

CASP14 Assembly Assessment

Ezgi Karaca

Computational Structural Biology Lab

Izmir Biomedicine and Genome Center (IBG)

ibg.edu.tr

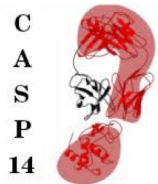
A big thanks to:



Burcu Ozden



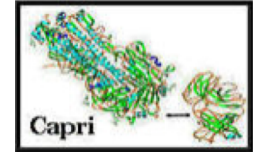
Andriy Kryshatafovych



CASP14 Organizing Team



Marc Lensink



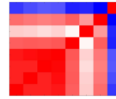
Shoshana Wodak

Outline

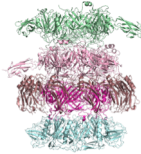


- Overview of the CASP14 assembly round

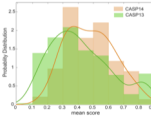
- CASP14 assembly scoring



- CASP14 target highlights



- CASP13 vs CASP14



- Challenges & Perspectives



Outline

- Overview of the CASP14 assembly round
- CASP14 assembly scoring
- CASP14 target highlights
- CASP13 vs CASP14
- Challenges & Perspectives



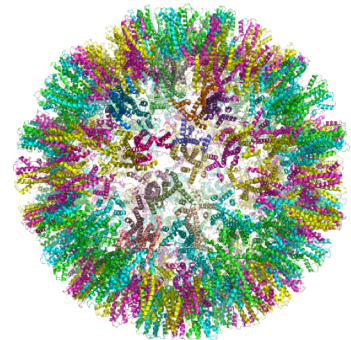
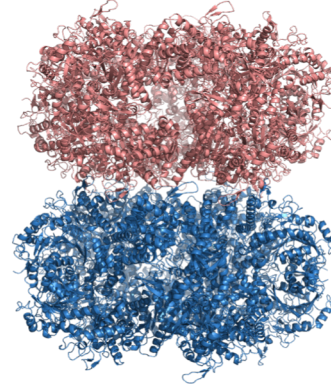
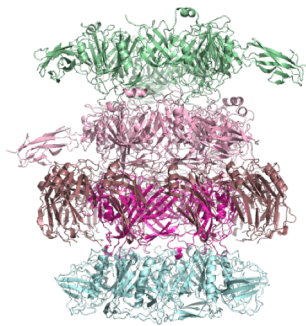
Target Classification

Easy → complete template for subunit AND assembly → **TBM**

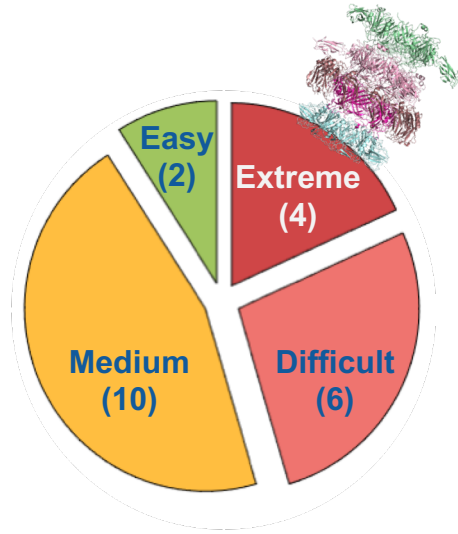
Medium → partial template for subunit AND/OR assembly → **TBM/FM**

Difficult → no template for subunit AND assembly → **FM**

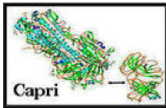
Extreme →



Target Classification

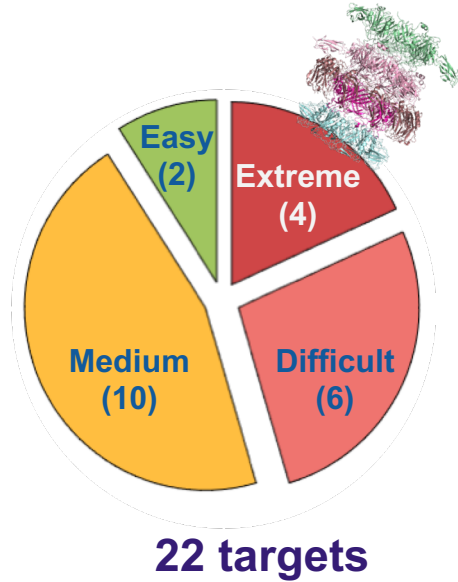


22 targets

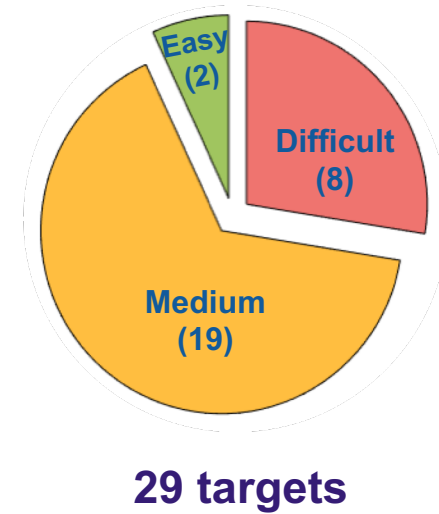


12 targets

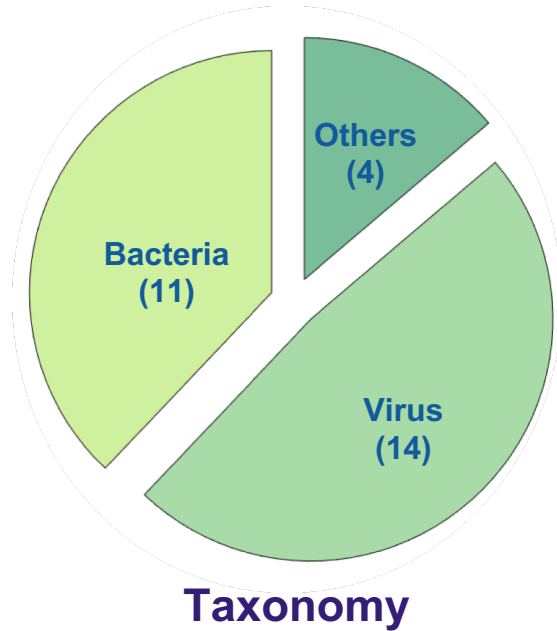
Target Classification



Introduction of sub-targets



Further statistics



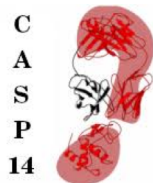
Over 29 targets

EM	52%
homomers	70%
>2-mer	62%
# groups	39
models/target	110

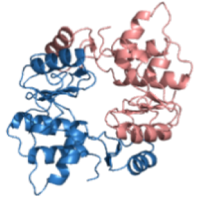
All stoichiometries were provided except for T1047.

Outline

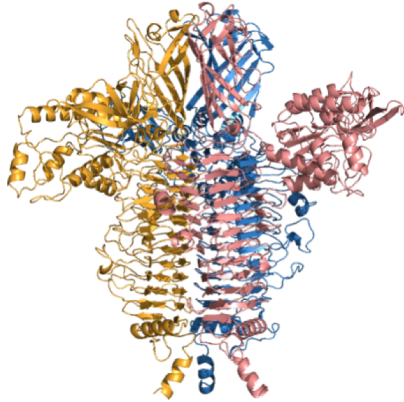
- CASP14 Assembly Targets
- Scoring metrics used
- CASP14 target highlights
- CASP13 vs CASP14
- Challenges & Perspectives



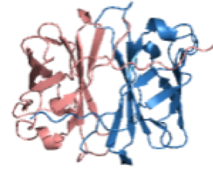
Homomers



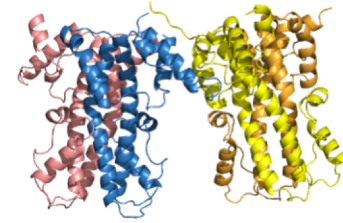
T1032o/6n64
Easy
C2



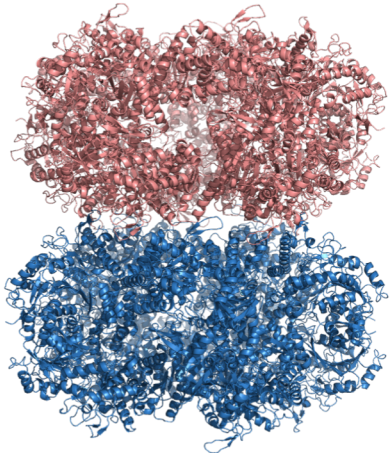
T1052o
Easy
C3



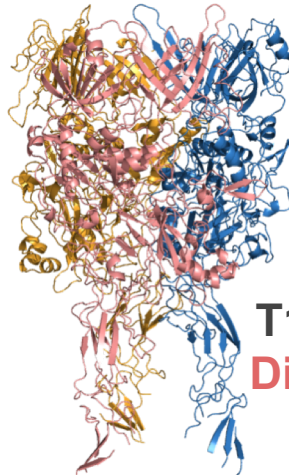
T1078o
Medium
C2



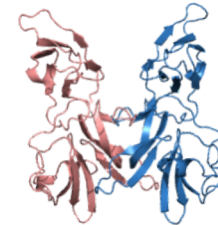
T1099o/6ygh
Medium
240-mer
T4 icosahedral



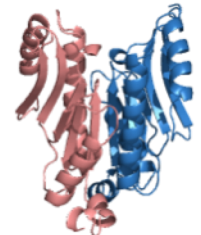
H1081
Medium
20-mer
D5



T1061o
Difficult
C3

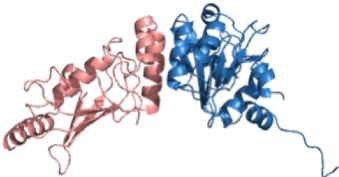


T1038o/6ya2
Difficult
C2

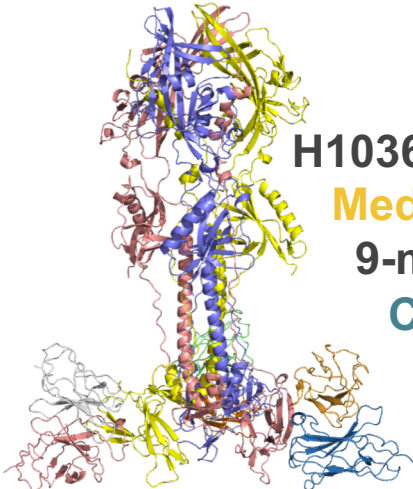


T1054o
Difficult
C2

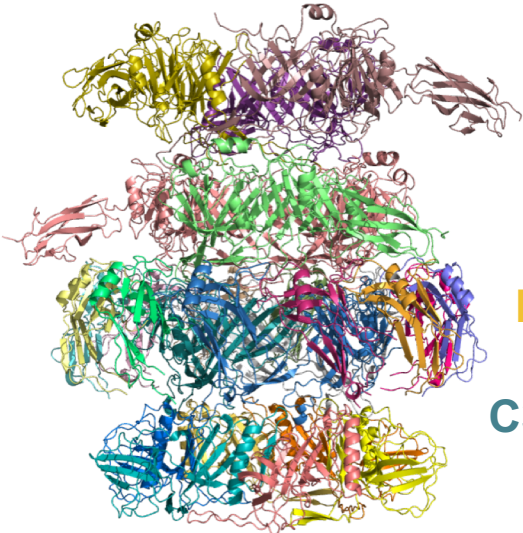
Heteromers



H1045
Medium



H1036/6vn1
Medium
9-mer
C3



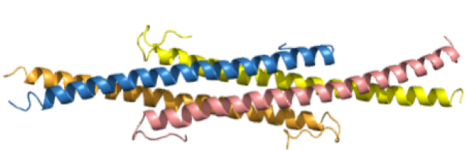
H1060
Medium
27-mer
C3/C12/C6



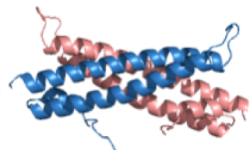
H1065
Difficult

H1047
Difficult
multi-mer

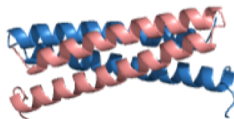
Coiled Coils



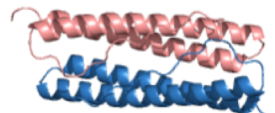
T1048o
Medium
D2



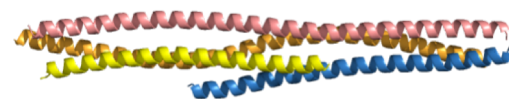
H1083
Medium
C2



H1084
Medium
C2

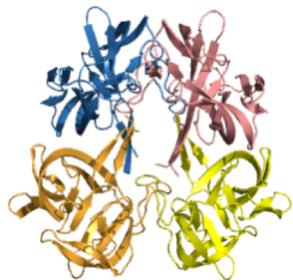


H1087
Medium
C2

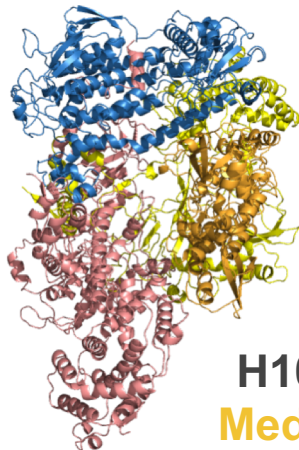


H1072/6r17
Medium
C2

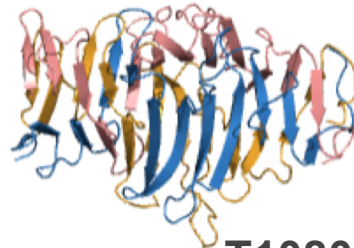
Intertwined Monomers



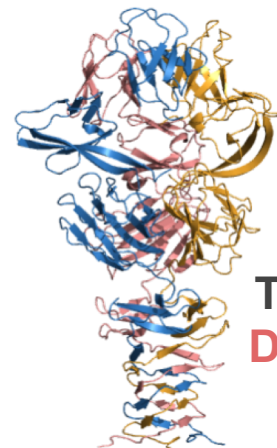
T1034o
Medium
D2



H1097
Medium



T1080o
Difficult
C3



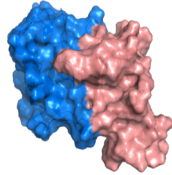
T1070o
Difficult
C3

Outline

- Overview of CASP14 assembly round
- CASP14 assembly scoring 
- CASP14 target highlights
- CASP13 vs CASP14
- Challenges & Perspectives

Assessment Scores

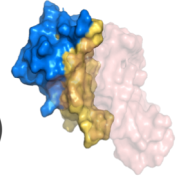
Assembly/Global:
Oligomeric IDDT



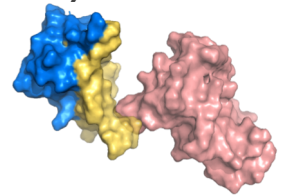
TM Score (Topological
similarity)

Interface/Local:

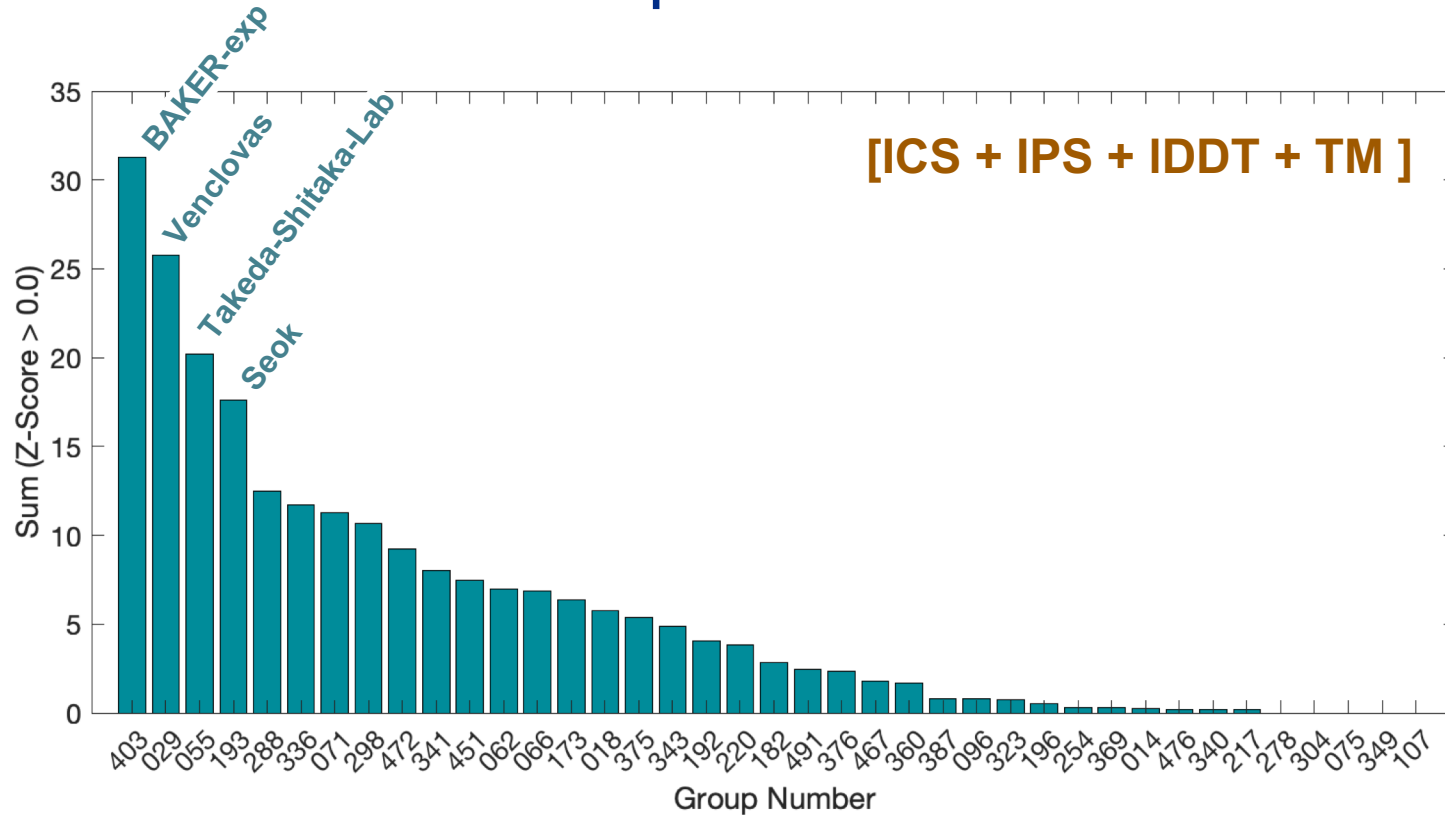
Interface Contact Score (ICS) - (F1)



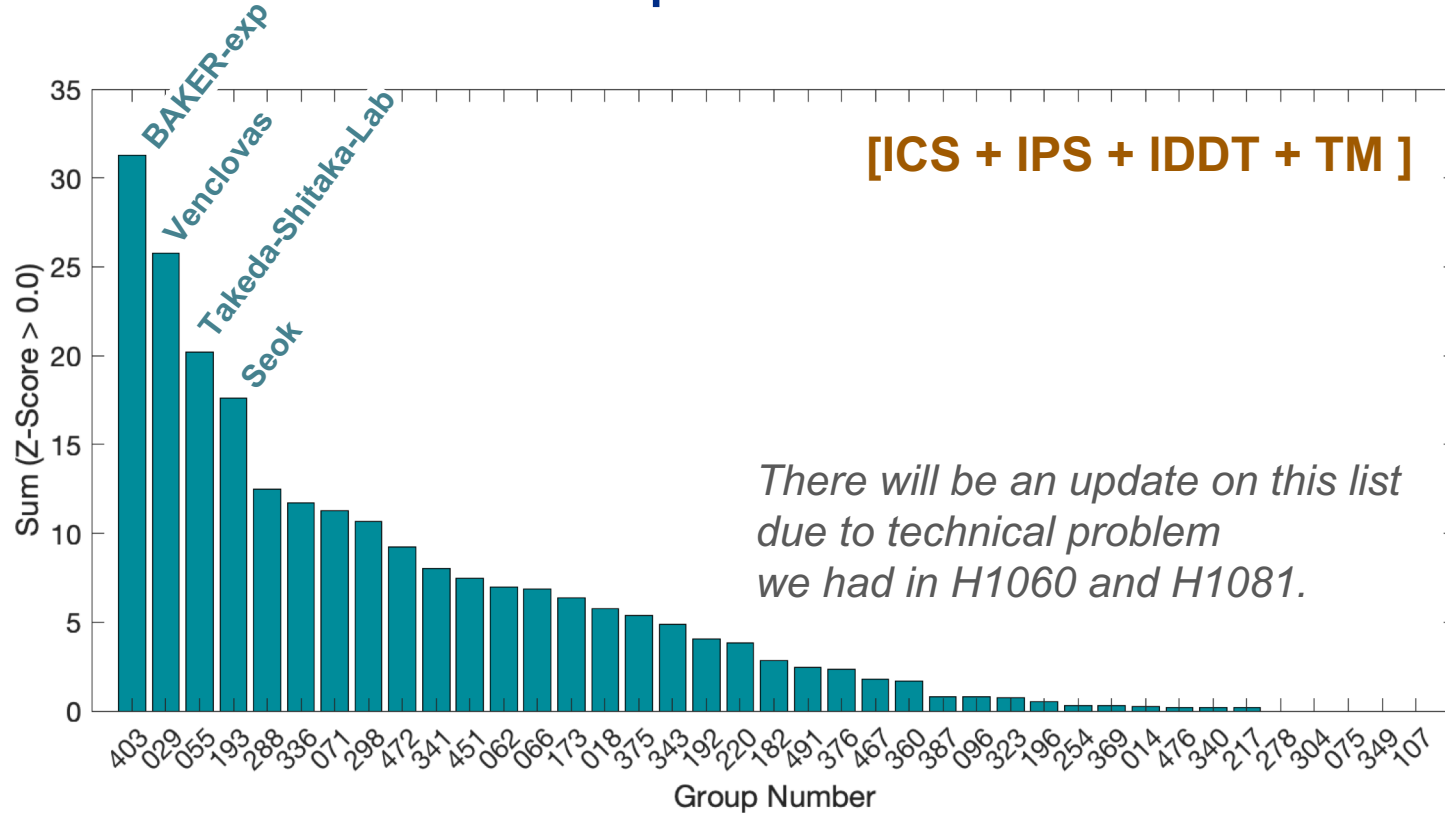
Interface Patch Score (IPS) - (Jaccard)



Z-score based overall performance

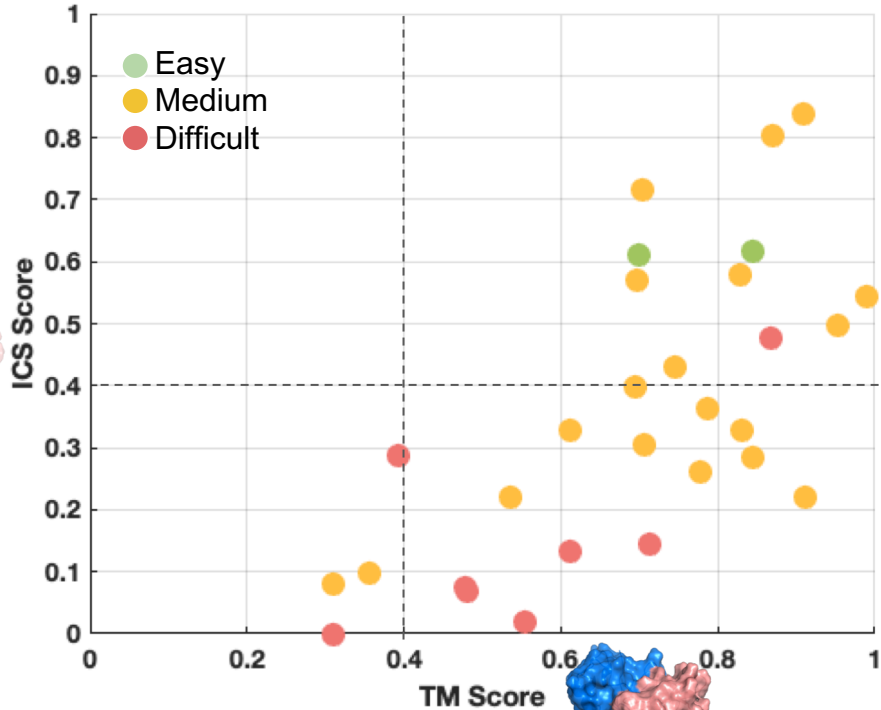


Z-score based overall performance



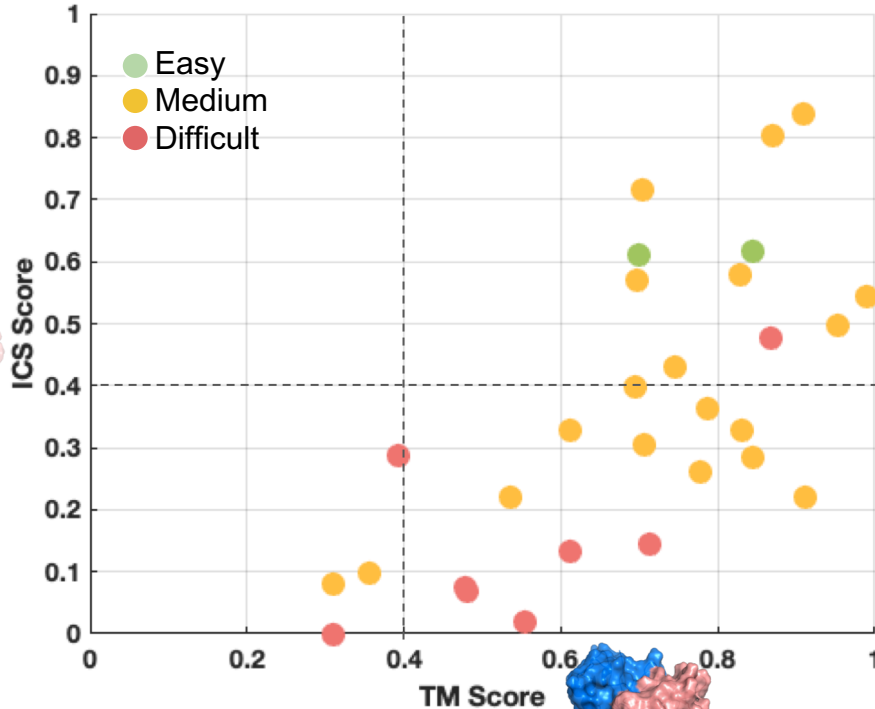
How well are the targets predicted?

ICS & TM distributions of the best models



How well are the targets predicted?

ICS & TM distributions of the best models



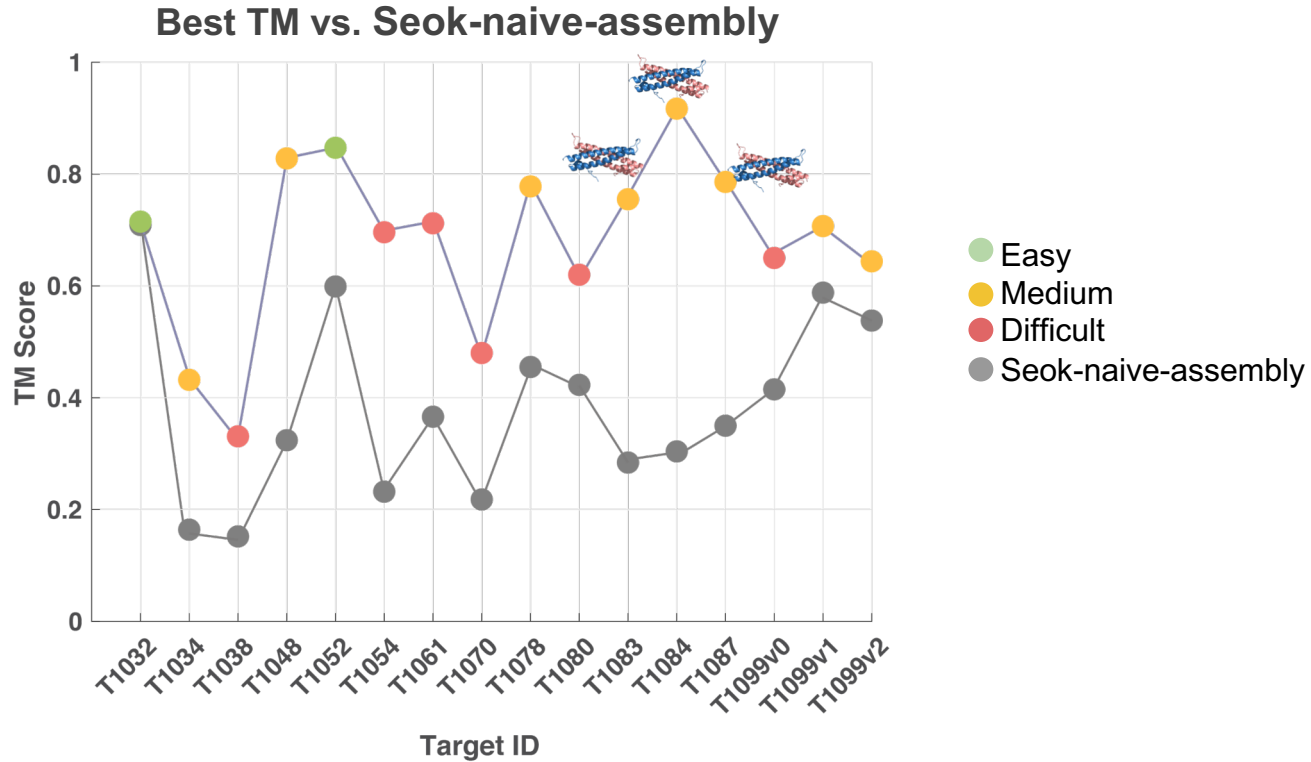
ICS > 0.4 → 38%

TM > 0.4 → 86%

Globular fold prediction ✓

Interface contact prediction !

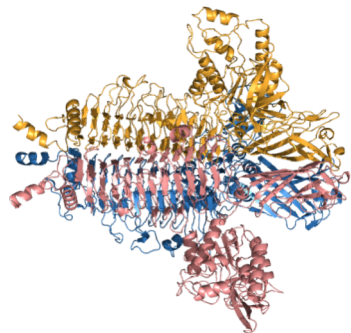
How well are the homomeric TMs predicted?



Outline

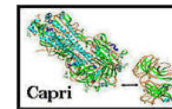
- Overview of the CASP14 assembly round
- CASP14 assembly scoring
- CASP14 target highlights 
- CASP13 vs CASP14
- Challenges & Perspectives

Easy Homomer

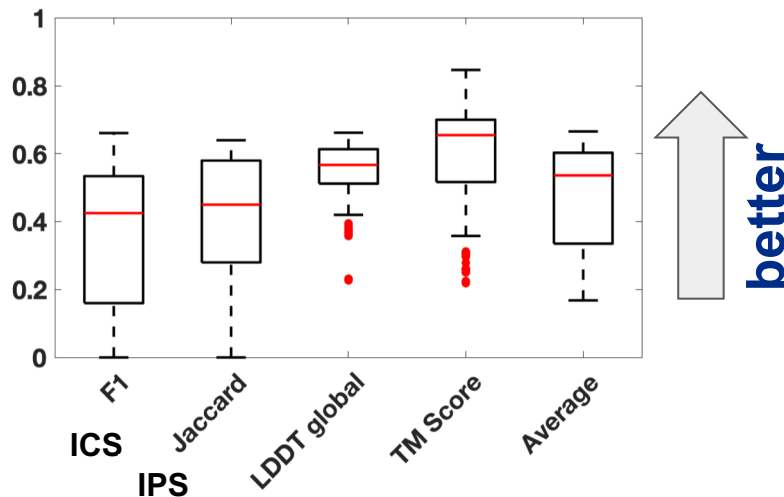


Salmonella virus
Host recognition

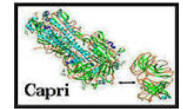
T1052o
C3



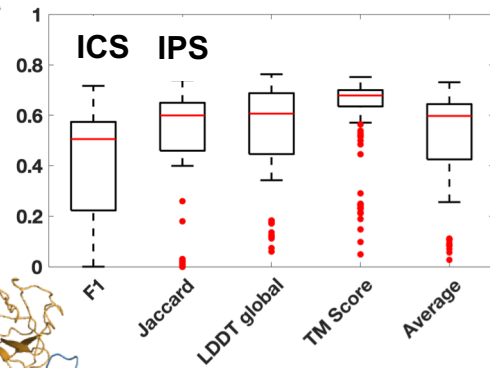
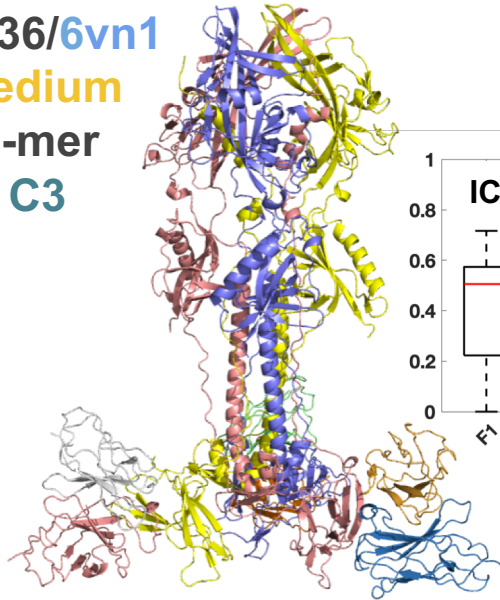
C
A
S
P
14



Best ICS (top 5)	Best TM (top 5)
Bates_BMM Venclovas Vakser Zou	Seok Vakser



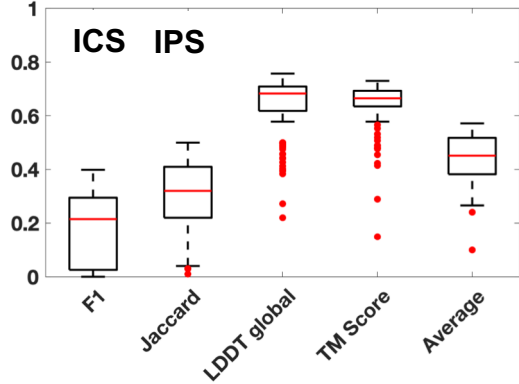
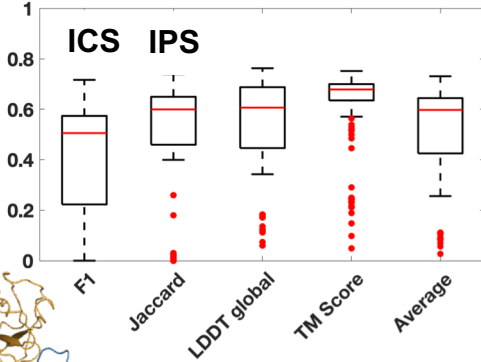
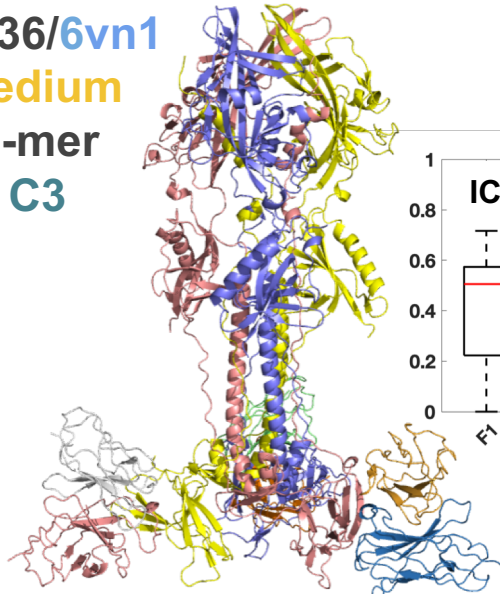
H1036/6vn1
Medium
9-mer
C3



Glycoprotein B-Neutralizing Antibody Complex with Virus

Best ICS (top 5)	Best TM (top 5)
BAKER-exp. Htjcadd	Zou Bio3d

H1036/6vn1
 Medium
 9-mer
 C3



Glycoprotein B-Neutralizing Antibody Complex with Virus

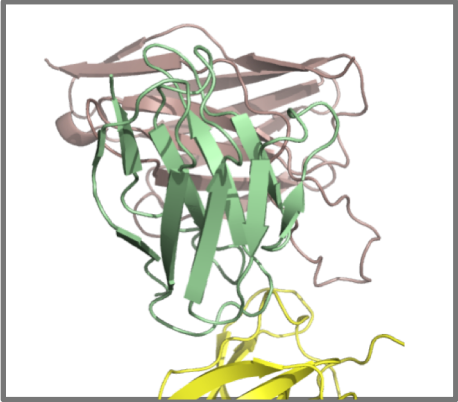
Best ICS (top 5)	Best TM (top 5)
BAKER-exp. Htjcadd	Zou Bio3d

Best ICS (top 5)	Best TM (top 5)
Htjcadd Fernandez-Recio Zou	Zou BAKER-exp. bio3d

H1036v0/6vn1

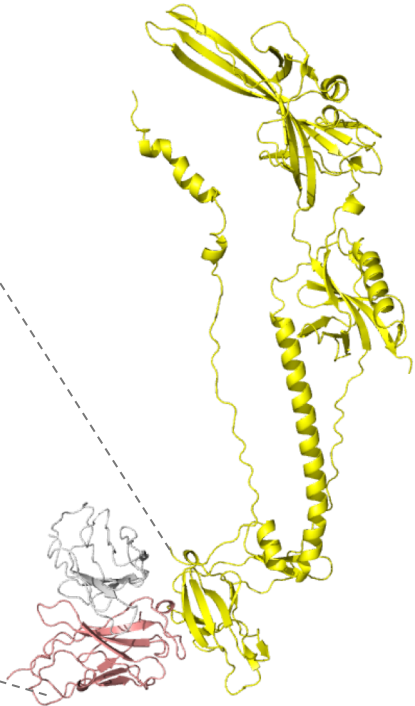
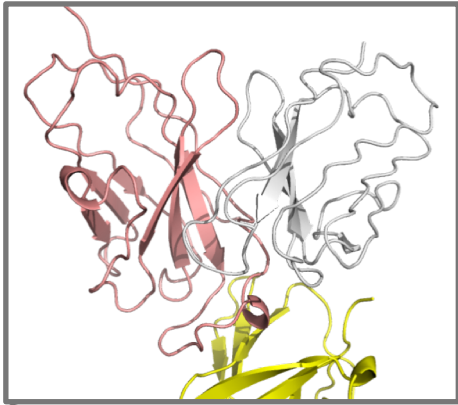
Medium

3-mer

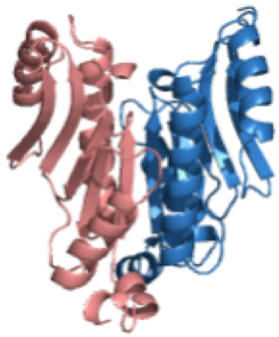


Best TM

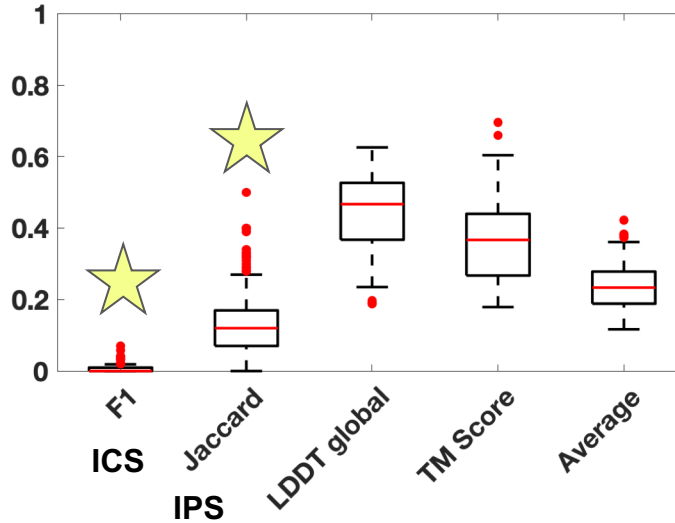
Zou



Size dependency of the scores!



T1054o
Difficult
C2



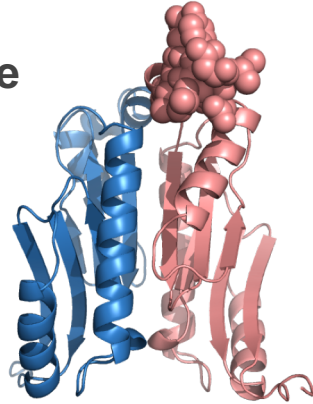
WHY?

Best ICS (top 5)	Best TM (top 5)
LamoureuxLab SBROD-Select Date Khiaralab Zou	VoroCNN-Select Kiharalab LamoureuxLab Date Venclovas

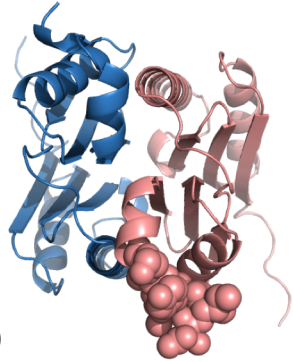
#	Name	10	20	30	40	50	60	70	80	90	100	110	120	130	140	15
1	T1054TS427_1	[Progress bar with yellow and green segments]														

reference

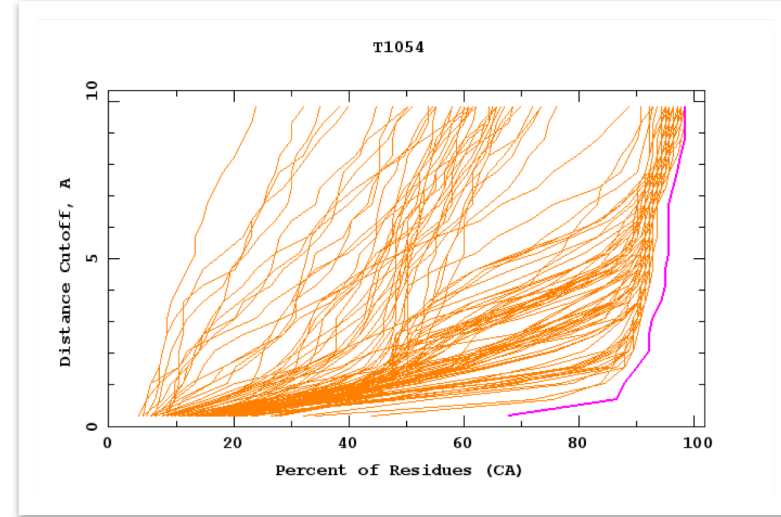
side



90°



top

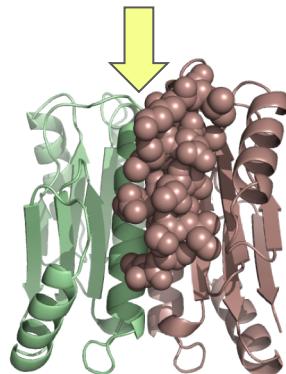


#	Name	10	20	30	40	50	60	70	80	90	100	110	120	130	140	15
1	T1054TS427_1															

reference

best model

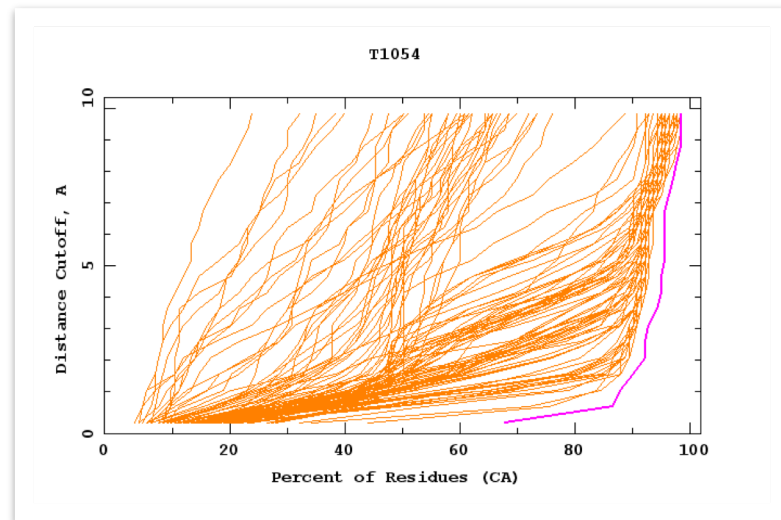
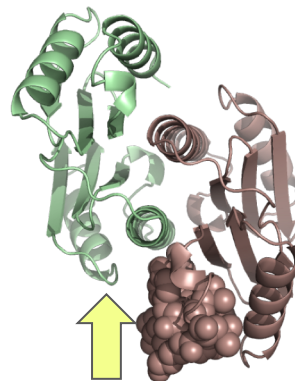
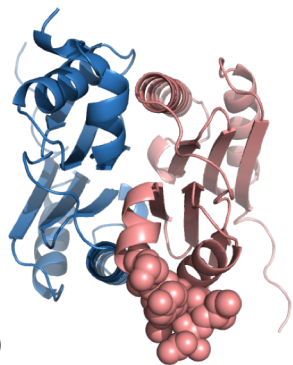
side



90°



top

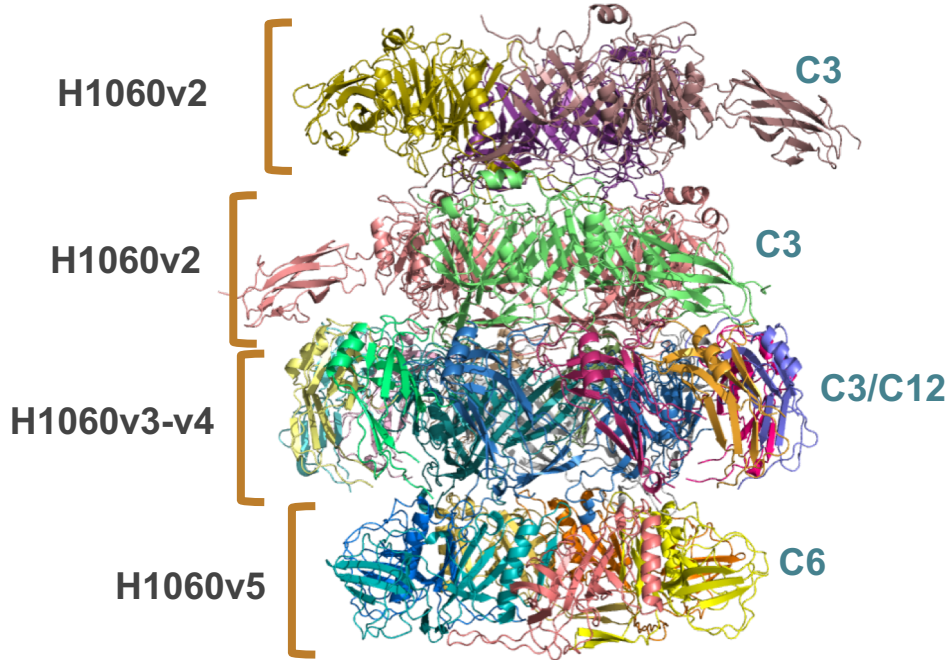


WHY?

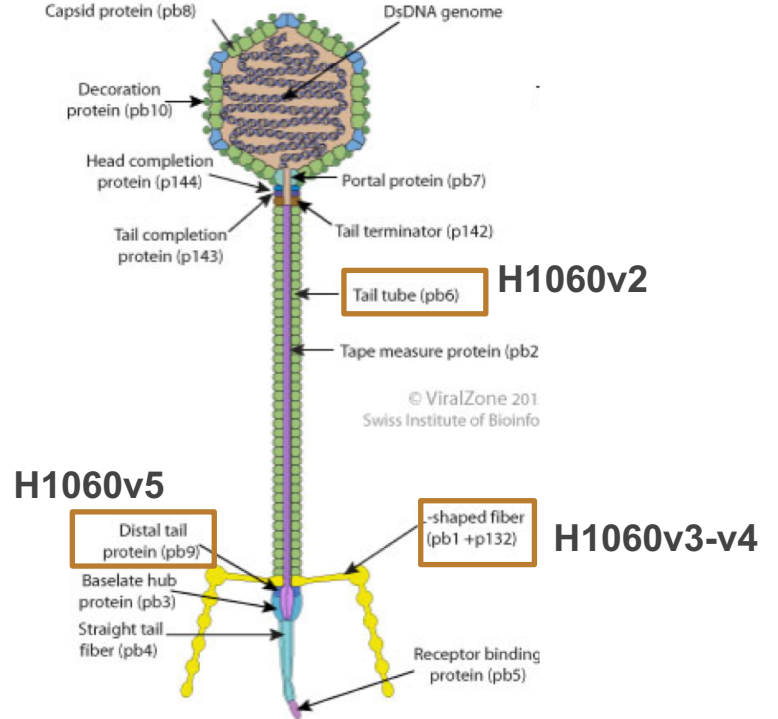
1st 15 amino acids!

H1060
Medium
27-mer

<https://viralzone.expasy.org/511>



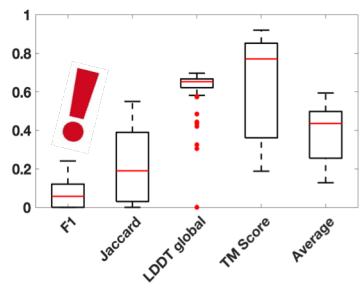
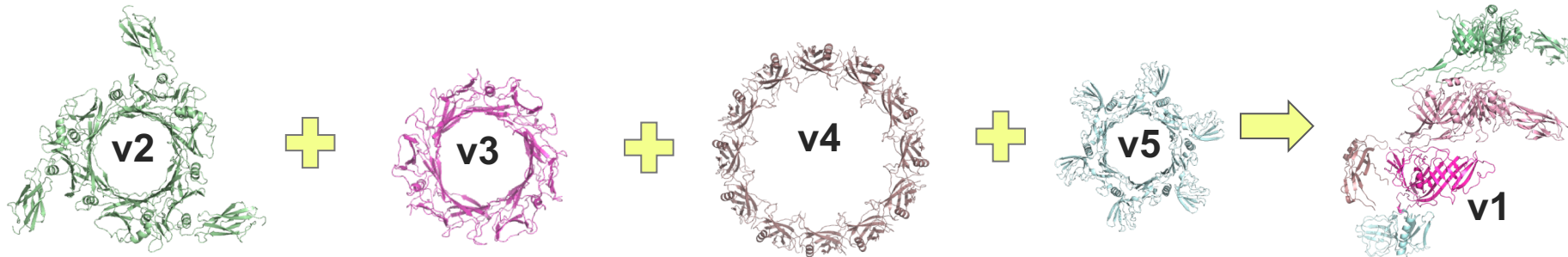
**Bacteriophage
T5 Phage Tail**



H1060 - 5 subtargets

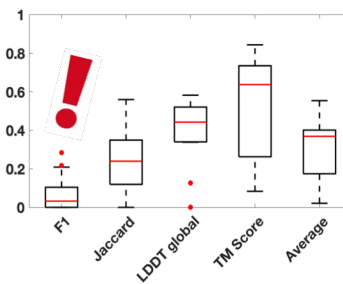
Medium

max contacting
min_zone



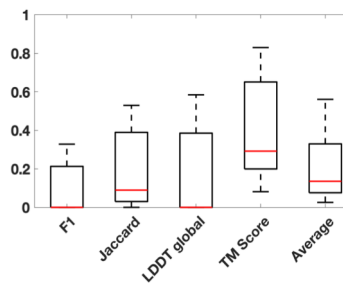
Best TM

Venclovas



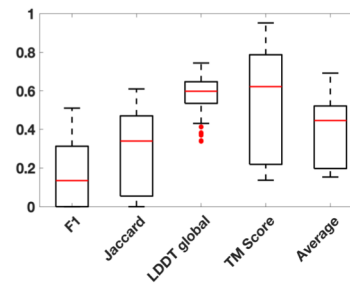
Best TM

Venclovas



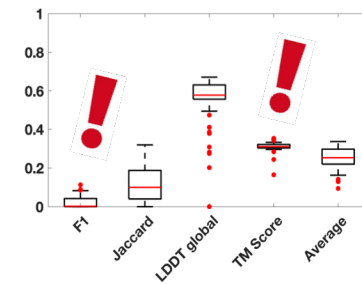
Best TM

Takeda-Shitaka-Lab



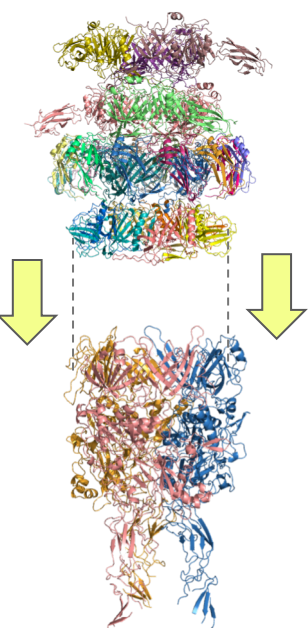
Best TM

BAKER-exp.

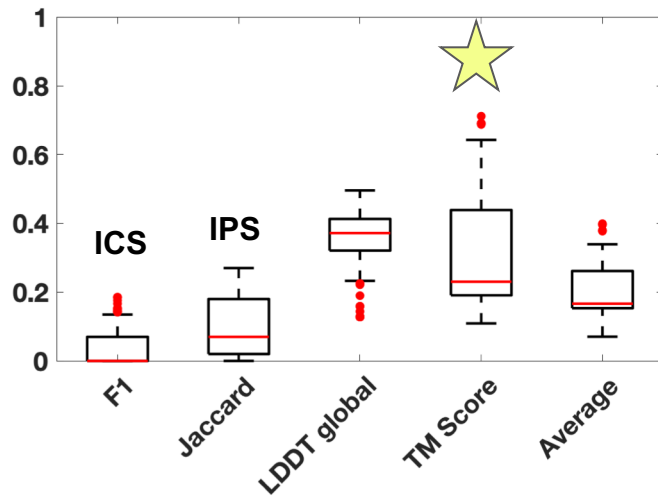


Best TM

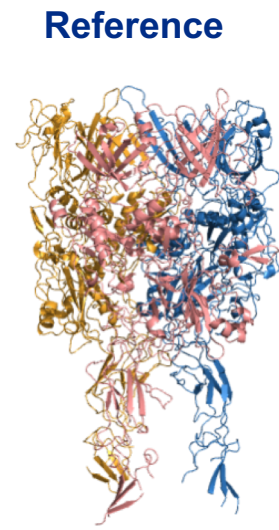
Risoluto



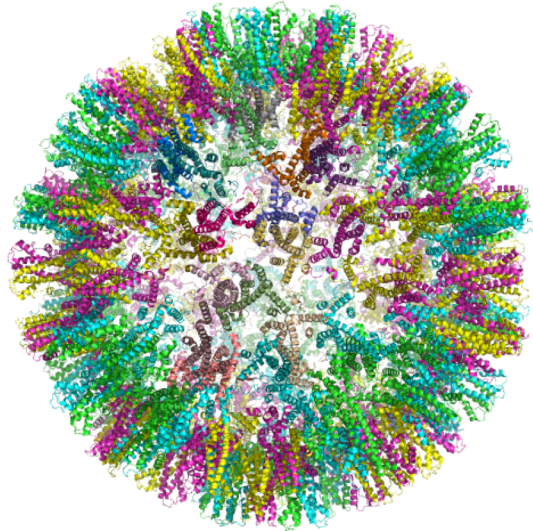
T1061o
Difficult
C3



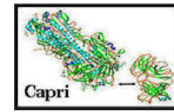
Best ICS (top 5)	Best TM (top 5)
Venclovas BAKER-exp.	BAKER-exp.



T1099o/6ygh
Medium
240-mer
T4 icosahedral

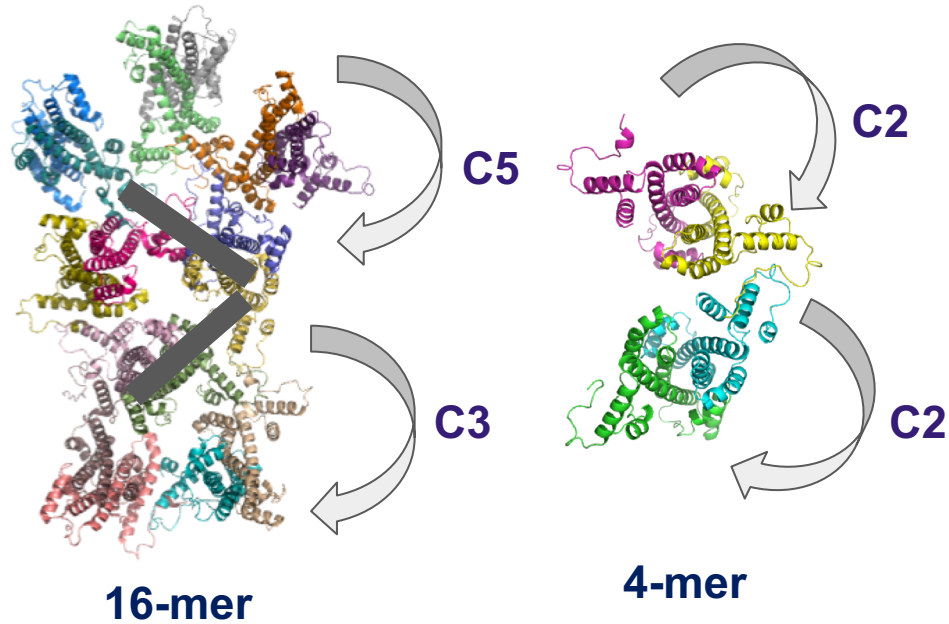


Duck hepatitis B virus



The target features $T=4$ icosahedral symmetry with a total of 240 subunits. Please submit the minimum number of subunits necessary to define the unique interfaces.

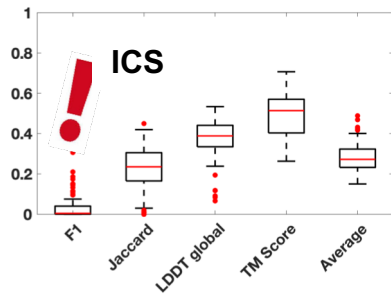
How we defined the subtargets:



T1099o/6ygh

Medium

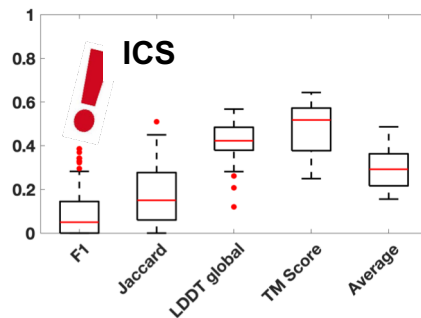
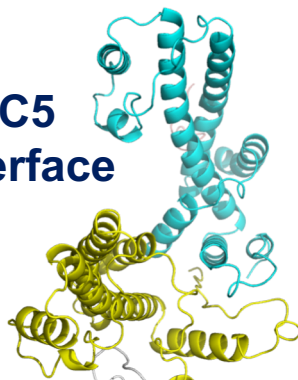
C2 dimer
v1



Best TM

Seok

C3/C5
Interface
v2

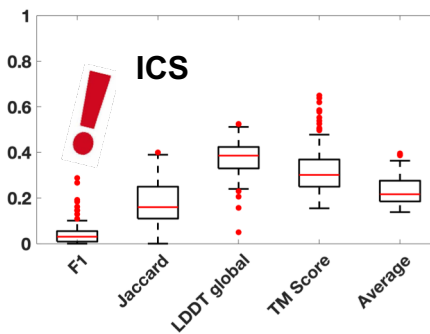
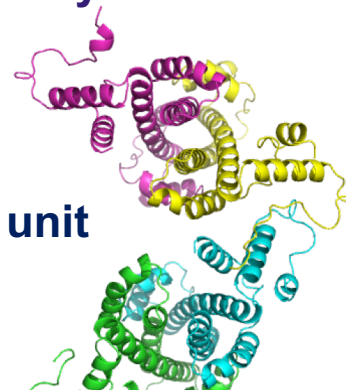


Best TM

CAPRI-Shen

selected by # of contacts

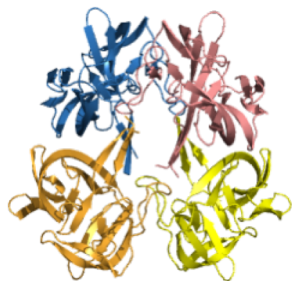
min unit
v0



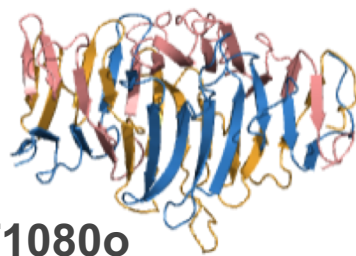
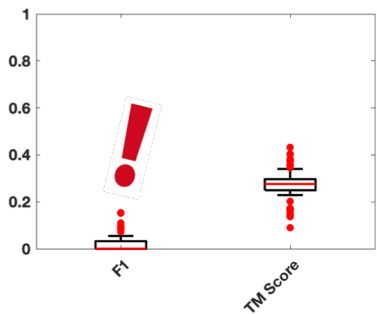
Best TM

CAPRI-Shen

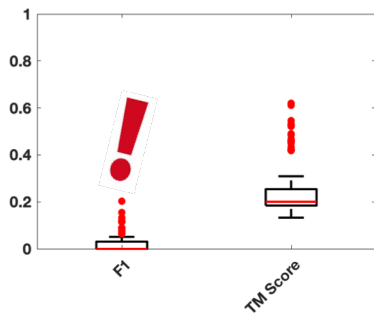
Intertwined Monomers



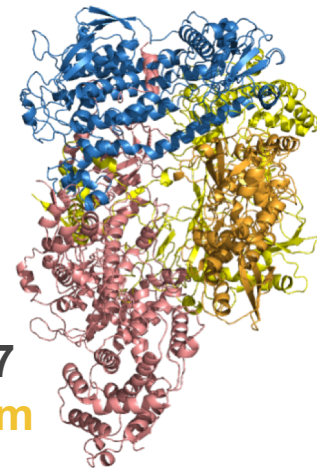
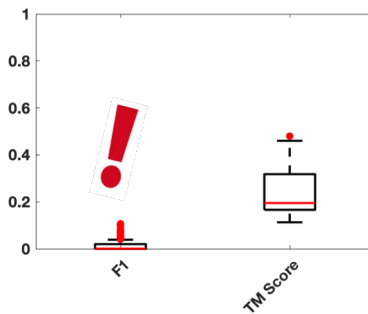
T1034o
Medium
D2



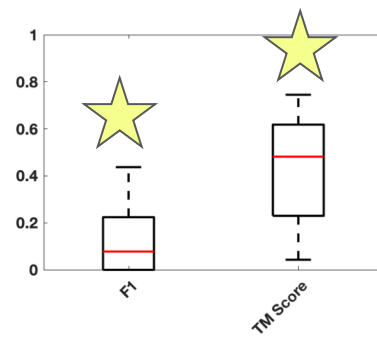
T1080o
Difficult
C3



T1070o
Difficult
C3




H1097
Medium



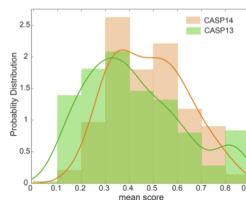
In summary

Why  - good template for the interface

Why  - partial/no template for the interface
Template of different conformation
Existence of domain swap
Too heterogeneous & too large systems

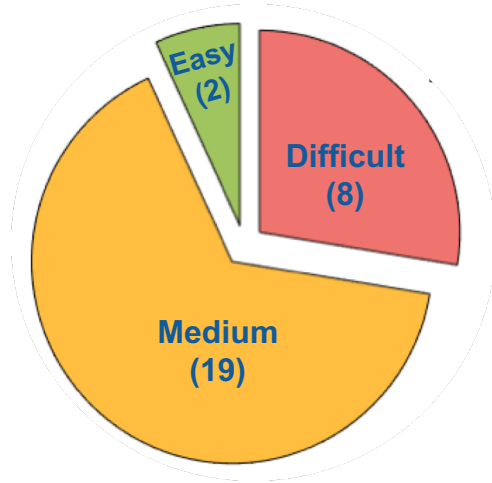
Outline

- Overview of the CASP14 assembly round
- CASP14 assembly scoring
- CASP14 target highlights
- **CASP13 vs CASP14**
- Challenges & Perspectives



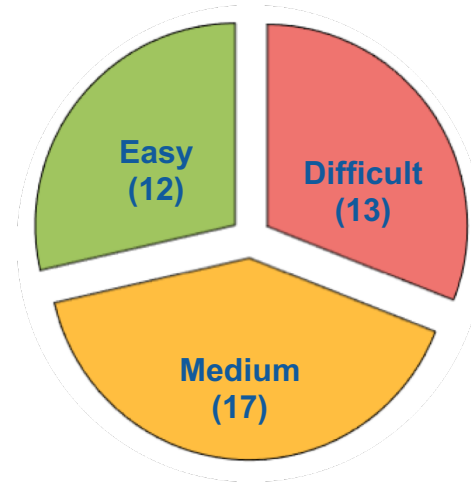
How was it in CASP13?

CASP14



29 targets

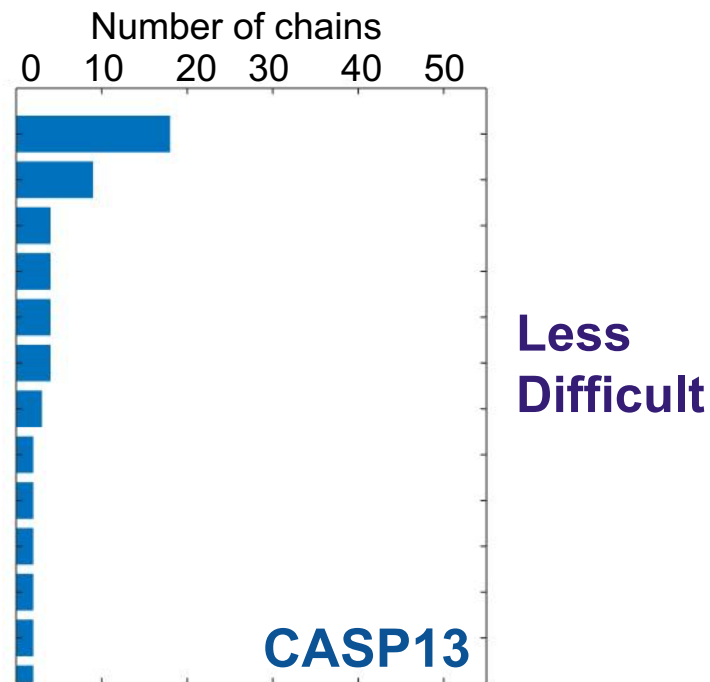
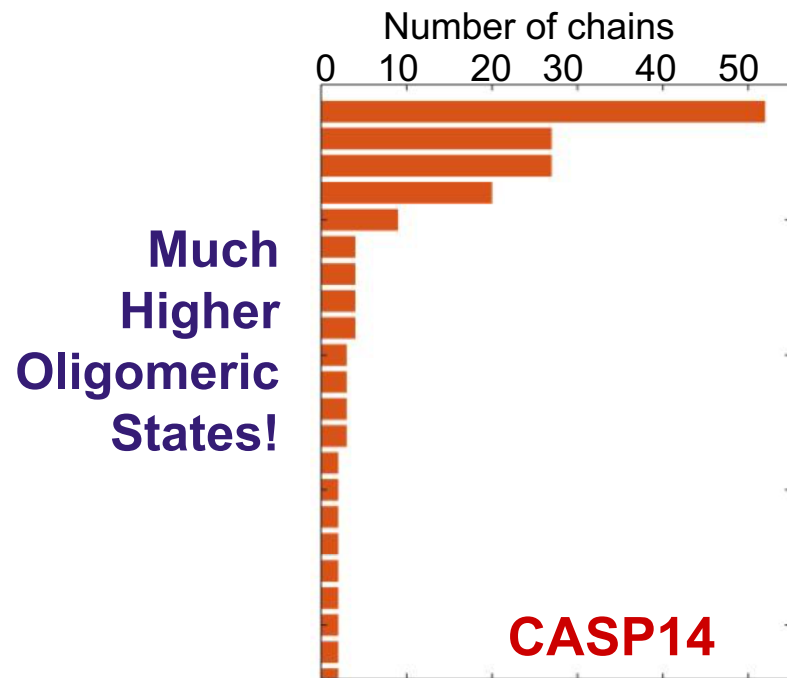
CASP13



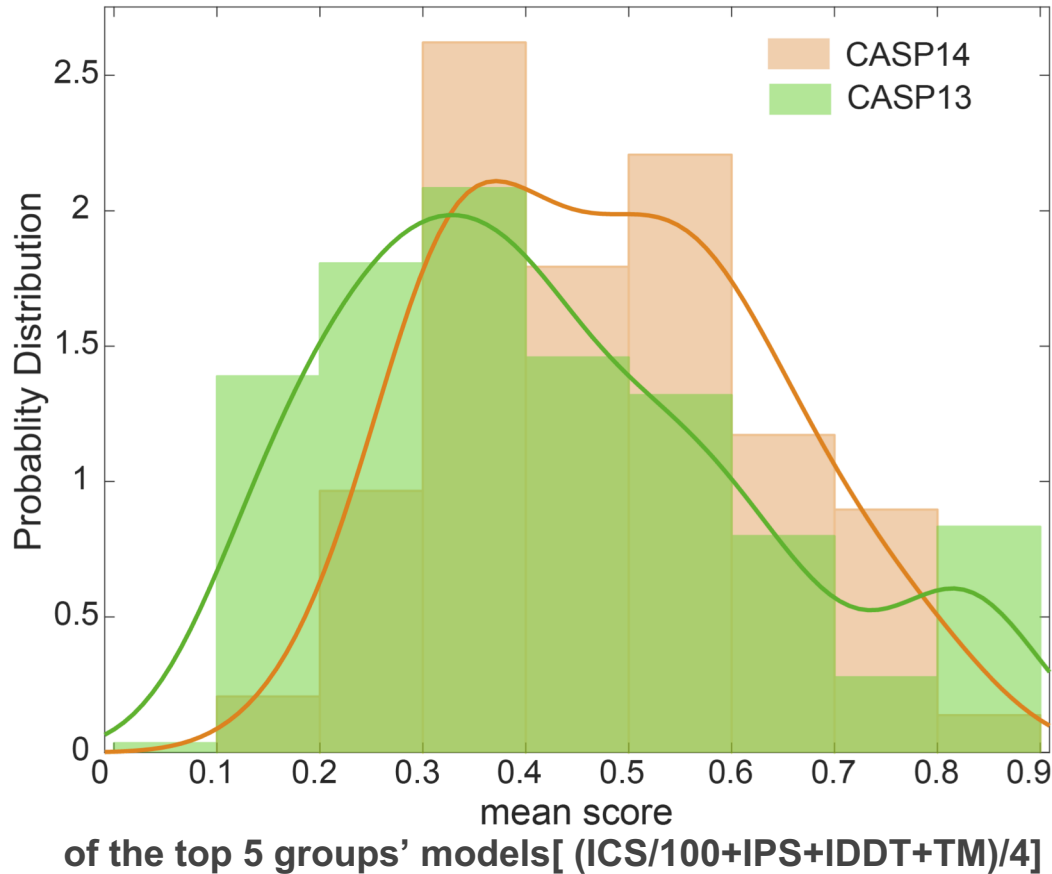
42 targets

**Almost
Equal
Share**

How difficult were CASP13's difficult cases?



Non-parametric fit on histogram



For CASP14:

**There is a shift to the right
for mean_score < 0.9**

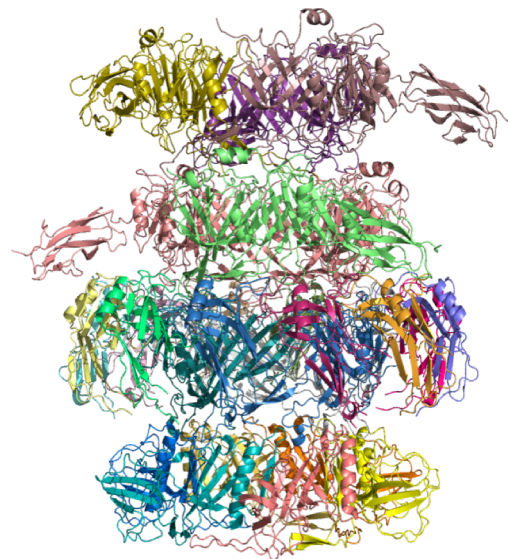
Outline

- Overview of the CASP14 assembly round
- CASP14 assembly scoring
- CASP14 target highlights
- CASP13 vs CASP14
- Challenges & Perspectives



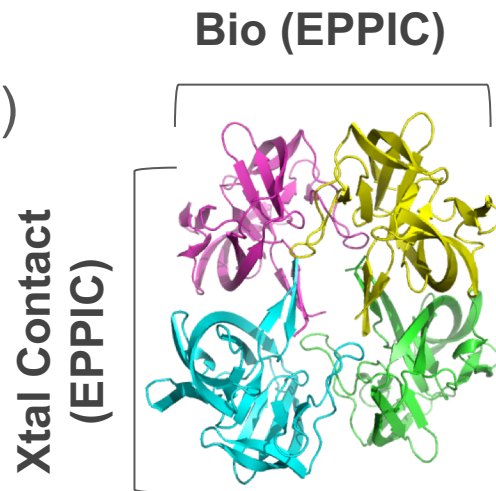
Challenges & Perspectives

- EM revolution!
- Heterogenous chain content!
 - Combinatorial chain mapping problem
- Calls for the development of **new metrics!**



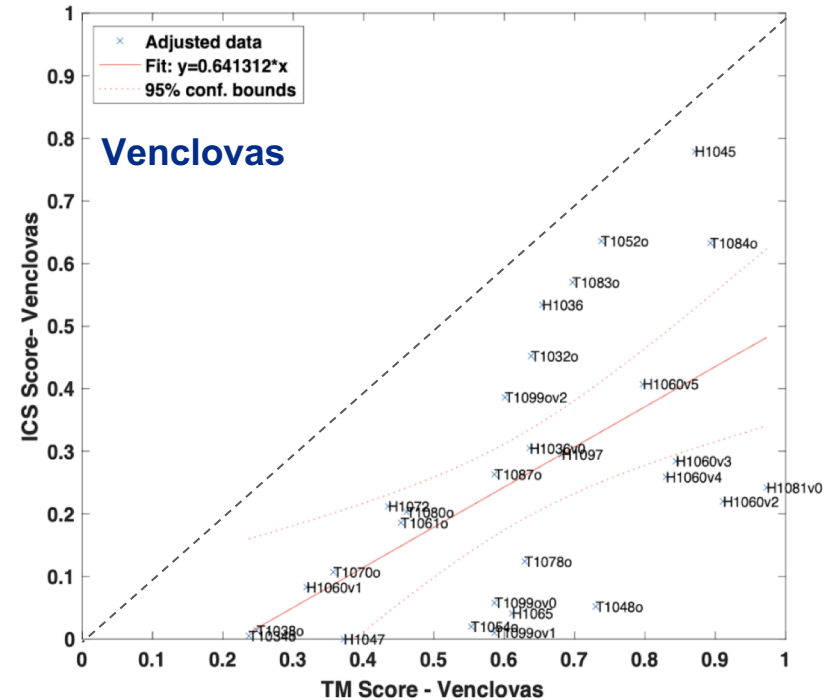
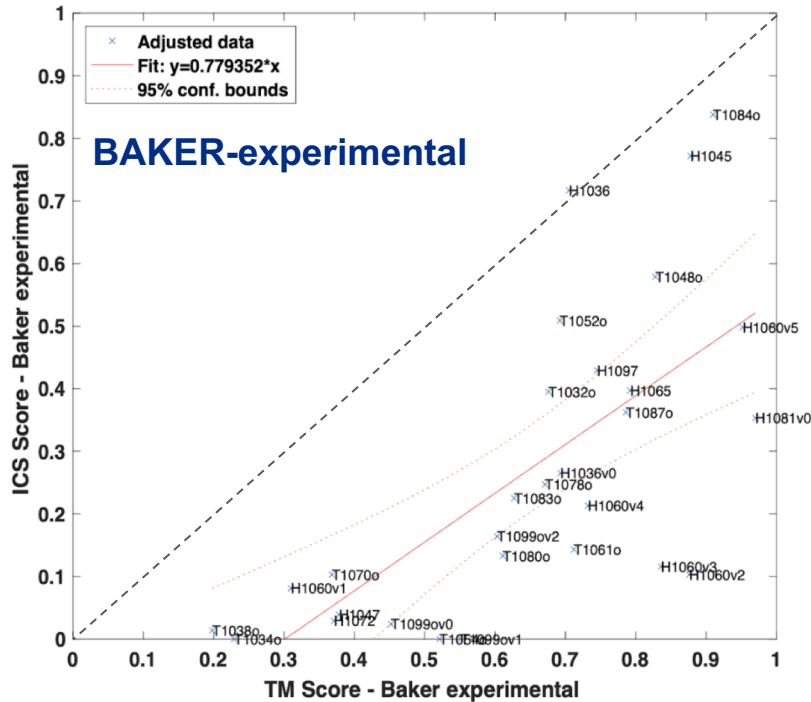
Challenges & Perspectives

- Stoichiometry ambiguity
 - Problematic cases (low F1 & TM):
 - T1034o, T1048o, T1073o (**cancelled**)



Challenges & Perspectives

- ICS vs. TM problem



Challenges & Perspectives

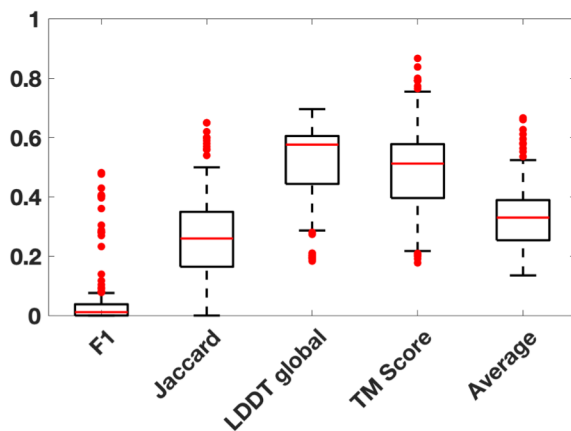
- How to improve the contact predictions?
 - Note -- we did not have AlphaFold in this round



H1065

Difficult

**Serratia marcescens
DNA Methyltransferase**



Best ICS (top 5)

**Elofsson
AILON
DATE
Ornate-select
E2E**

Best TM (top 5)

**AILON
Elofsson
Ornate-select
DATE
E2E**

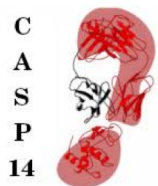
Thanks! CASP14 Assembly Predictors



Burcu Ozden



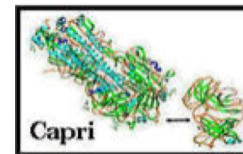
Andriy Kryshchak



CASP14 Organizing Team

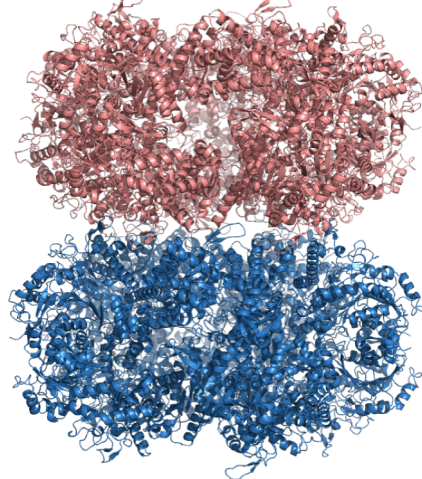


Marc Lensink



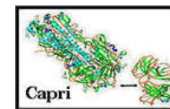
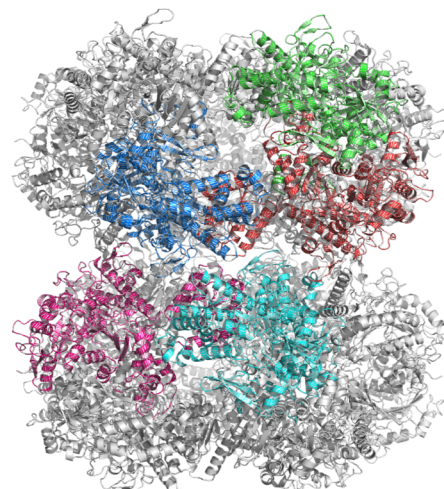
Shoshana Wodak

H1081
Medium
20-mer
D5

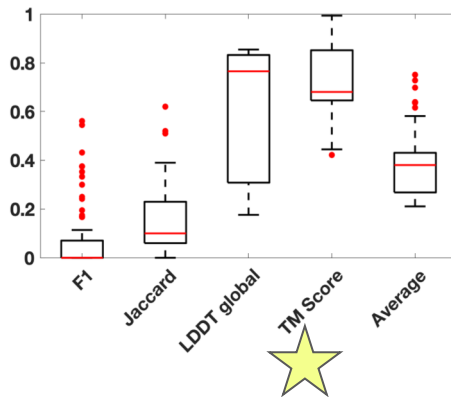


Bacteria
Arginine decarboxylase

“2 stacked decamers, each adopting D5 symmetry. The main challenge for this target is to **predict the stacking interface.**”



H1081v0



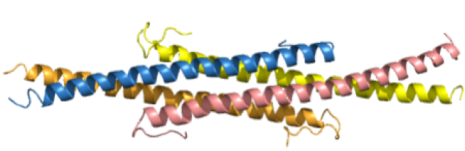
Best ICS (top 5)

Best TM (top 5)

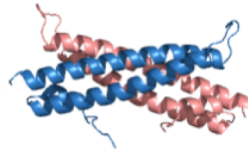
Zou
Zhang-assembly

Zou
Venclovas

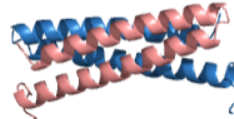
Coiled Coils



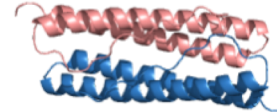
T1048o
Medium
D2



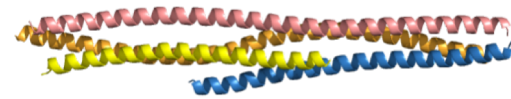
T1083
Medium
C2



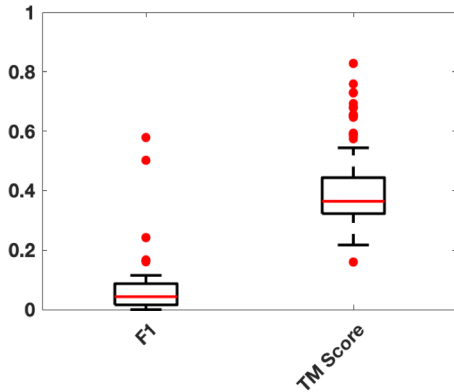
T1084
Medium
C2



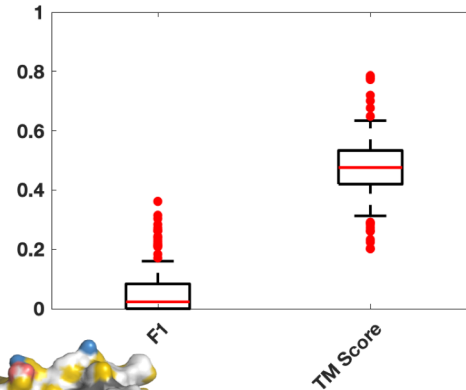
T1087
Medium
C2



H1072/6r17
Medium
C2

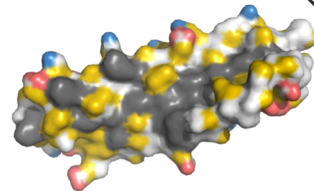


Best TM
BAKER-exp.



Best TM
BAKER-exp.

“~5% tetramer for the protein in solution”



Histidines in black

Appendix: Score Correlation Matrix

