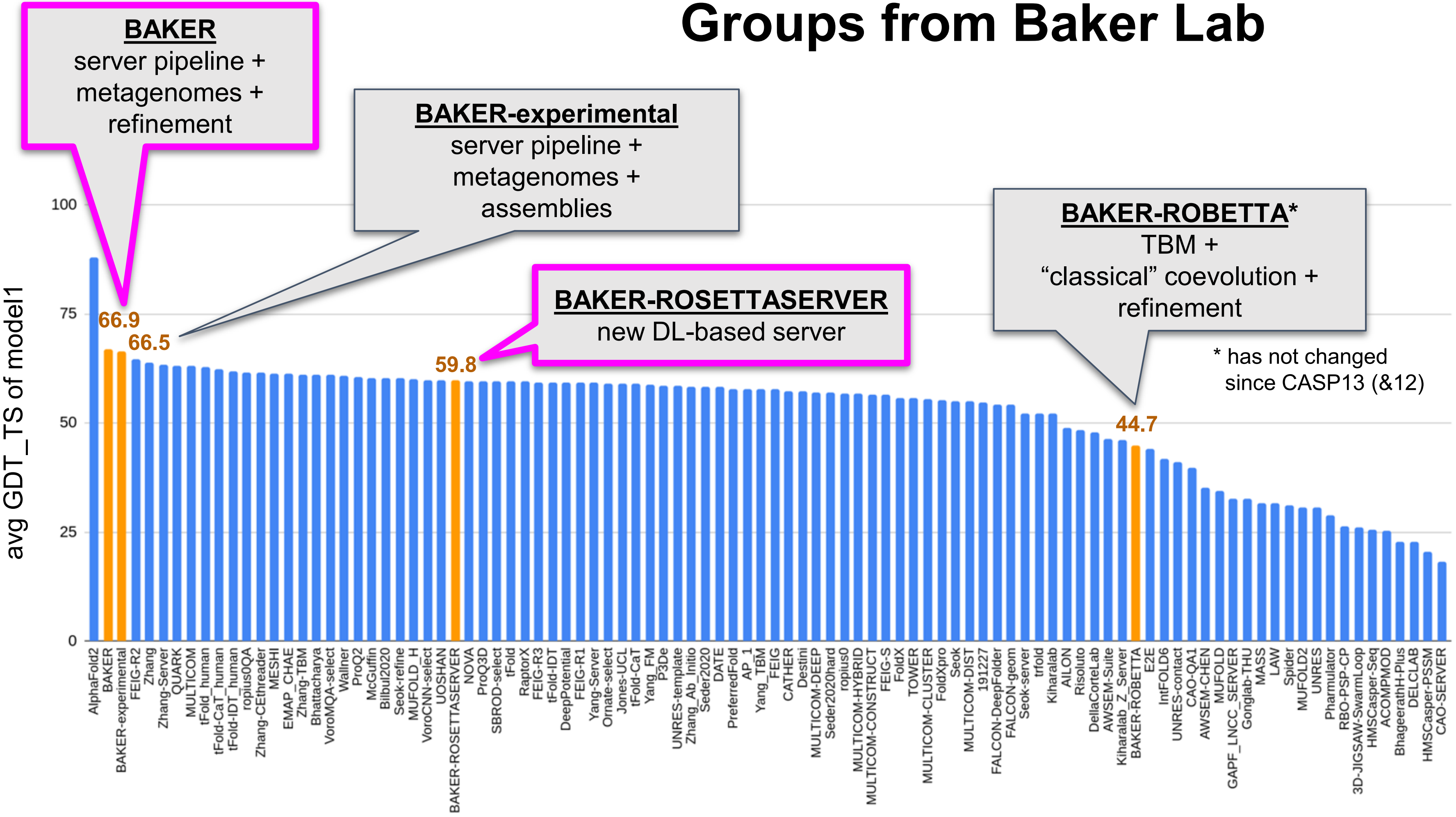
Tertiary structure (TS) prediction and refinement from Baker groups

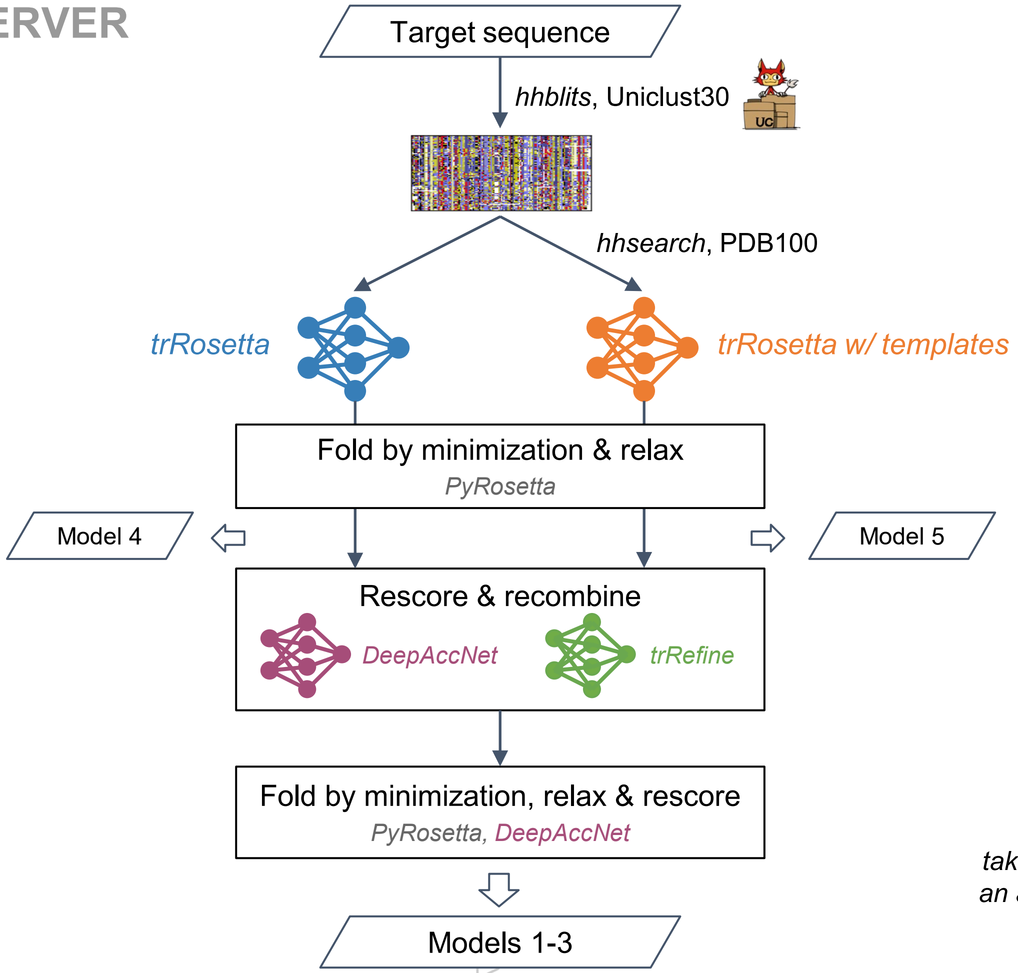
Ivan Anishchenko, Minkyung Baek, Hahnbeom Park, Justas Dauparas, Naozumi Hiranuma, Sanaa Mansoor, Ian Humphrey, and David Baker

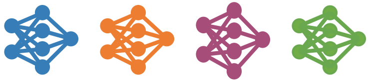
INSTITUTE FOR Protein Design
UNIVERSITY of WASHINGTON

Groups from Baker Lab



BAKER-ROSETTASERVER

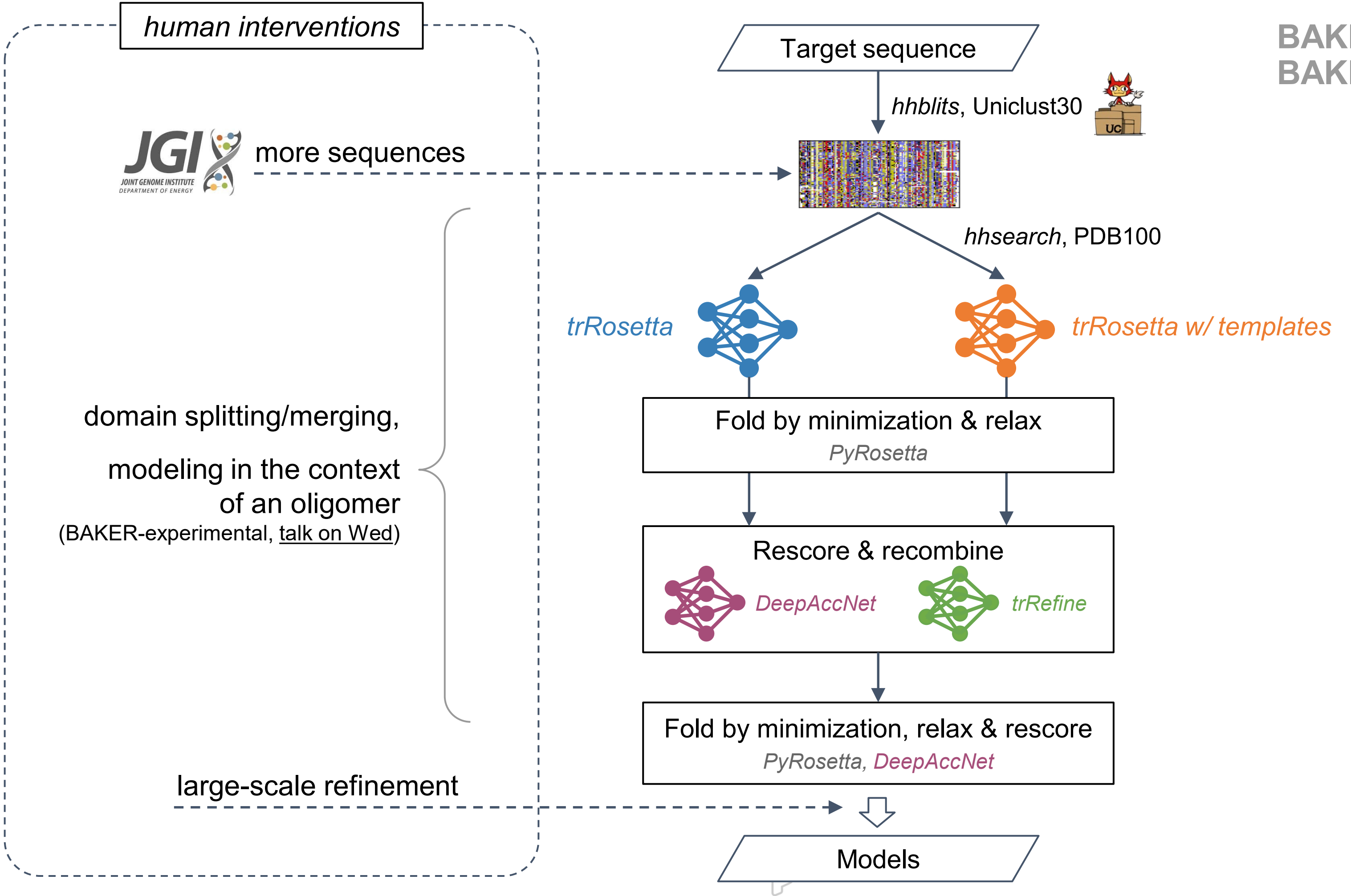




2D ResNets

takes several hours for an average size protein

BAKER
BAKER-experimental

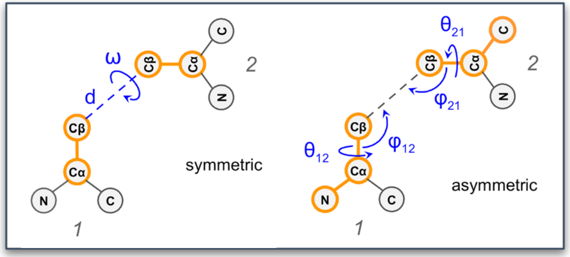


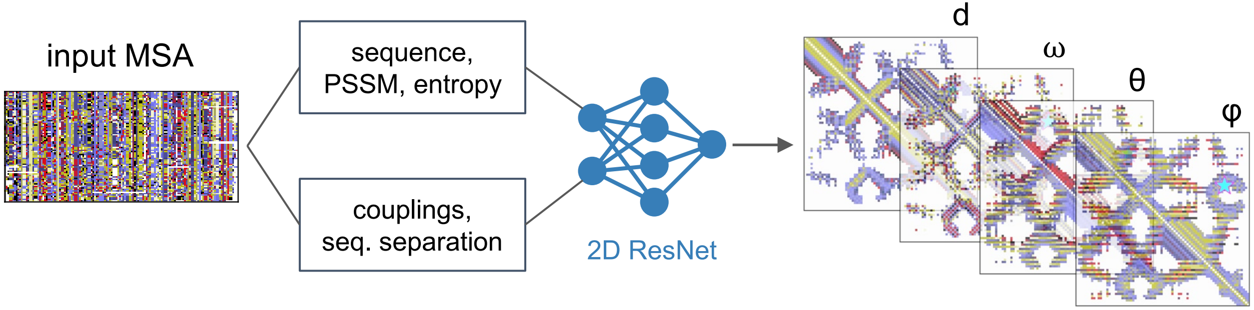
BAKER-ROSETTASERVER

trRosetta

transform-restrained Rosetta

residue-residue geometries (6 DoF)



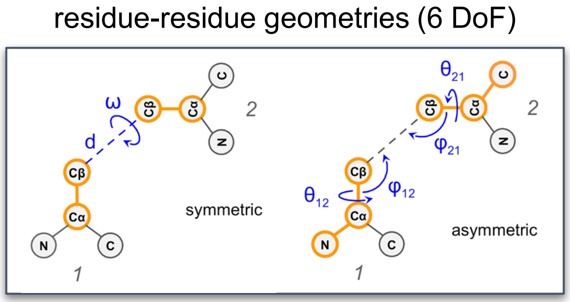


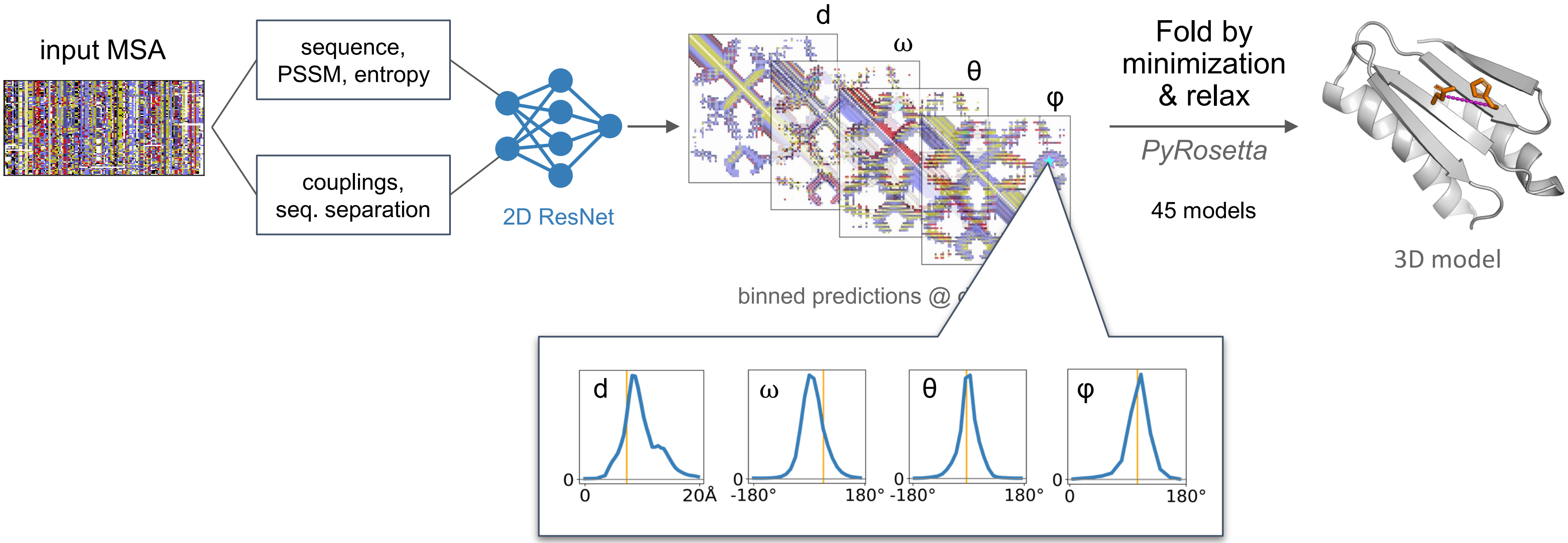
binned predictions @ d<20Å

J Yang, I Anishchenko, H Park, Z Peng, S Ovchinnikov, D Baker,
Improved protein structure prediction using predicted interresidue orientations,
PNAS, 117: 1496-1503 (2020)

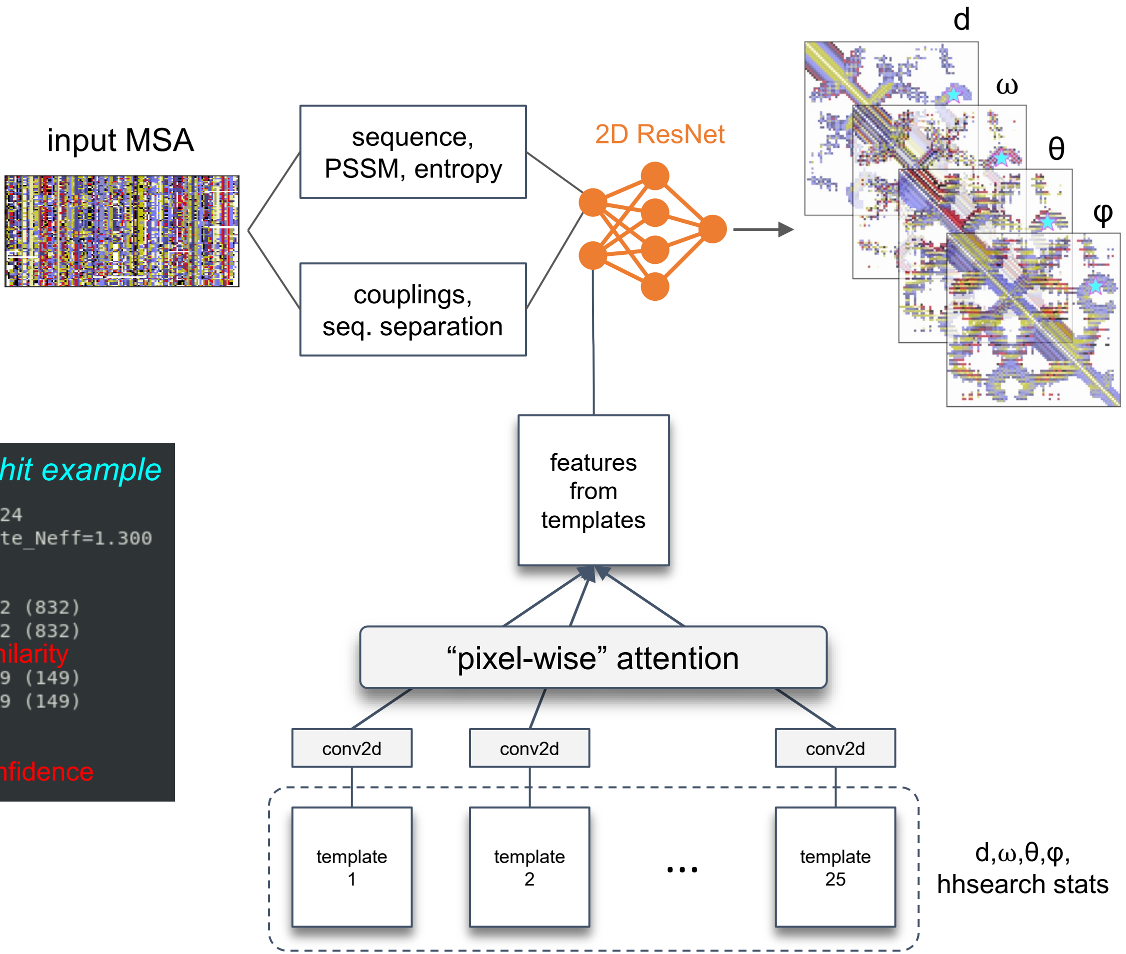
trRosetta

transform-restrained Rosetta





trRosetta
with templates



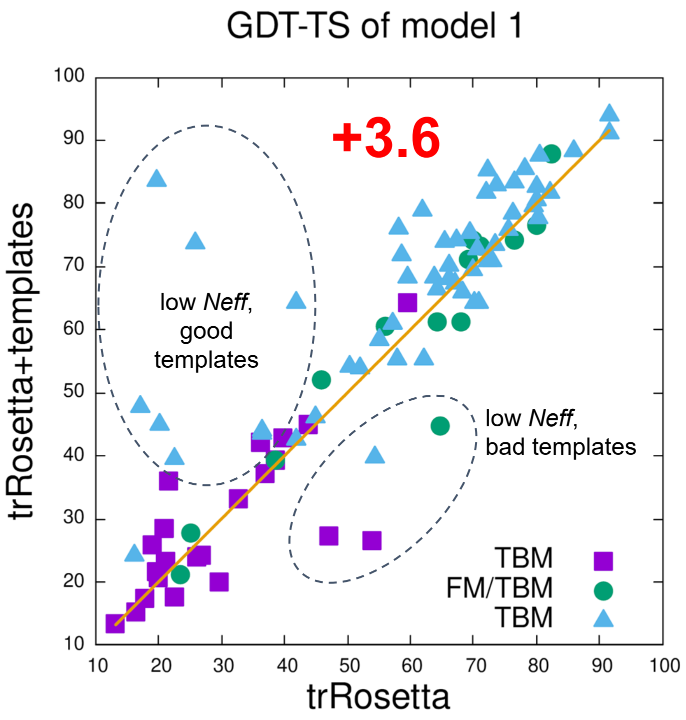
hhsearch hit example

```
No 470
>1o22_A
Probab=5.04  E-value=1.3e+03  Score=26.24  Aligned_cols=24
Identities=33%  Similarity=0.667  Sum_probs=10.7  Template_Neff=1.300

Q ss_pred                ceecCcE------ECHHHHHHHHHcCCCC
Q T1052              809 GLYIGAK------VNADFIAQFIKSKGWGG  832 (832)
Q Consensus          809 ~l~~g~~------vn~~f~~~fi~~kgw~~  832 (832)
                         --|+-|+      +-+||+.-|..+|||..
T Consensus          110 vpyssvk~knr~e~v~efmkyffeskgw~p  139 (149)
T 1o22_A             110 VPYSSVKKKNRNELVEEFMKYFFESKGWNP  139 (149)
T ss_dssp                EEGGGSTTCCHHHHHHHHHHHHHHHHHTCCG
T ss_pred                EEhHHccccCHHHHHHHHHHHHHHHhcCCCC
Confidence               0011111      23566666666777741
```

similarity

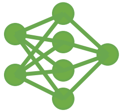
confidence



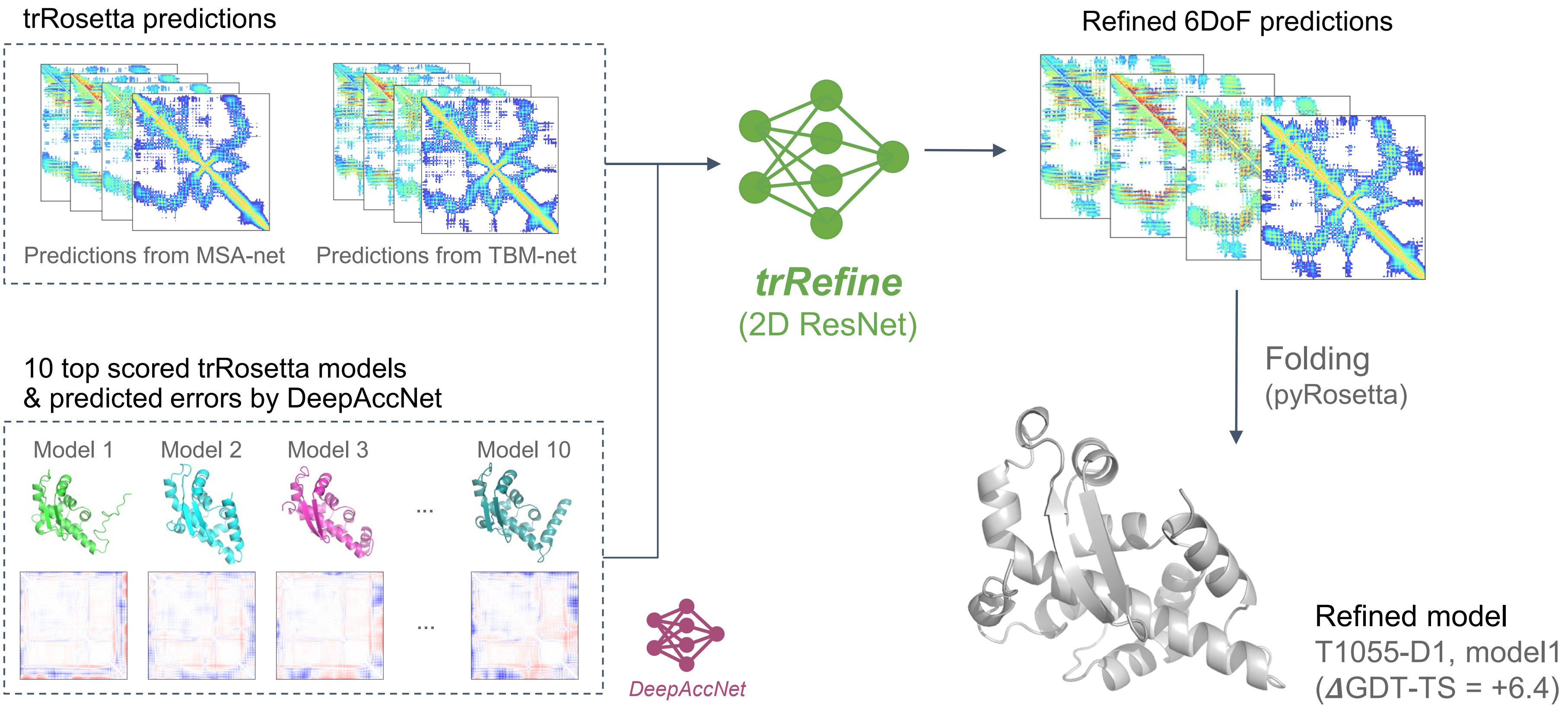
- network with templates typically gives better predictions than the MSA-only net
- in low *Neff* regime unrelated templates may misguide predictions

merge predictions from both networks

trRefine

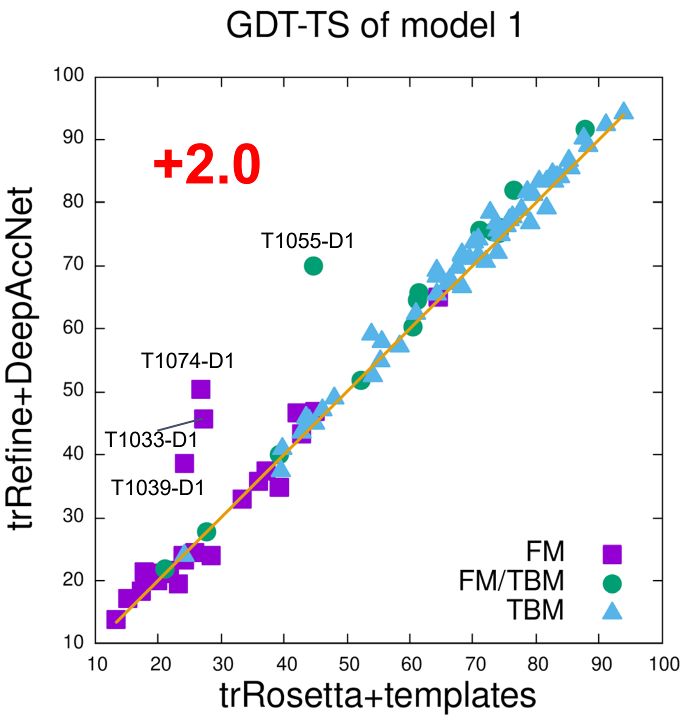


Model rescoring and recombination

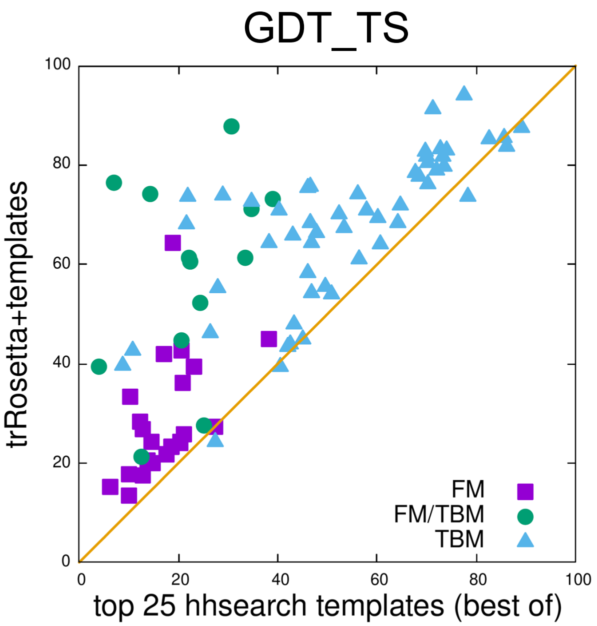


EMA roundtable on Thur

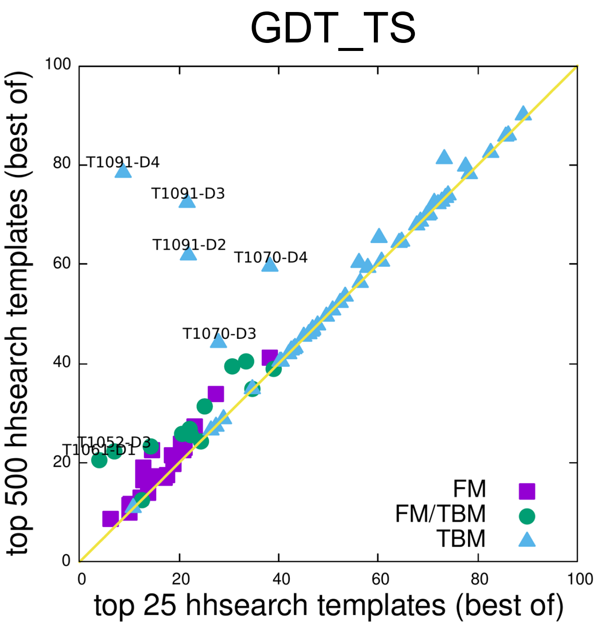
Recombining models with *trRefine* generally improves their quality



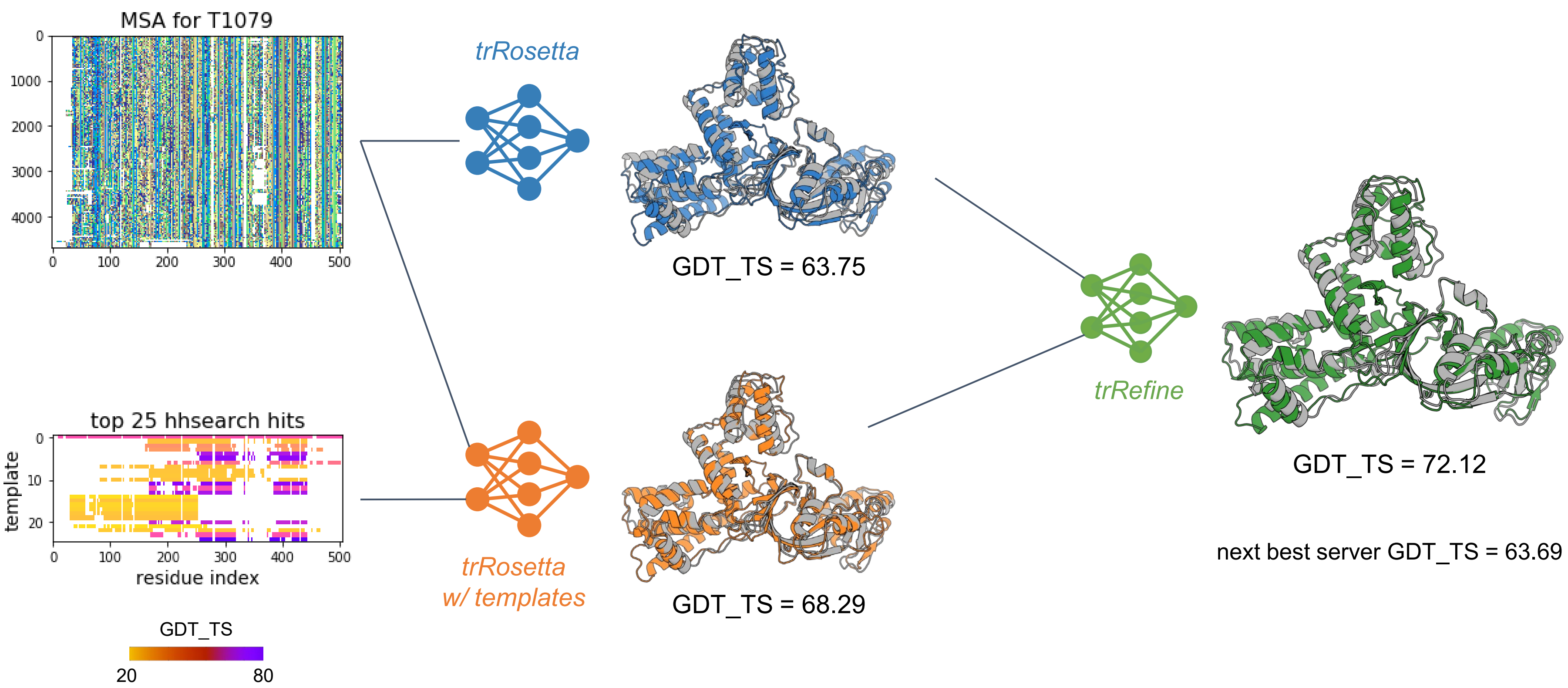
Submitted models are generally of better quality than the best selected *hhsearch* template ...



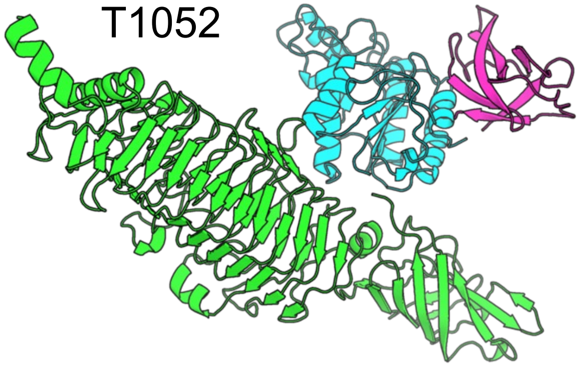
… but some good templates were missed

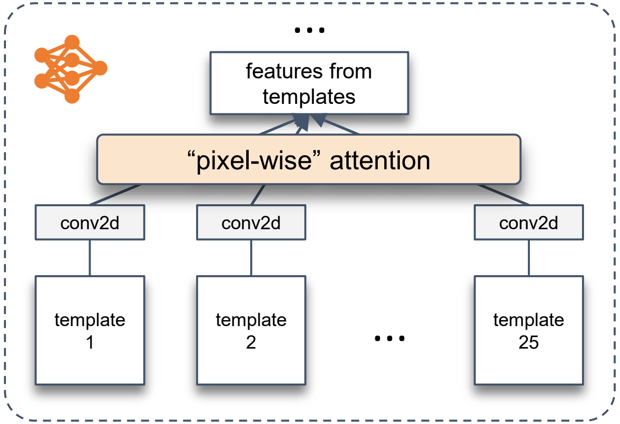


Prediction workflow by the example of T1079

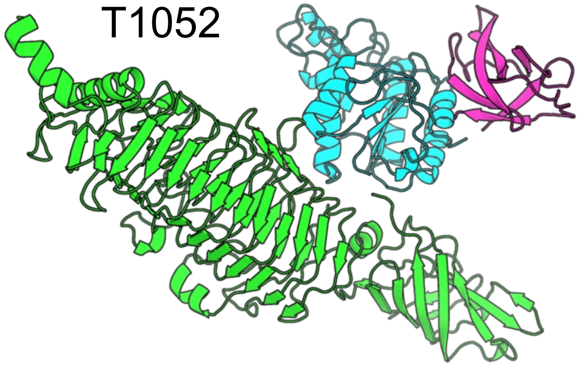


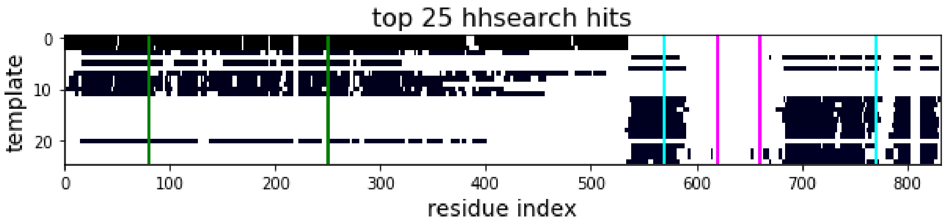
Recombination of templates

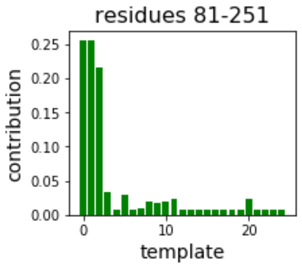


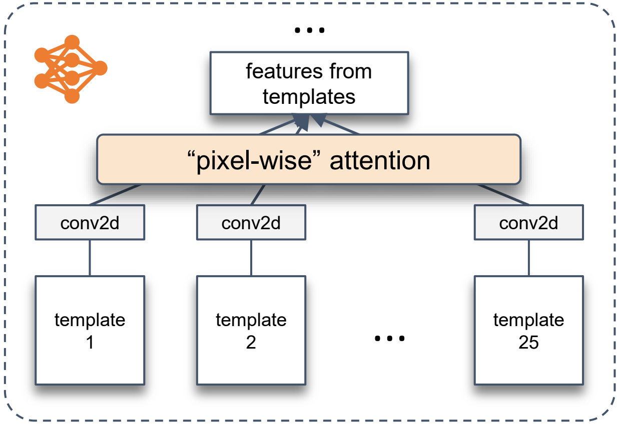


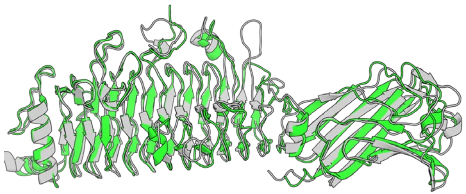
Recombination of templates





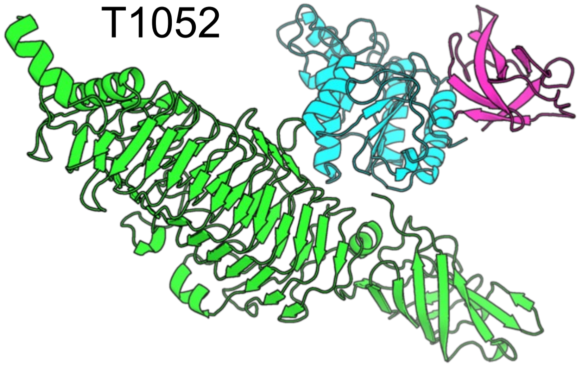


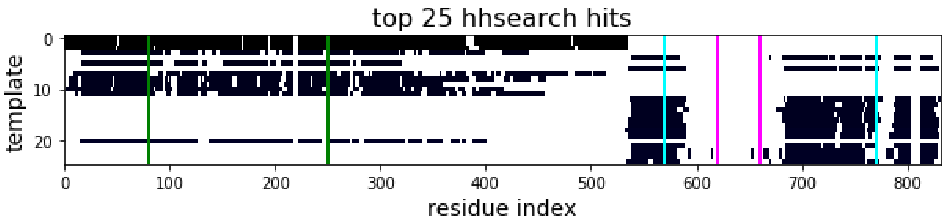


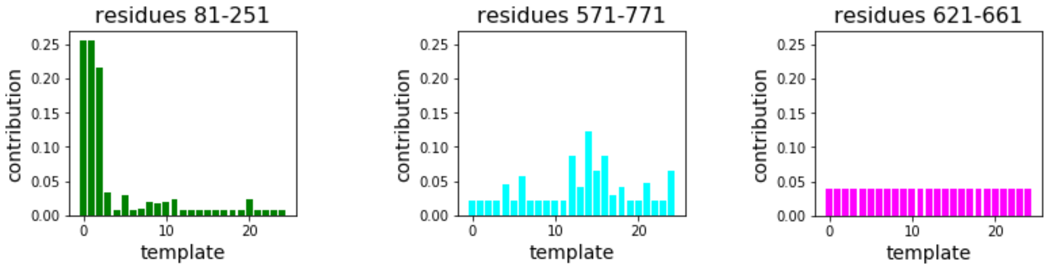


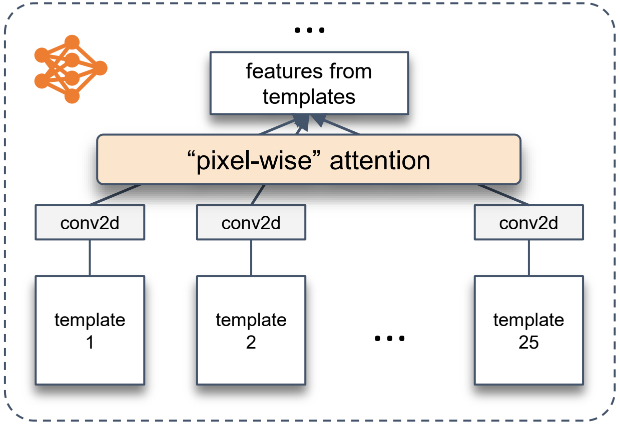
GDT_TS = 75.60
best tmplt = 78.15

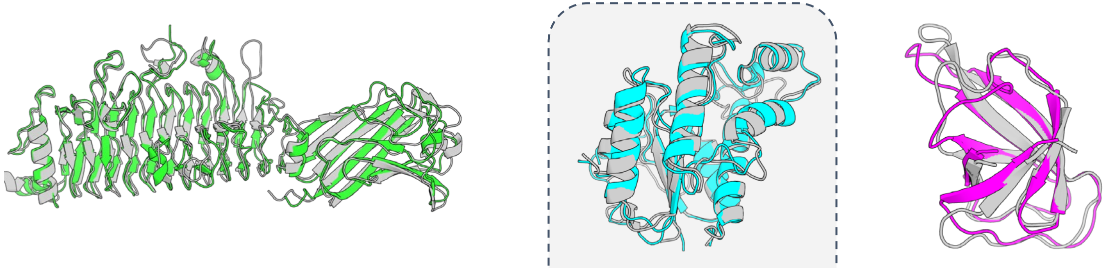
Recombination of templates







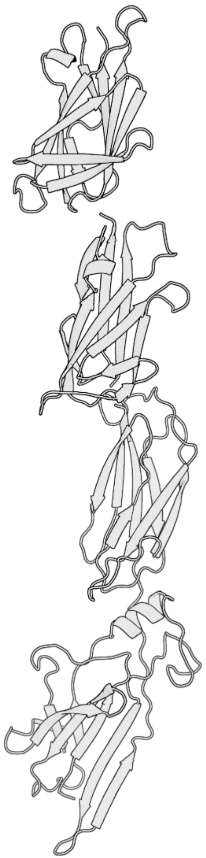


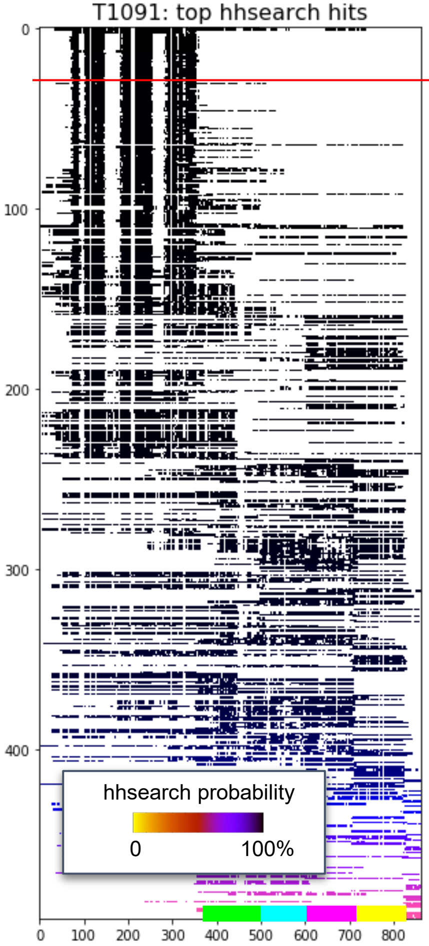


GDT_TS =	75.60	78.05	78.44
best tmplt =	78.15	52.23	6.88

next best ΔGDT_TS = -4.11

T1091:
templates were missed





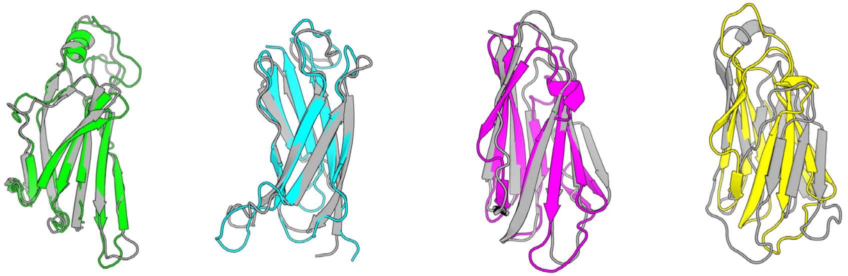
no good templates in top 25

plenty of templates down the list

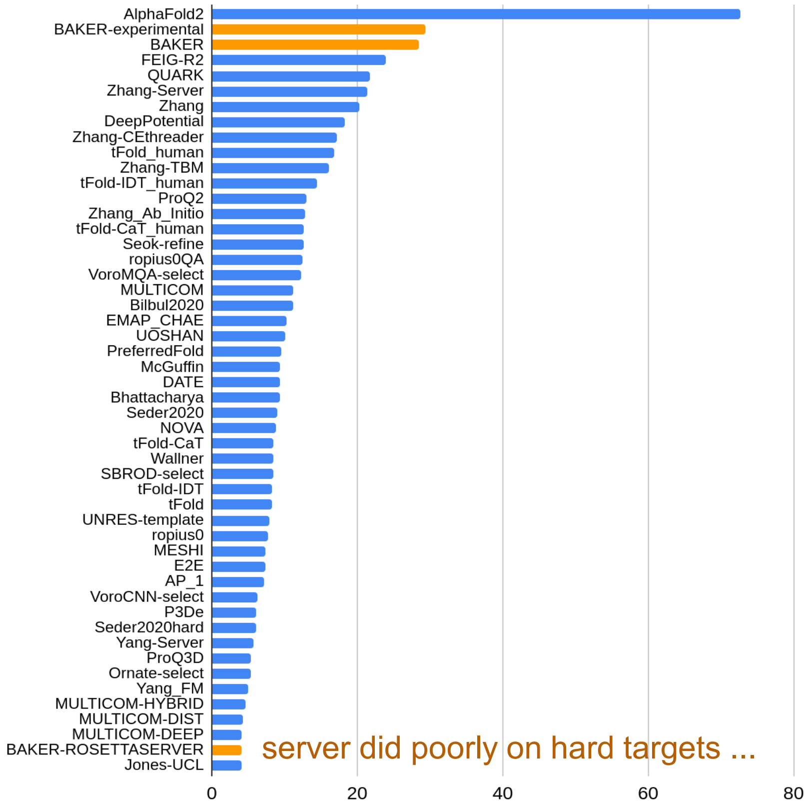
used for network predictions

	D1	D2	D3	D4
best tmplt, top25	46.04	21.96	21.70	8.71
best tmplt, top500	46.04	61.92	72.41	78.57
model1	**76.08**	**77.34**	**65.80**	**43.97**

could have been modeled better

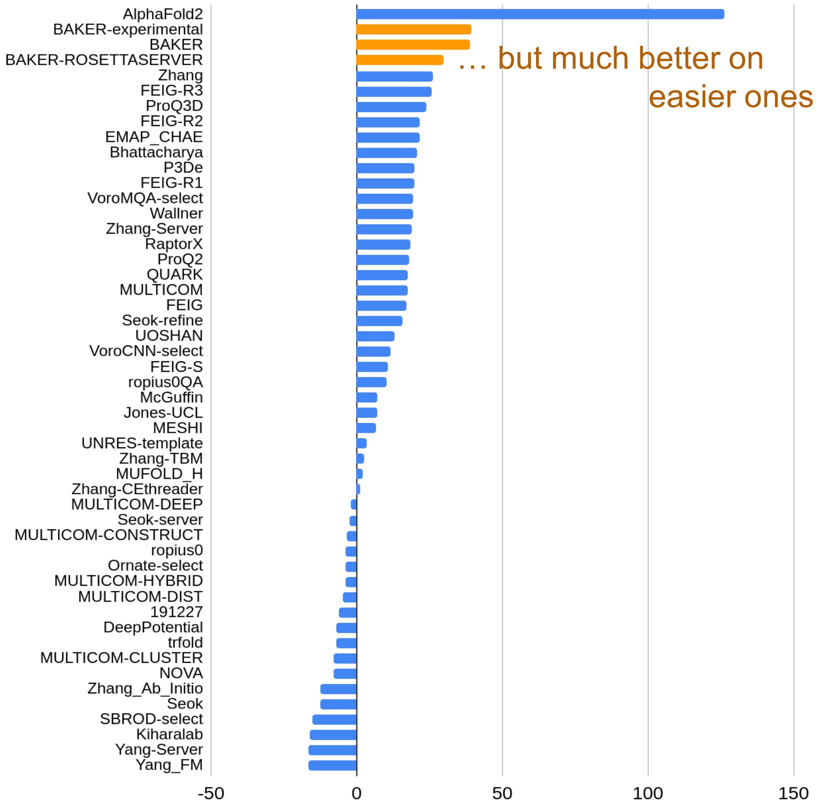


FM targets



$Z\_TS + Z\_QCS$

FM/TBM and TBM targets



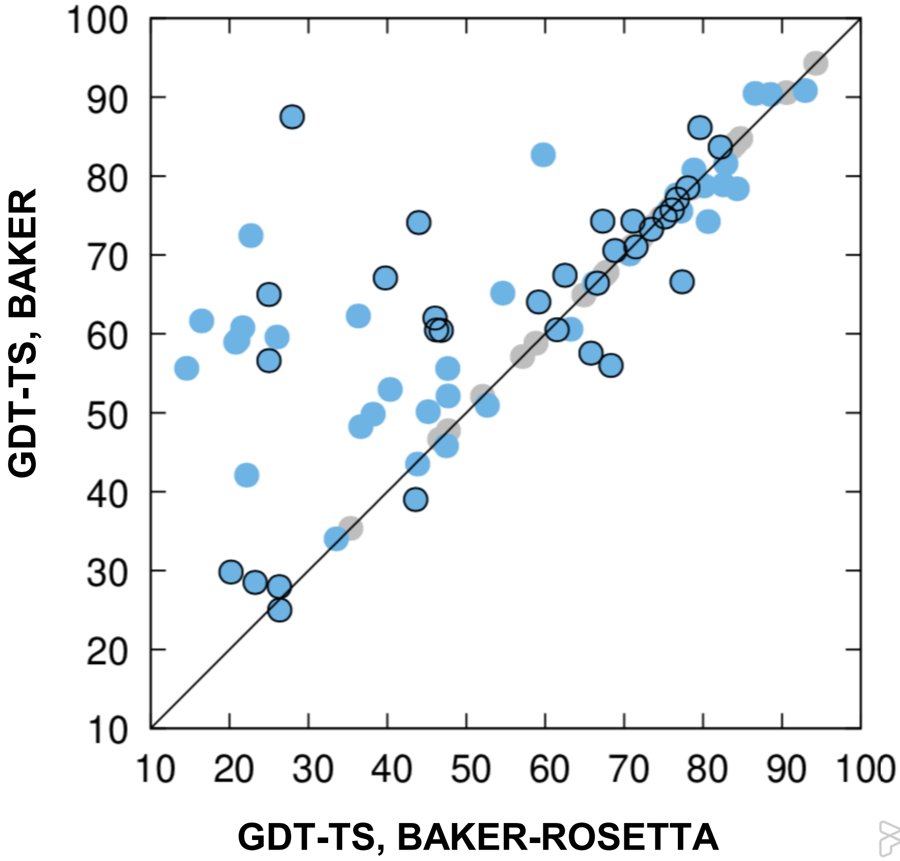
$Z\_HA + (Z\_SG + Z\_lDDT + Z\_CAD)/3 + Z\_ASE$

Server Summary

- Joint use of templates and MSA worked well
- Templates could have been selected and used better
- No good reason for not using more sequences (metagenomes)
- *trRefine* does improve model quality but not dramatically

Human Tertiary structure prediction & Refinement (BAKER)

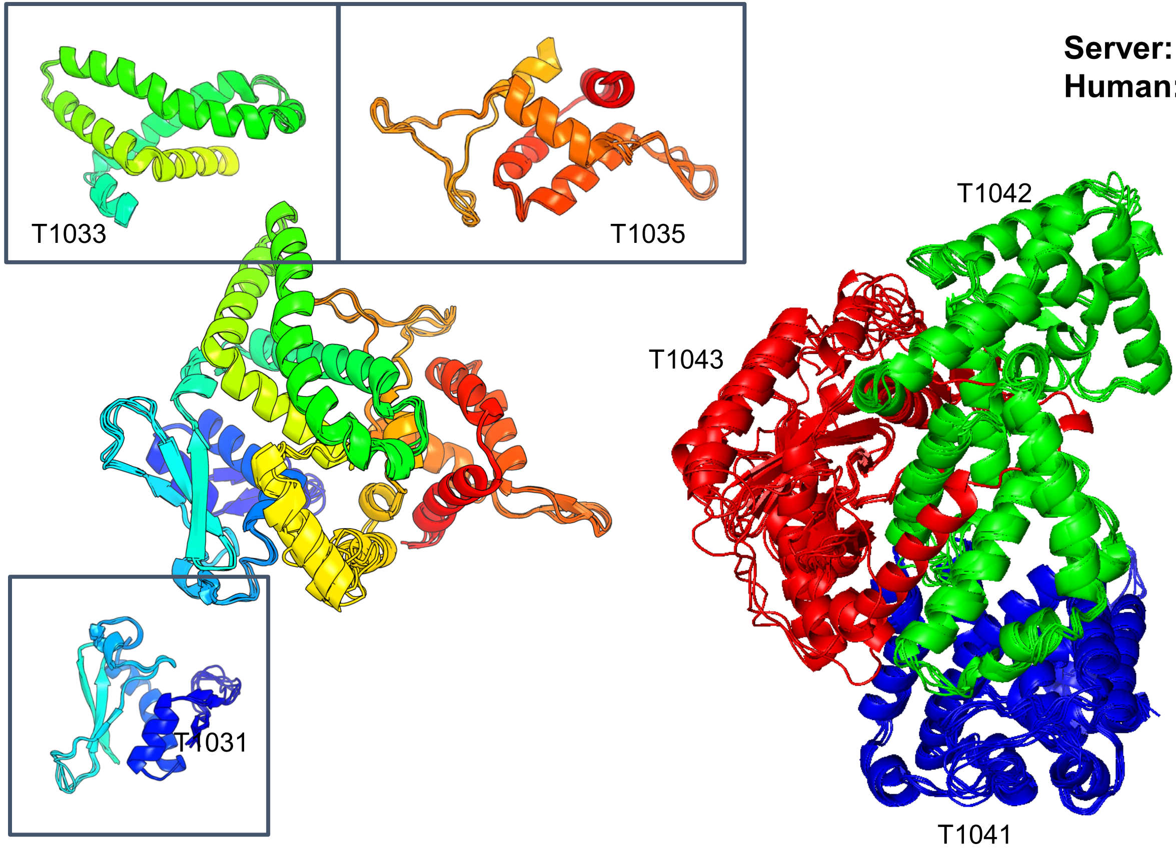
Human vs Server:
Contribution from Additional Sequence search for trRosetta

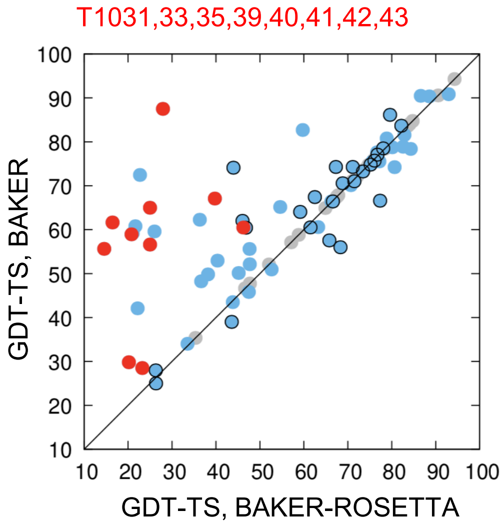


Human modeling
26 domains: Submitted as server models
67 domains: Remodeled with alternative MSAs
(open circles: starting points for refinement)

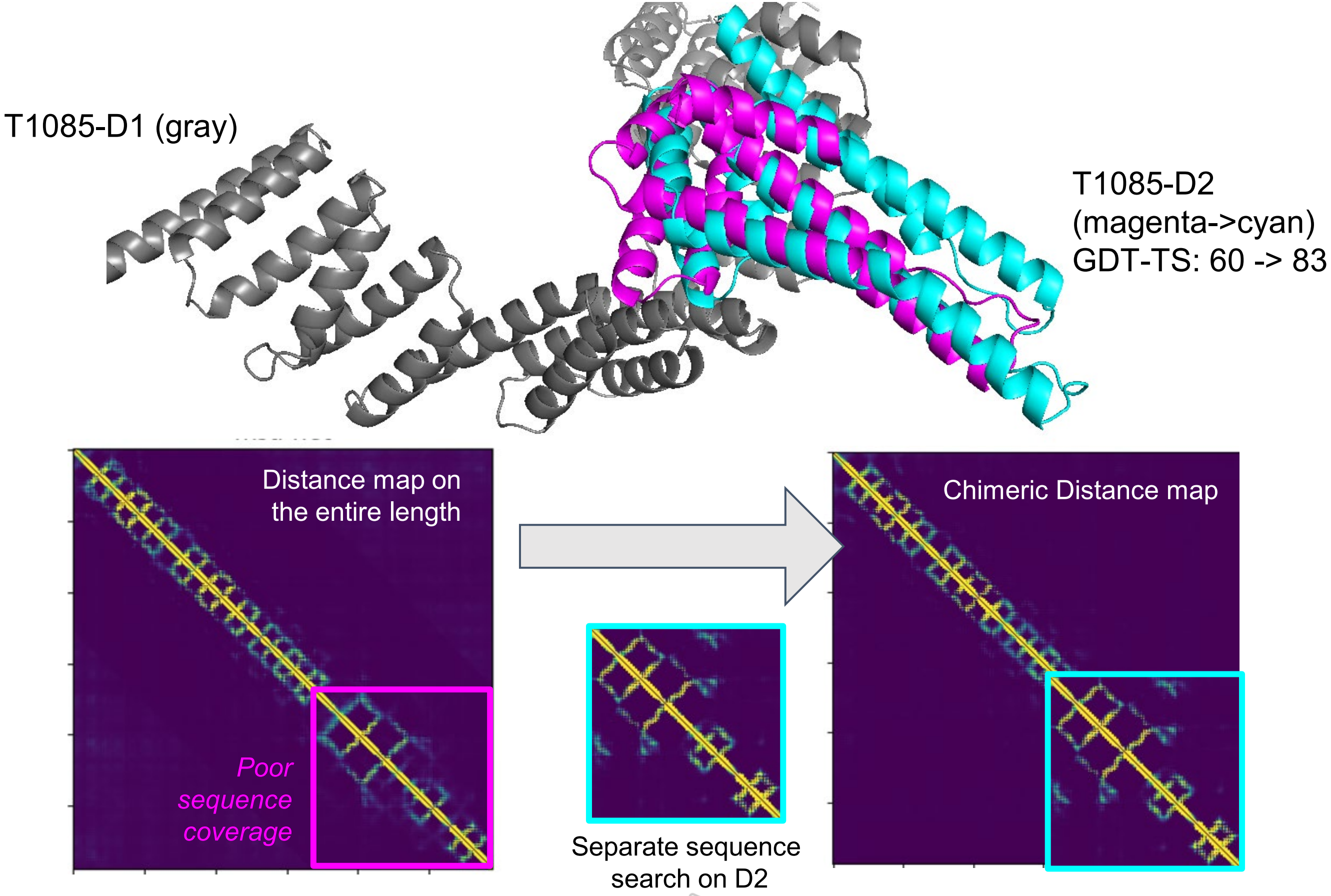
Example 1: Sequence search & Modeling as a whole Protein

Server: per-target MSA, Nseq=1~3 (UniRef30)
Human: MSA & modeling on entire protein,
Nseq>2000 (+MetaGenome, IMG/VR DB)





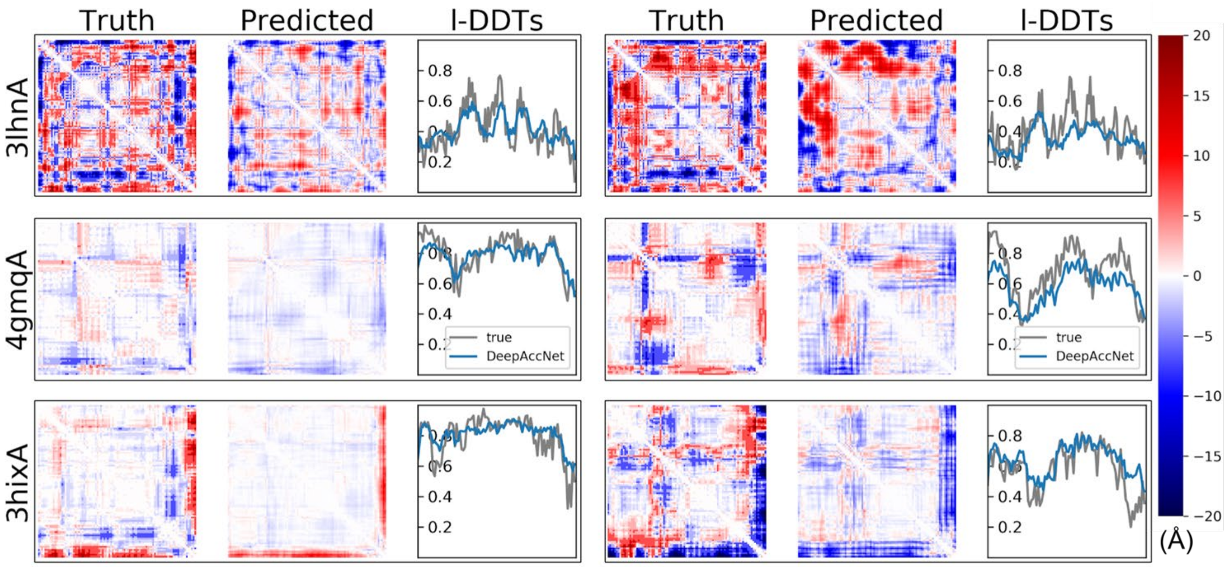
Example 2: Chimeric distance map for multi-domain targets



Refinement guided by DeepAccNet

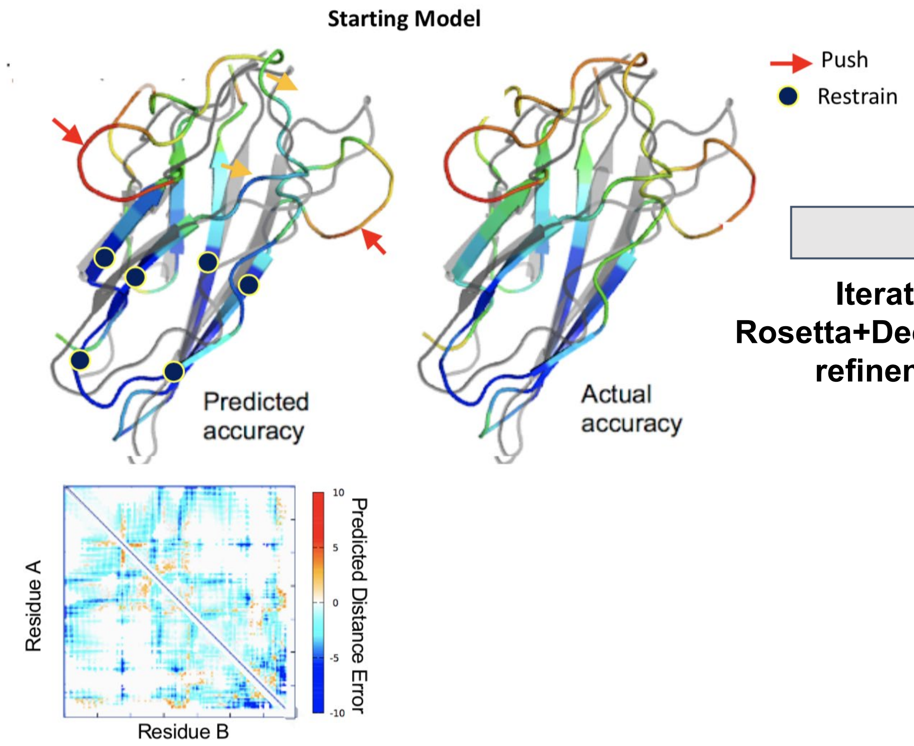
Key idea of refinement in CASP14: To use EMA to guide Refinement search

Signed distance error predictions from DeepAccNet

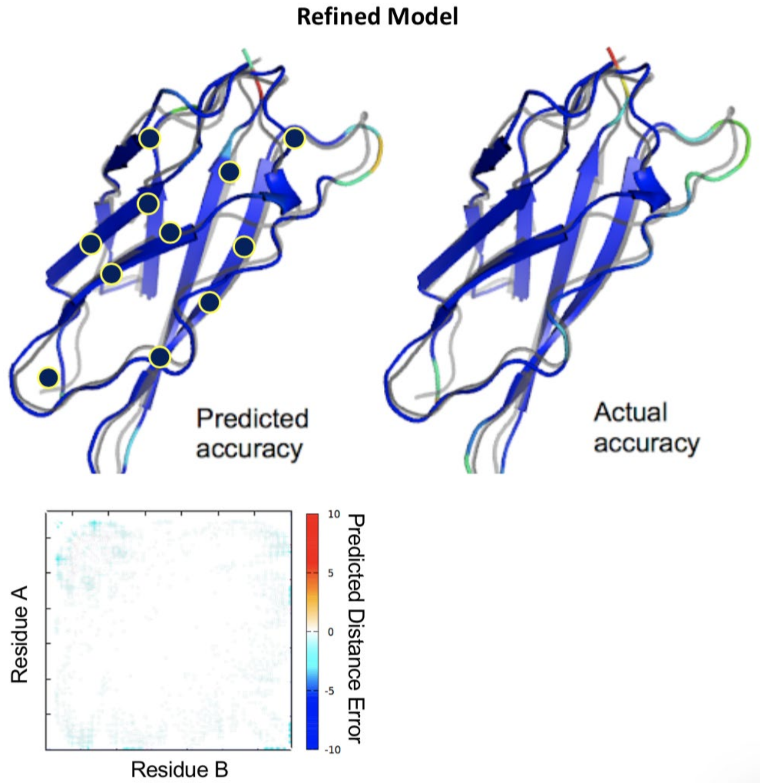


N Hiranuma et al,
Improved protein structure refinement guided by
deep learning based accuracy estimation,
bioRxiv.

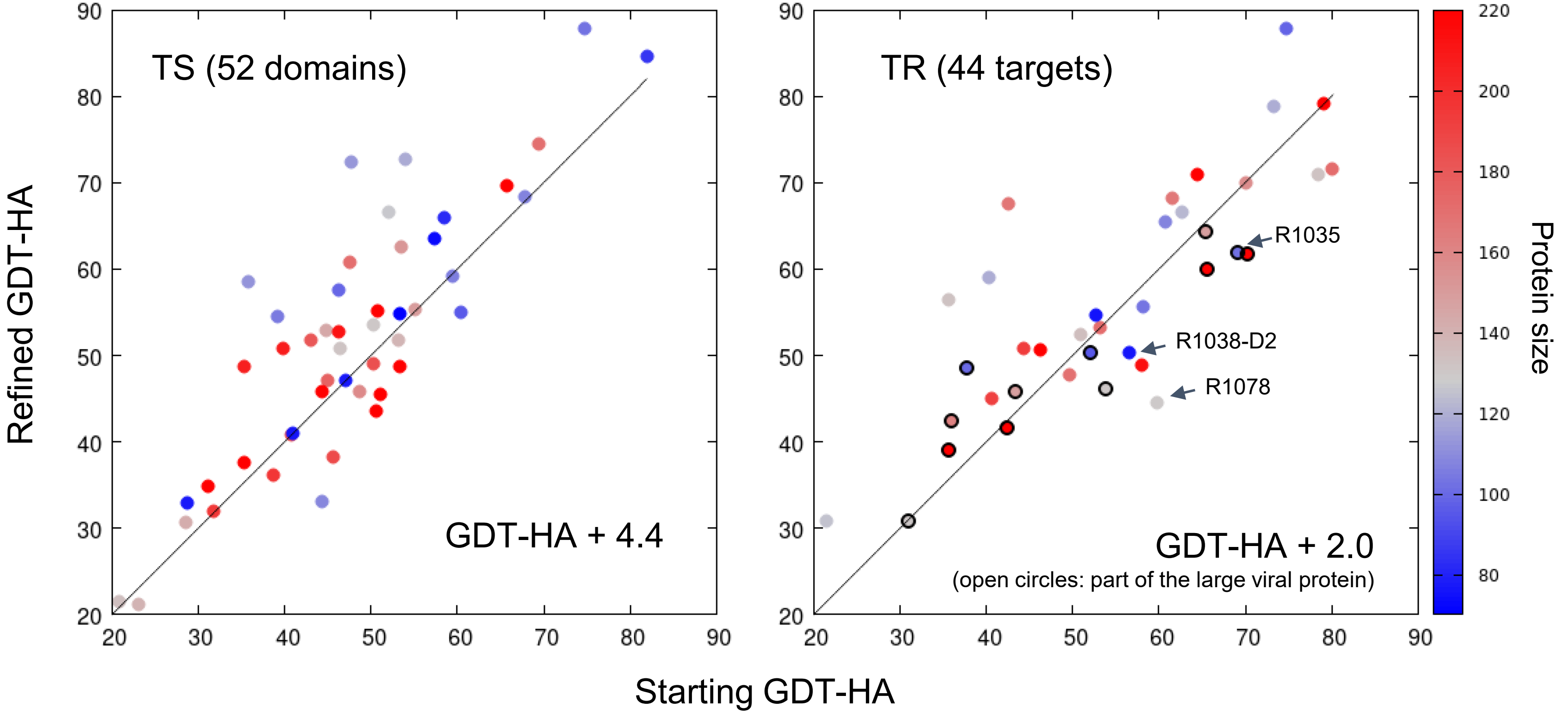
Refinement protocol



Iterative Rosetta+DeepAccNet refinement

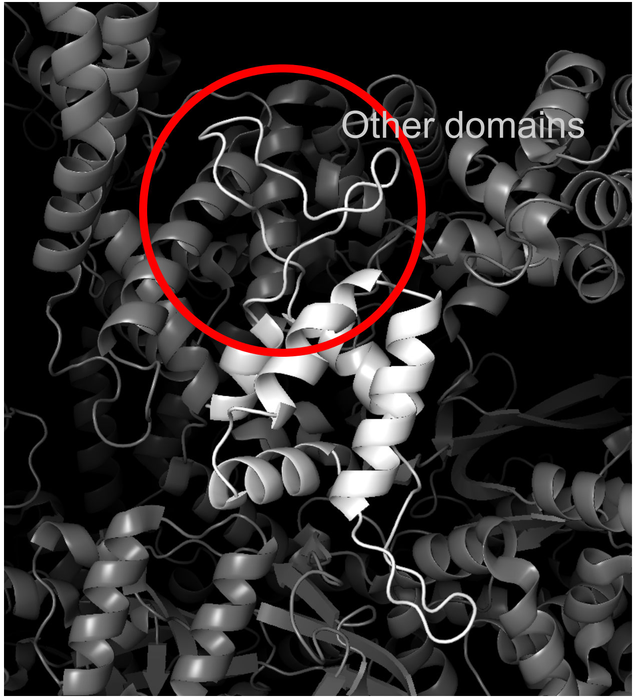


Refinement results (colored by size)



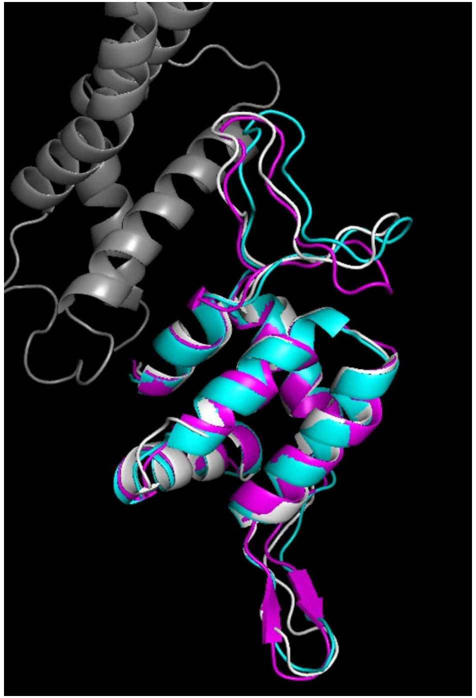
What went wrong -- R1035 Δ(GDT-HA): -7

Xtal-structure



Starting/Refined as domain

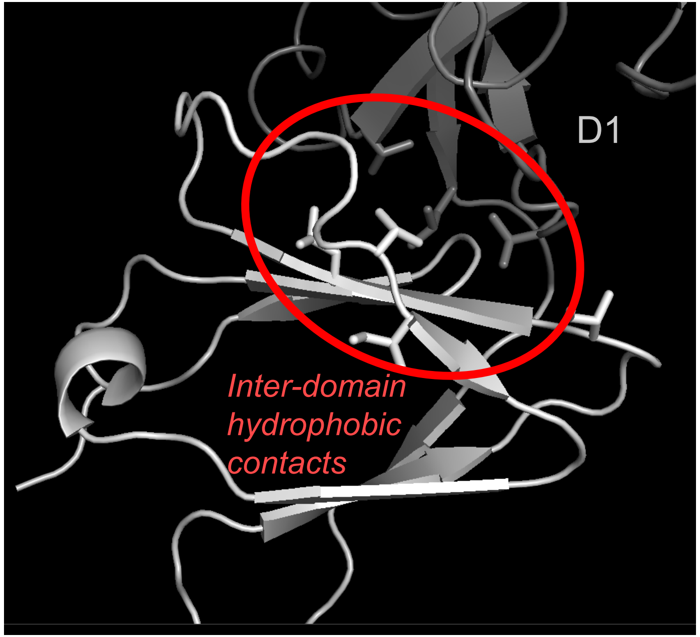
T1035: Starting/
Refined together with T1033



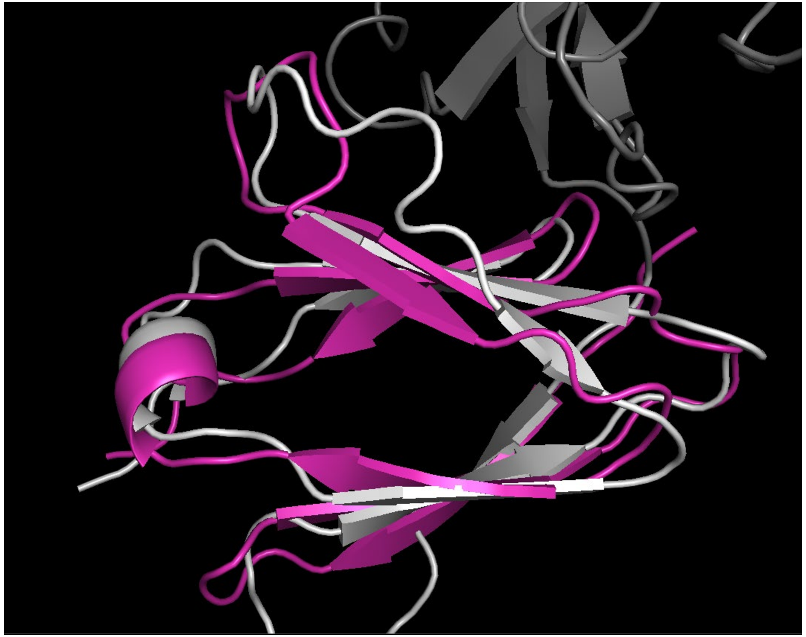
What went wrong -- R1038-D2

Δ(GDT-HA): -6

Xtal-structure



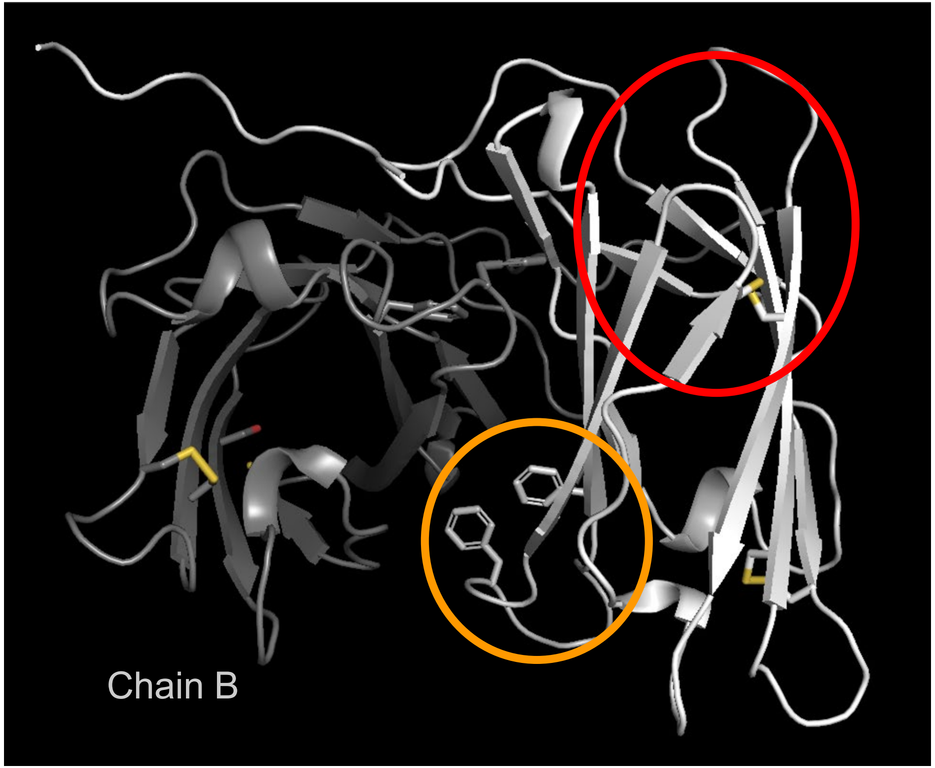
Refined as domain



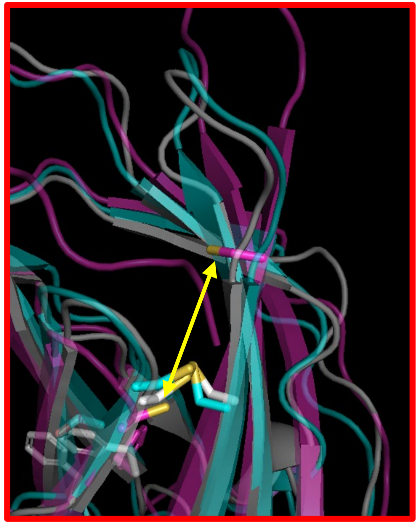
What went wrong -- R1078

Δ(GDT-HA): -15

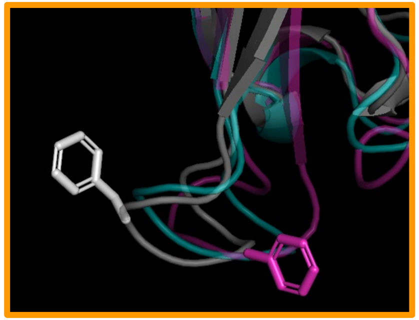
Xtal-structure (homo-dimer)



Starting/Refined as monomer



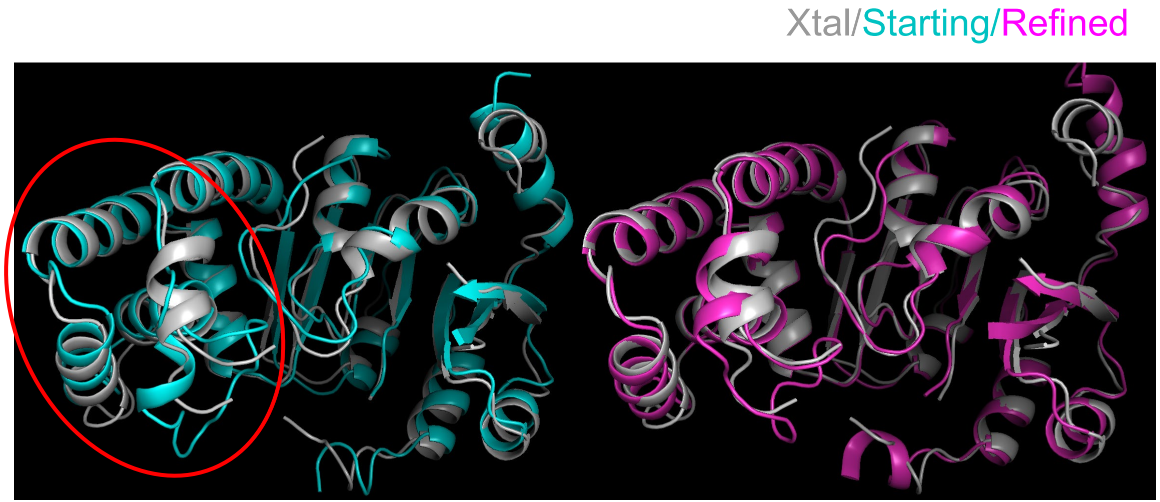
Broken Disulfide bond



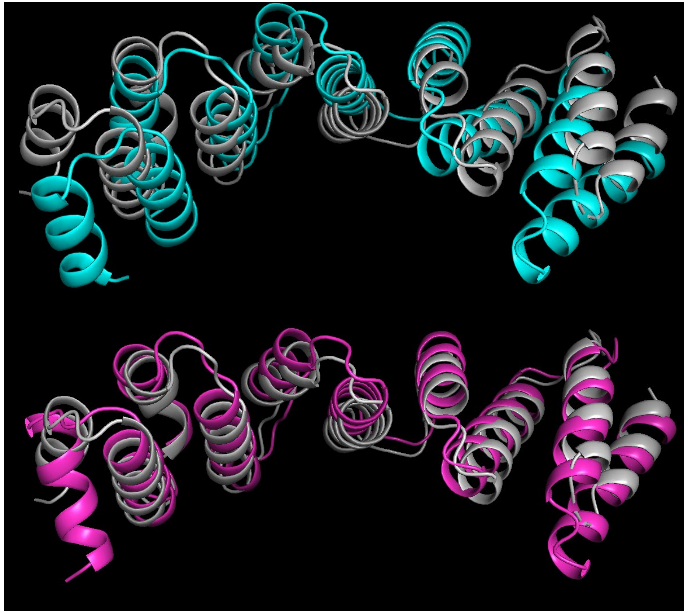
Dimer interface

Bright side: EMA-guided refinement enables improving relatively **larger proteins**

R1067

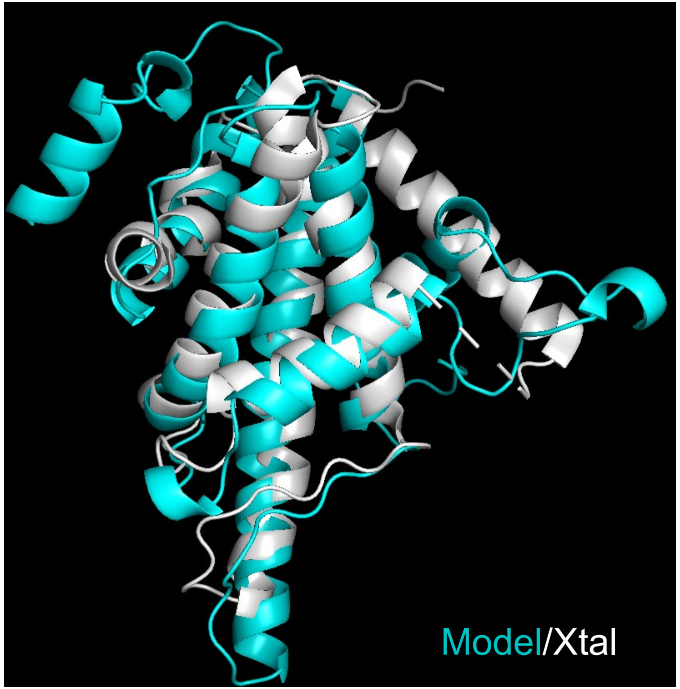


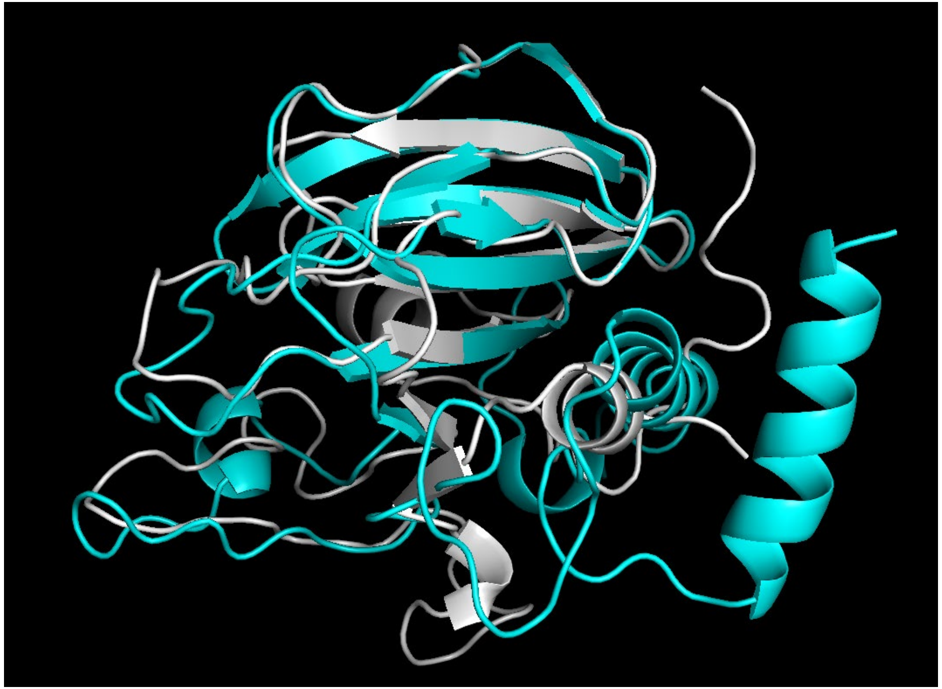
R1085-D1



Still challenging -- R1068, R1056, etc.

Could not improve significantly when topologies were complicated & sizes were big





Take home messages

- **EMA-guided refinement tested in CASP14; general challenge remains**
- **Refinement in a monomeric context fails;
Information is often more critical than principles
in real practice modeling scenarios**

Acknowledgements

BAKER group CASP14 team

Ivan Anishchenko
Minkyung Baek
Hahnbeom Park
David Kim
Justas Dauparas
Naozumi Hiranuma
Sanaa Mansoor
Ian Humphrey
Luki Goldschmidt
David Baker

trRosetta developers

Jianyi Yang
Sergey Ovchinnikov

CASP organizers & participants