Protein Assembly by Server and Human

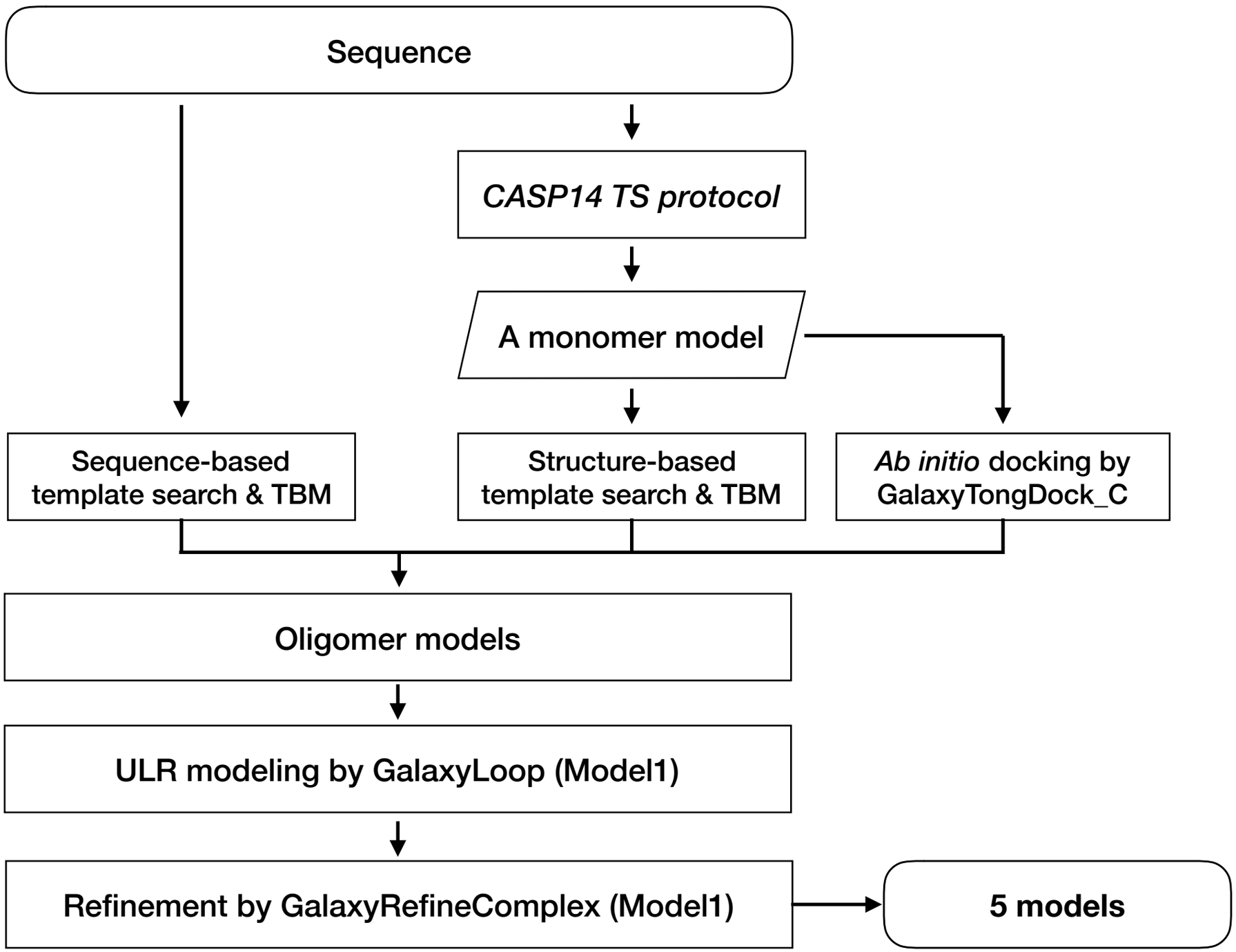
Taeyong Park, Hyeonuk Woo, Jinsol Yang, Sohee Kwon,
Jonghun Won and Chaok Seok

Contents

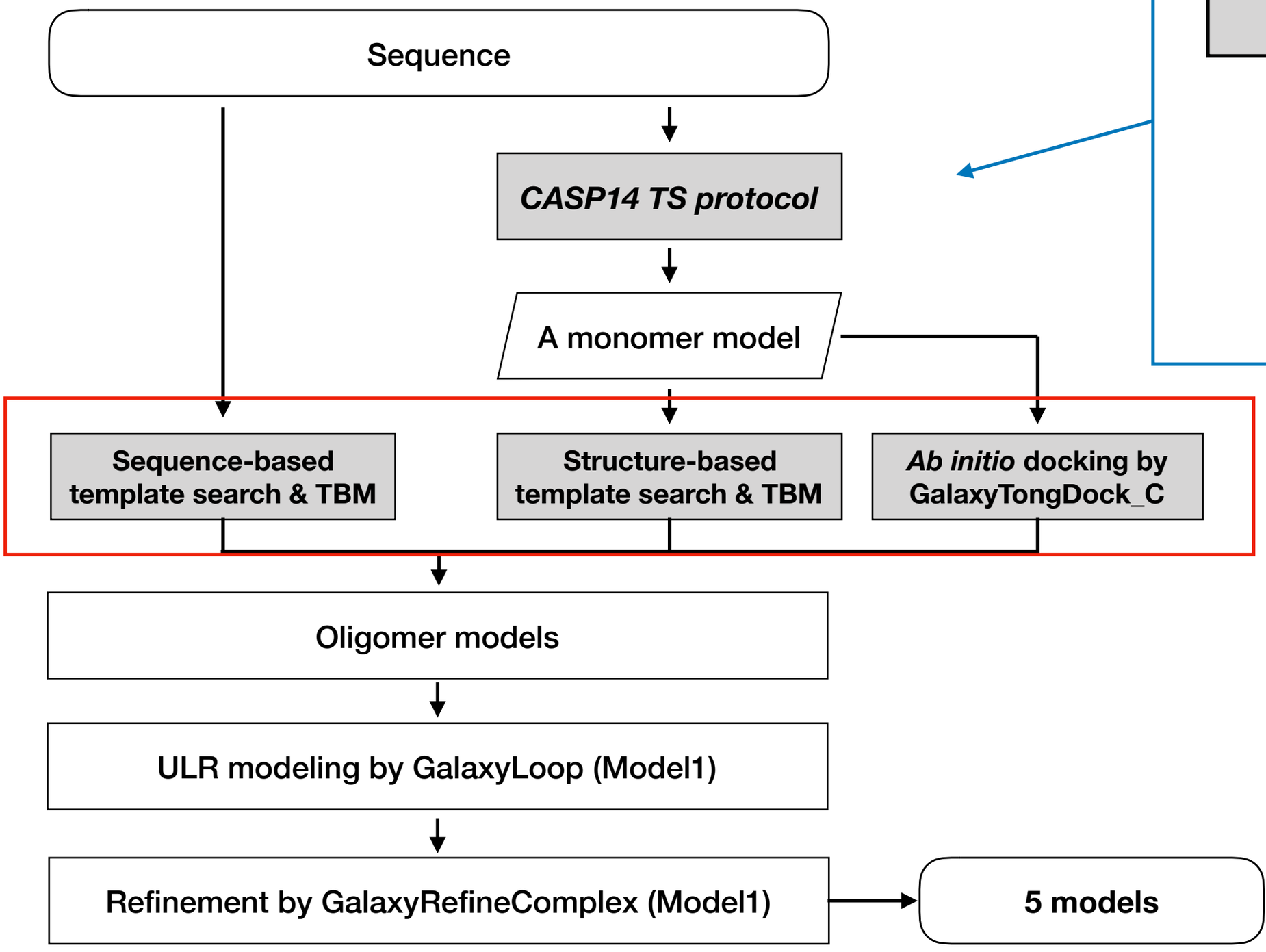
Seok assembly pipeline

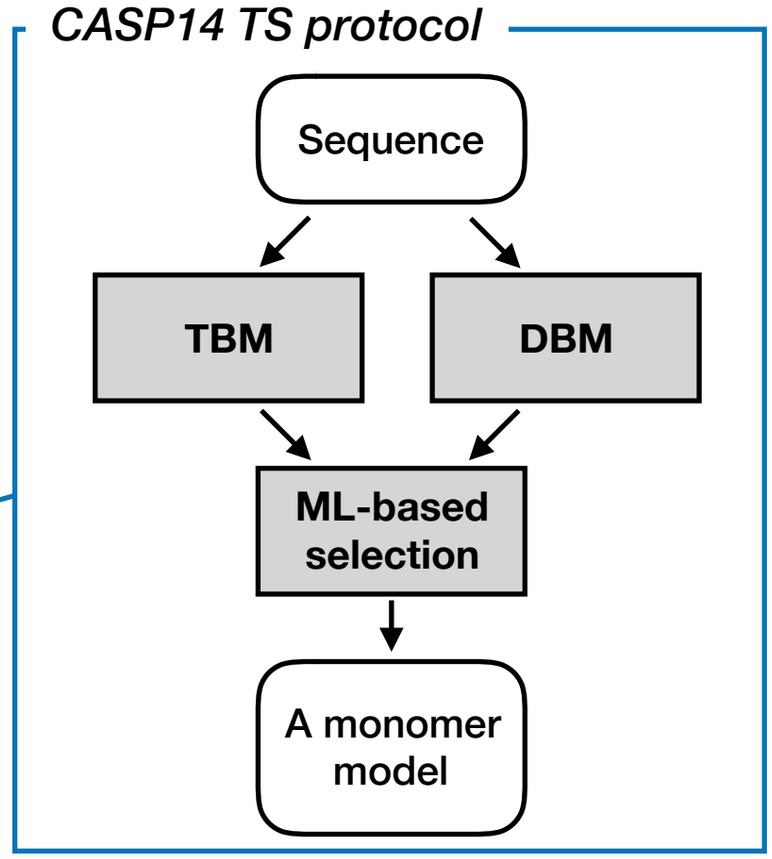
Server pipeline on A_n target

GalaxyHomomer2



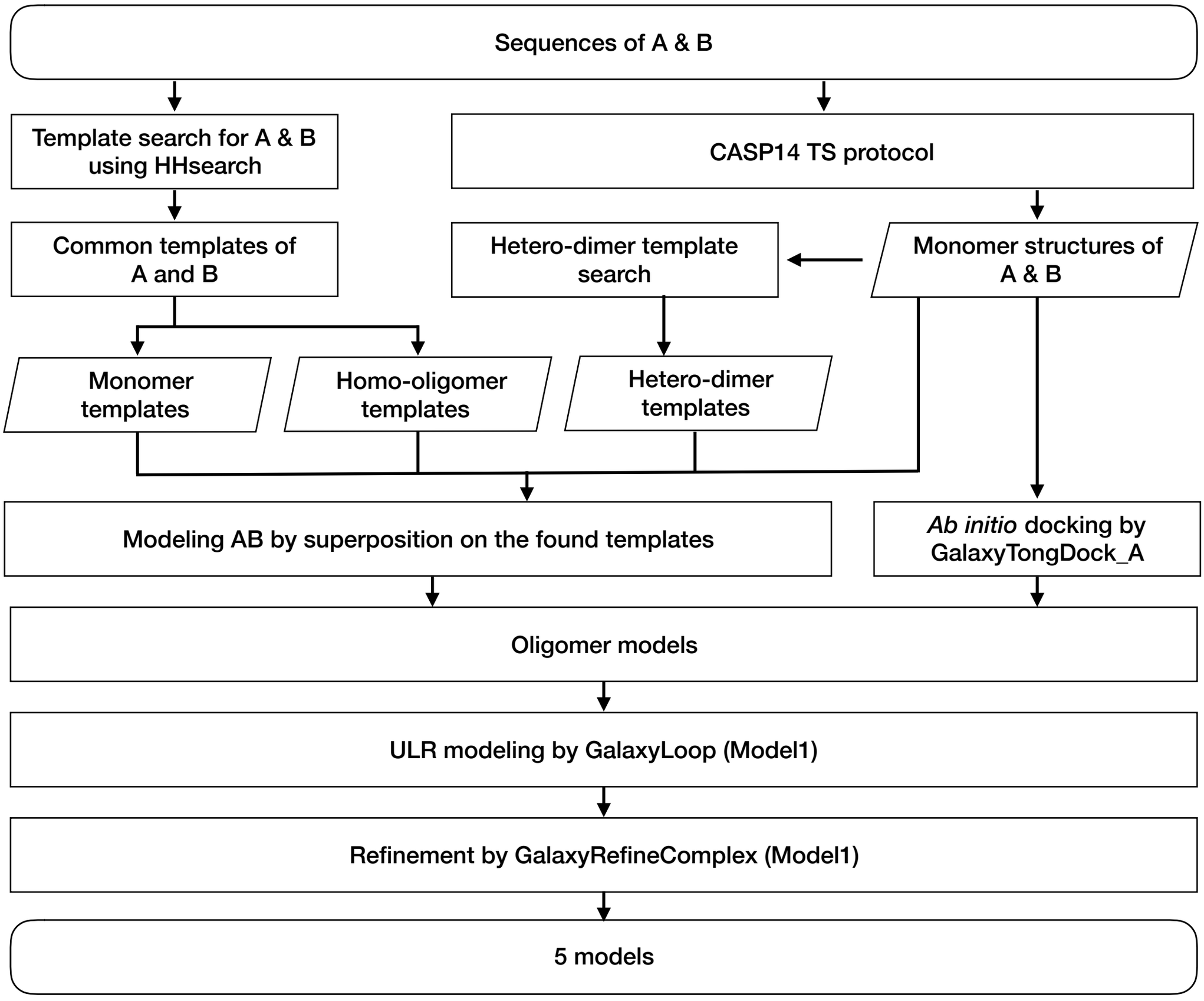
Server pipeline on A_n target



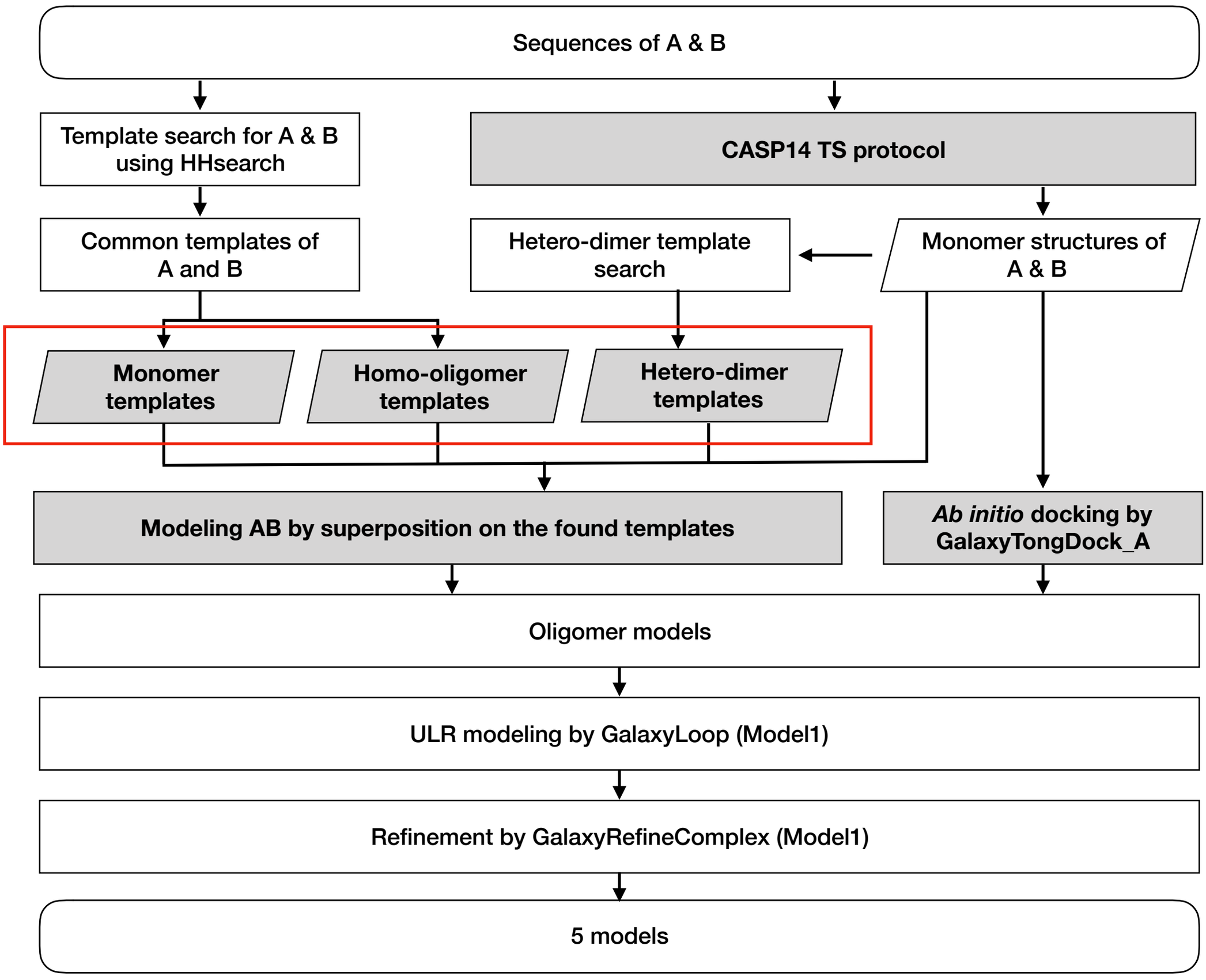


Server pipeline on A_1B_1 target

GalaxyHeteromer



Server pipeline on A_1B_1 target



Server prediction on the targets of more complicated stoichiometry ($A_lB_mC_n$ ...)

- **Individual interfaces were** predicted by *GalaxyHeteromer* and *GalaxyHomomer2*.

✓ In the case of H1072 (A_2B_2), A_2 and B_2 interfaces were predicted by *GalaxyHomomer2*. AB interface was predicted by *GalaxyHeteromer*.

Human

1. Available information from the literature and human insight were employed for model generation and selection.

2. Monomer models from other servers were also utilized to predict oligomer structures.

Server vs. Human

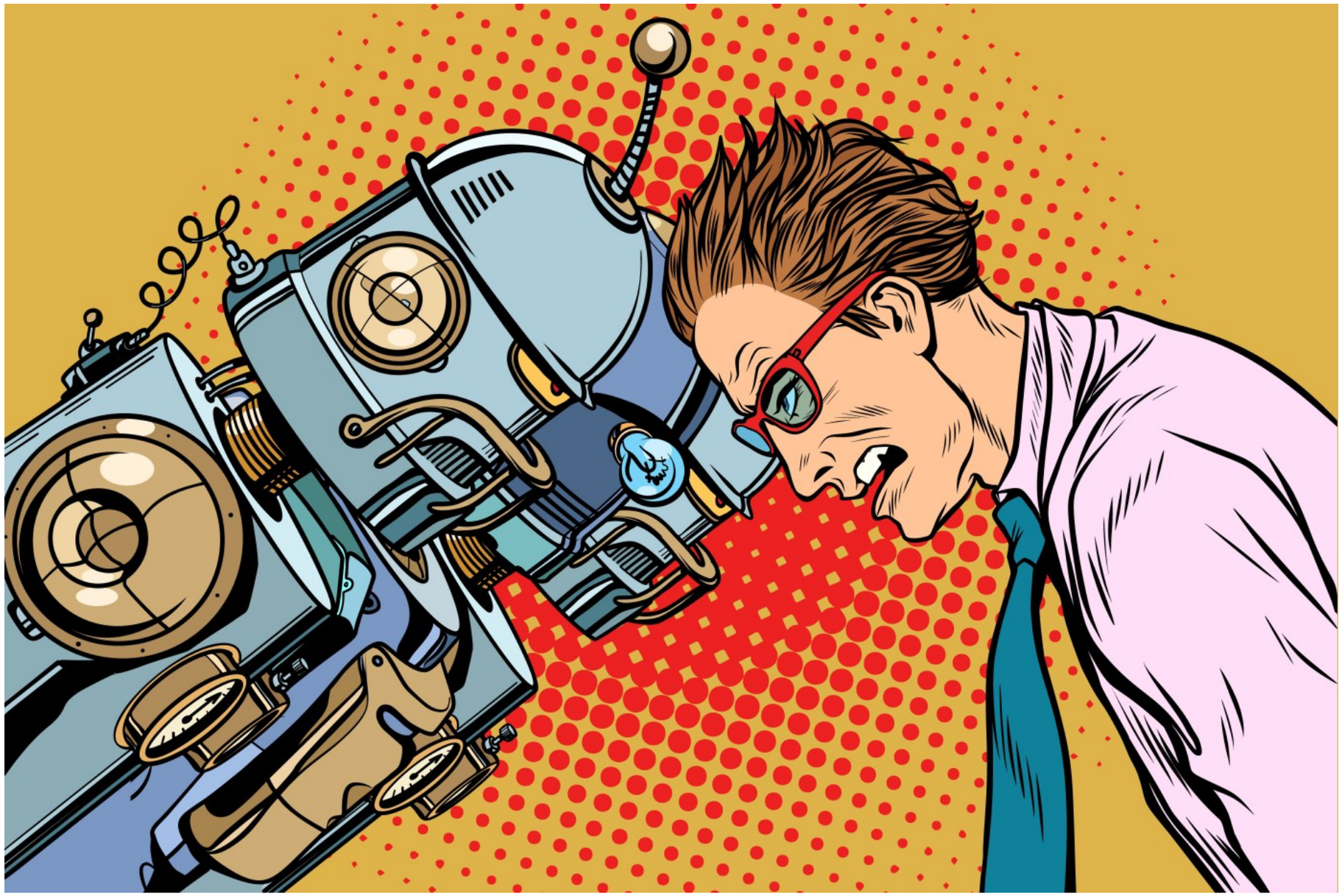
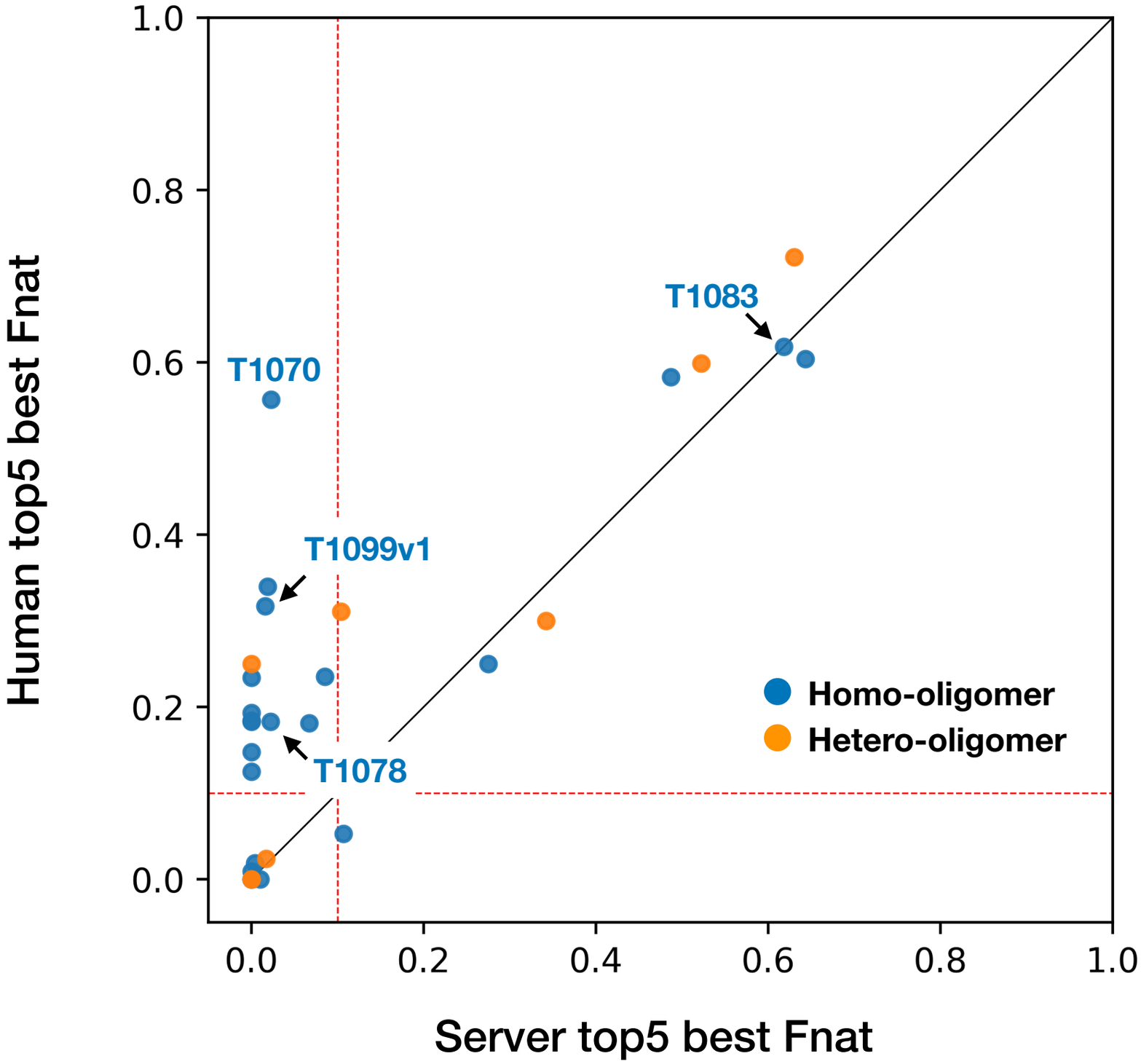


Image credit: depositphotos.com

Server vs. Human



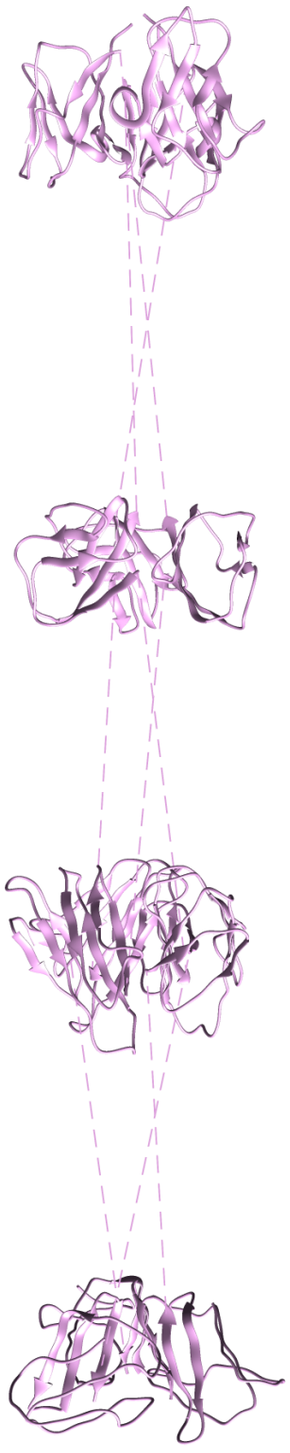
What went right (four targets)

1. T1070

(Ab initio docking)

T1070 (A_3)

Manually separated into 4 domains based on server models for monomer.



domain 4: 264-335

domain 3: 190-249

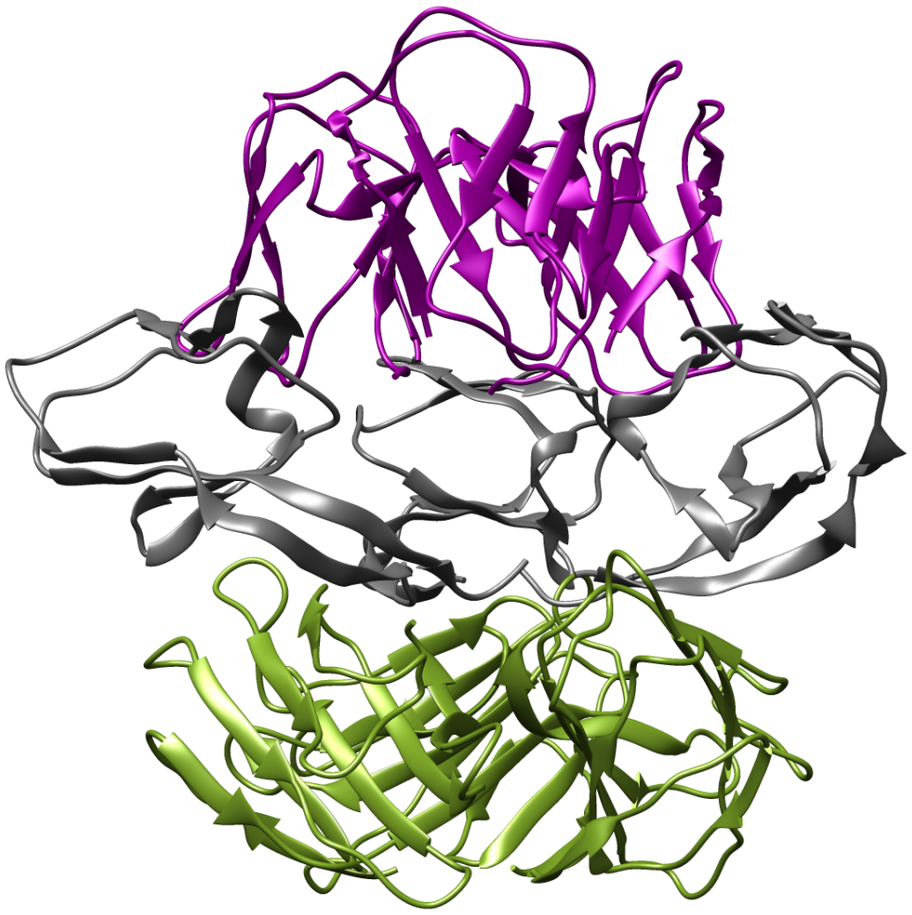
domain 2: 80-165

domain 1: 1-76

Crystal
Model

T1070 (A_3)

The crystal structure also can be divided into similar 4 domains



domain 4: 265-335 (264-335)

domain 3: 181-256 (190-249)

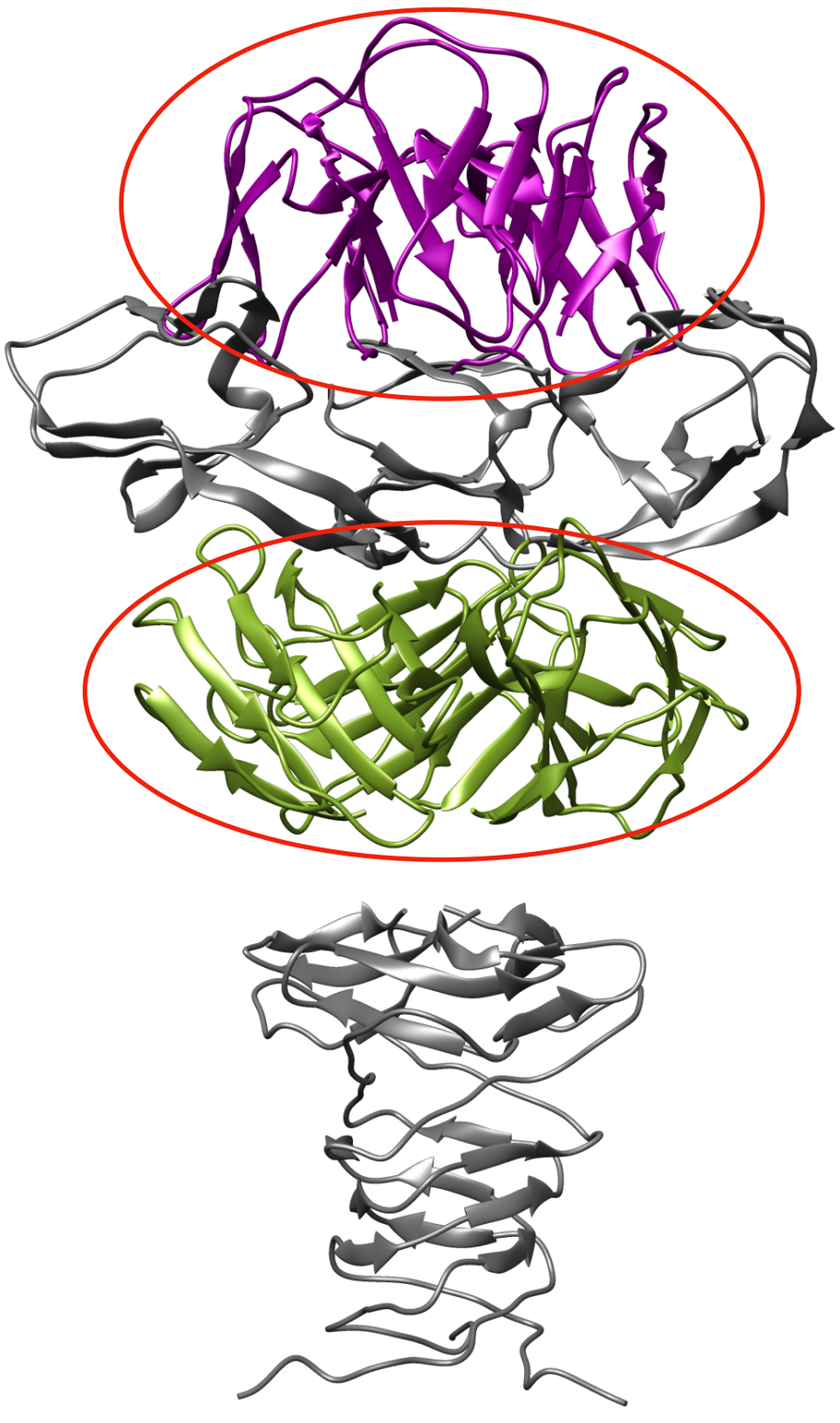
domain 2: 80-180 (80-165)

domain 1: 1-76 (1-76)

Domain 4
Domain 3
Domain 2
Domain 1

T1070 (A_3)

The crystal structure also can be divided into similar 4 domains



domain 4: 265-335 (264-335)

domain 3: 181-256 (190-249)

domain 2: 80-180 (80-165)

domain 1: 1-76 (1-76)

Domain 4
Domain 3
Domain 2
Domain 1

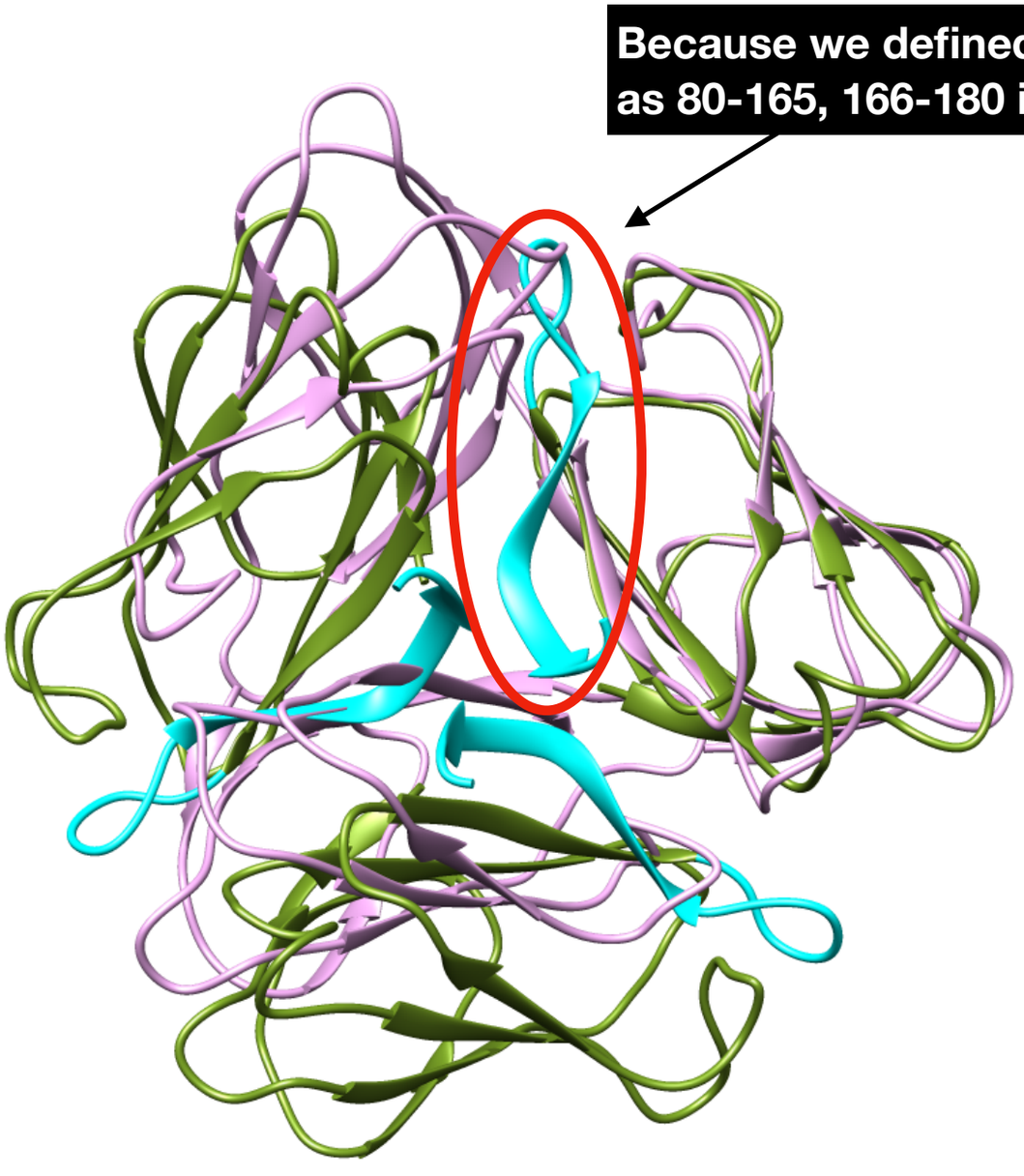
T1070 (A_3) domain 2 (80-180)

Model 5: little bit closer each other because of lack of 166-180 due to mis-spliting of the domain

Crystal
Model

T1070 (A_3) domain 2 (80-180)

Model 5: little bit closer each other because of lack of 166-180 due to mis-spliting of the domain



Model 5

F_{nat}: 0.18

IRMSD: 5.26Å

LRMSD: 5.31Å

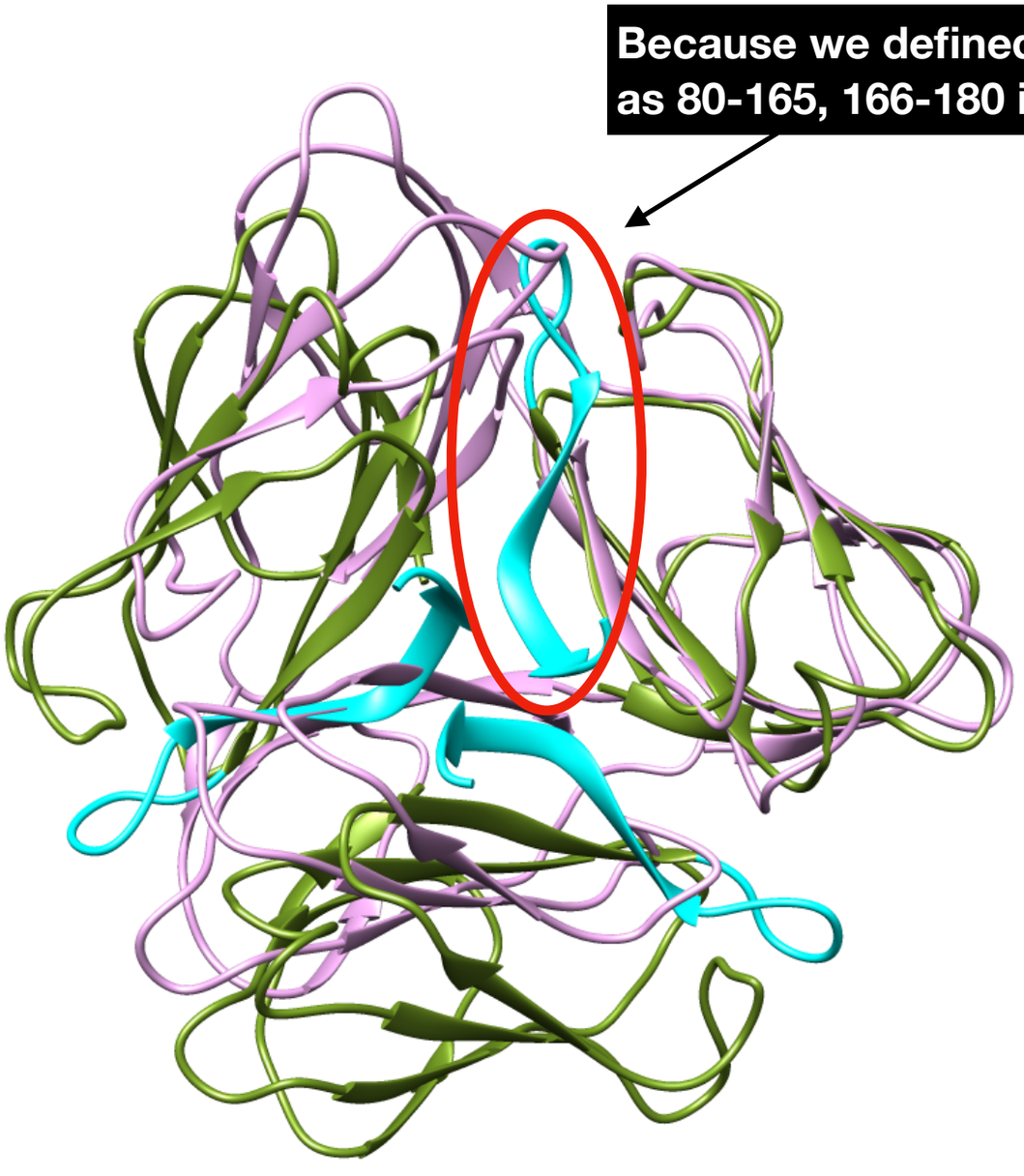
"Acceptable quality"

Crystal

Model

T1070 (A_3) domain 2 (80-180)

Model 5: little bit closer each other because of lack of 166-180 due to mis-spliting of the domain



Model 5

F_{nat}: 0.18

IRMSD: 5.26Å

LRMSD: 5.31Å

"Acceptable quality"

GalaxyTongDock top 1 prediction by using Seok-server_TS1 monomer structure

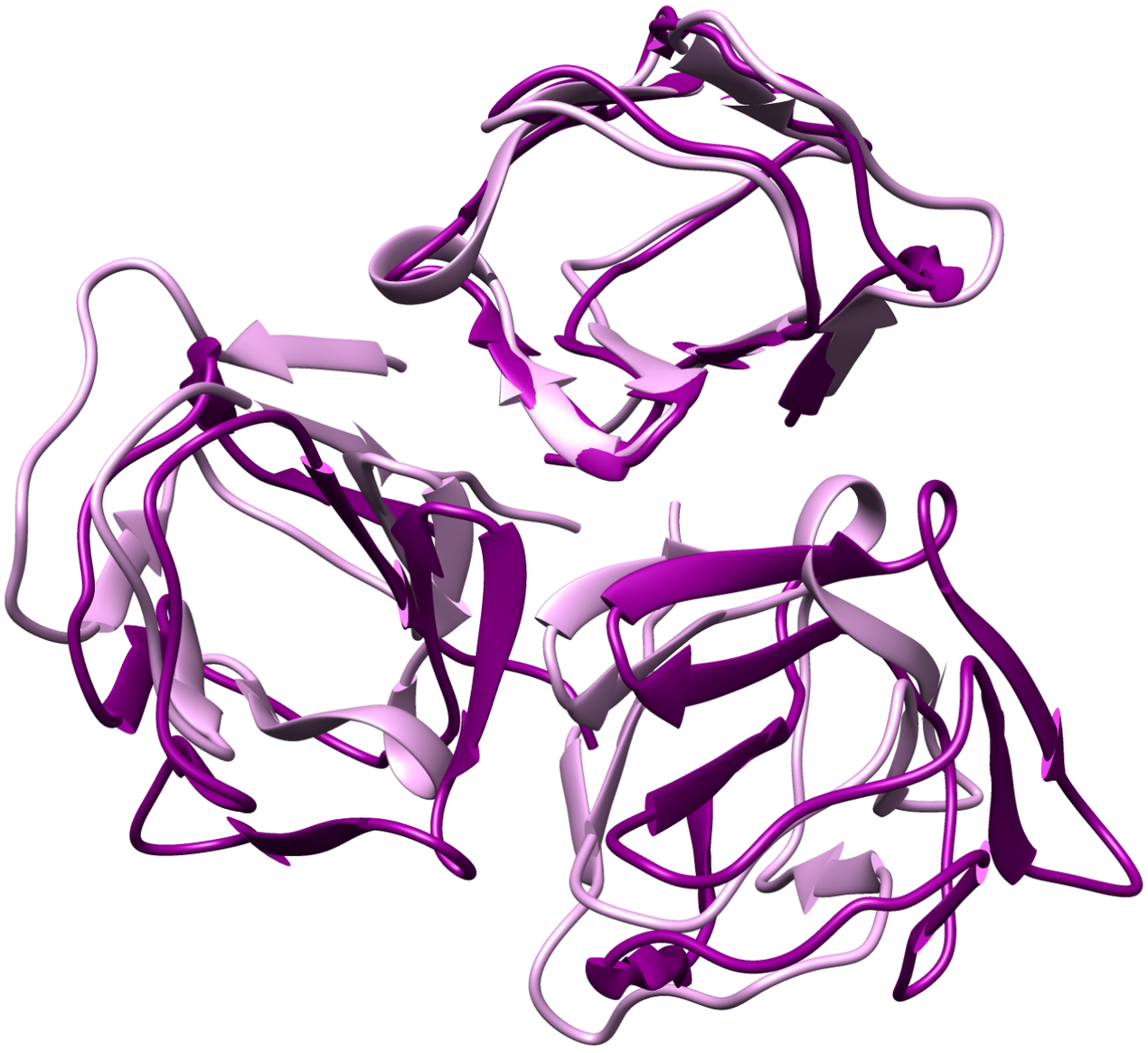
Scoring based on human intuition didn't work

Crystal

Model

T1070 (A_3) domain 4 (265-335)

Domain-split was successful 265-335 vs. 266-335.



Model 1

F_{nat}: 0.42

IRMSD: 3.22Å

LRMSD: 4.06Å

"Medium quality"

GalaxyTongDock top <u>16 prediction</u> by using Zhang-CEthreader_TS1 (The best docking pose generated)

Scoring based on human intuition worked well

Crystal

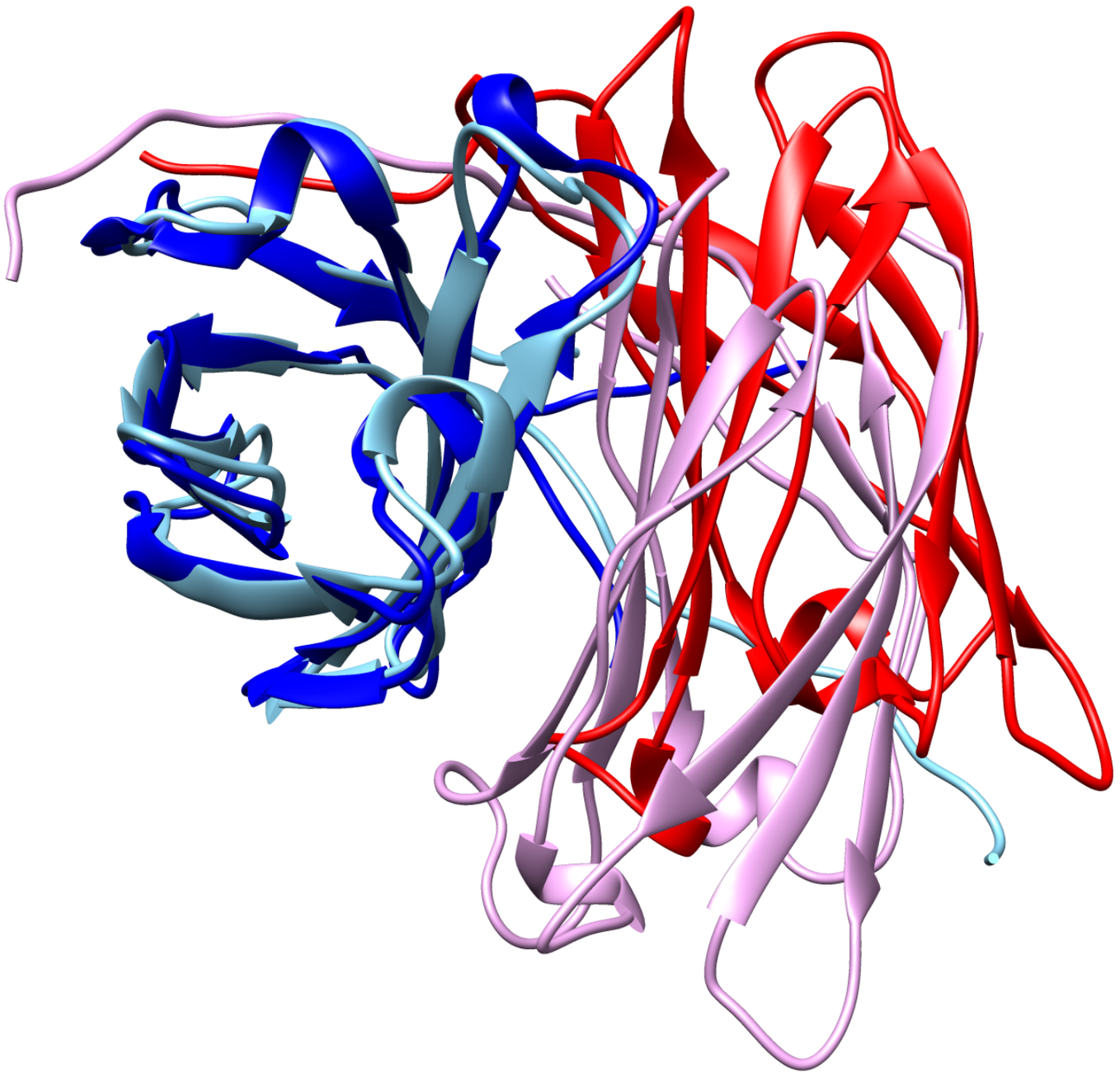
Model

2. T1078

(Ab initio docking)

T1078 (A_2)

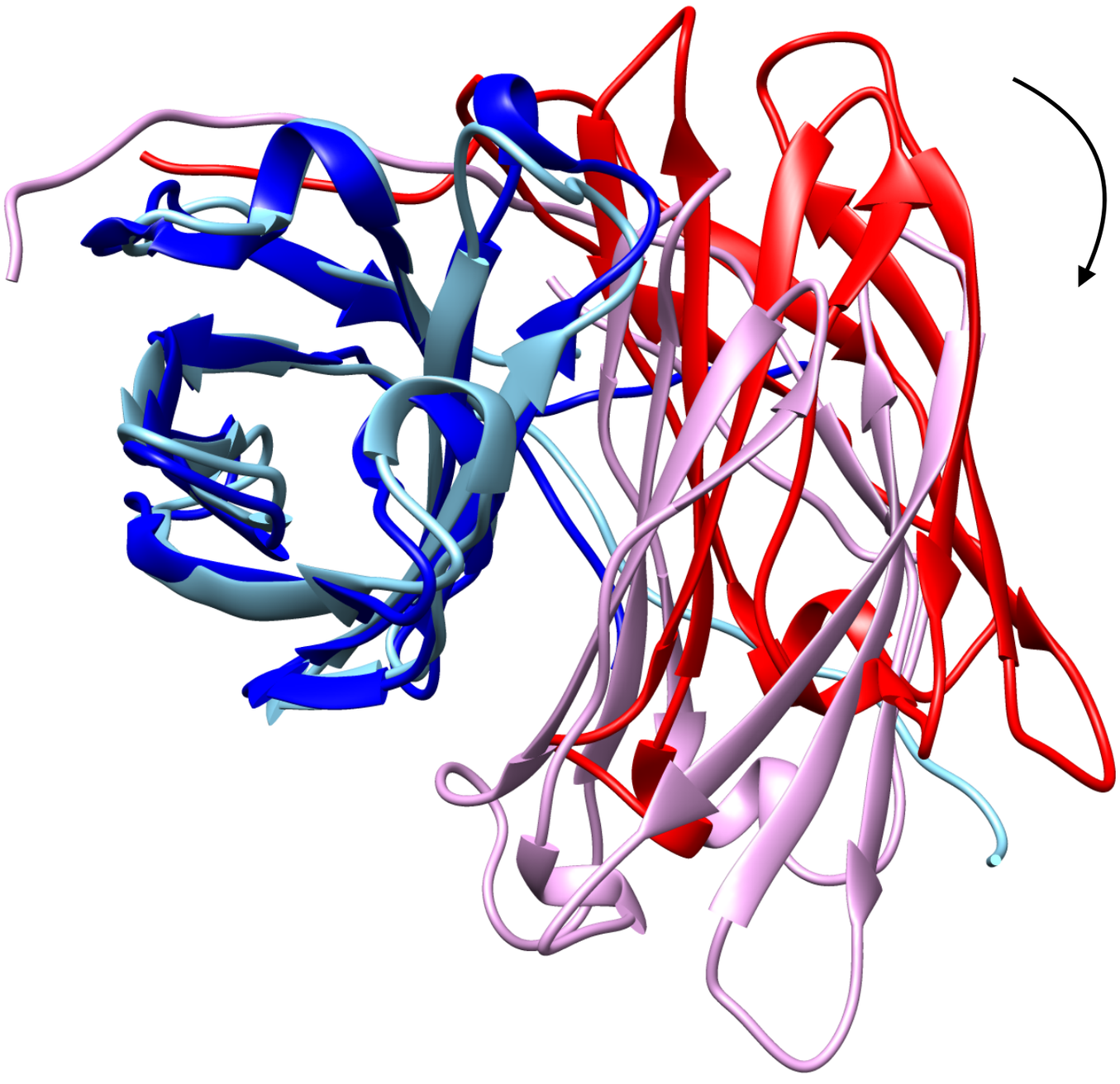
Model 3 is one of the four acceptable quality models.



Crystal s1
Model s1
Crystal s2
Model s2

T1078 (A_2)

Model 3 is one of the four acceptable quality models.

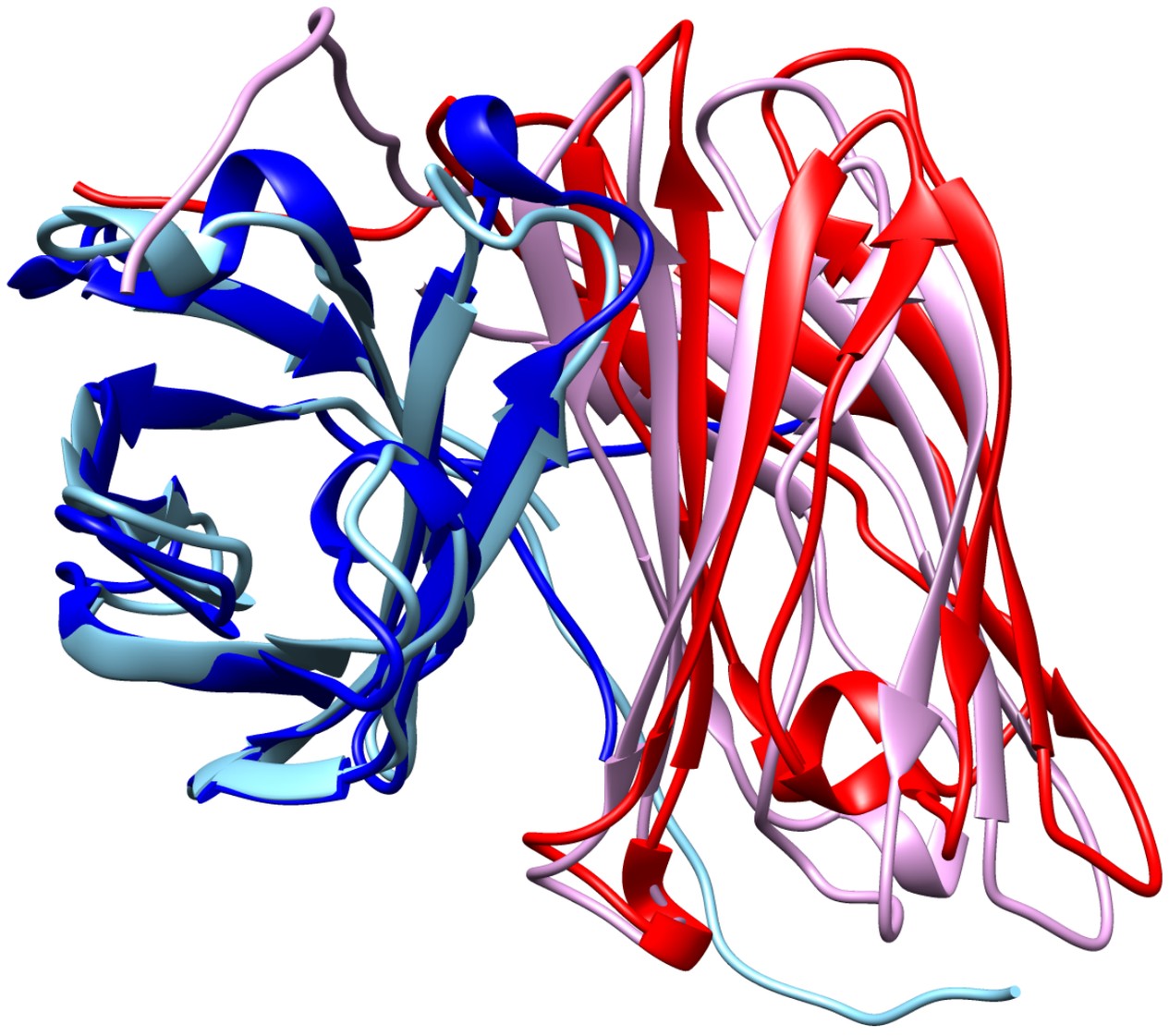


F_{nat}: 0.18
IRMSD: 4.28Å
LRMSD: 6.95Å

Crystal s1
Model s1
Crystal s2
Model s2

T1078 (A_2)

Model 7 has right orientation, but is of lower quality.



F_{nat}: 0.15
IRMSD: 5.50Å
LRMSD: 7.80Å

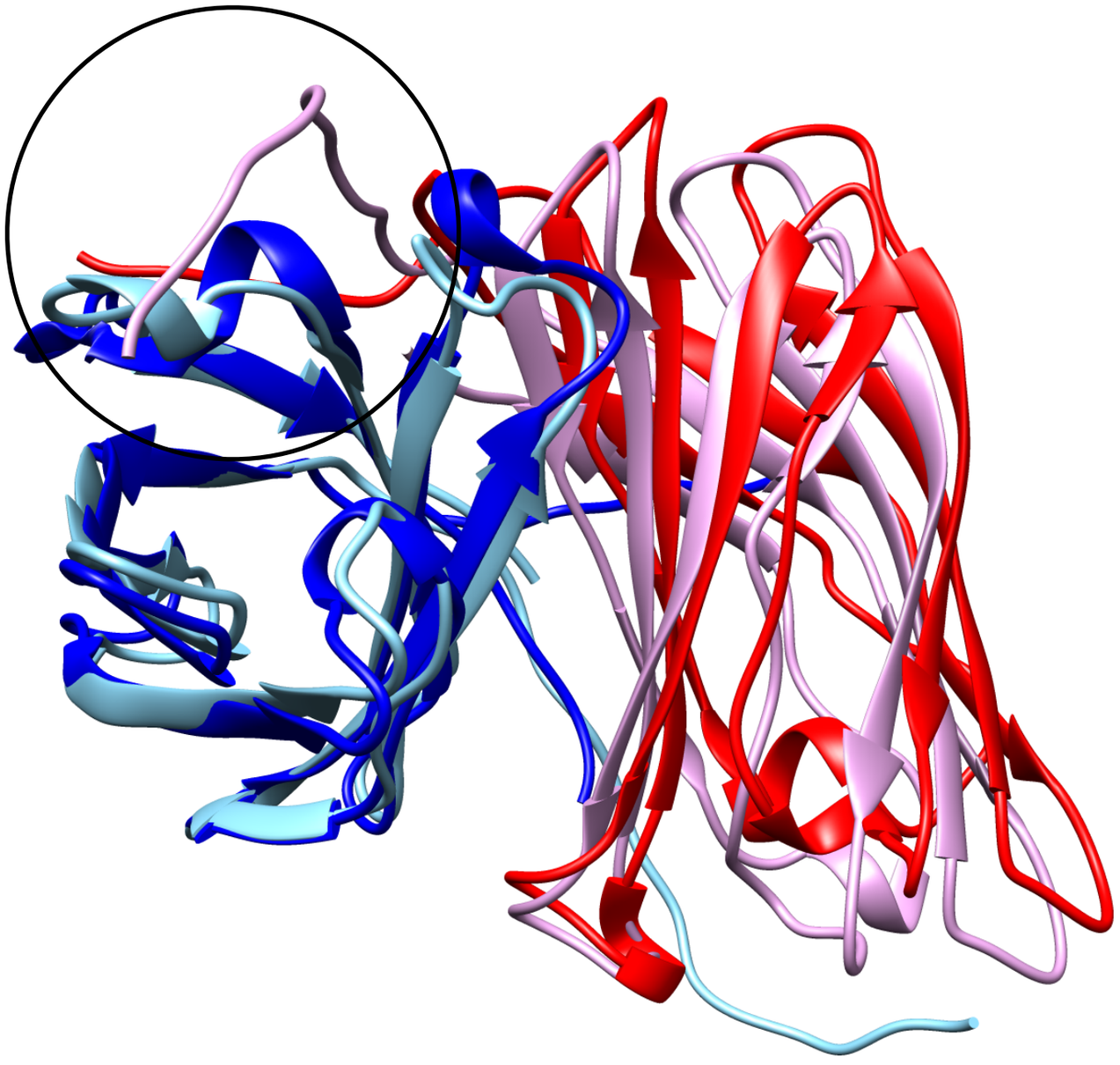
"Acceptable quality"

Crystal s1
Model s1
Crystal s2
Model s2

The two termini affected the model quality.

T1078 (A_2)

Model 7 has right orientation, but is of lower quality.



F_{nat}: 0.15
IRMSD: 5.50Å
LRMSD: 7.80Å

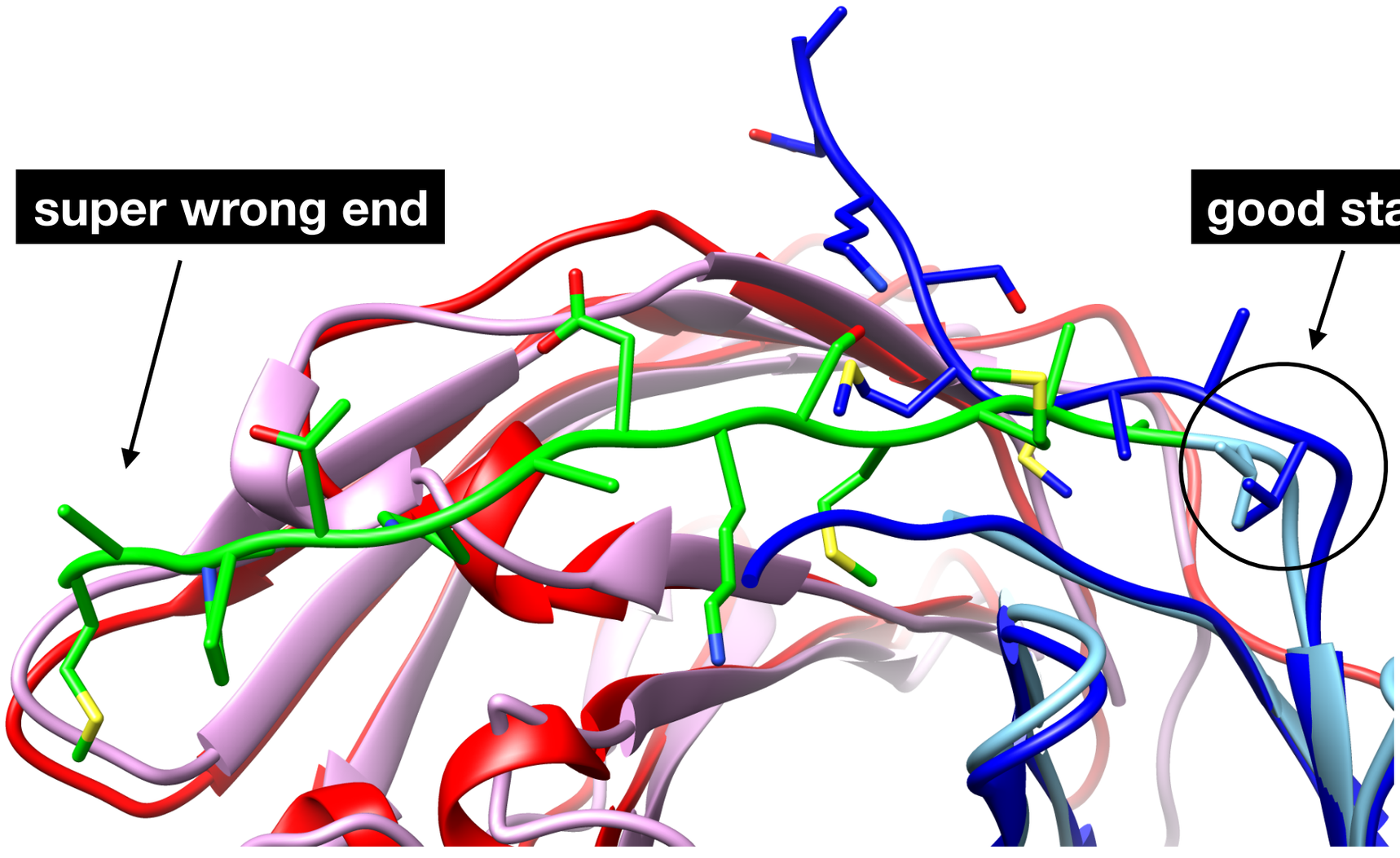
"Acceptable quality"

Crystal s1
Model s1
Crystal s2
Model s2

The two termini affected the model quality.

T1078 (A_2)

Unsuccessful N-terminal modeling (1-14)



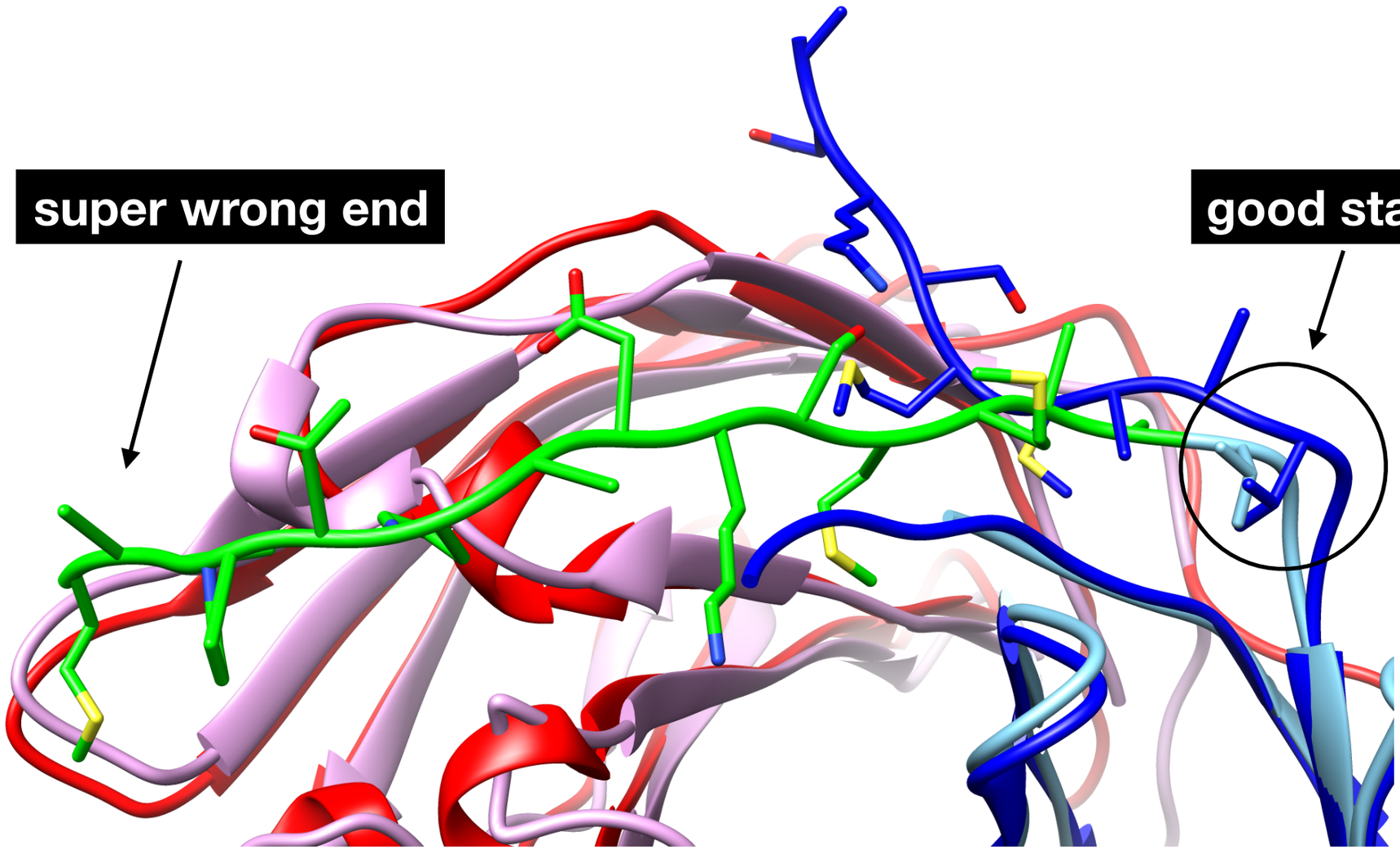
F_{nat}: 0.15
IRMSD: 5.50Å
LRMSD: 7.80Å

GalaxyTongDock was used with only 15-128 region.

Crystal s1
Model s1
Crystal s2
Model s2

T1078 (A_2)

Unsuccessful N-terminal modeling (1-14)



F_{nat}: 0.15
IRMSD: 5.50Å
LRMSD: 7.80Å

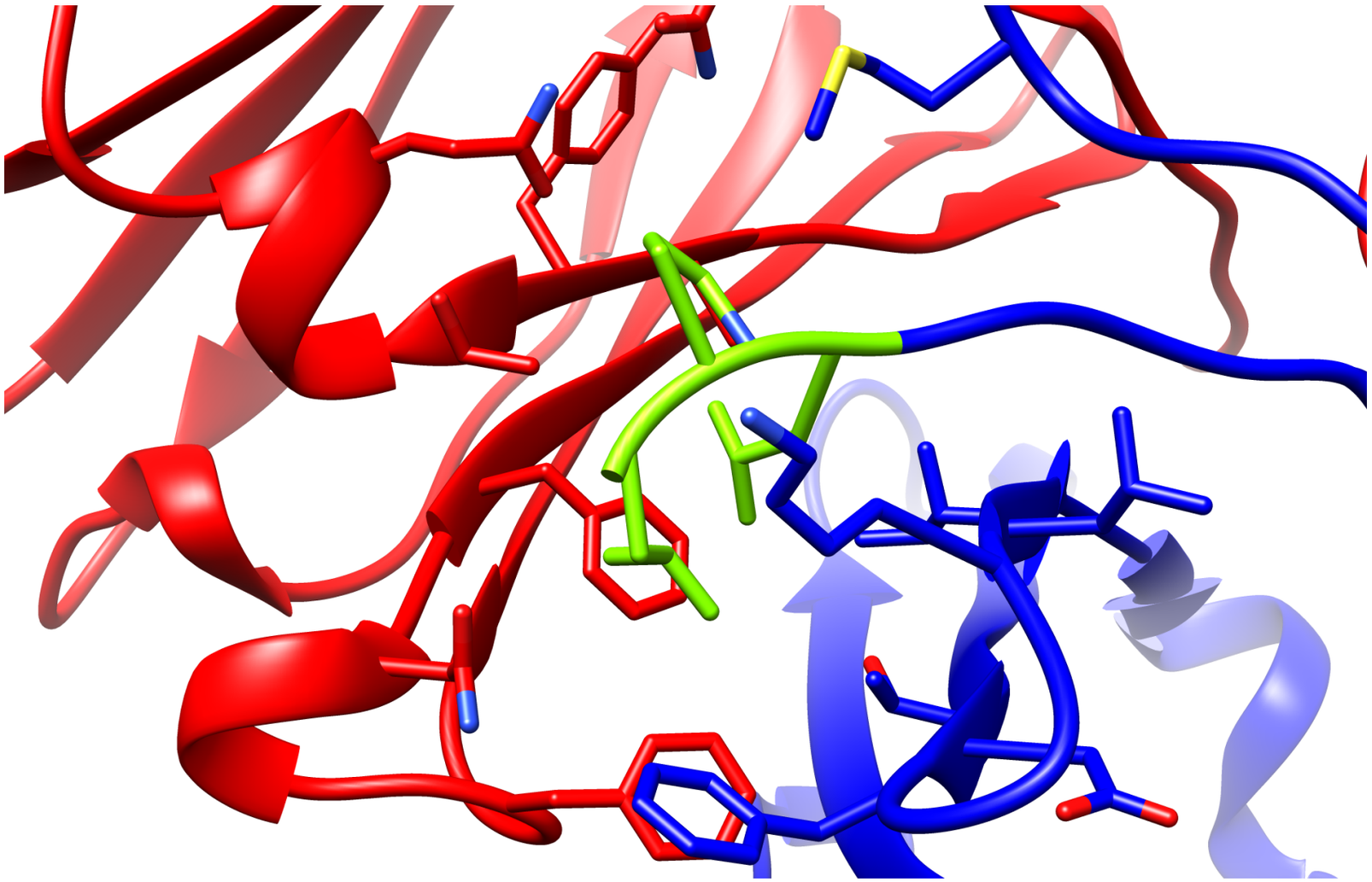
Evaluation without N-terminal (1-14)
F_{nat}: 0.25
IRMSD: 1.72Å
LRMSD: 3.65Å

GalaxyTongDock was used with only 15-128 region.

Crystal s1
Model s1
Crystal s2
Model s2

T1078 (A_2)

Hexahistidine tag (LPLEHHHHHH, 129-138)



Evaluation without N-terminal (1-14)

F_{nat}: 0.25

IRMSD: 1.72Å

LRMSD: 3.65Å

Part of the exp tag forms strong interaction.

Crystal s1

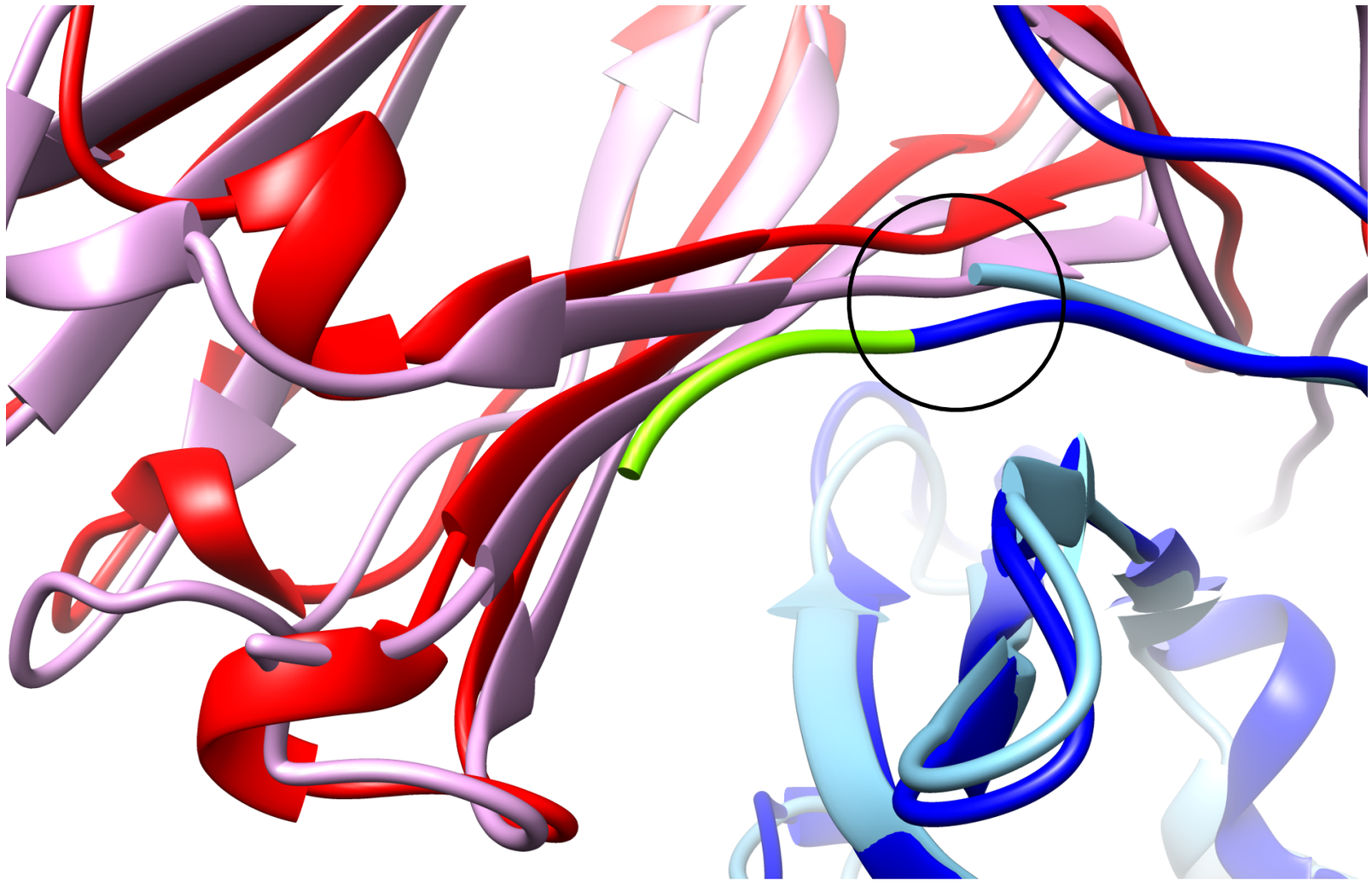
Model s1

Crystal s2

Model s2

T1078 (A_2)

Hexahistidine tag (LPLEHHHHHHH, 129-138)



Evaluation without N-terminal (1-14)
F_{nat}: 0.25
IRMSD: 1.72Å
LRMSD: 3.65Å

Also without exp tag (129-138)
F_{nat}: 0.43
IRMSD: 1.72Å
LRMSD: 3.69Å

"Medium quality"

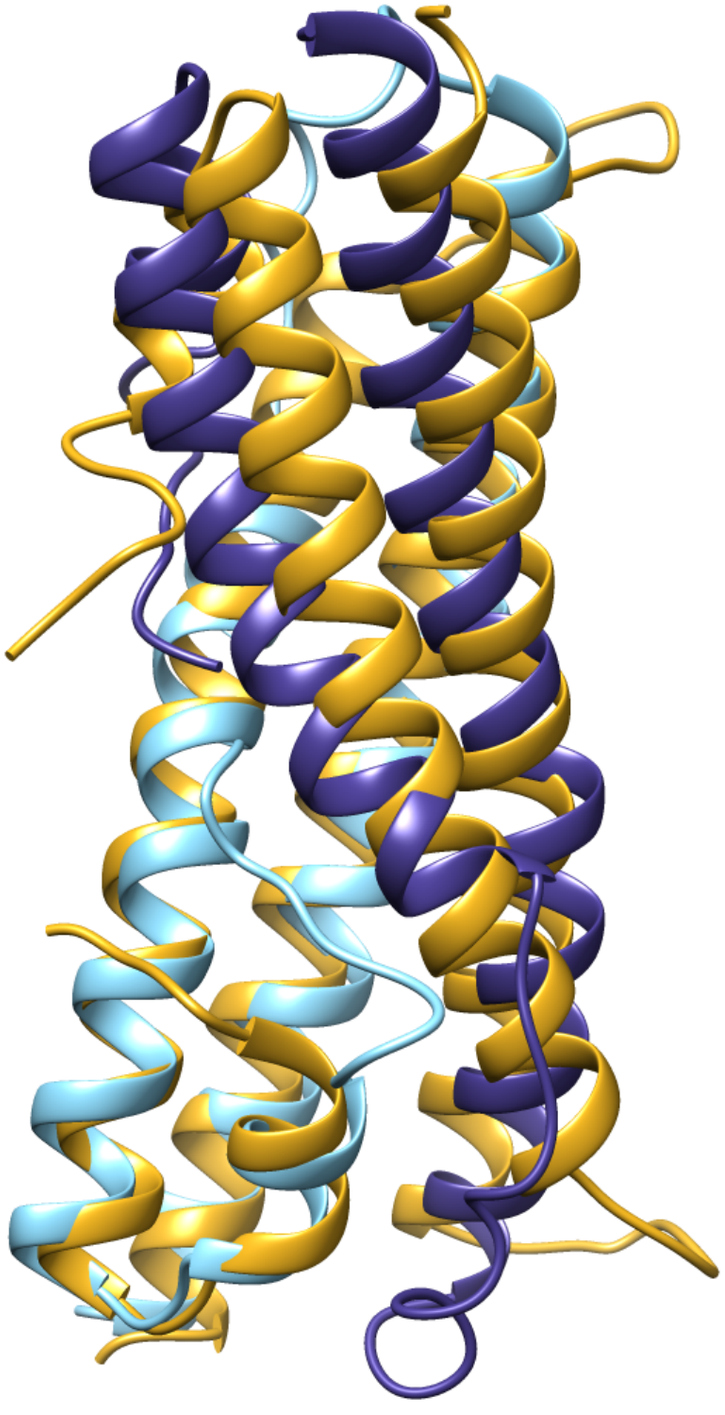
Monomer quality (FEIG)
RMSD: 1.63Å

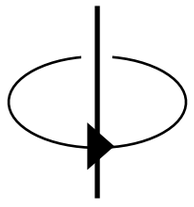
We didn't model the expression tag.

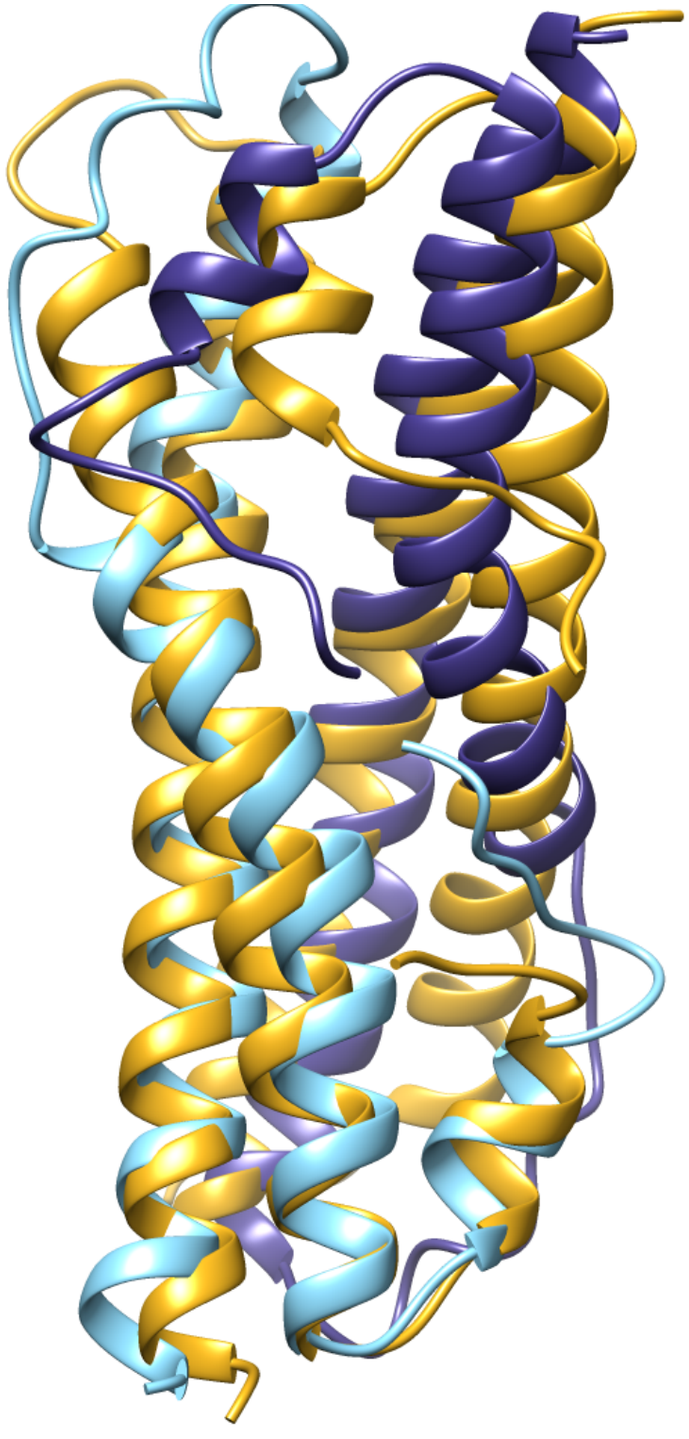
Crystal s1
Model s1
Crystal s2
Model s2

3. T1083 (Template-based docking and refinement)

T1083 (A_2)







Crystal
Model s1
Model s2

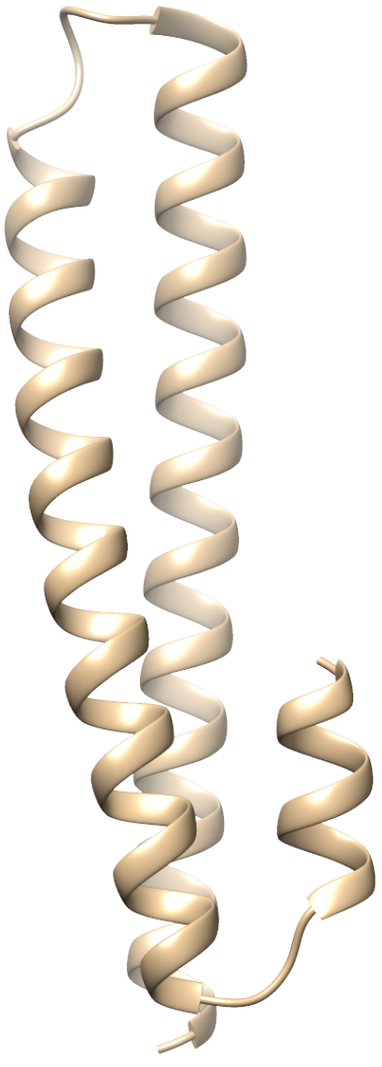
Model 1
F_{nat}: 0.6078
IRMSD: 2.51Å
LRMSD: 4.83Å

The only "Medium quality" model by server

Monomer quality (Seok_assembly) RMSD: 2.82Å

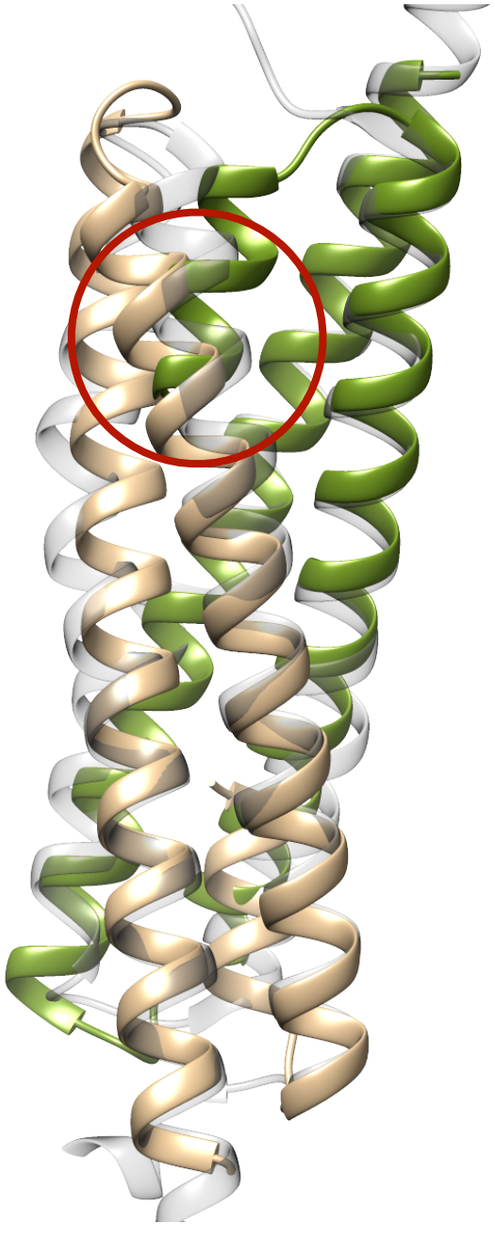
T1083 (A_2)

Monomer model



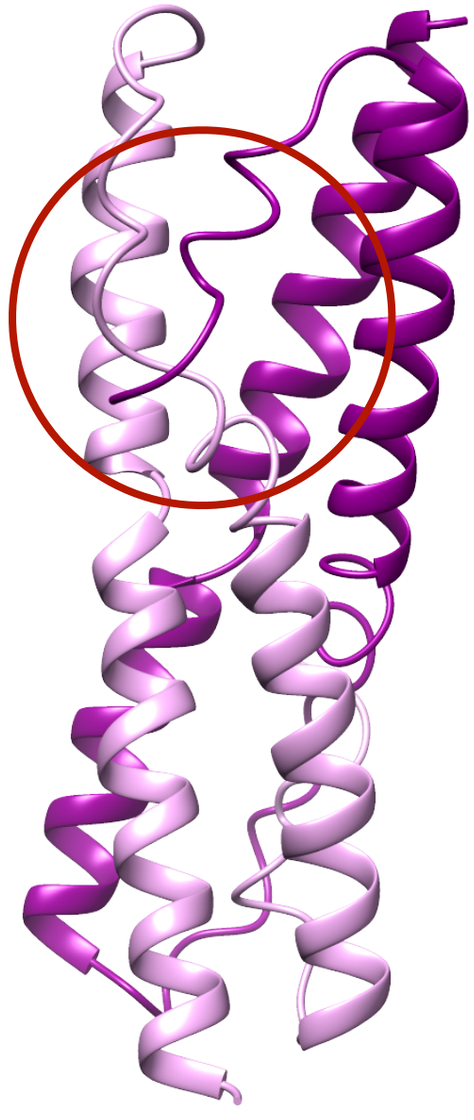
Monomer quality
(Seok_assembly)
RMSD: 2.82Å

Structure-based template



Superposed model s1
Superposed model s2

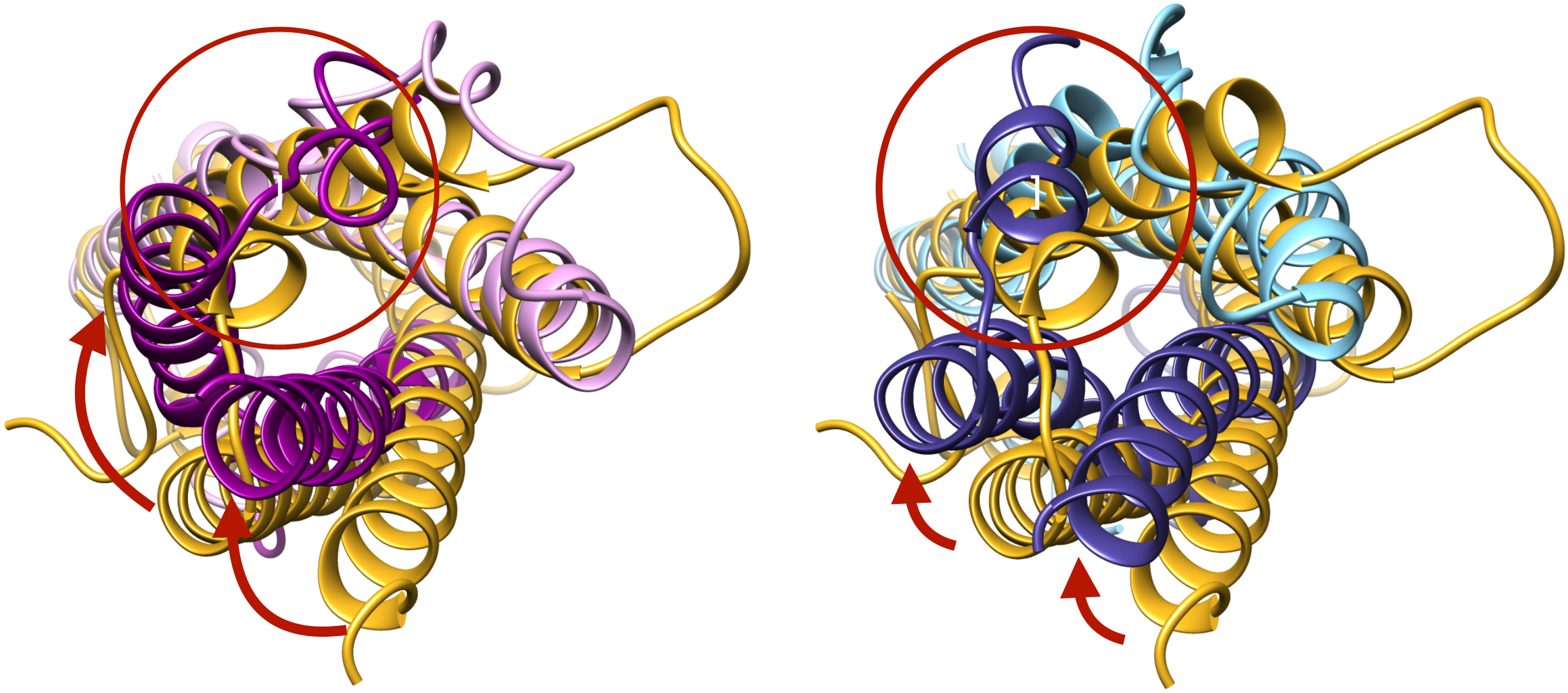
Relaxed structure



Relaxed model s1
Relaxed model s2

T1083 (A_2)

GalaxyRefineComplex



Before refinement

F_{nat}: 0.3922
IRMSD: 4.09Å
LRMSD: 10.147Å

"Incorrect quality"

Crystal
Model s1
Model s2
Model s1
Model s2

After refinement

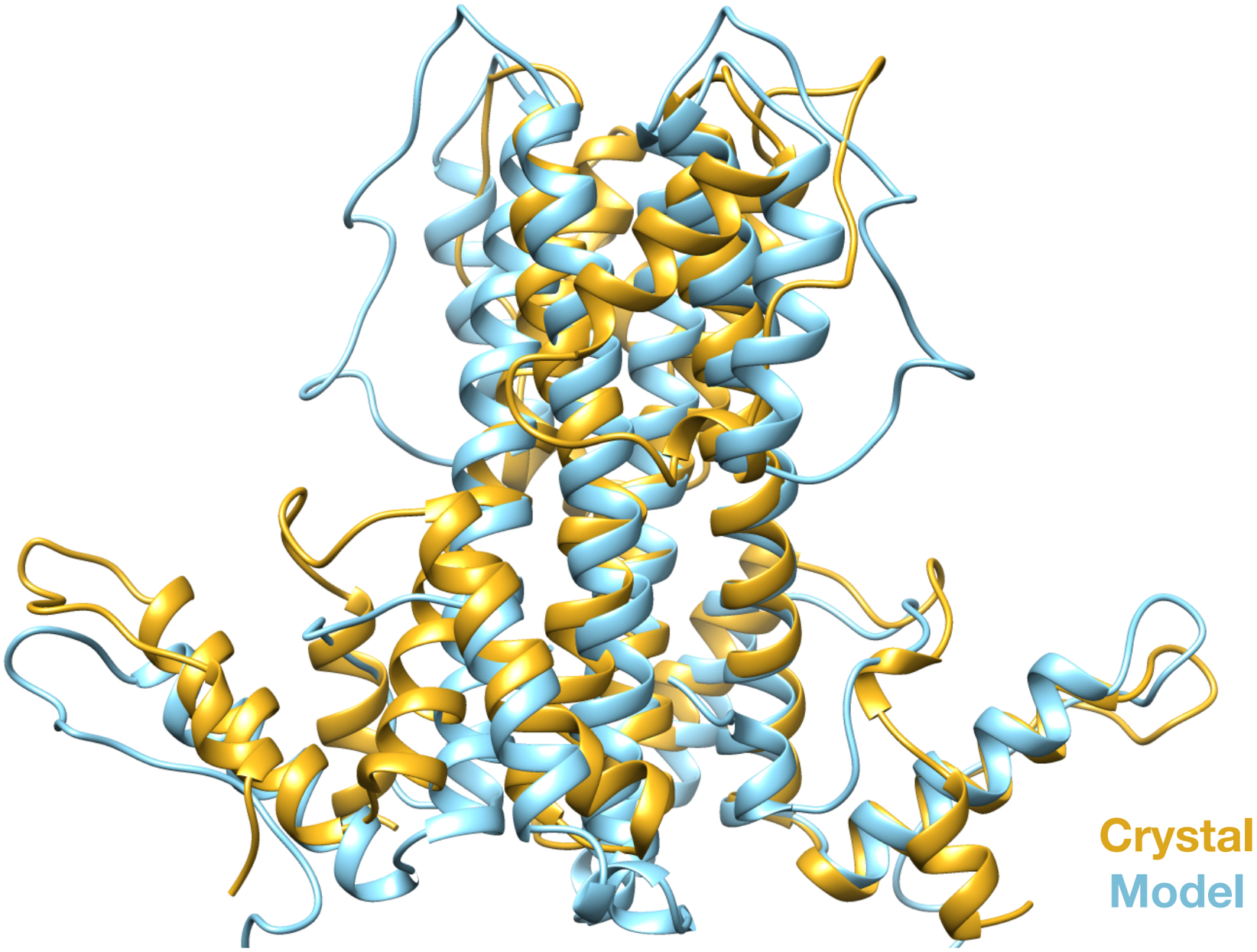
F_{nat}: 0.6078
IRMSD: 2.51Å
LRMSD: 4.83Å

"Medium quality"

4. T1099

(Template-based docking and data-assisted refinement)

T1099v1 (A_2)



Model 2

F_{nat}: 0.3398

IRMSD: 3.90Å

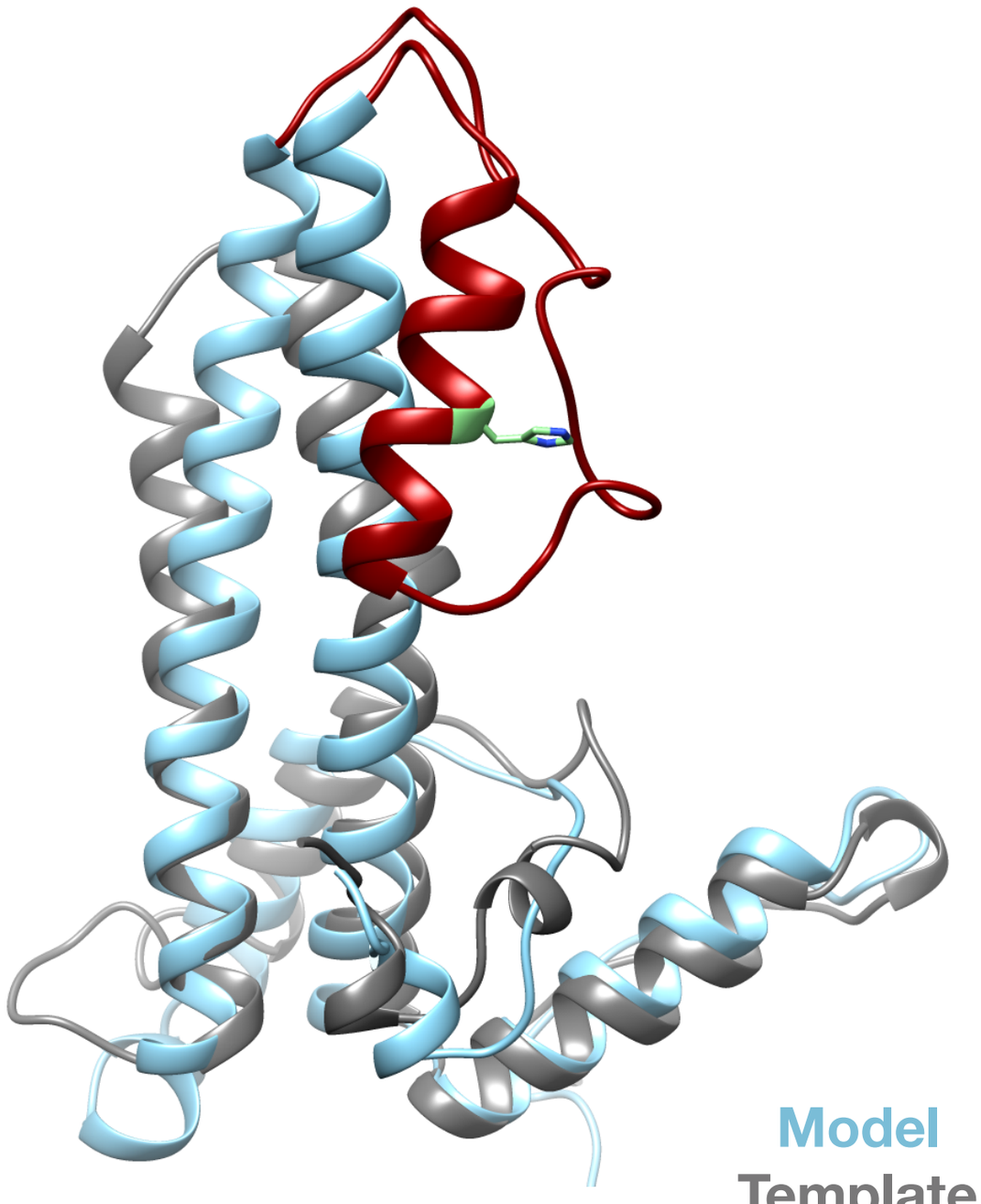
LRMSD: 8.04Å

The only "Acceptable quality"

Monomer modeling starts from (FALCON-TBM TS1) RMSD: 9.5Å

T1099v1 (A_2)

Relative orientation & insertion region for DHBc.



Template (PDB ID:3J2V) is available.
seqID*coverage = 12.1.

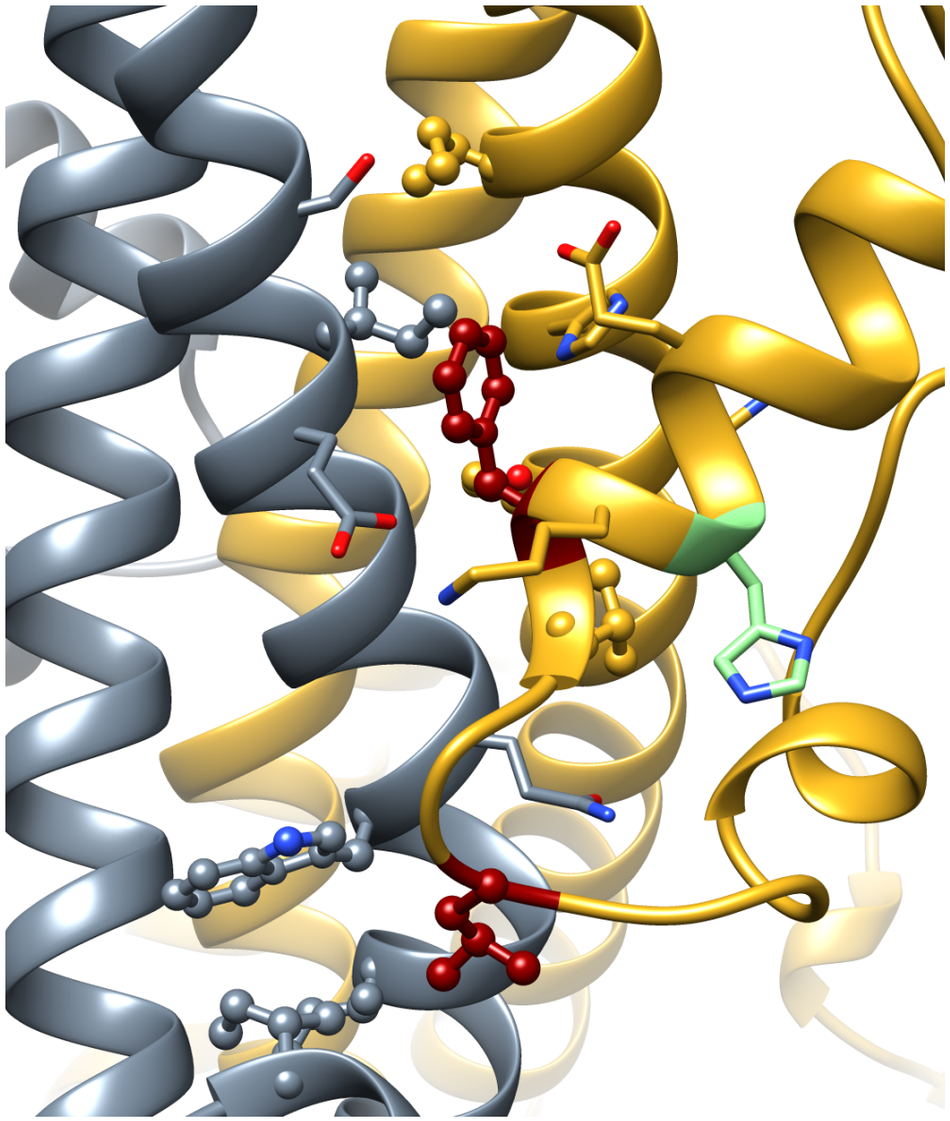
Template(HBV) vs Crystal(DHBc)

- Long insertion region at DHBC
- Different relative orientation of main helix.

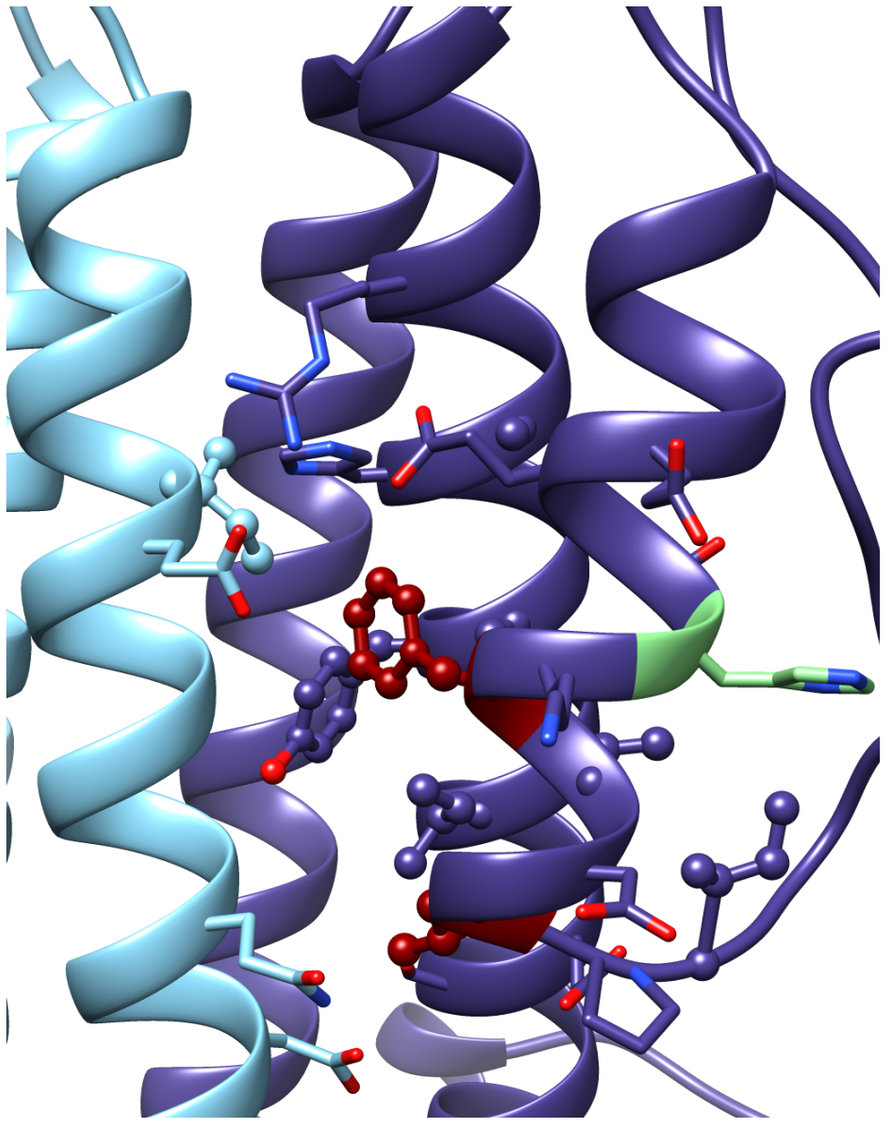
From literature search, information about the role and binding mode of insertion region for assembly was obtained.

T1099v1 (A_2)

Insertion region for DHBc.



Crystal structure



Our model

F_{nat}: 0.3398
IRMSD: 3.90Å
LRMSD: 8.04Å

"Acceptable quality"

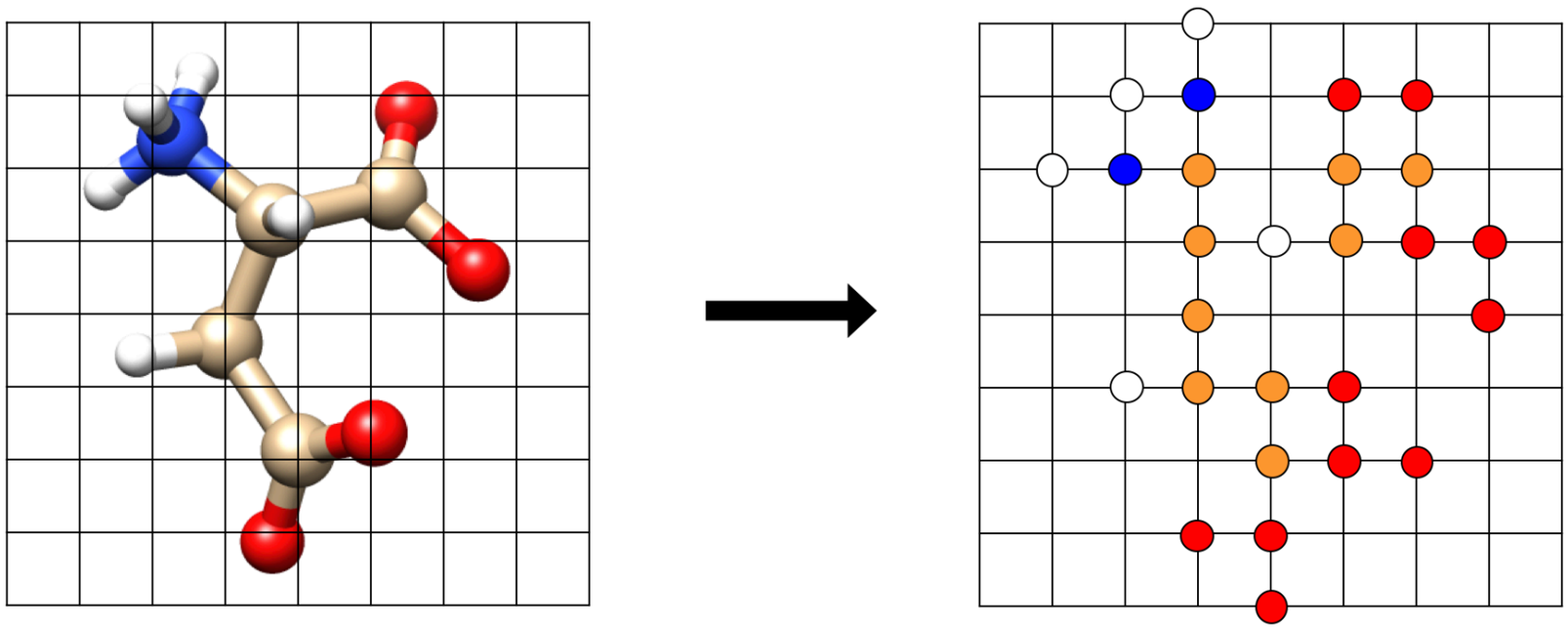
Monomer modeling starts from (FALCON-TBM TS1) RMSD: 9.5Å

Crystal s1
Crystal s2
Model s1
Model s2

A preliminary study of applying deep learning to ab initio docking

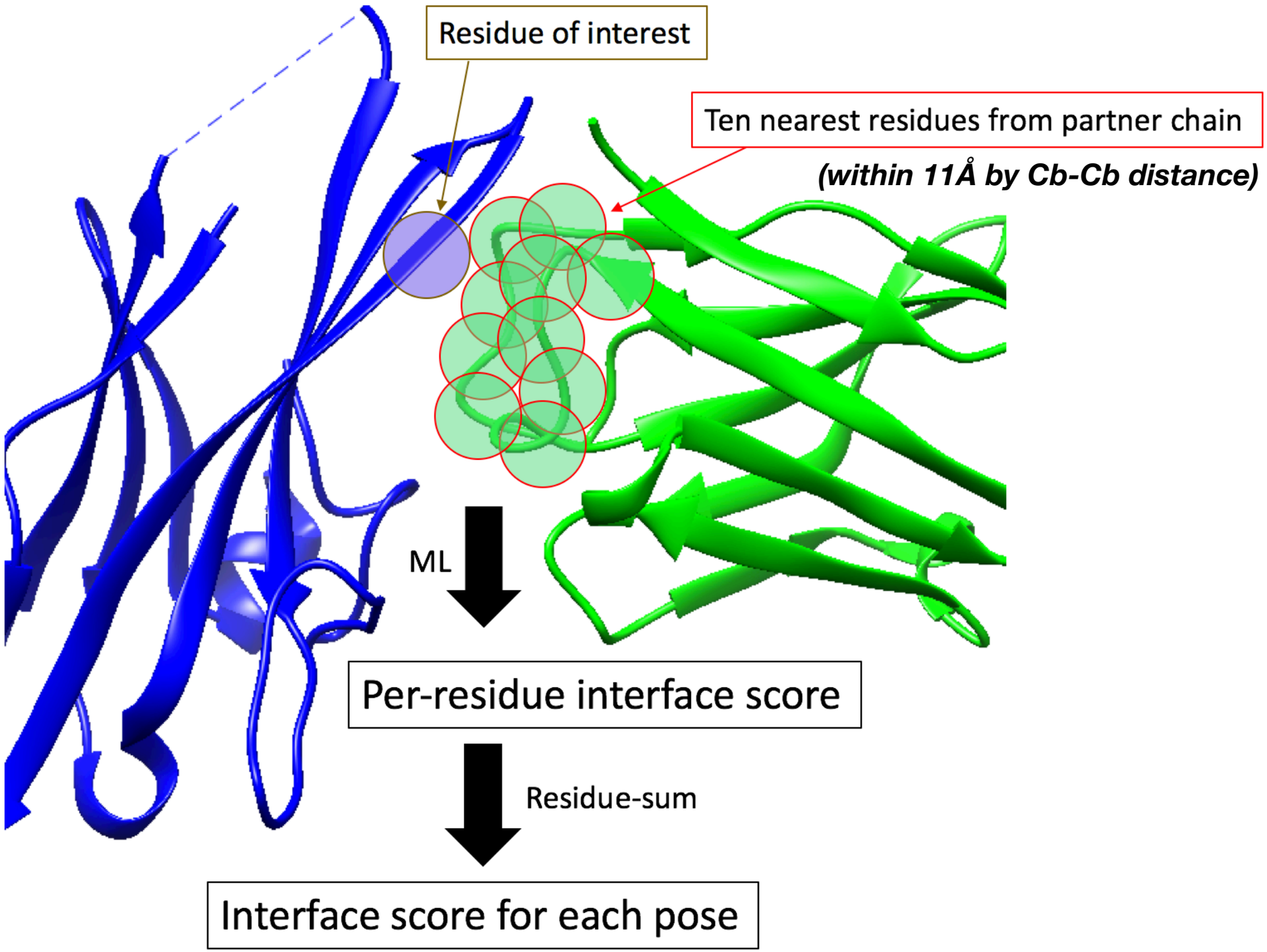
Re-scoring docking poses by deep learning

- **GalaxyTongDock**, a grid-based rigid-body docking, has limited scoring power.



- Deep learning-based re-scoring method has been developed.

Simple interface scoring by deep learning



Deep learning-based re-scoring improves performance of *ab initio* docking

Non-redundant 1,568 hetero-dimers
> 1,156 (training set) and 412 (test set)

Performance on 412 hetero-dimers in test set

TongDock	***	**	*
Top 1	0.49	6.1	11.2
Top 10	1.70	12.9	23.5
Top 100	2.91	22.8	50.5
Top 1,000	3.40	38.6	81.1

DL	***	**	*
Top 1	**0.73**	**10.9**	**25.5**
Top 10	1.70	**22.3**	**41.0**
Top 100	2.91	**36.4**	**60.9**
Top 1,000	**4.37**	**49.0**	**85.2**

Take-home message

1. ***Improved monomer quality improved performance of ab initio docking.***

2. ***Interface refinement can improve oligomer model quality.***

3. ***Advent of deep learning is changing everything.***

Thank you for listening