Does inclusion of residue-residue contact information boost protein threading?
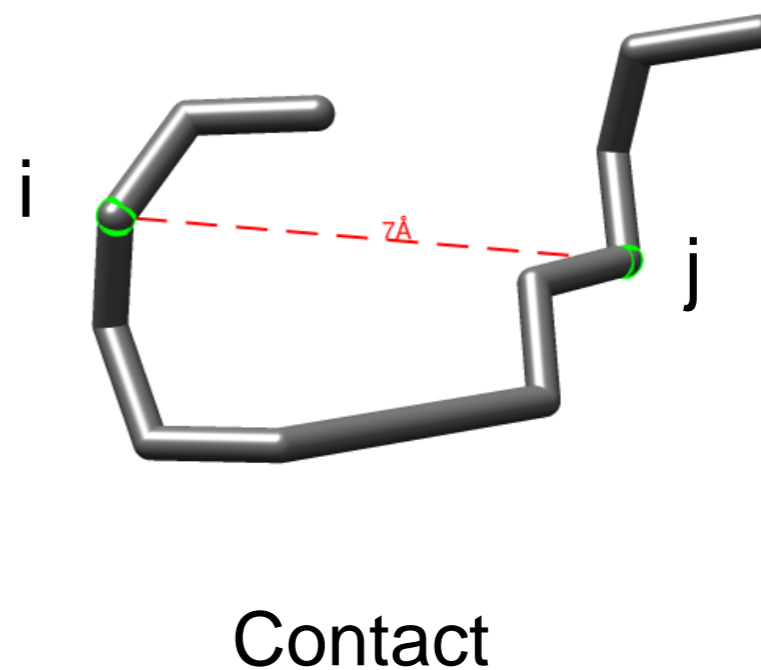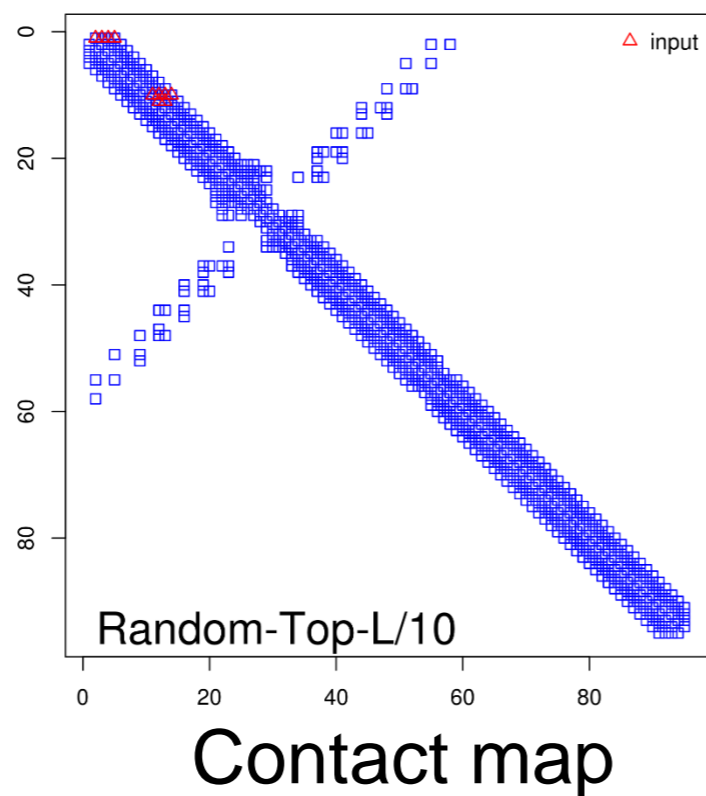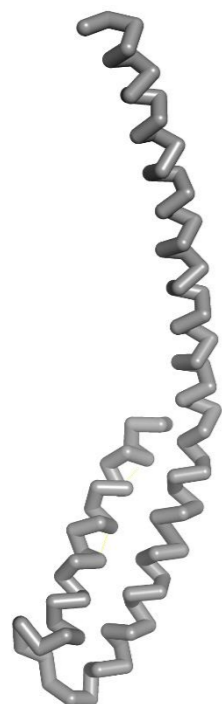
Sutanu Bhattacharya
Auburn University

MCBIOS 2019

Bhattacharya
Laboratory

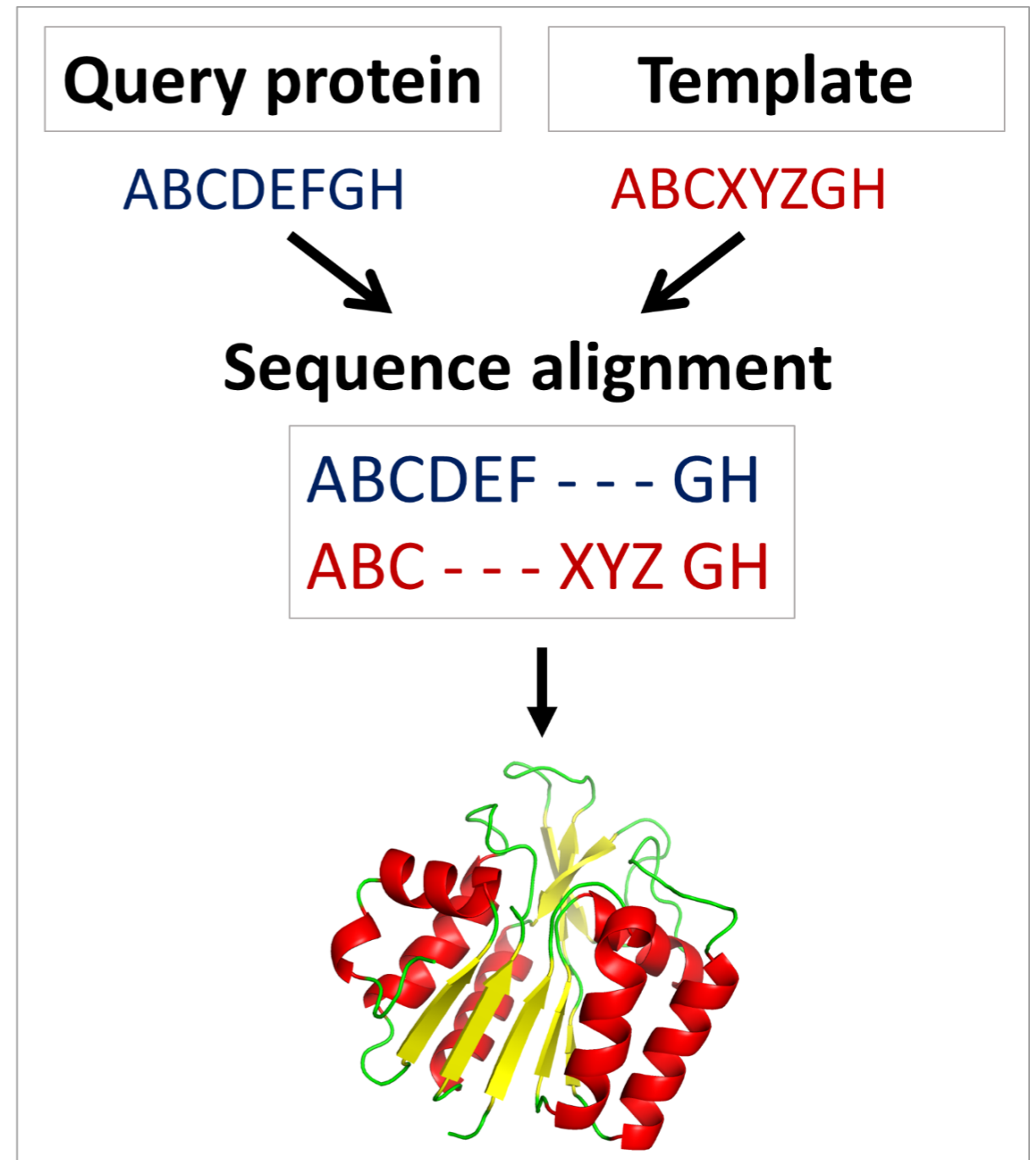March 29, 2019

Residue-residue contact map

Contact map

Random-Top-L/10

Contact

Bhattacharya
Laboratory

Research question

Does contact map boost protein threading?

What is protein threading?

– align sequence to template

Query protein   **Template**

ABCDEFGH   ABCXYZGH

Sequence alignment

ABCDEF - - - GH
ABC - - - XYZ GH

Challenges in protein threading

- Challenge 1:

   Finding **TOP 1 template** from the template library (**fold recognition**)

- Challenge 2:

   Getting optimal **query-template alignment**

Our hypothesis

Contact information on top of protein **threading boosts** the **performance** of purely threading-based methods.
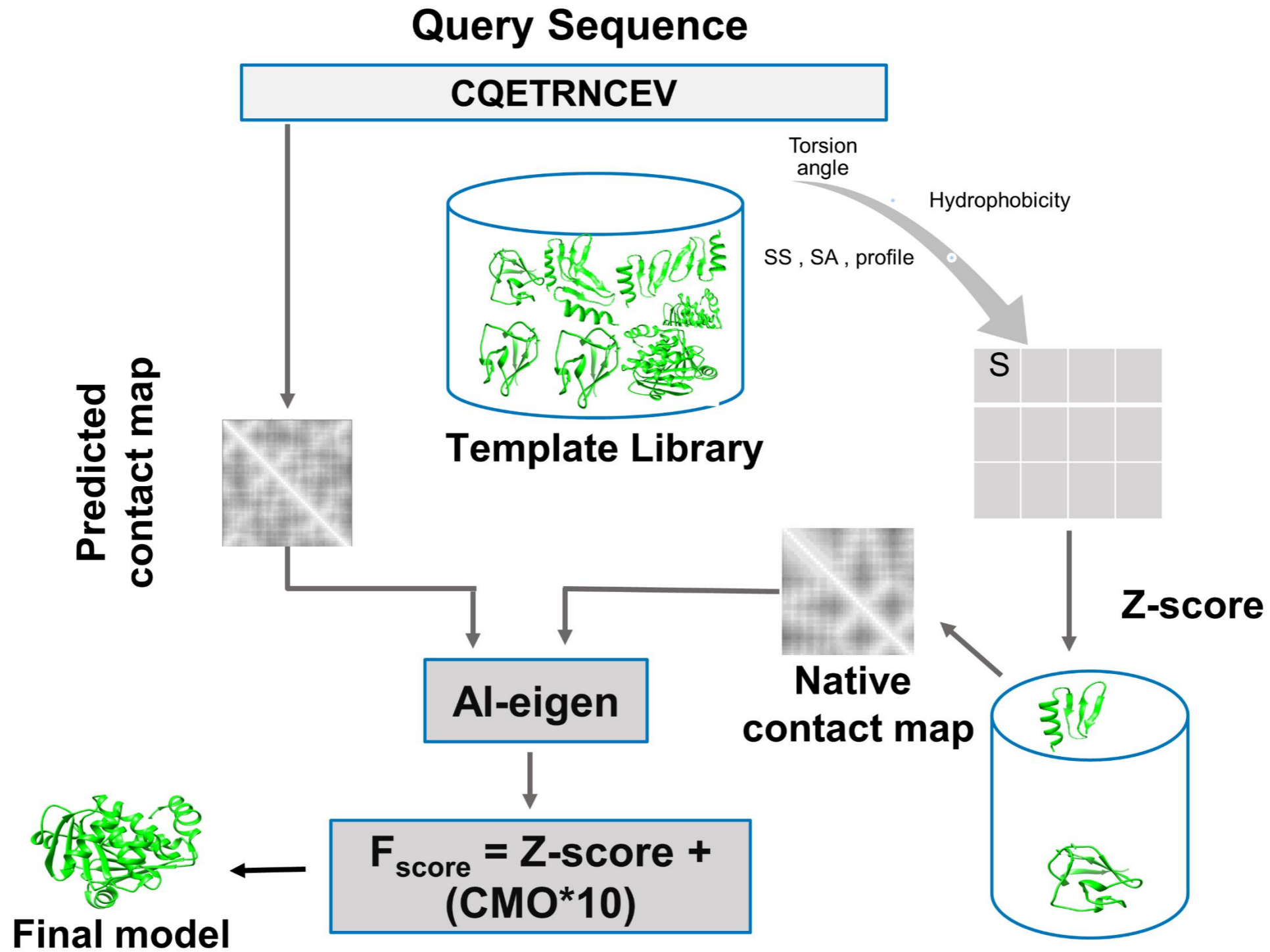
Bhattacharya
Laboratory

Research Article | 🔓 **Full Access** |

Does Inclusion of Residue-Residue Contact Information Boost Protein Threading?

Sutanu Bhattacharya, Debswapna Bhattacharya ✉

Bhattacharya
Laboratory

Flow chart of our work

(Lena and coworkers, 2010
Bhattacharya and coworkers, 2019)

Results

benchmark on 500 protein targets

Bhattacharya
Laboratory

(Zhang and coworkers, 2008)

Results (1)

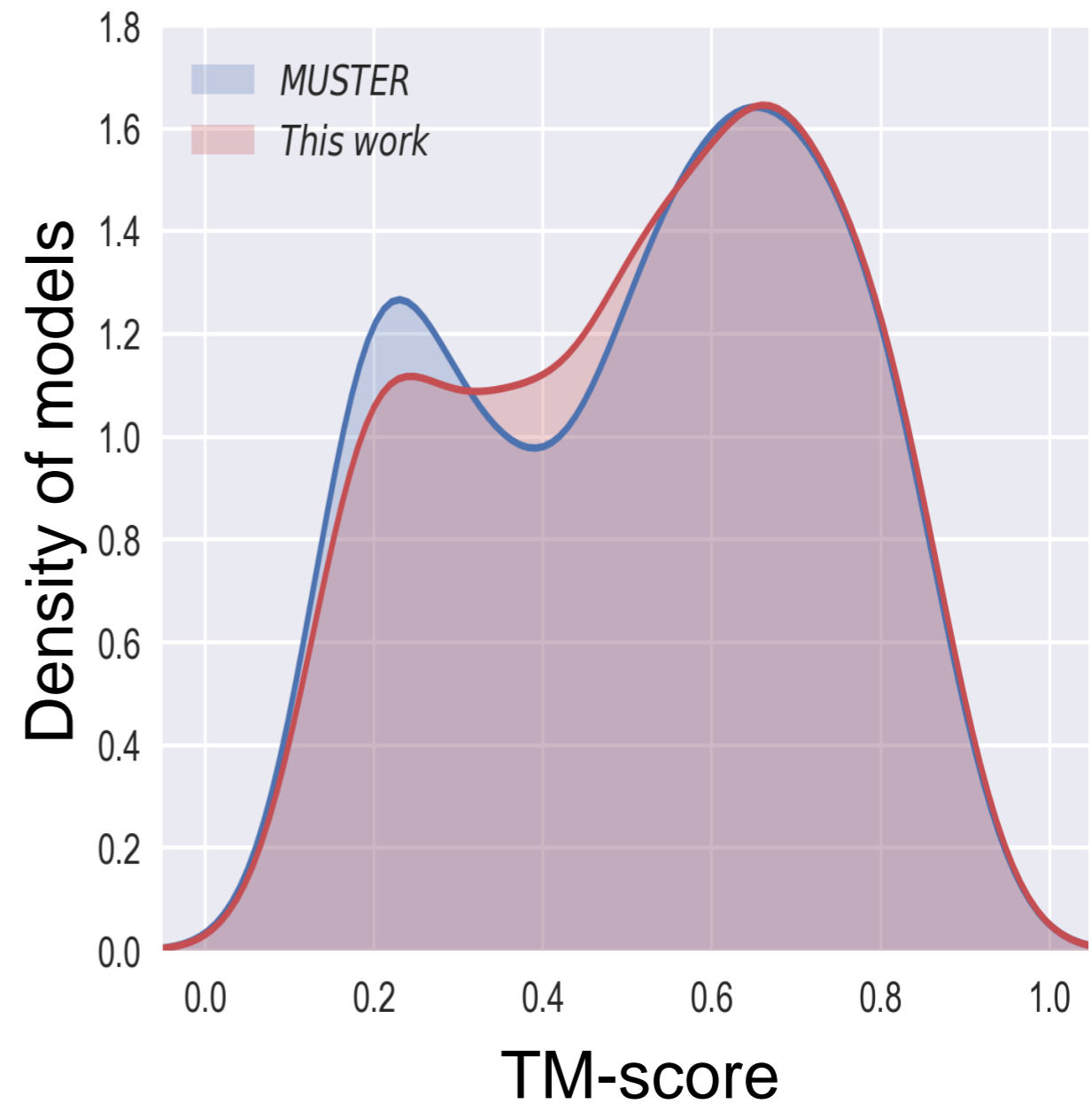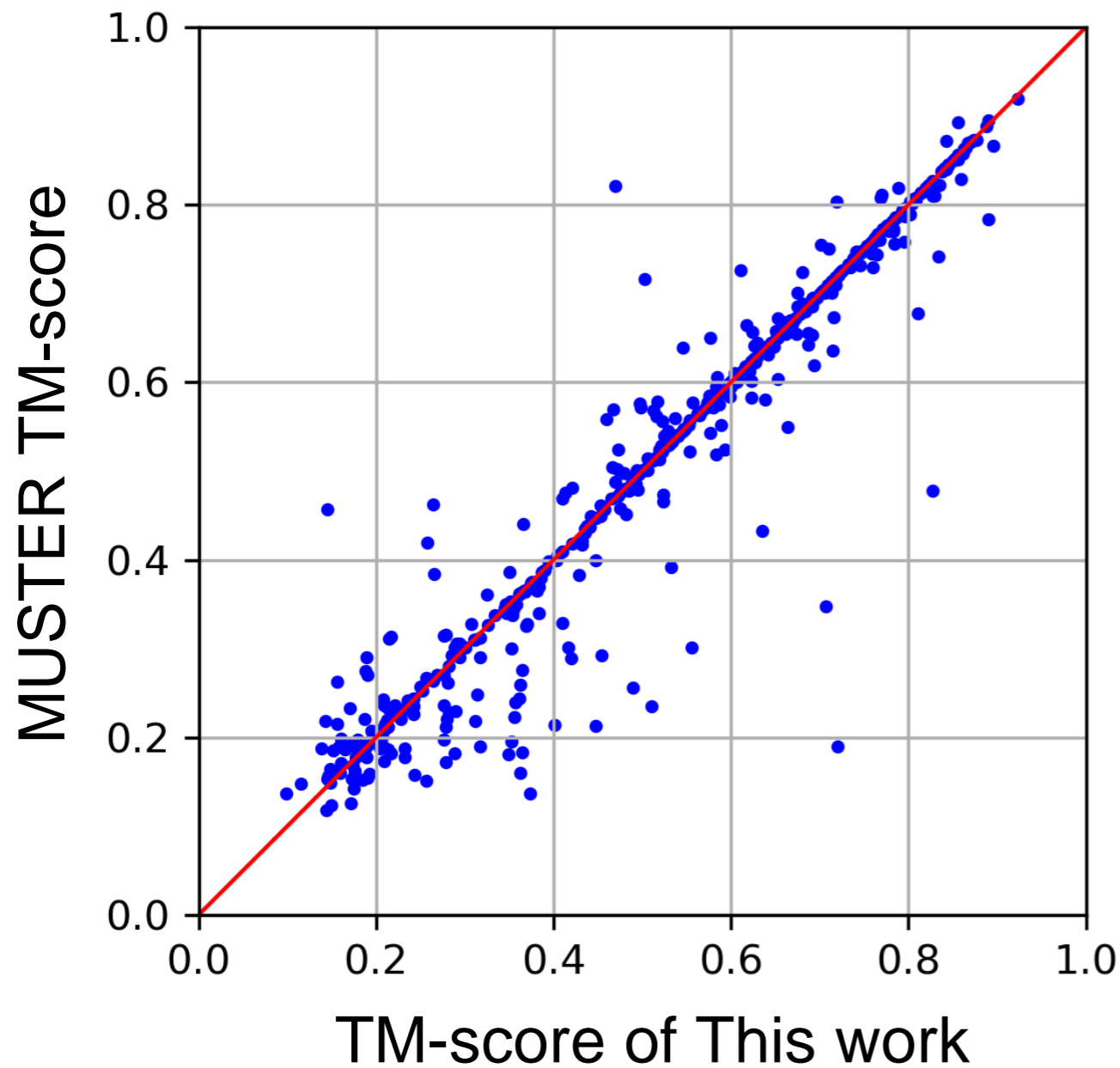| dataset	comparisons with threading method	measures
Test500	MUSTER	TM-score of top-ranked model

Test500 (MUSTER vs. This work)

- **better** average TM-score than MUSTER
- statistically significance (p-value < 0.05)

Bhattacharya
Laboratory

Results (2)

benchmark on 150 protein targets

Bhattacharya
Laboratory

(Jones and coworkers, 2012)

Results (2)

dataset	comparisons with ab-initio folding method	measures
PSICOV150	CONFOLD2	TM-score of top-ranked model

Bhattacharya
Laboratory

(Cheng and coworkers, 2018)

PSICOV150 (CONFOLD2 vs. This work)
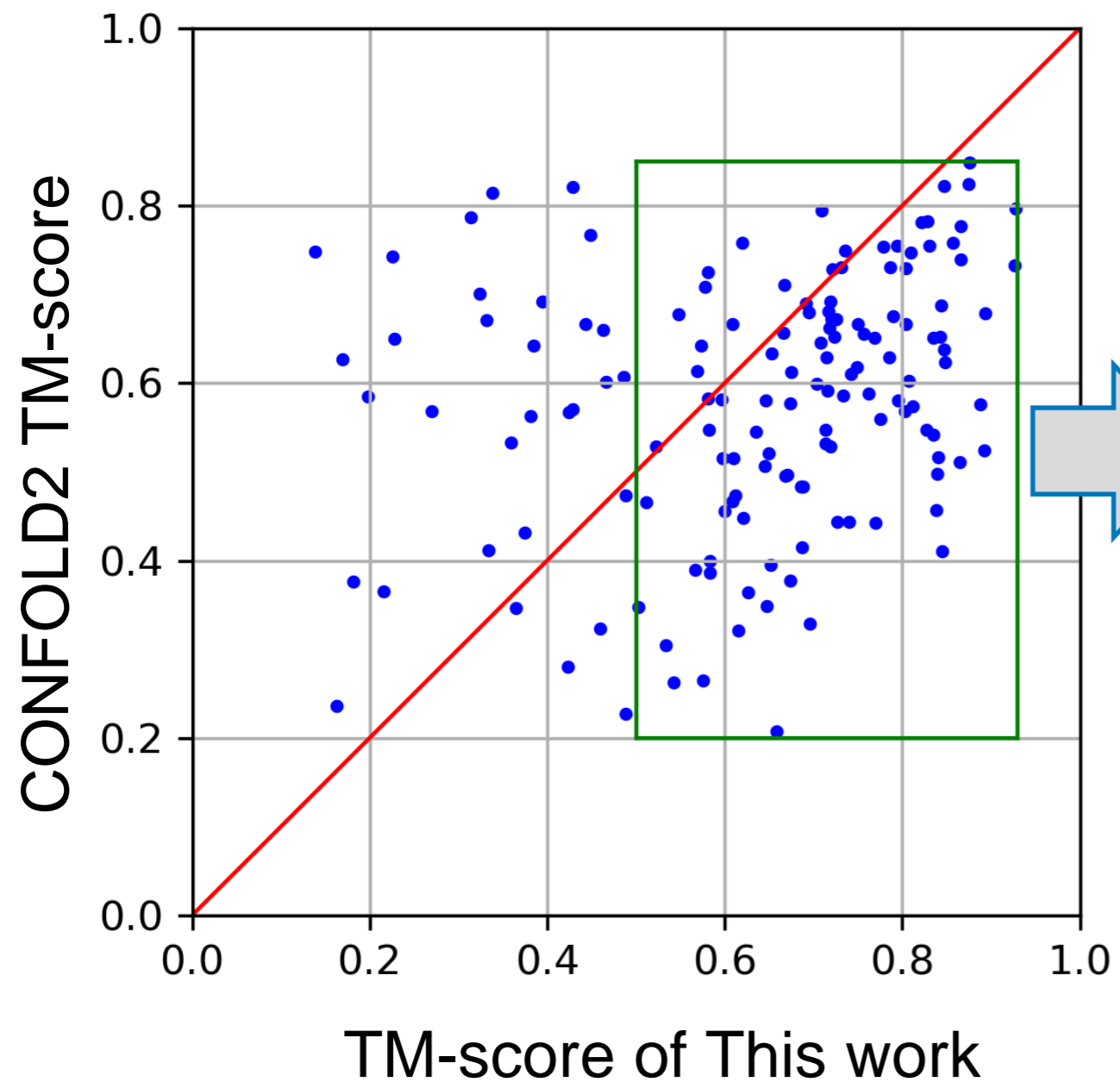
- **better** average TM-score than CONFOLD2
- statistically significance (p-value < 0.05)
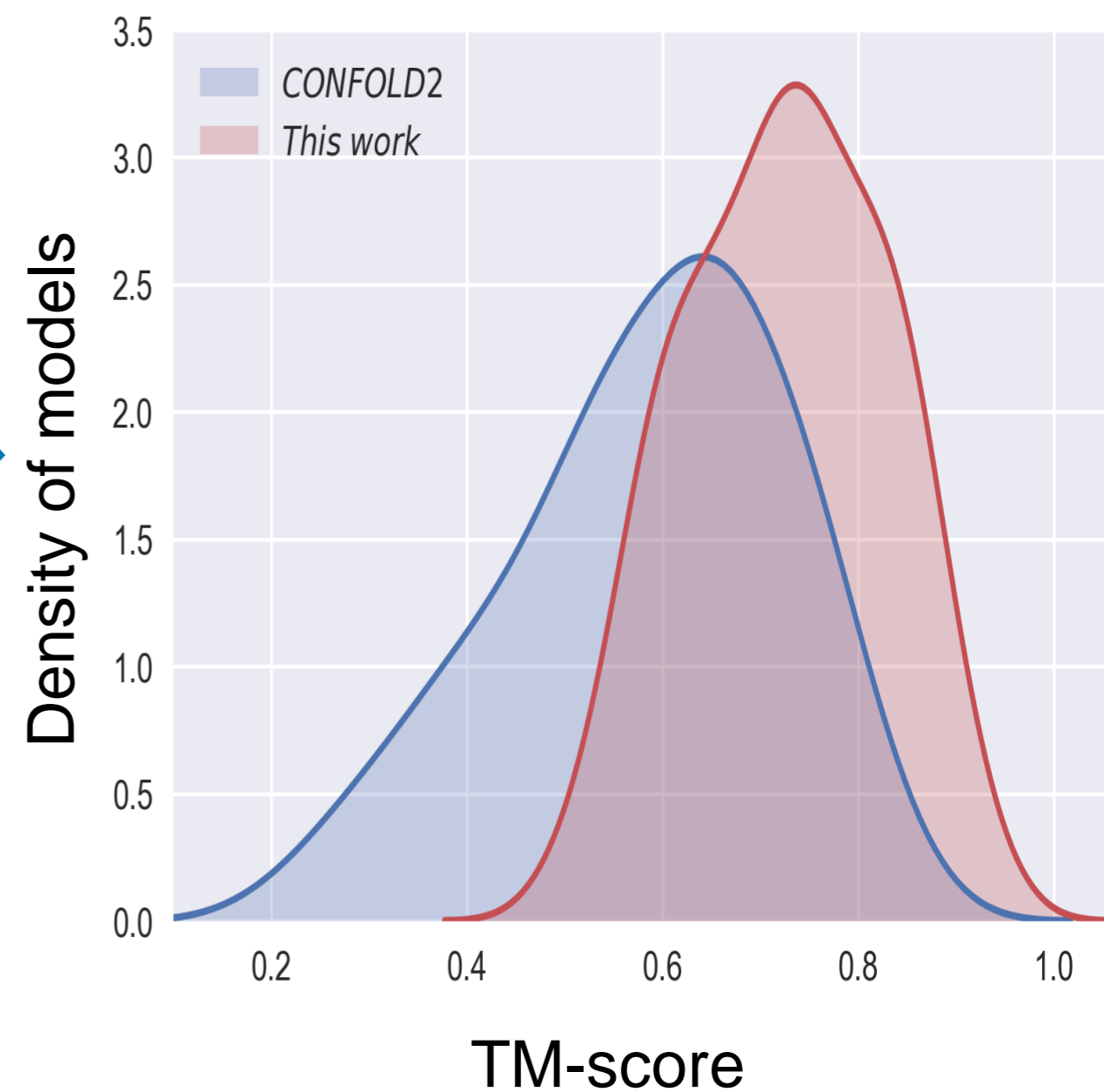
Results (3)

benchmark on 20 CASP13 protein targets

Bhattacharya
Laboratory

(http://predictioncenter.org/casp13)

Results (3)

dataset	comparisons with state-of-the-art contact-assisted threading methods	measures
CASP13 (**20 full-length** proteins in total of **32 domains**)	- EigenTHREADER - map-align	TM-score of top-ranked model

(Jones and coworkers, 2017
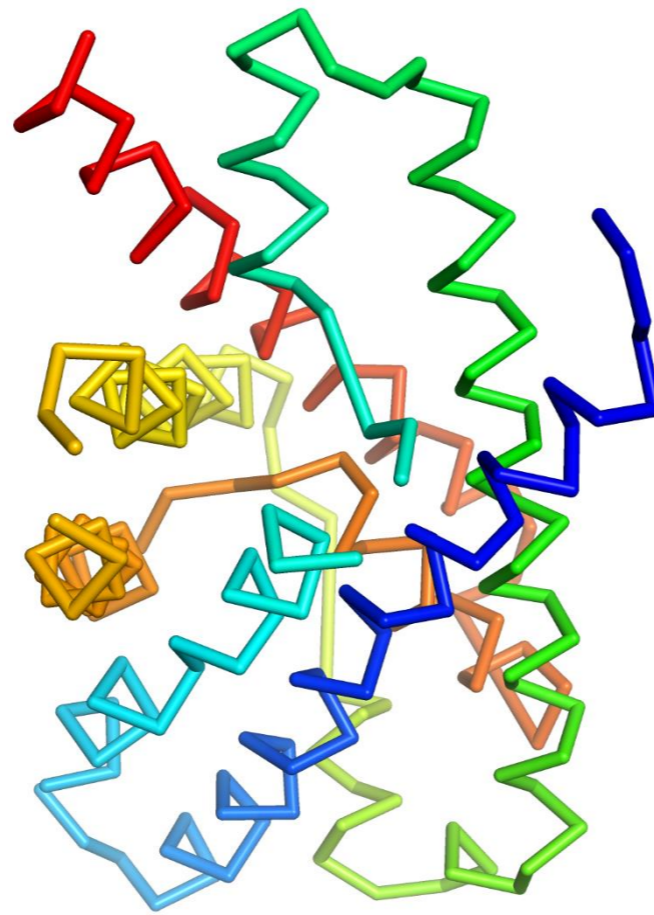Baker and coworkers, 2017)

CASP13 performance

Bhattacharya
Laboratory
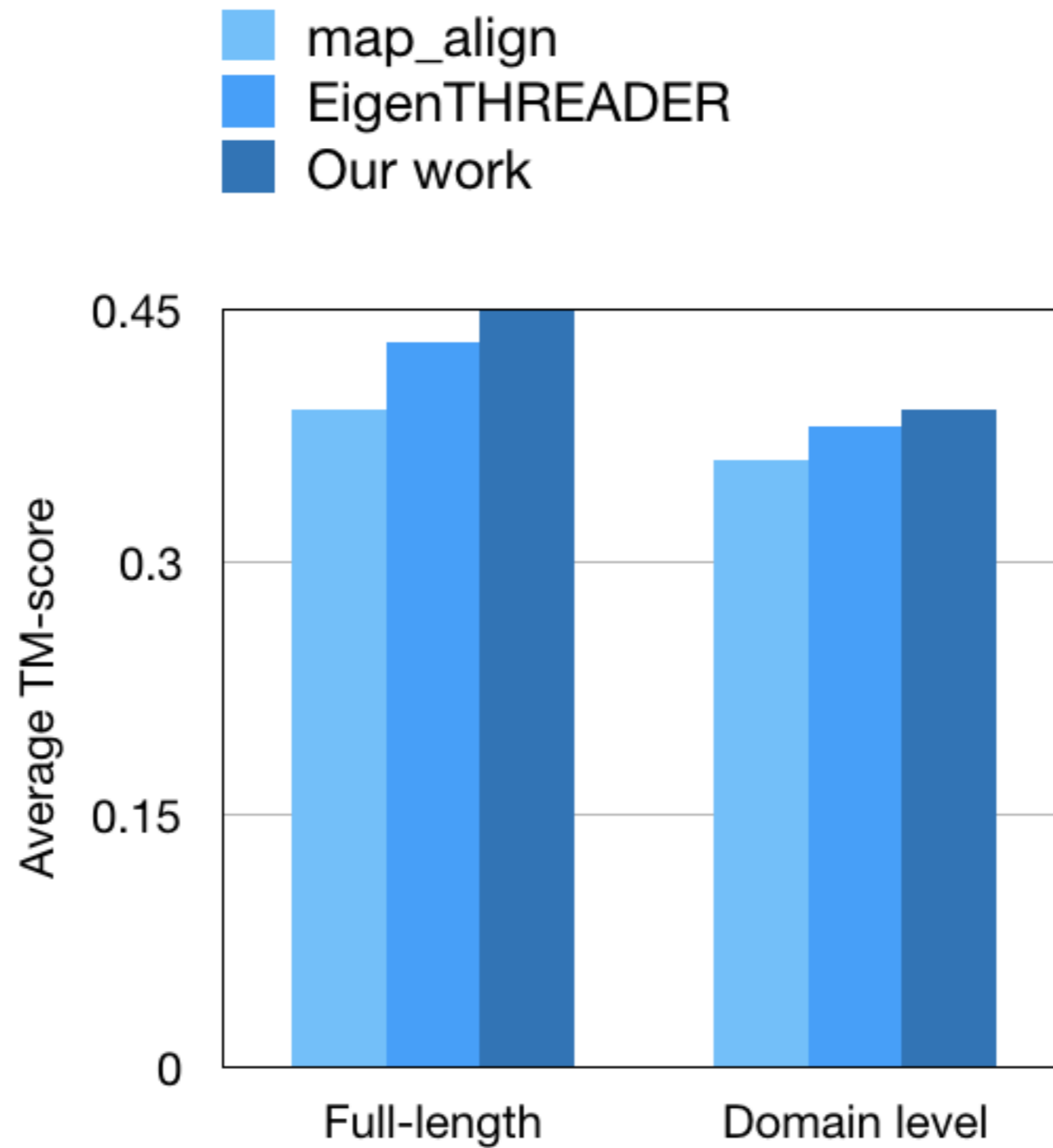
Case study: 2f2ba (245 residues)

Native 3D structure
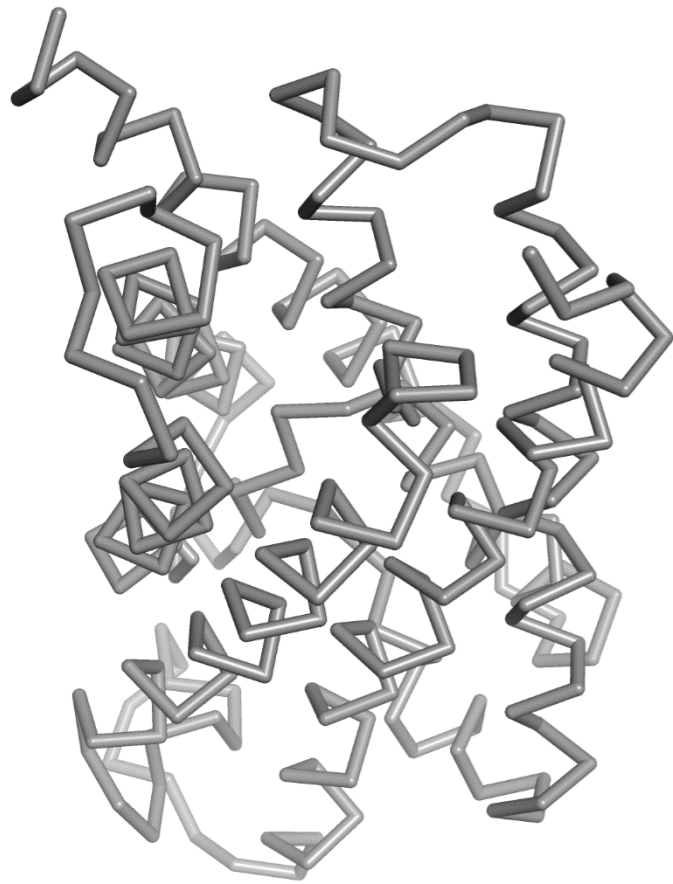
Predicted model
Contact + Threading

Predicted model
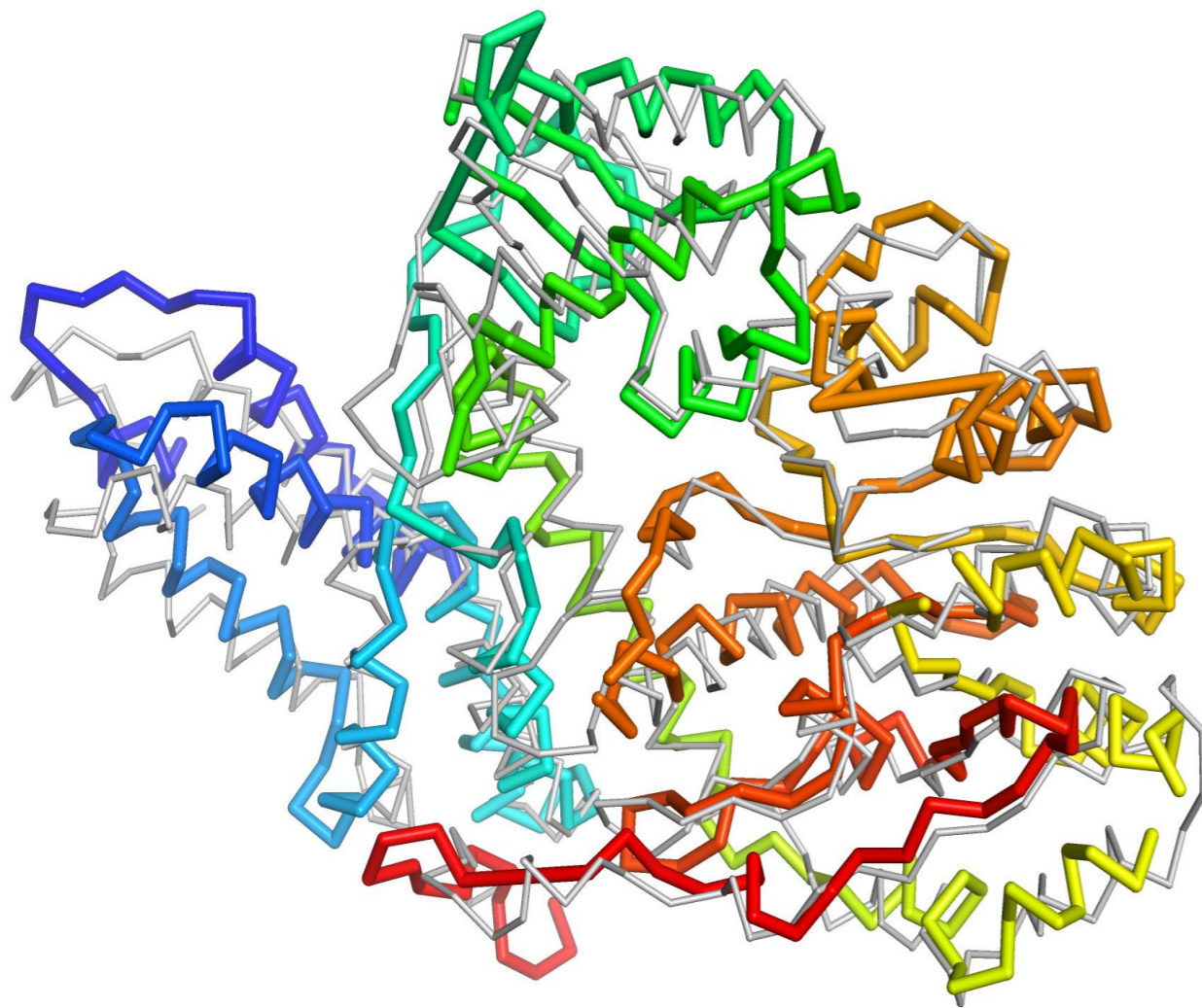**Only Threading**

TM-score: **0.59**

TM-score: **0.28**

Bhattacharya
Laboratory
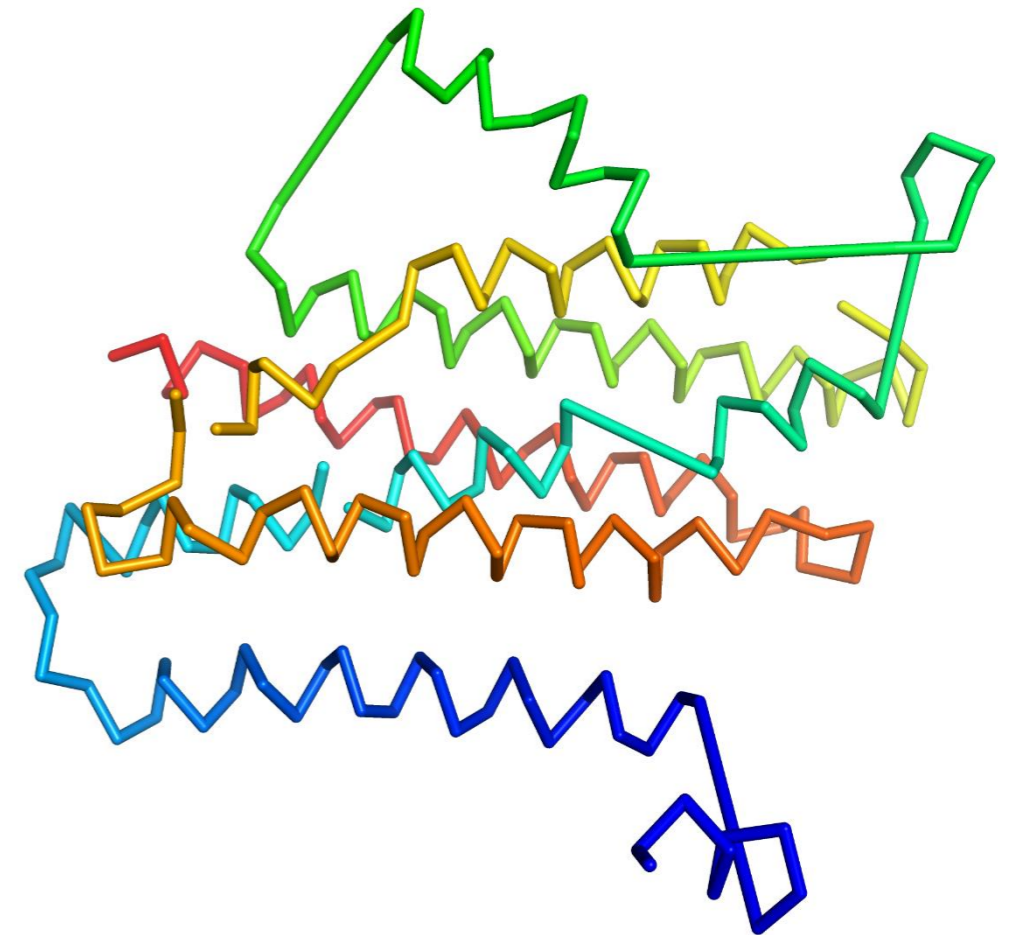
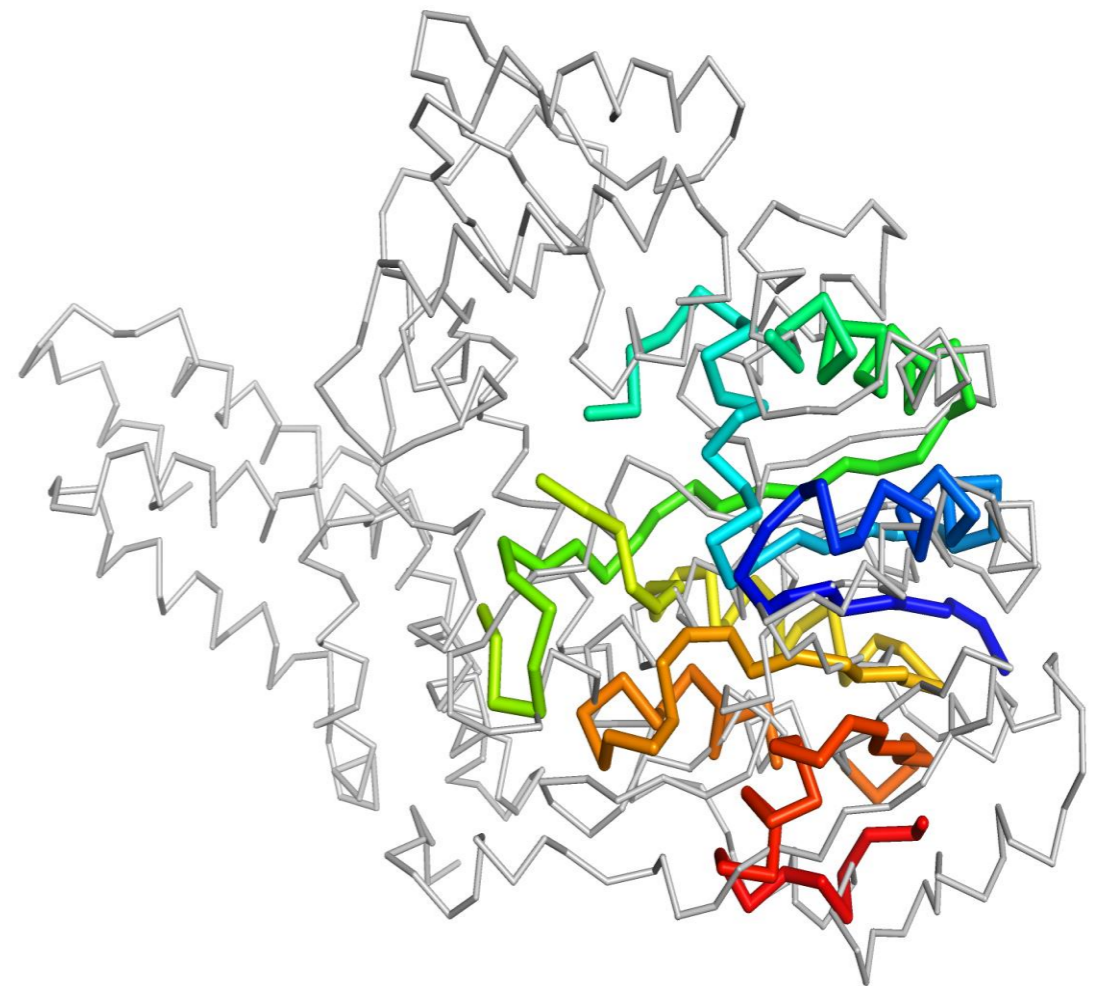# Case study: T0966 (494 residues)

Our predicted **model** (rainbow) superimposed to the native (gray)

EigenTHREADER predicted model (rainbow) superimposed to the native (gray)

TM-score: **0.8**

TM-score: 0.19

Conclusions and Future work

– **Test500**: **contact + threading better** than purely threading-based methods

– **PSICOV150**: **contact + threading better** than contact-assisted ab initio folding methods

– **Contact-assisted threading** is a **promising** avenue for **fold recognition**.

– What about getting best fit **query-template alignment** (**Challenge 2**)?

– What about residue-residue **distance** instead of contacts?

Bhattacharya
Laboratory

Acknowledgements

Dr. Debswapna
Bhattacharya

Auburn
University

Query Sequence

CQETRKKCTEMKKKFKNCEV

Hydrophobicity match

SS & SA match

profile match

Torsion Angles match

Score (S)

Template Library

Predicted contact map

Template sequence

$S_{C,E}$

Query sequence

DP table

AI-eigen

CMO

Z_{score}

Native contact map

$F_{score} = Z_{score} + (CMO*10)$

Final model

1

20