Zhang Groups: Integration of threading and deep-learning for protein structure prediction

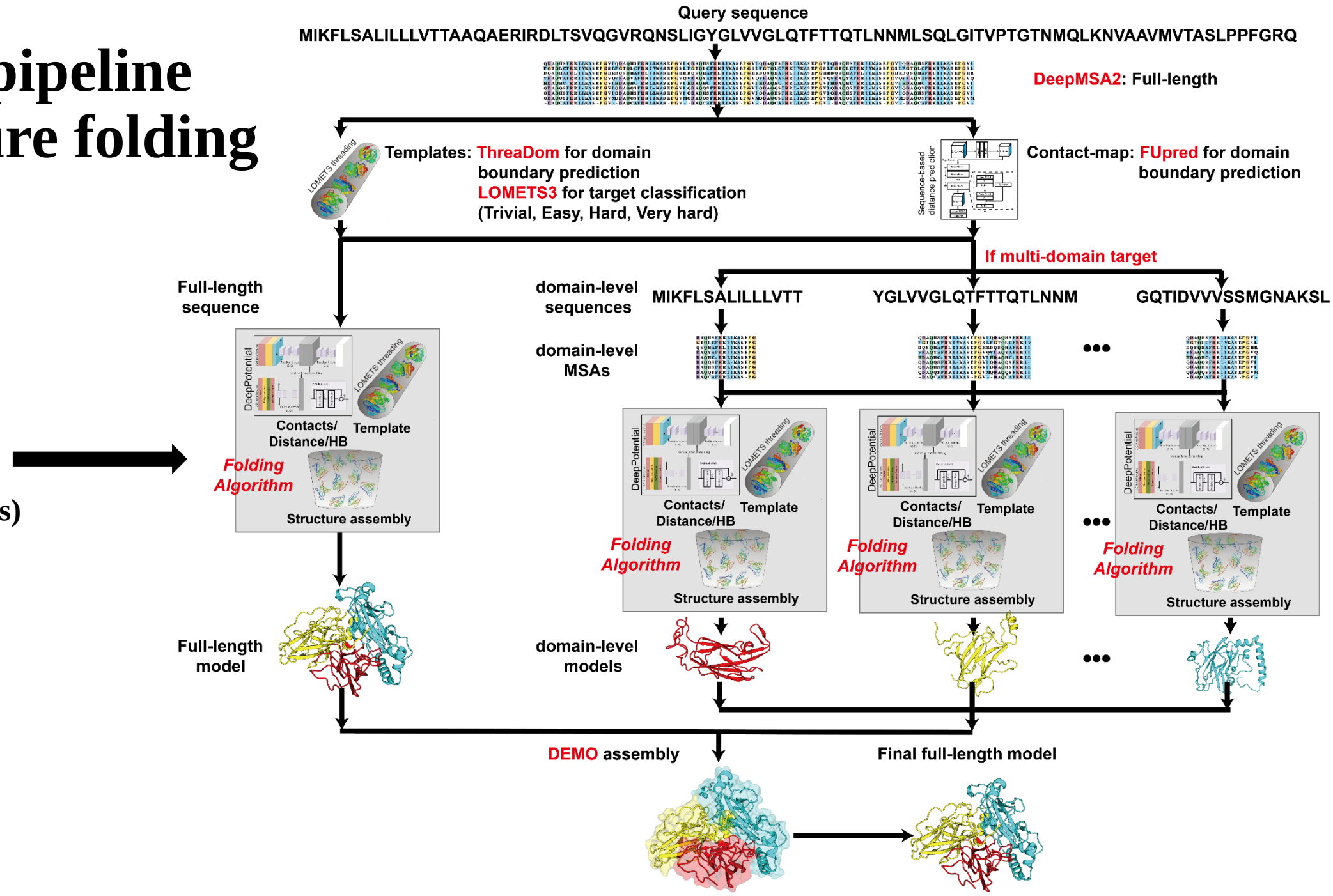
Wei Zheng, Yang Li, Chengxin Zhang, Xiaogen Zhou,
Xiaoqiang Huang, Robin Pearce, Eric Bell, Yang Zhang

Department of Computational Medicine and Bioinformatics
Department of Biological Chemistry
University of Michigan

Methods

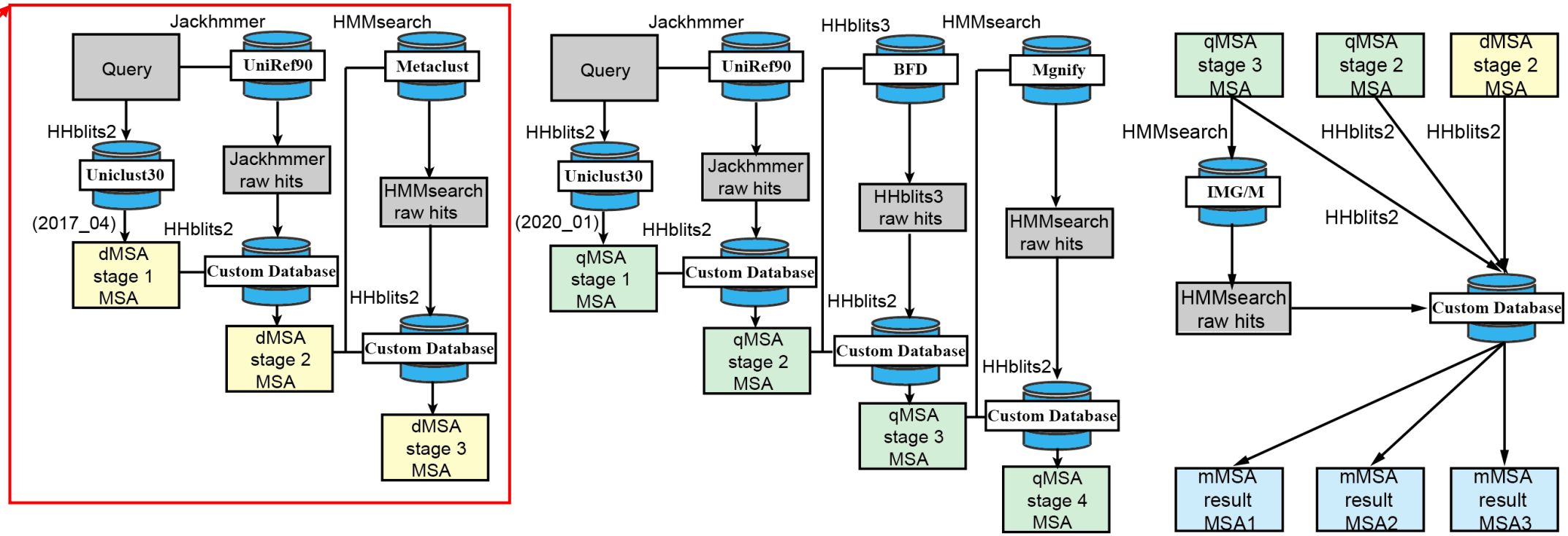
Overall pipeline for structure folding

Zhang (h)
Zhang-Server (s)
QUARK (s)
Zhang-CEthreader (s)

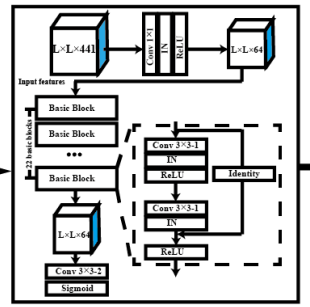


DeepMSA2 for MSA construction

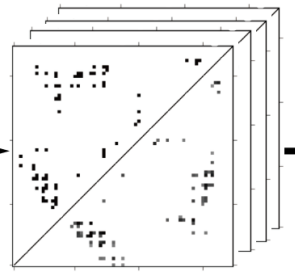
DeepMSA pipeline



(D) MSA selection



TripletRes

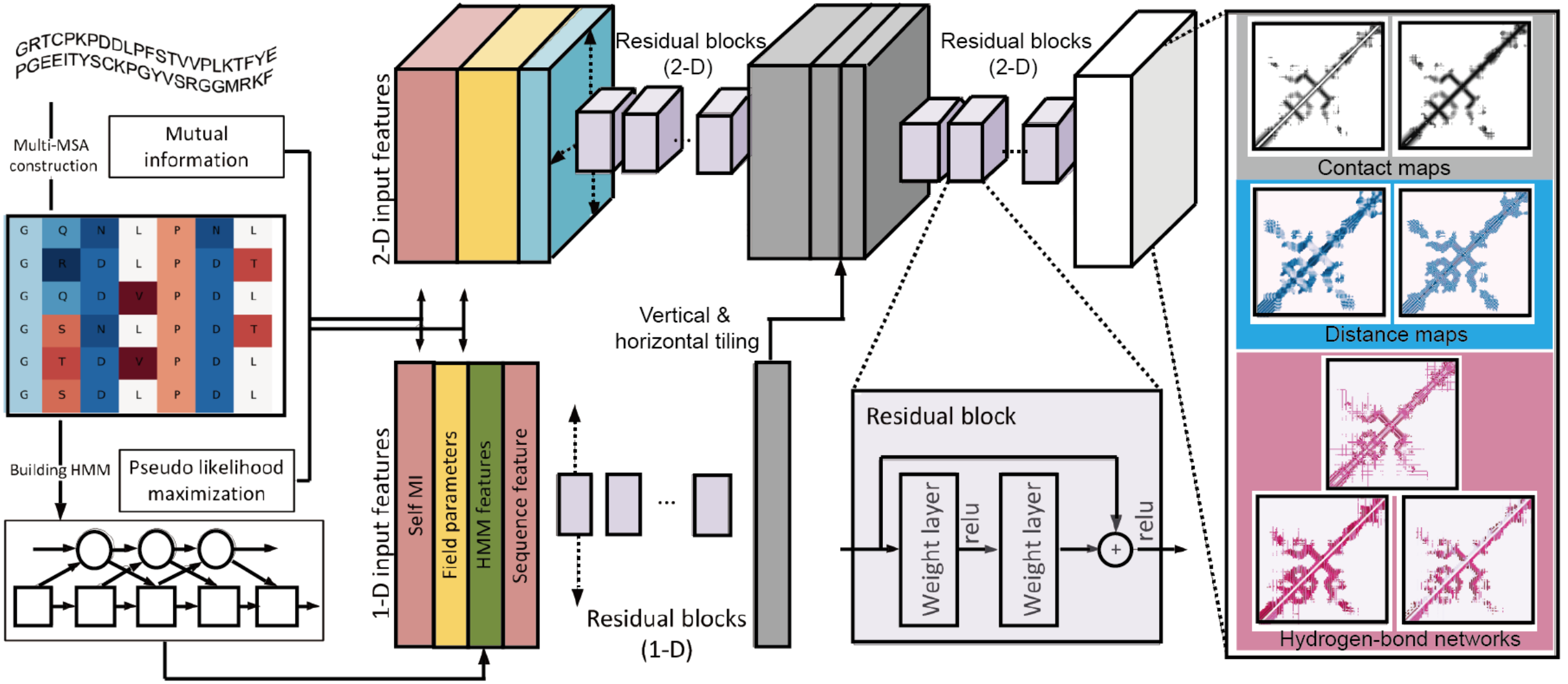


Ranking MSAs
by the sum of top 10L
predicted contact probablities

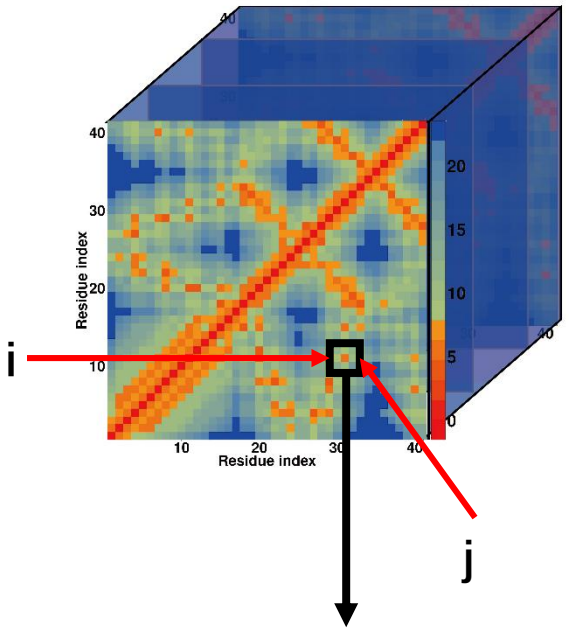
Final MSA

$$Neff = \frac{1}{\sqrt{L}} \sum_{n=1}^{N} \frac{1}{1 + \sum_{m=1, m \neq n}^{N} I[S_{m,n} \geq 0.8]}$$

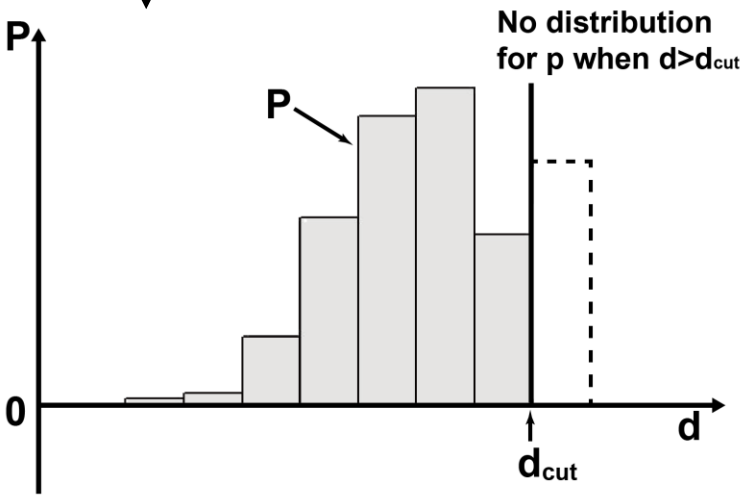
DeepPotential for contact/distance/HB-network prediction



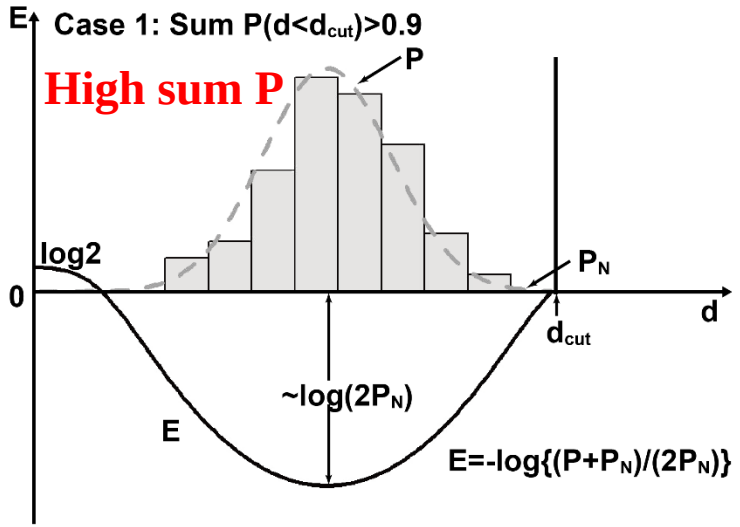
Distance potential in D-I-TASSER

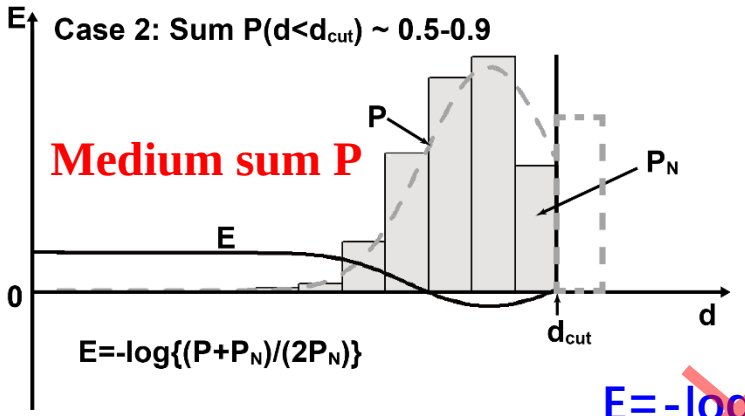


Distance distributions prediction from DeepPotential

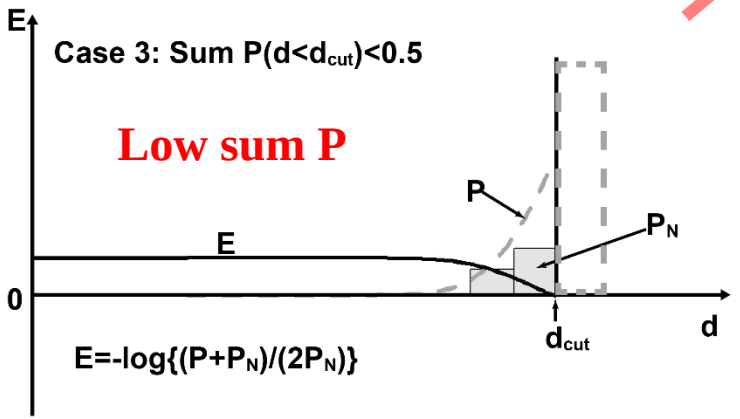


Potential

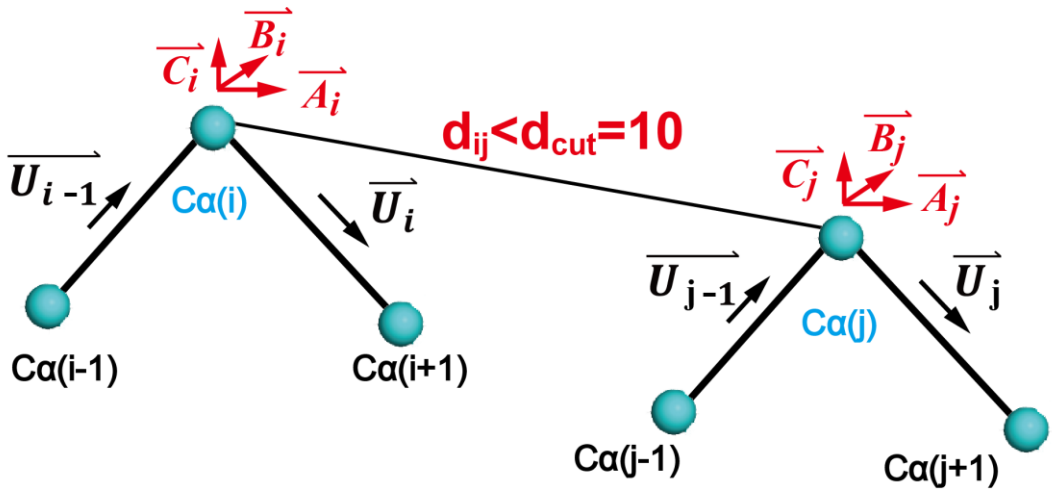




E=-log(P)



Hydrogen-bond potential in D-I-TASSER



$$\overrightarrow{A_i} = \frac{\overrightarrow{U_{i-1}} + \overrightarrow{U_i}}{\left|\overrightarrow{U_{i-1}} + \overrightarrow{U_i}\right|}$$

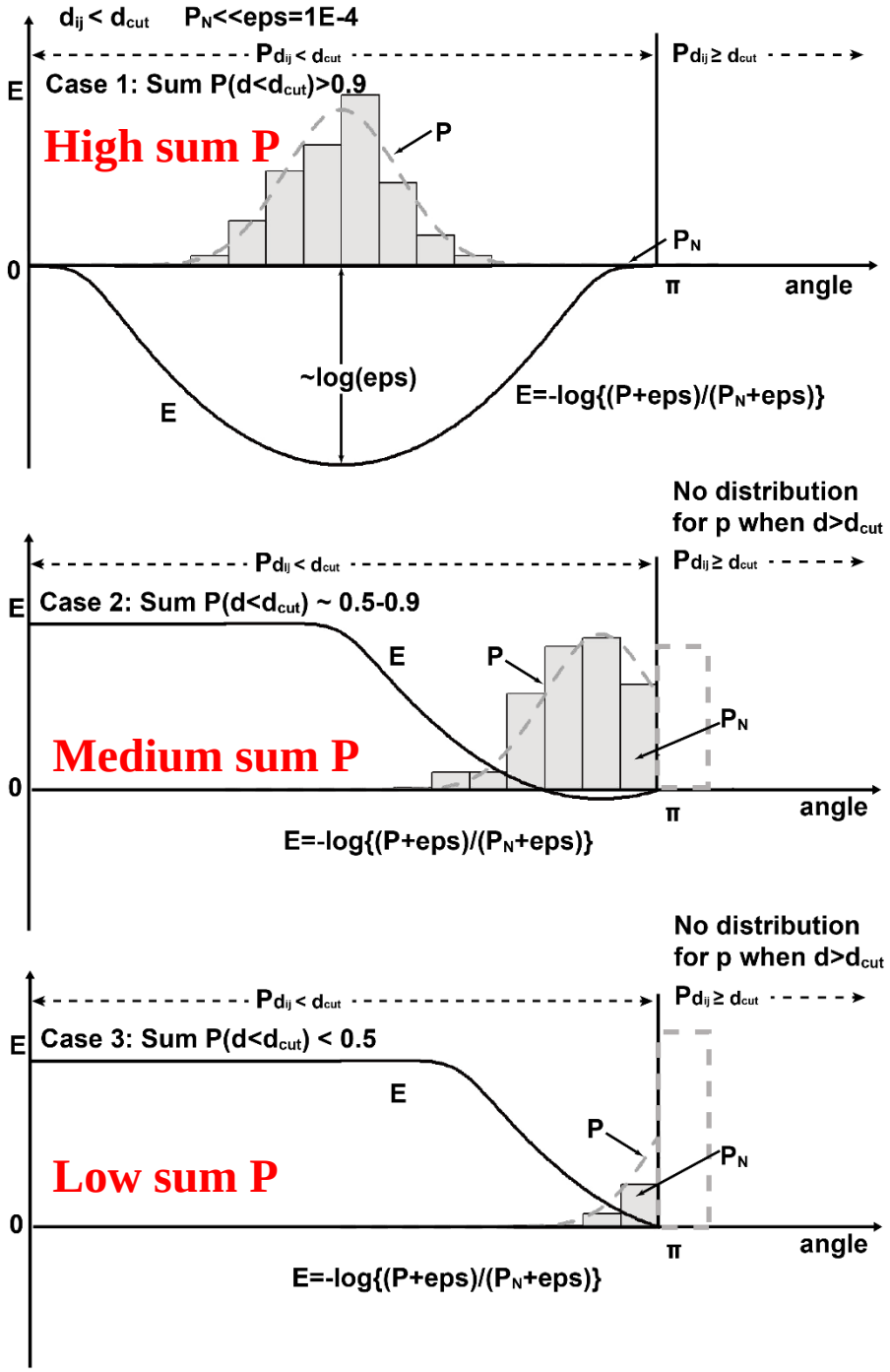
$$\overrightarrow{B_i} = \frac{\overrightarrow{U_{i-1}} \times \overrightarrow{U_i}}{\left|\overrightarrow{U_{i-1}} \times \overrightarrow{U_i}\right|}$$

$$\overrightarrow{C_i} = \frac{\overrightarrow{U_{i-1}} - \overrightarrow{U_i}}{\left|\overrightarrow{U_{i-1}} - \overrightarrow{U_i}\right|}$$

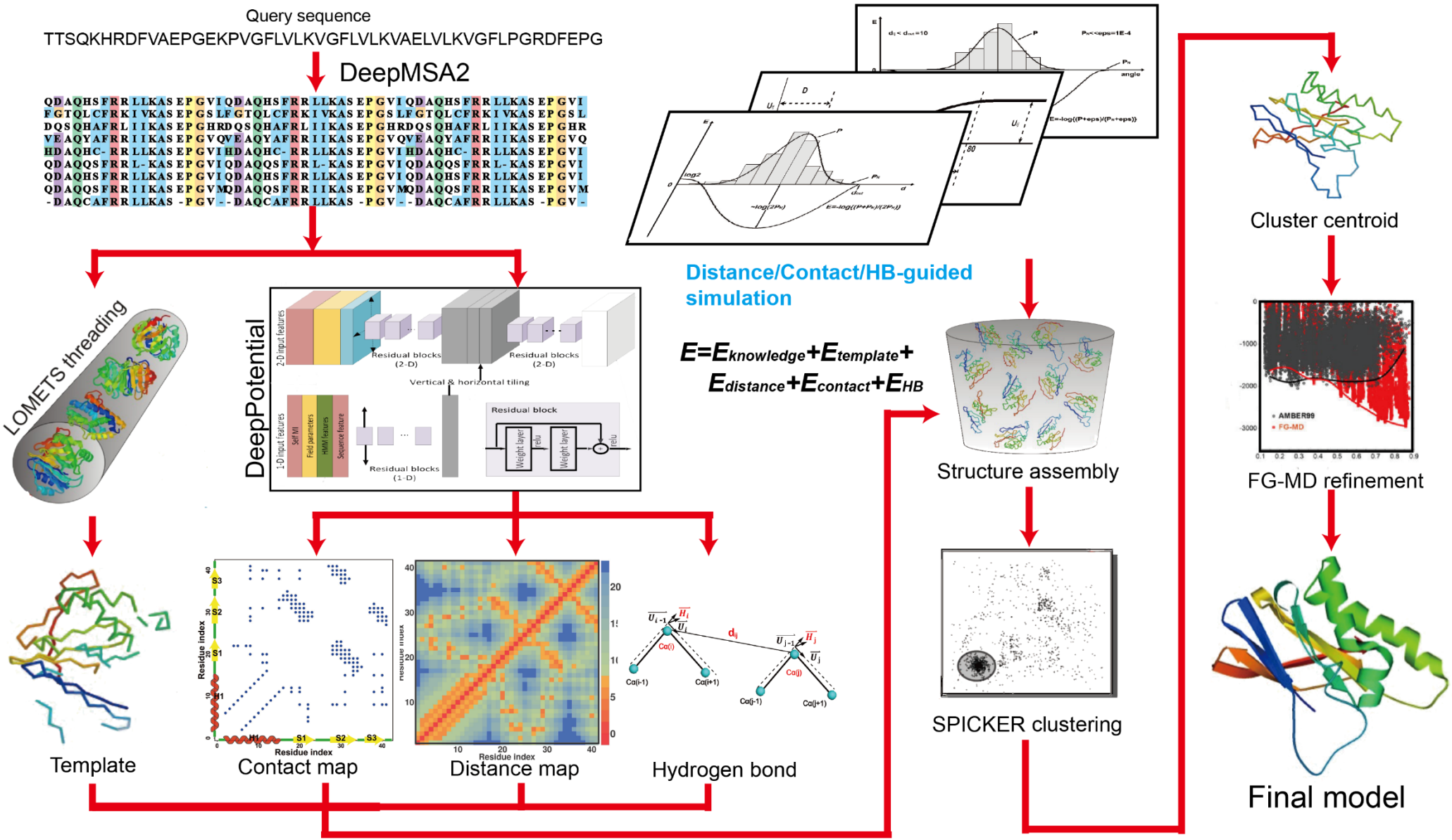
$$AA = arccos(\overrightarrow{A_i} * \overrightarrow{A_j})$$

$$BB = arccos(\overrightarrow{B_i} * \overrightarrow{B_j})$$

$$CC = arccos(\overrightarrow{C_i} * \overrightarrow{C_j})$$

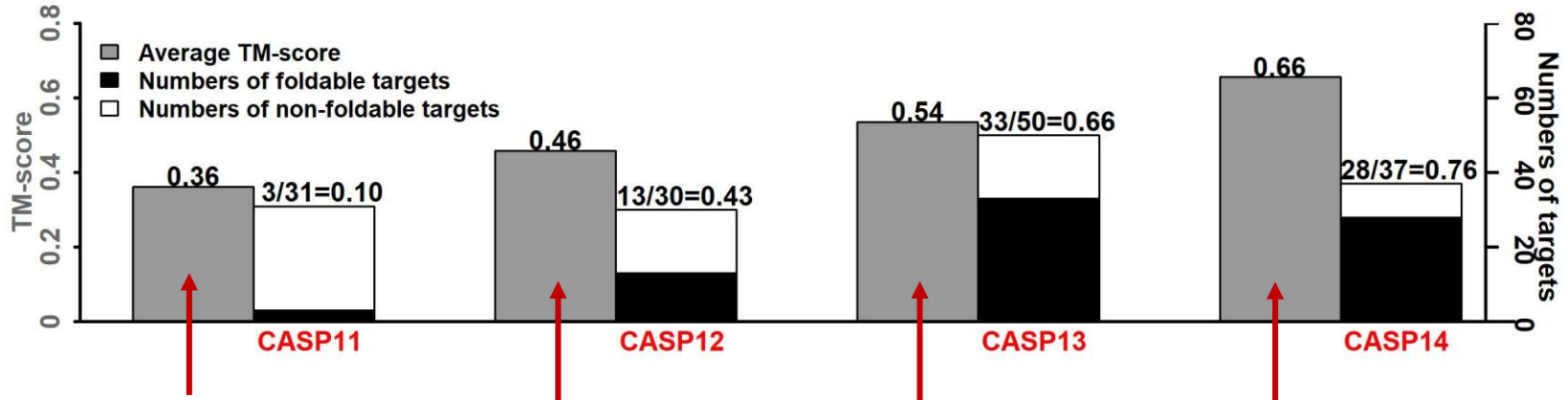


Zhang-Server pipeline built from D-I-TASSER



Results

Summary of FM targets folded by Zhang group



Fragment assembly-based

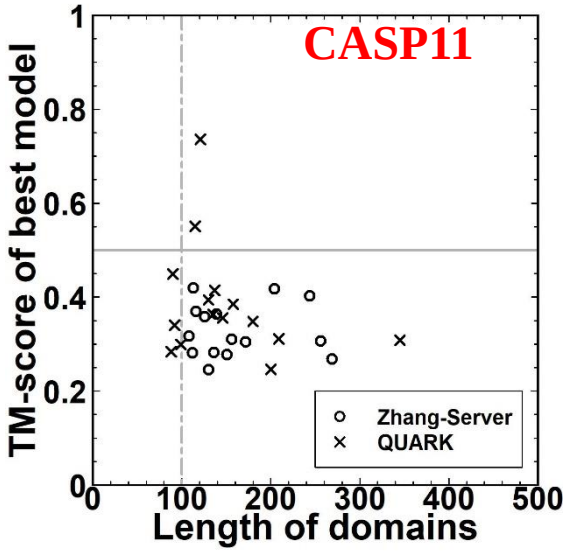
Direct coupling analysis-based contacts added

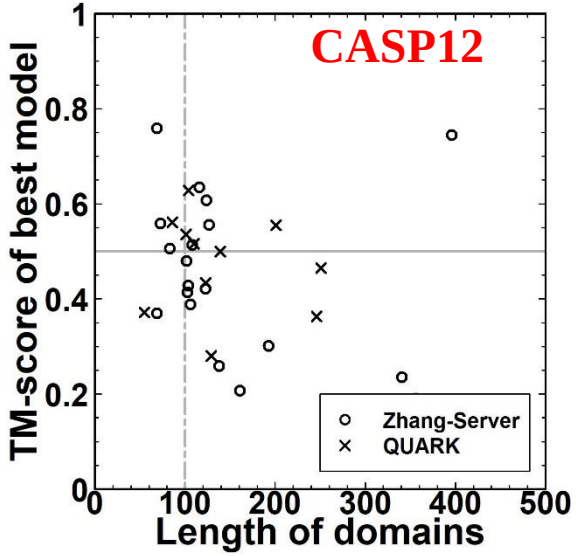
Deep-learning-based contacts added

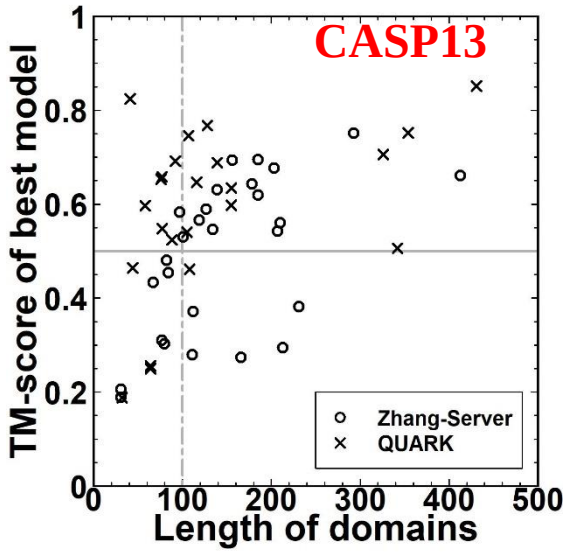
Deep-learning-based distances added

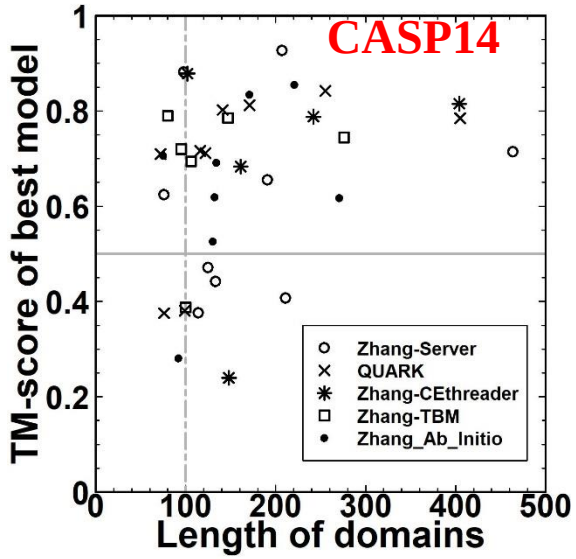
23 FM targets and 14 FM/TBM targets are classified as "FM" targets in the following analysis

76% targets are foldable based on best server models from Zhang group in CASP14

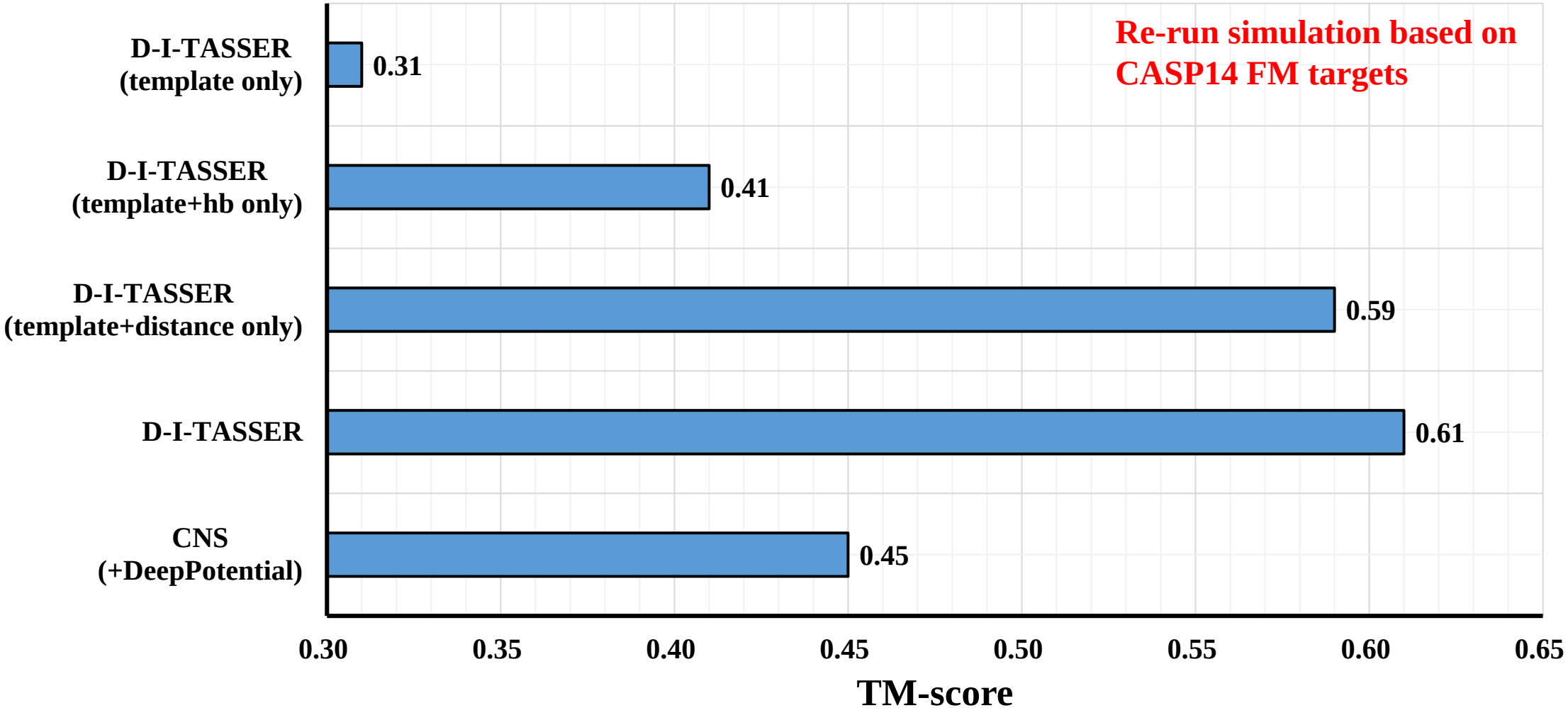




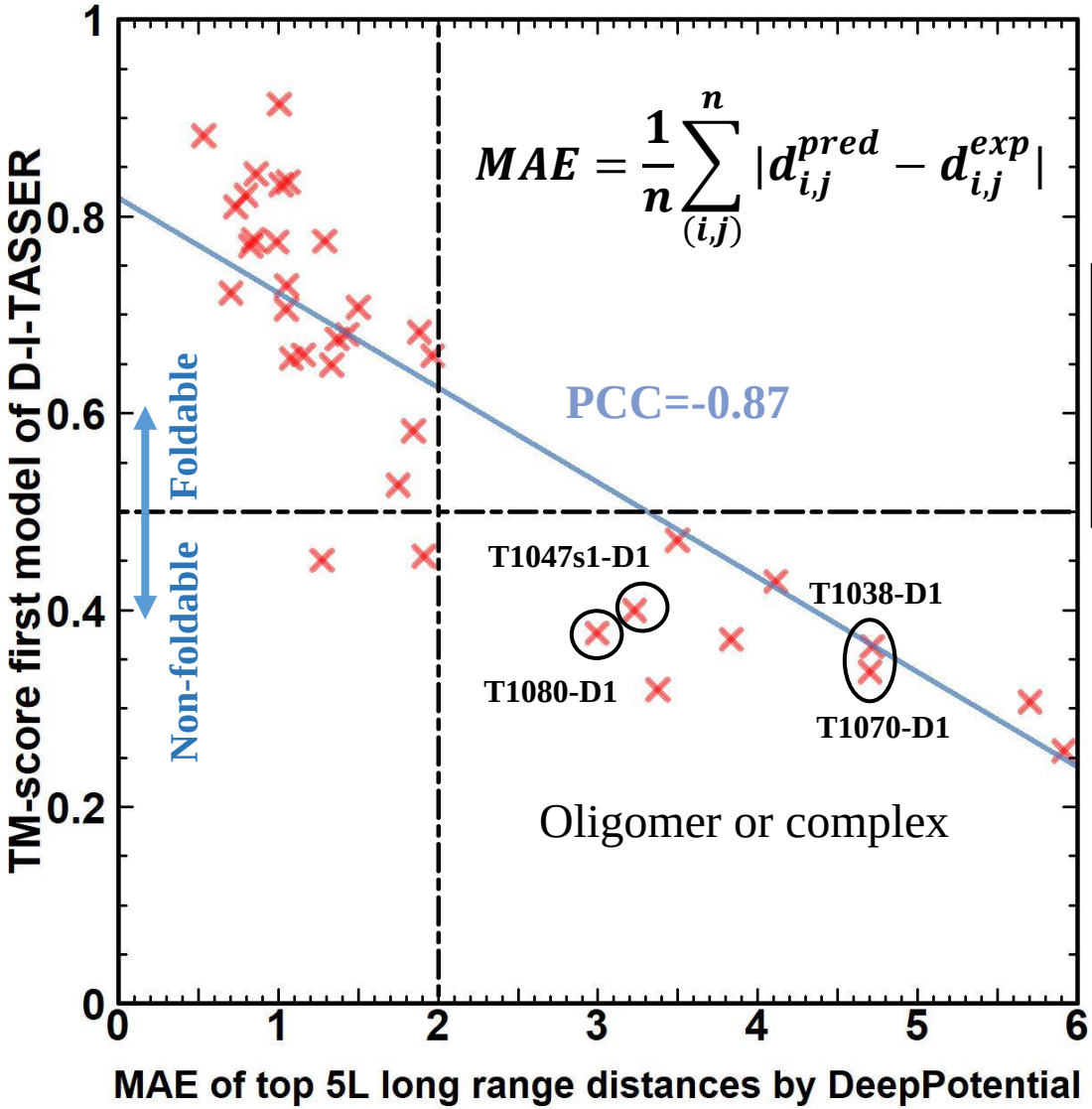




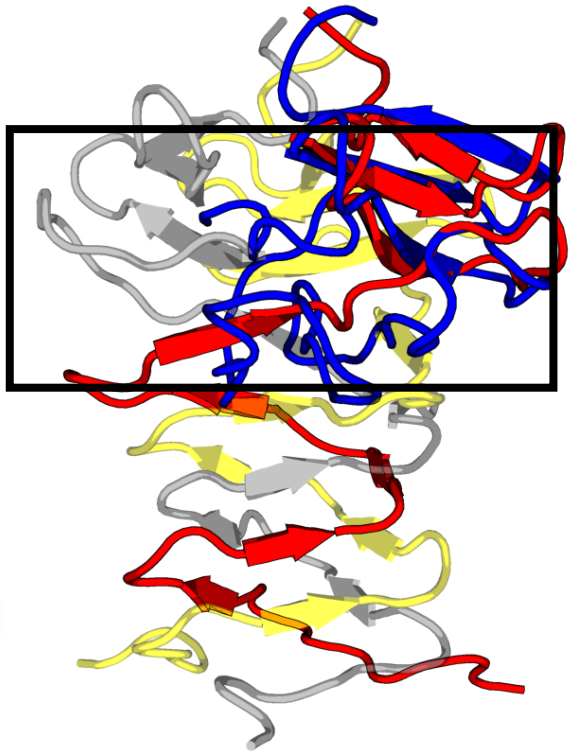
Impact of different energy components on D-I-TASSER for FM targets



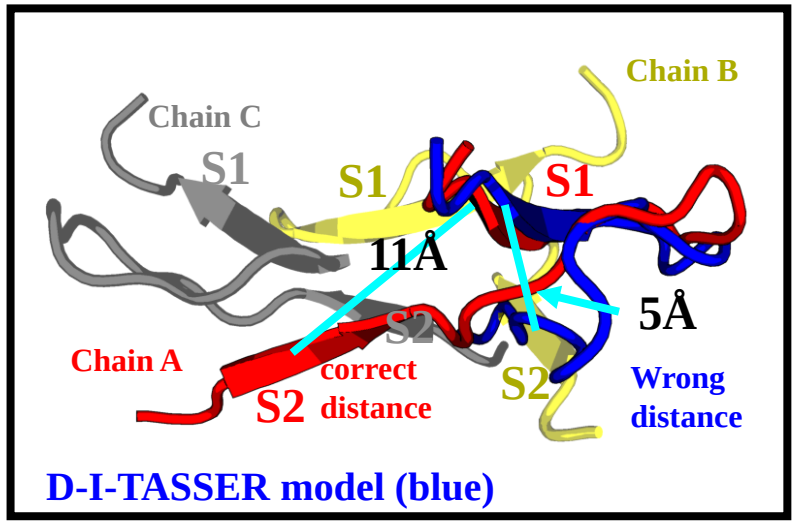
Impact of distance prediction on D-I-TASSER for FM targets

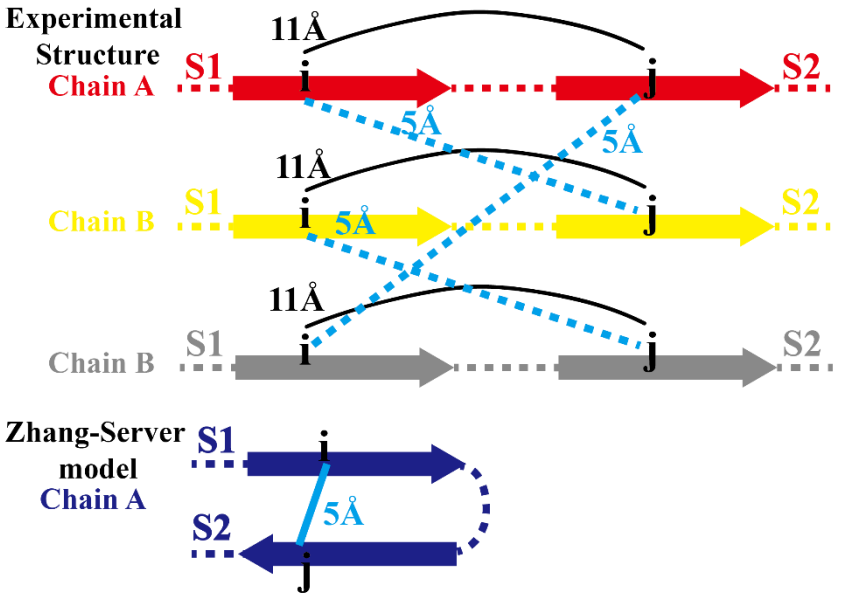


T1070-D1 oligomer (red, yellow, gray)
D-I-TASSER model (blue)
TM-score=0.34



Alignment of MSA is sparse

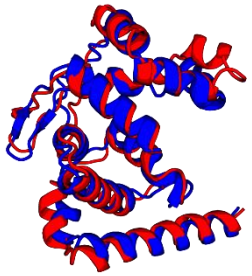




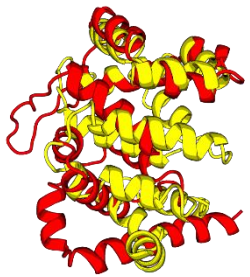
Impact of DeepMSA2 on the D-I-TASSER for FM targets

Experimental structure (red)

T1096-D2, FM target, L=171AA

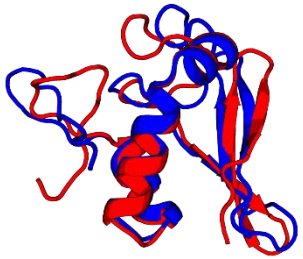


TM-score=0.83
Neff=28.11
DeepMSA2

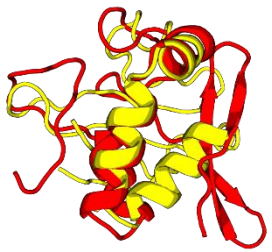


TM-score=0.32
Neff=0.99
DeepMSA

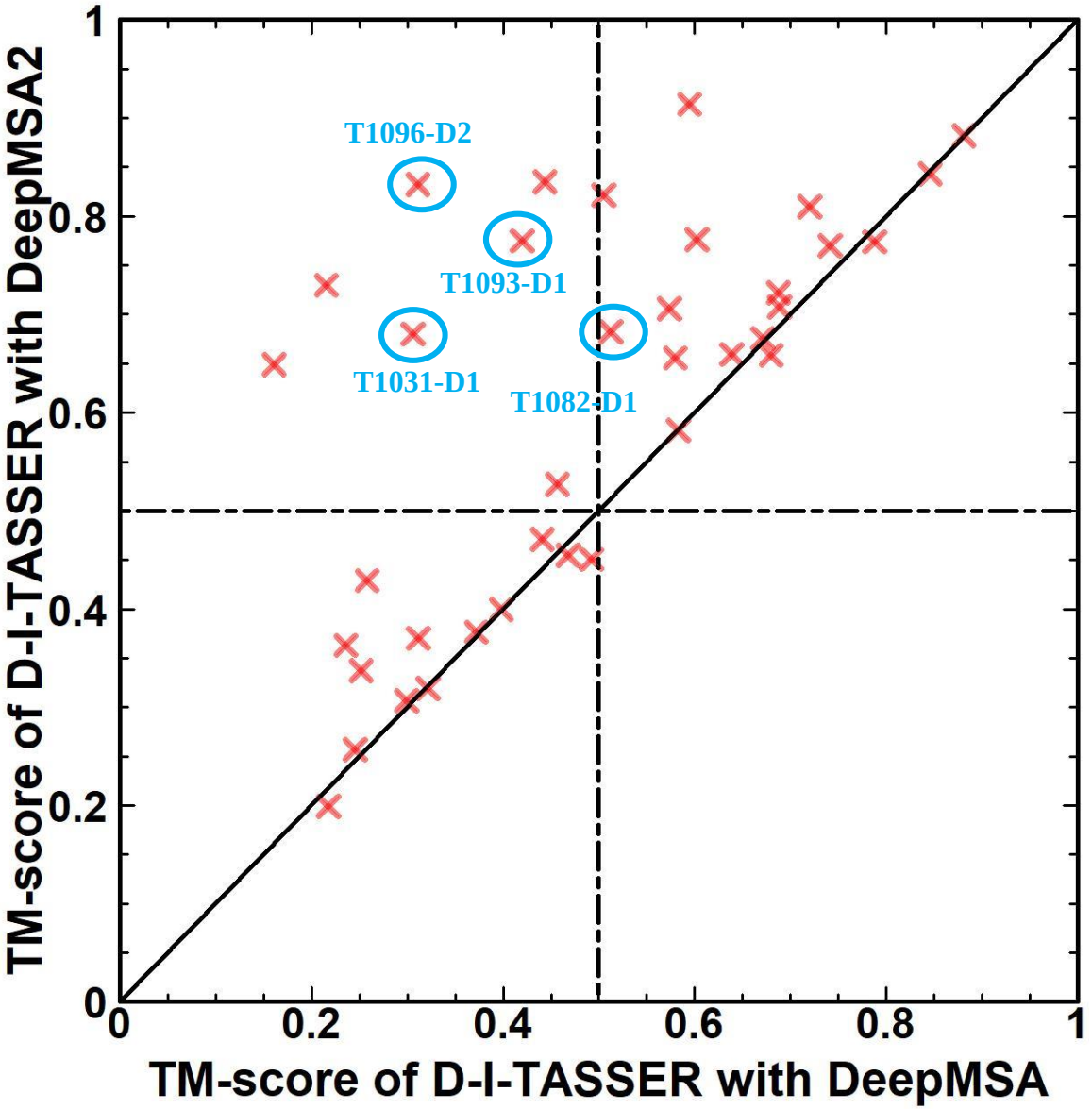
T1031-D1, FM target, L=95AA



TM-score=0.68
Neff=45.41
DeepMSA2

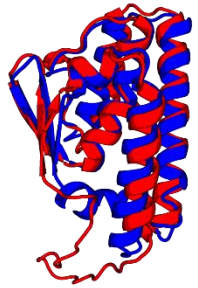


TM-score=0.31
Neff=1.03
DeepMSA

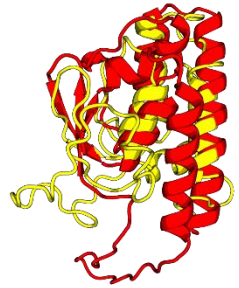


Experimental structure (red)

T1093-D1, FM target, L=141AA

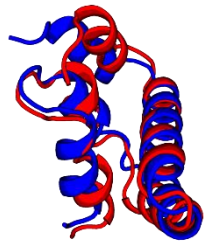


TM-score=0.77
Neff=10.87
DeepMSA2

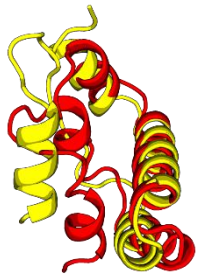


TM-score=0.42
Neff=0.42
DeepMSA

T1082-D1, FM/TBM target, L=75AA



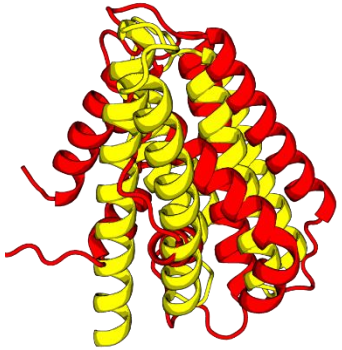
TM-score=0.68
Neff=1.78
DeepMSA2



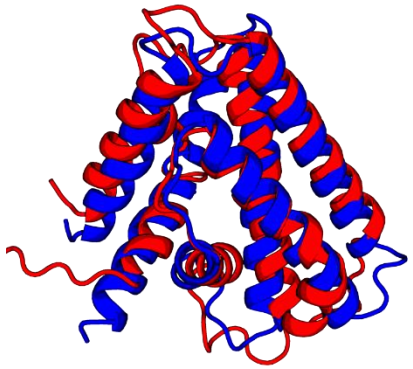
TM-score=0.51
Neff=1.32
DeepMSA

T1039-D1 by Zhang-Server and Zhang-CEthreader

T1039-D1 is an FM target with 161 residues

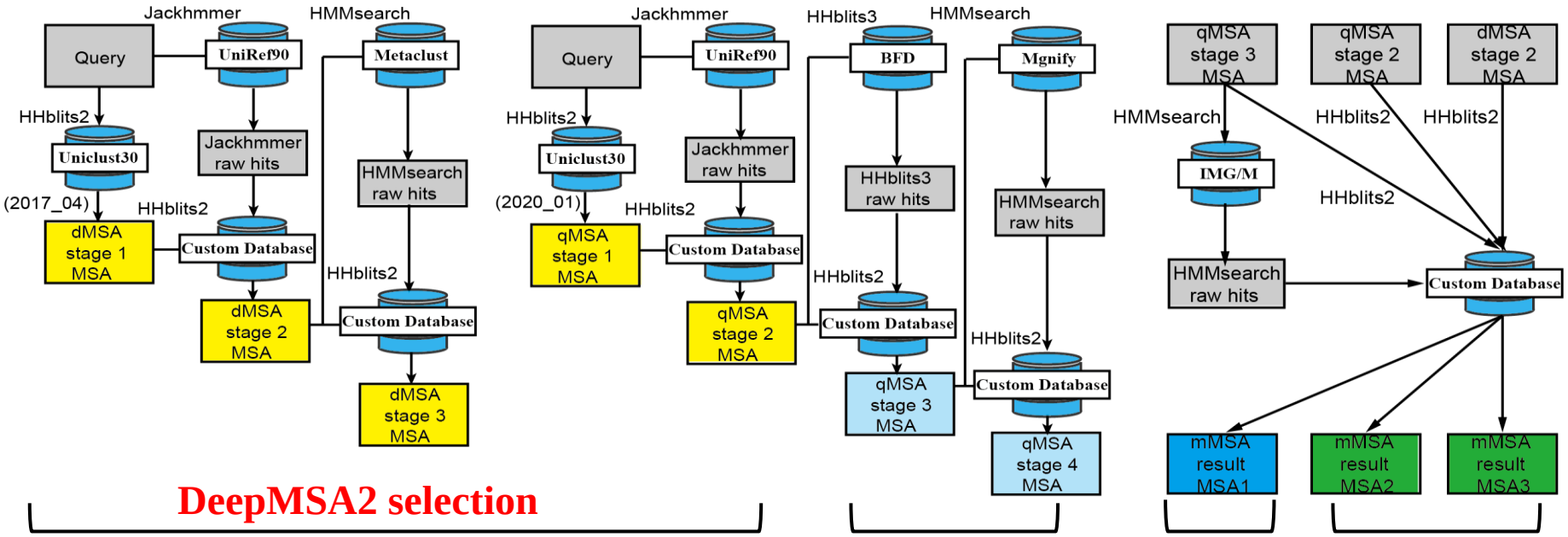


Zhang-Server model1
TM-score=0.31



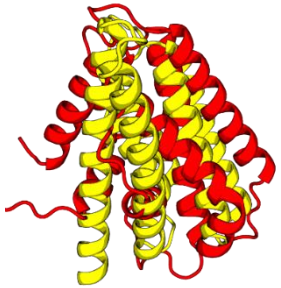
Zhang-CEthreader model1
TM-score=0.68

Experimental structure (red)
Models (blue, cyan, yellow, green)



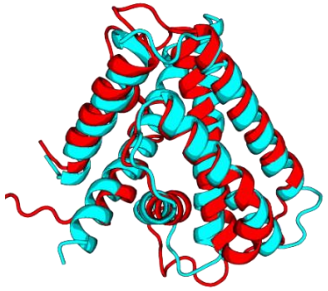
DeepMSA2 selection

Neff=0.08
Contact probability score=614.3



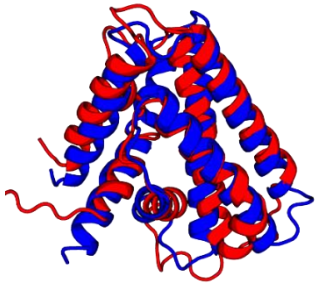
TM-score=0.31

Neff=1.10
Contact probability score=614



TM-score=0.681

Neff=1.34
Contact probability score=613



TM-score=0.683

Neff=0.39
Contact probability score=600.6

TM-score=0.52

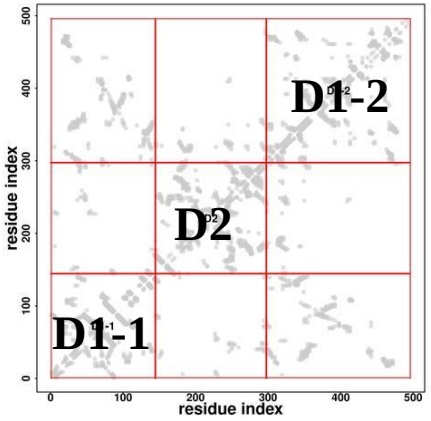
Impact of domain partition

T1094, two domain target

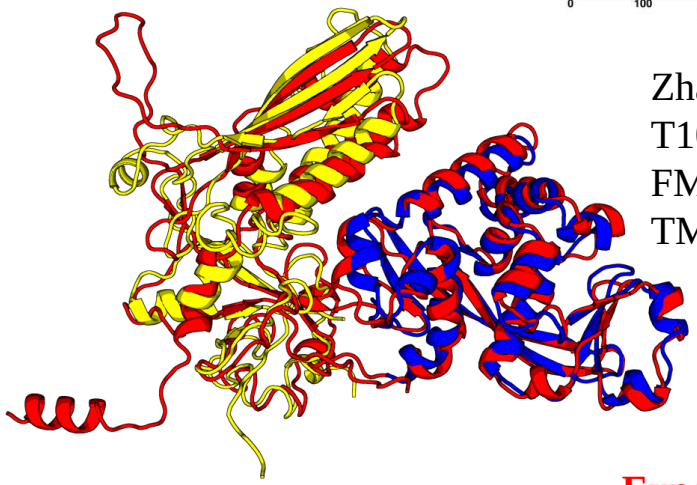
Domain from CASP

T1094-D1 T1094-D2

Zhang group domain by FUpred



Zhang-Server model1
T1094-D1
TBM-hard
TM-score=0.64



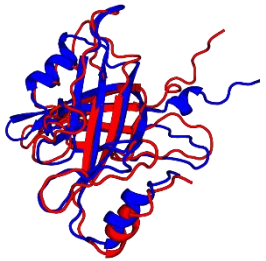
Zhang-Server model1
T1094-D2
FM
TM-score=0.91

T1061-D1, FM/TBM target, discontinuous domain

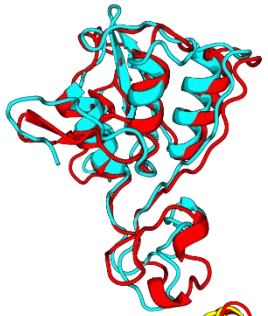
Domain from CASP

Zhang group domain

T1061-D1_f1
TM-score=0.76

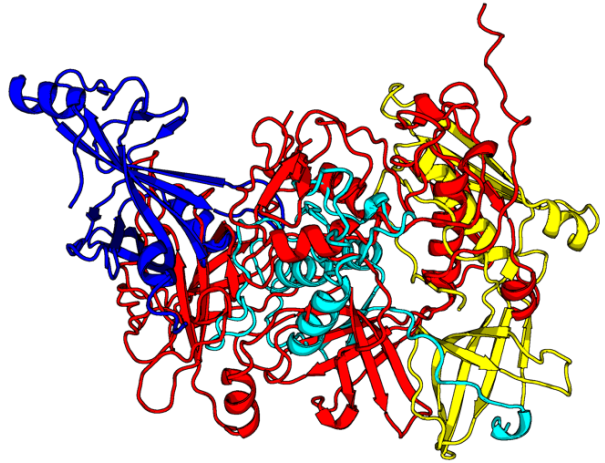


T1061-D1_f2
TM-score=0.67



T1061-D1_f3
TM-score=0.74

Zhang-Server model1
TM-score=0.45



Experimental structure (red), Models (blue, yellow, cyan)

Summary

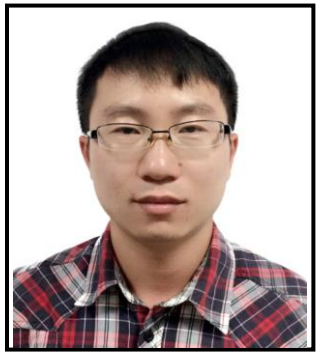
- **What works**
  - Distance and hydrogen-bond prediction by DeepPotential
  - Advanced structure assembly simulations by I-TASSER/QUARK
  - DeepMSA2 for deeper MSA generation
- **What needs to be improved**
  - Modeling oligomers
  - Selecting good MSA
  - Partitioning domains

Acknowledgements

Yang Zhang

Yang Li

Chengxin Zhang



Xiaogen Zhou

Xiaoqiang Huang

Robin Pearce

Eric W. Bell

Jonathan Poisson

National Science Foundation

National Institutes of Health

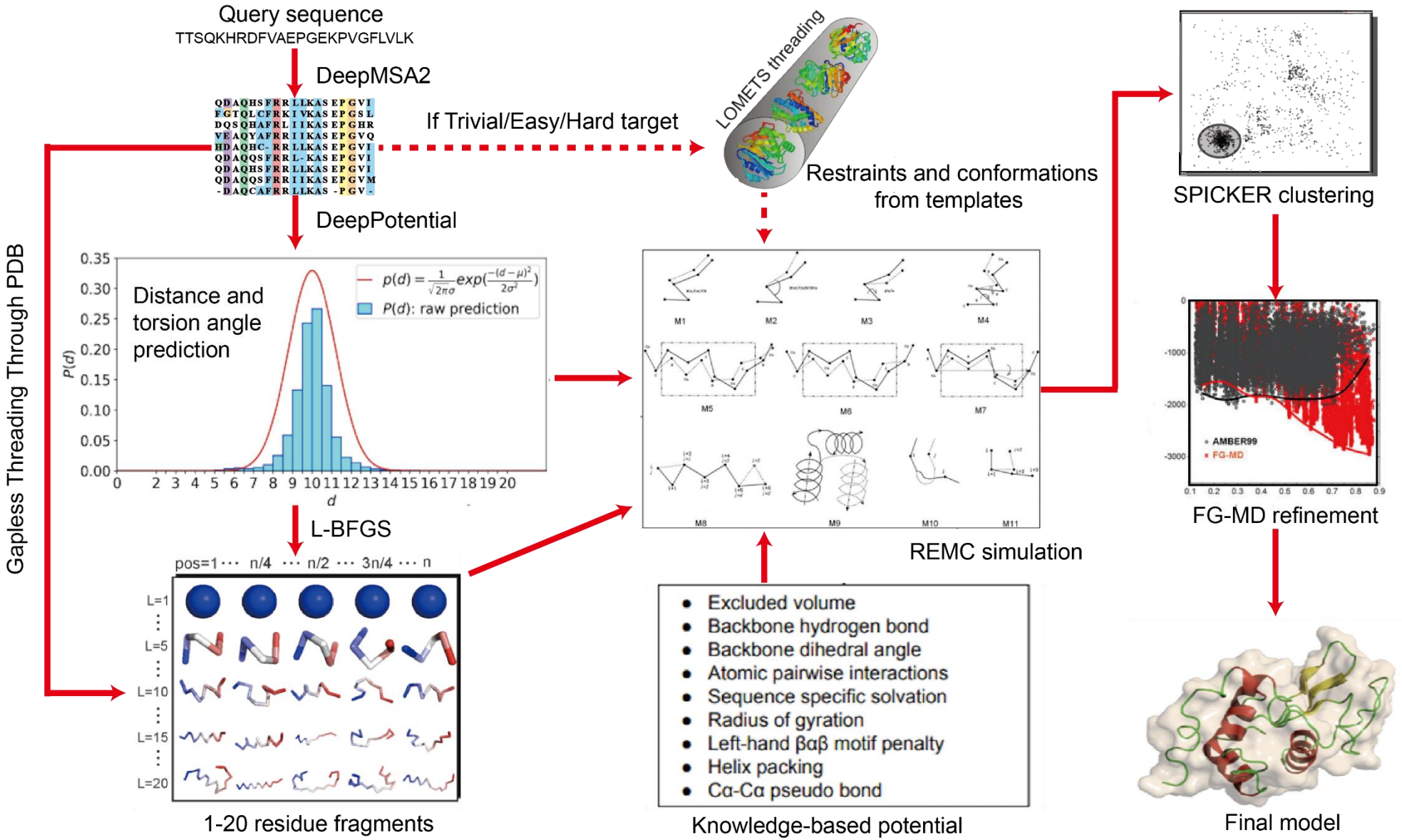
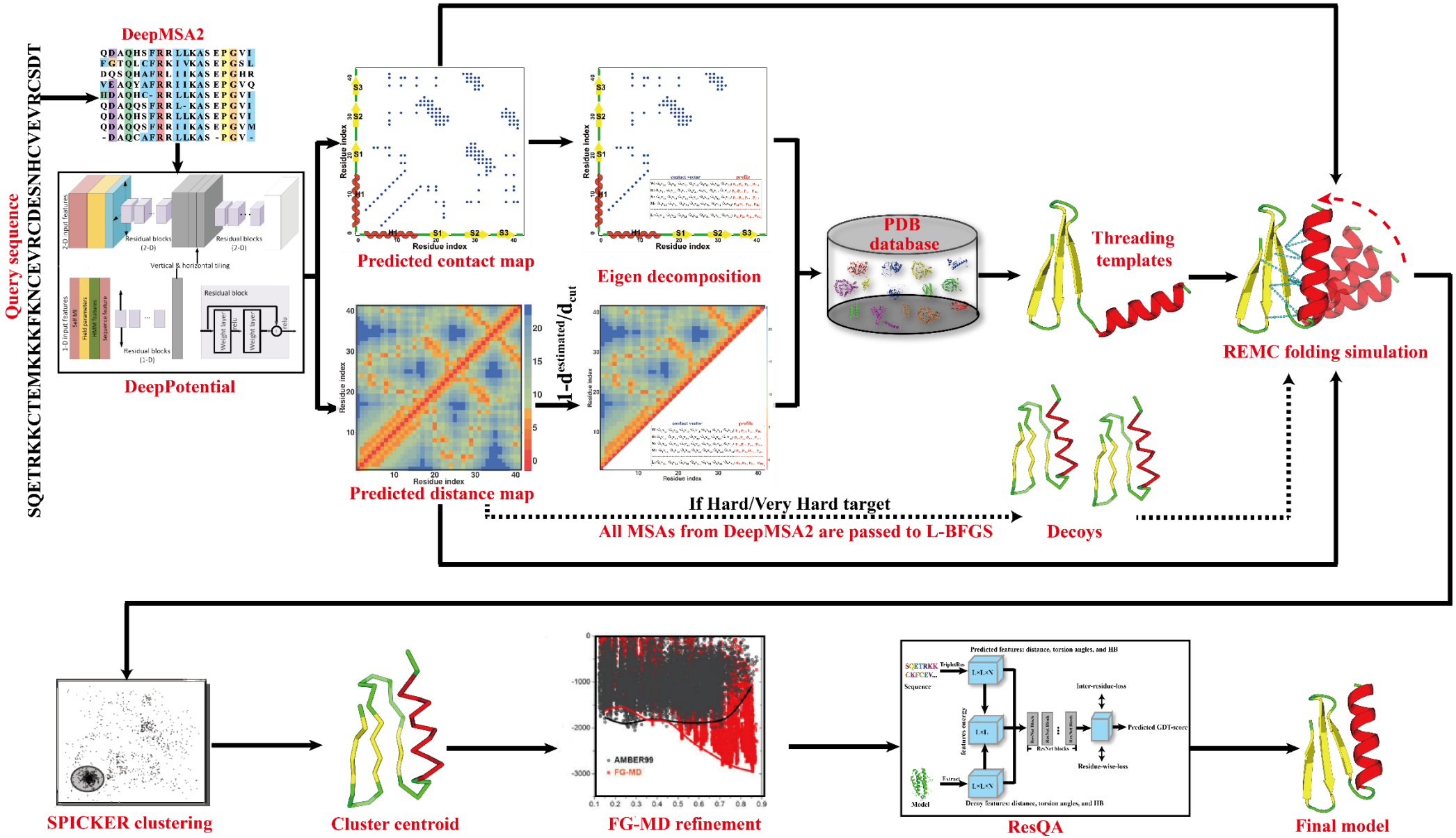
POWERED BY XSEDE
Extreme Science and Engineering Discovery Environment

- CASP Organizers
- Experimentalists that shared their structures

Thank you!

Appendix

QUARK pipeline built from D-QUARK



Zhang-CEthreader pipeline built from DEthreader and D-I-TASSER simulation