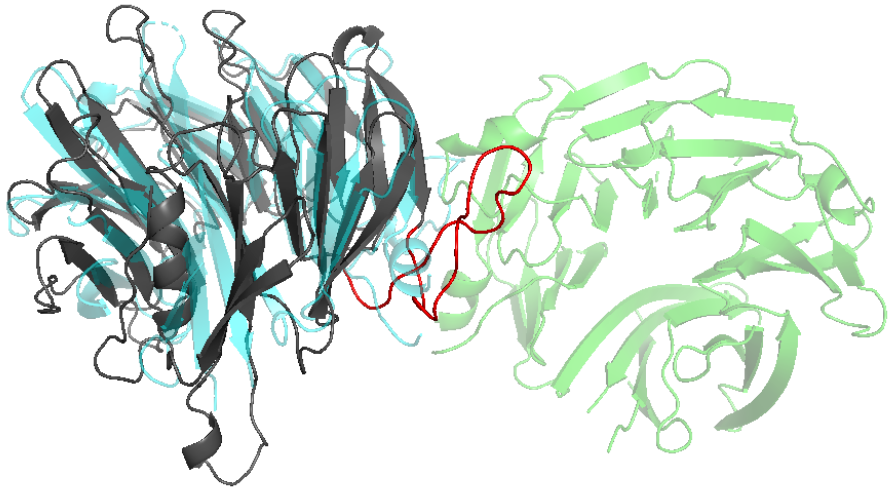




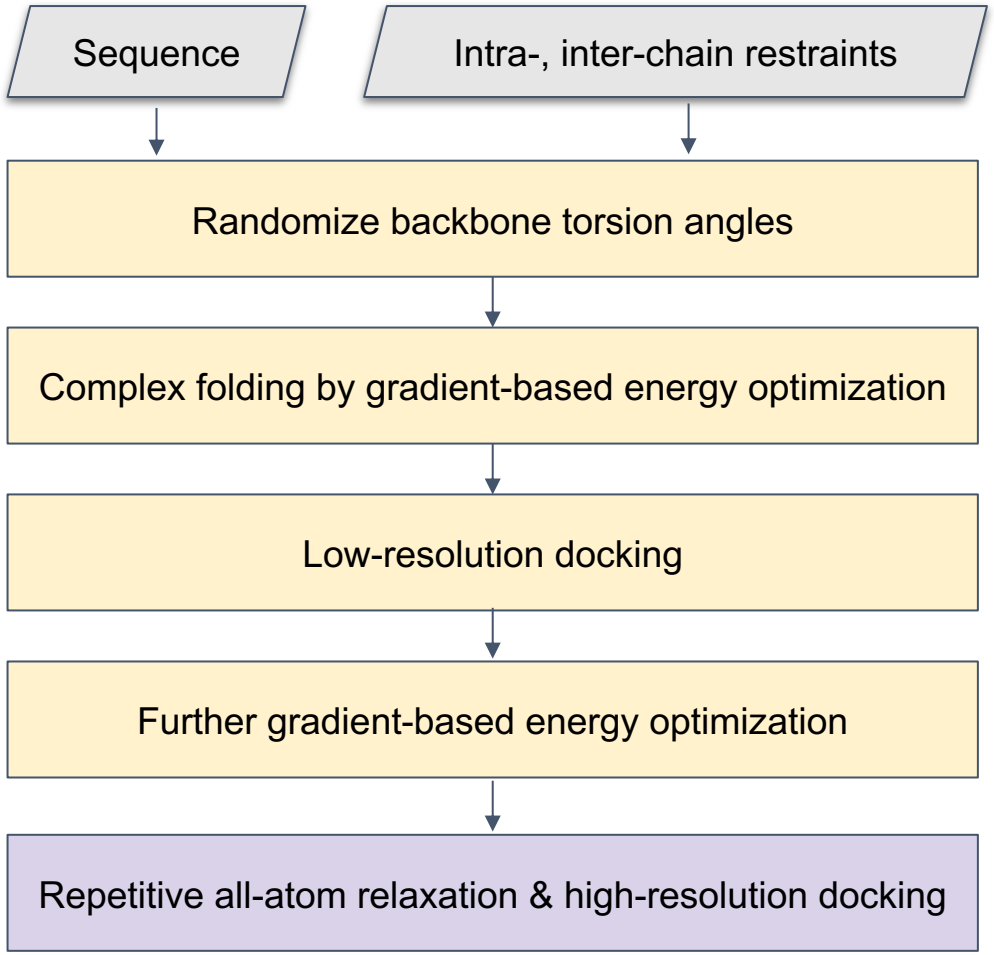
Gradient-based fold-and-dock protocol

- Why simultaneous folding and docking?
  - Local inaccuracy at the interface → hard to predict correct oligomer structures by ab initio docking
  - Induced fit effect → hard to be captured by typical docking after subunit modeling approach

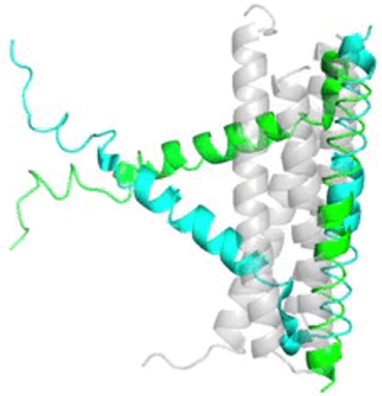




Gradient-based fold-and-dock protocol



Entire process is symmetry-aware for homo-oligomers



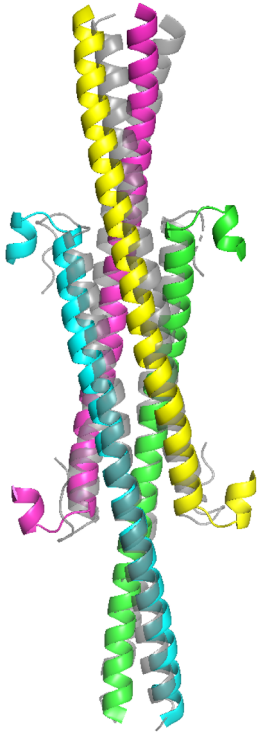
Example: T1084 (C2, medium)
oligoTM of model1: 0.91

Centroid representation

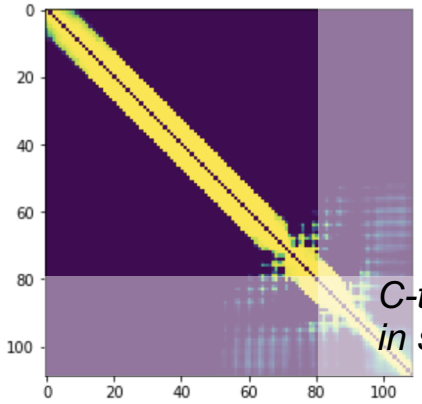
all-atom representation

Successful examples w/ fold-and-dock approach

***T1048* (D2, medium)**

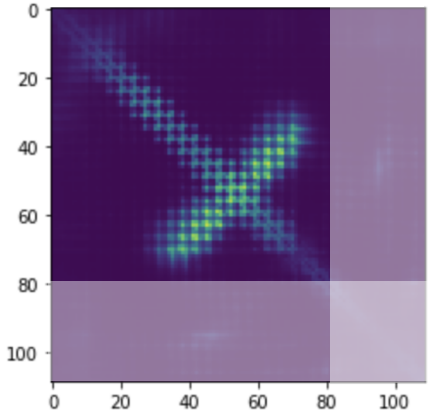


Intra-chain contacts



C-term is missing in structure

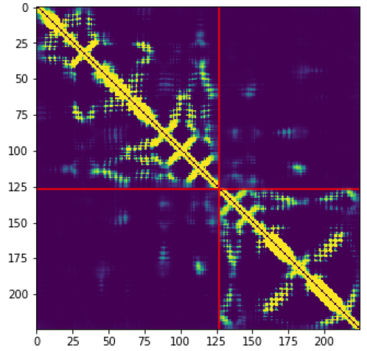
Inter-chain contacts



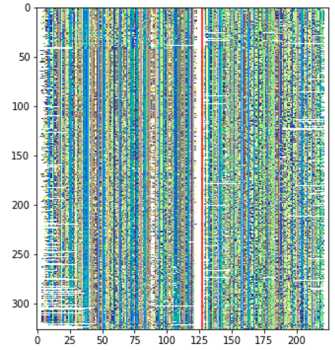
Model 3 **fold-and-dock**
(F1: 57.9 / TM: 0.828)
Next best group - (F1: 16.1 / TM: 0.326)

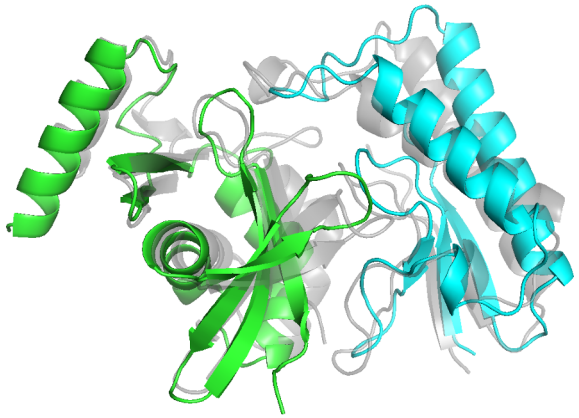
***H1065* (A2, hard)**

trRosetta-discont



Joint MSA

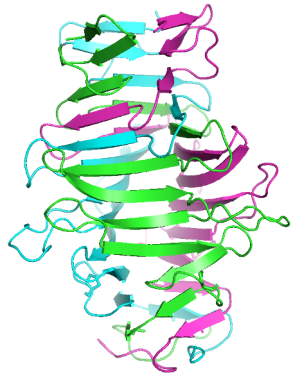




Model 1 **fold-and-dock**
(F1: 39.7 / TM: 0.792)

What we did relatively well, but could do better in the future

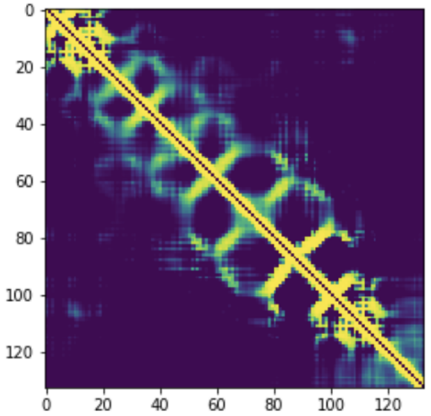
T1080 (C3, hard)



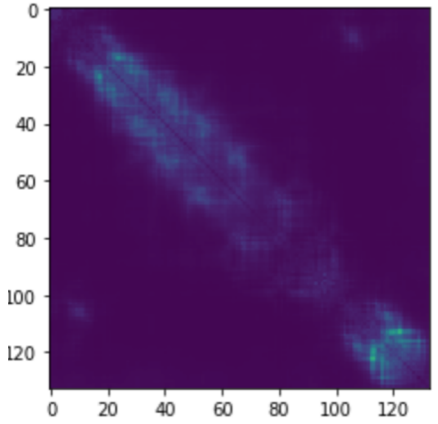
Experimental structure

Model 4 **fold-and-dock**
(Jacc: 0.38 / F1: 13.3 / TM: 0.611)

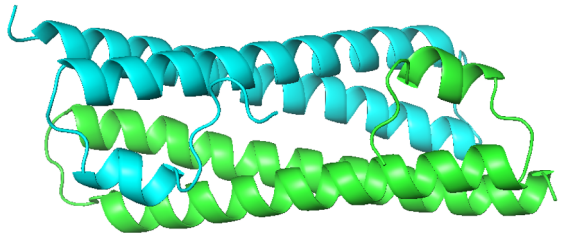
Intra-chain contacts



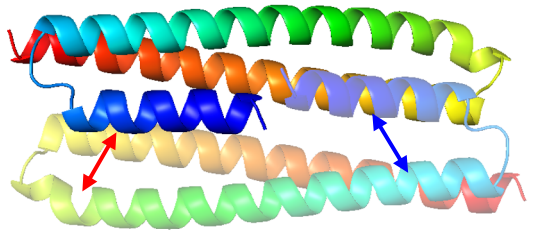
Inter-chain contacts



T1083 (C2, medium)

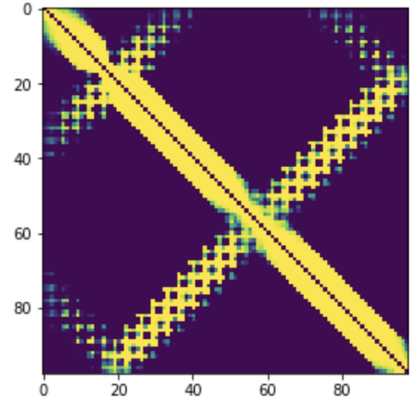


Experimental structure

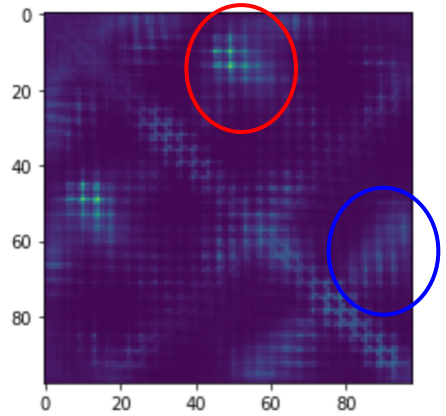


Model 1 **fold-and-dock**
(Jacc: 0.69 / F1: 22.5 / TM: 0.627)

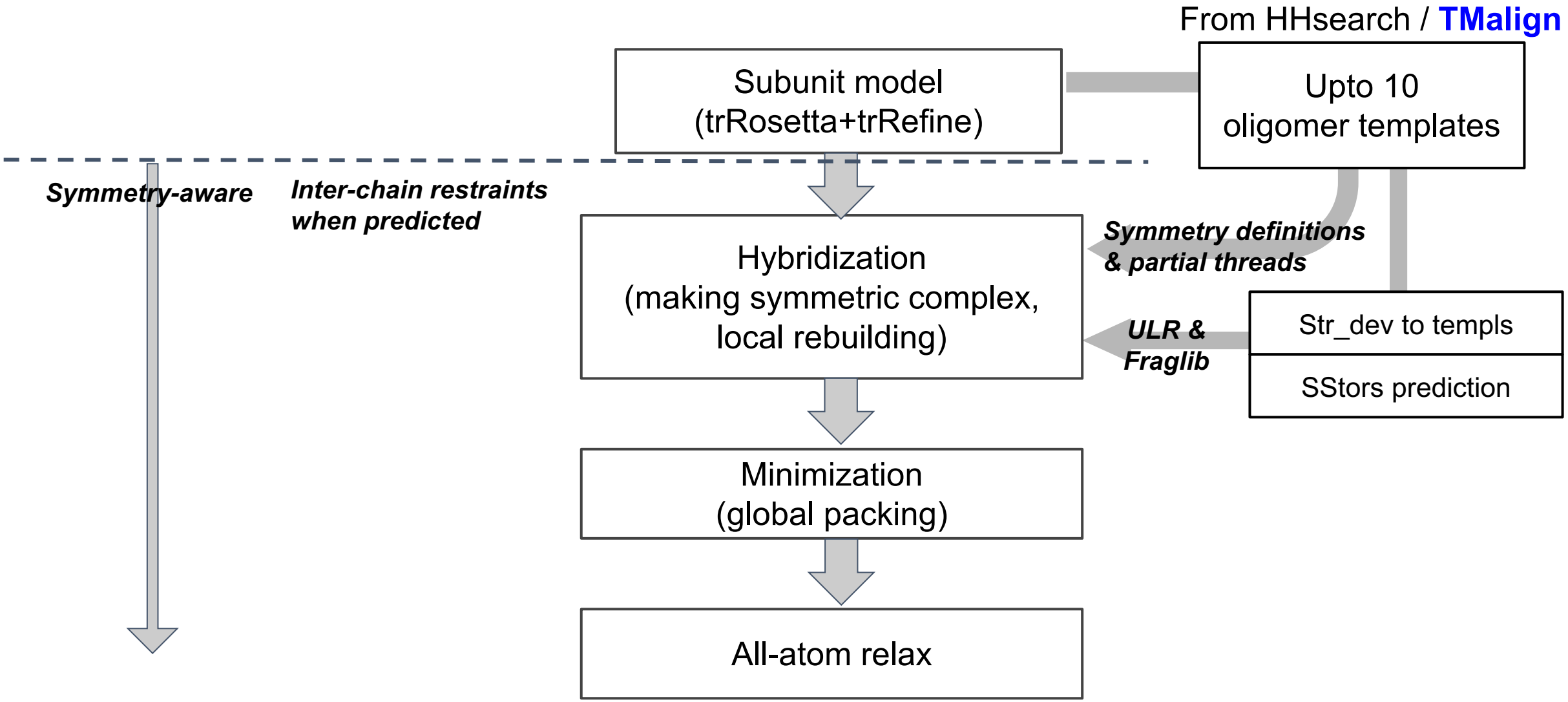
Intra-chain contacts



Inter-chain contacts

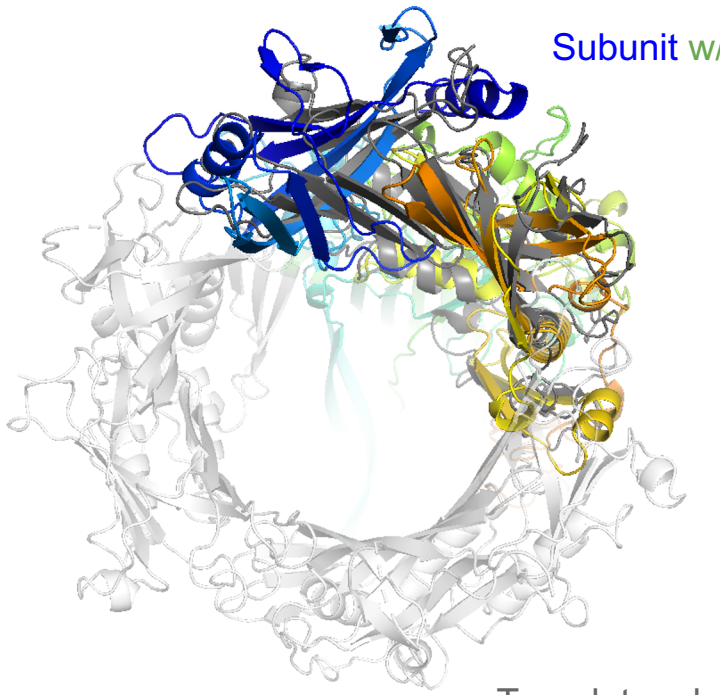


trRefineSymm: Refining trRosetta models with symmetry from templates



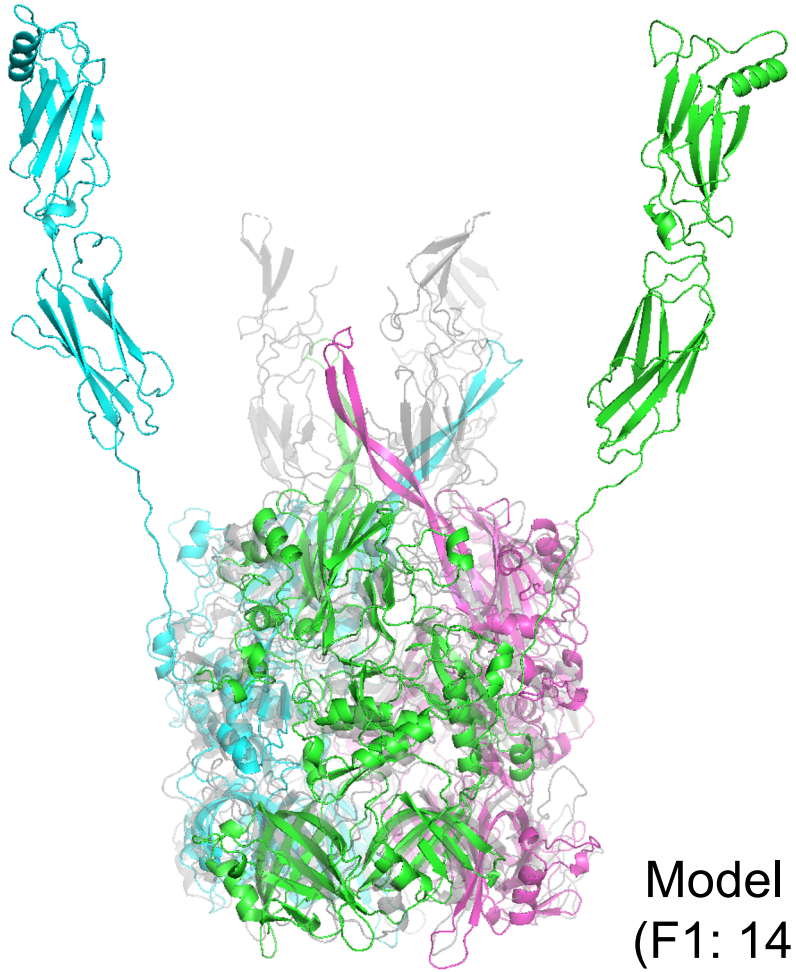
Examples: found distant templates

T1061 **(C3, hard)**



Subunit w/ metagenome search

Templates detected by TM-align
(PDB ID: 3cdd, 4mtk)



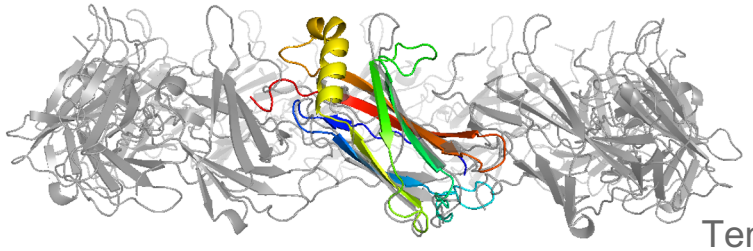
C-term domains
are totally off
(not covered by template /
no predicted contacts)

Model 3 **trRefineSymm**
(F1: 14.3 / TM: 0.712)

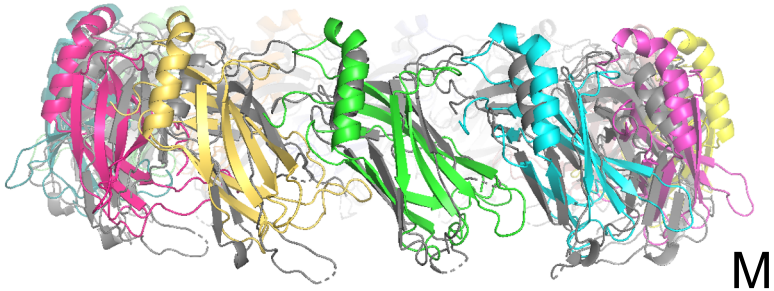
Hard to make good interface contacts by refinement due to its size & qualities of initial subunit and templates

Examples: found distant templates

H1060v4 **(C12, medium)**

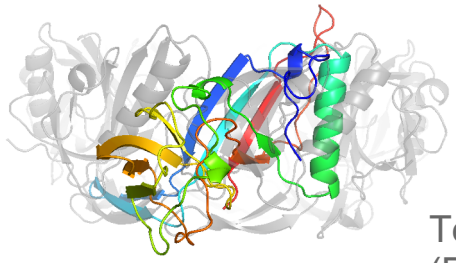


Templates detected by TM-align
(PDB ID: 6v8i, 4v96)

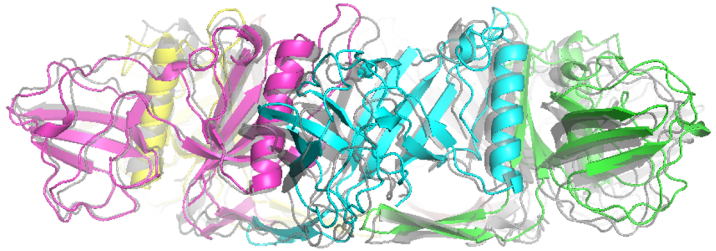


Model 1 **trRefineSymm**
(F1: 22.3 / TM: 0.746)

H1060v5 **(D6, medium)**



Template detected by TM-align
(PDB ID: 4v96)

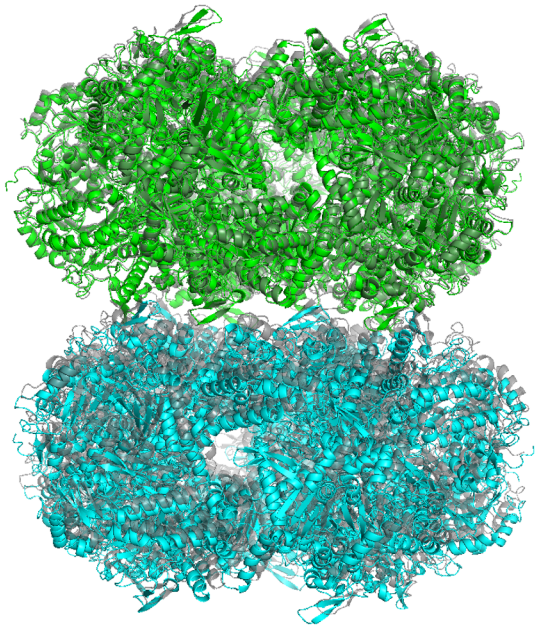


Model 1 **trRefineSymm**
(F1: 47.6 / TM: 0.951)

Hard to make good interface contacts by refinement due to its size & qualities of initial subunit and templates

ab initio docking w/ reasonable building blocks

H1081 **(D5+D5, medium)**



Template-based (2vyc, 5xx1)

Model 1
(F1: 35.3 / TM: 0.970)

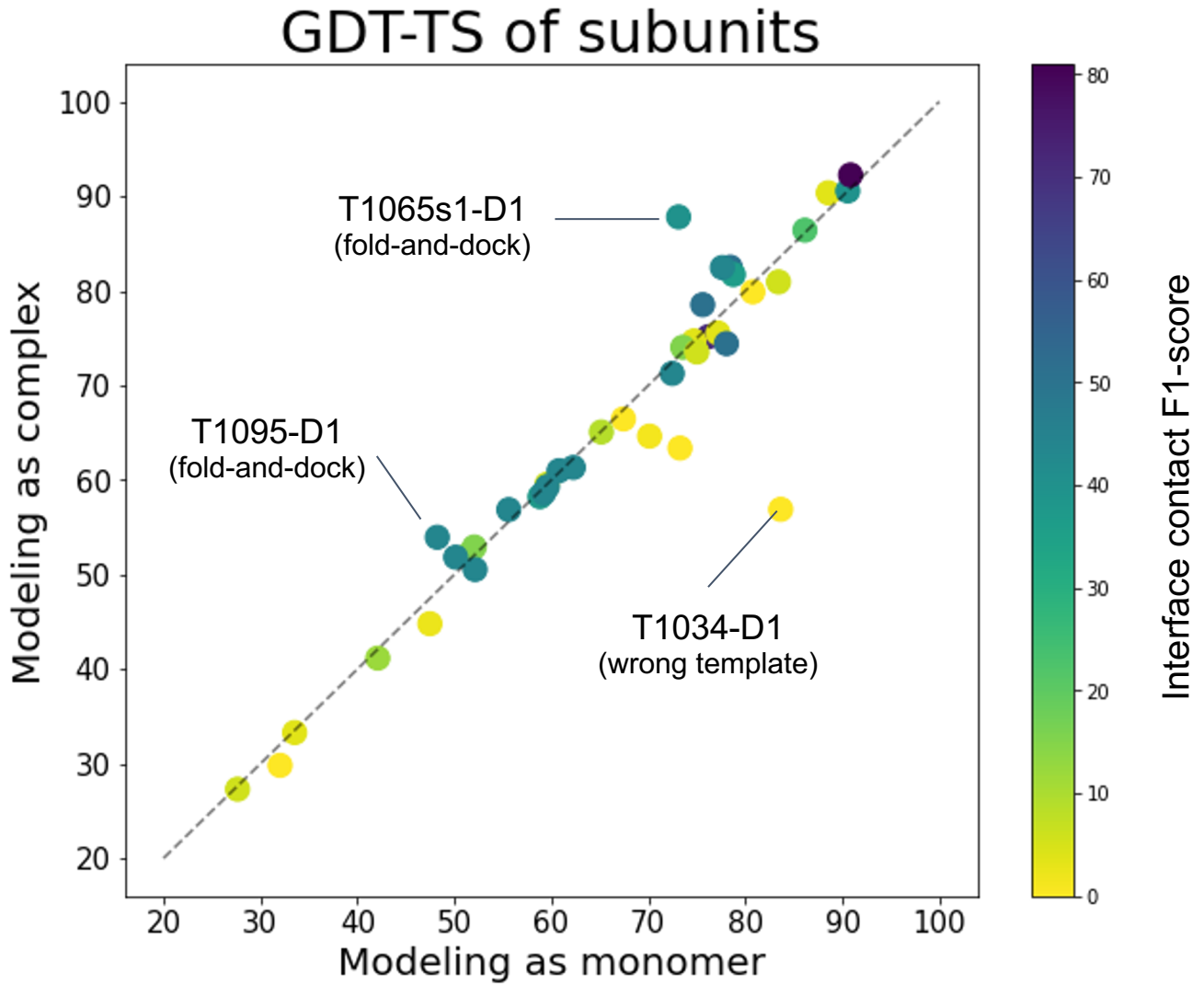
Stacking decamers by
rigid-body sampling along symmetry axes

T1078 **(C2, hard)**

Model 5 SymmDock
(F1: 24.7 / TM: 0.671)

Could do better if we use better subunit models (GDT-TS of our subunit: 63.9; best server model: 81.2)

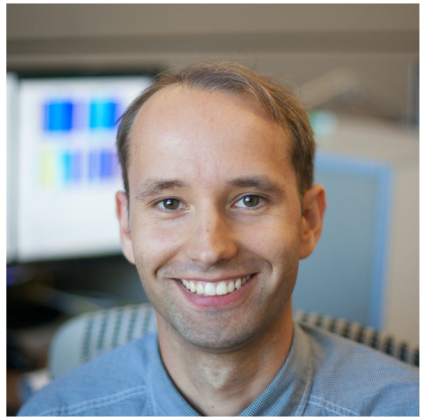
Did assembly modeling improve subunit qualities?



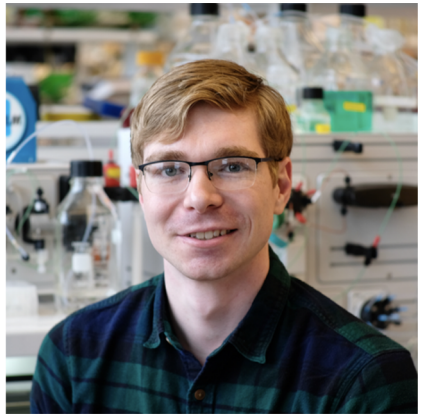
Summary

- New simultaneous fold-and-dock approach
  - Reasonable complex structures were predicted even if there were no templates
  - Inter-chain contact/distance prediction is the key
  - In some cases, it improved subunit qualities too
- Progress in template-based approach
  - Mainly thanks to the better subunit structures with deep learning
  - distant templates were detected based on structural similarity
- Ab initio docking was rarely successful
  - Only when we have good subunit structures, restrictions in sampling space, etc.
- Huge progress on TS modeling (AlphaFold2) → What would happen next time?
  - Higher F1-score / Easier to make complex structure by *ab initio* docking / Even end-to-end complex modeling?

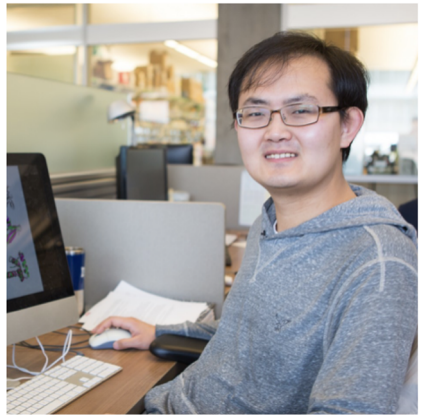
Acknowledgements



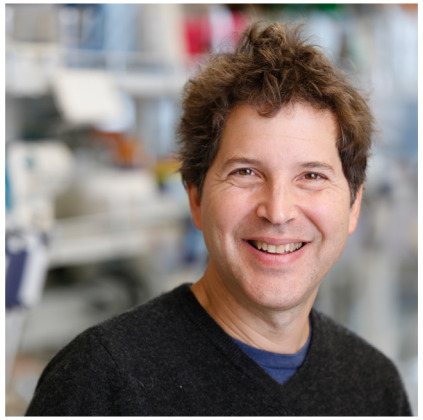
Ivan Anishchenko



Ian Humphrey



Hahnbeom Park



David Baker

CASP14/CAPRI organizers, target providers, assessors, and participants!

INSTITUTE FOR Protein Design
UNIVERSITY of WASHINGTON