

Supplementary Information

for

DisCovER: distance- and orientation-based covariational threading for weakly homologous proteins

Sutanu Bhattacharya¹, Rahmatullah Roche², Bernard Moussad², Debswapna Bhattacharya^{2,*}

¹Department of Computer Science, Florida Polytechnic University, Lakeland, FL 33805, USA

²Department of Computer Science, Virginia Tech, Blacksburg, VA 24061, USA

*Corresponding author: Debswapna Bhattacharya

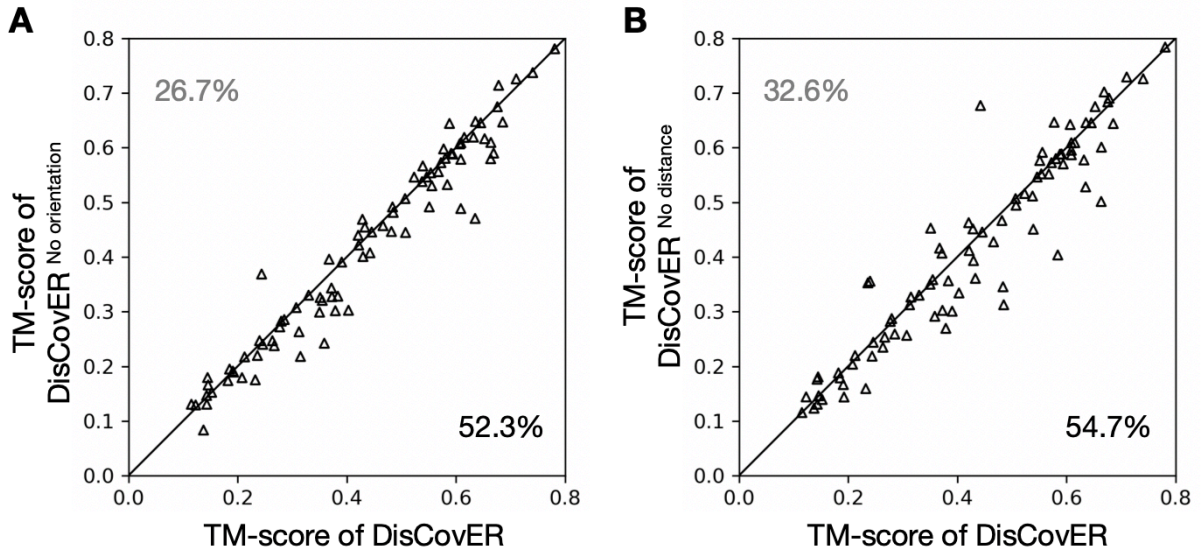
2160N Torgersen Hall, 620 Drillfield Drive, Blacksburg, VA
24061, USA

Phone: (540) 231-2865

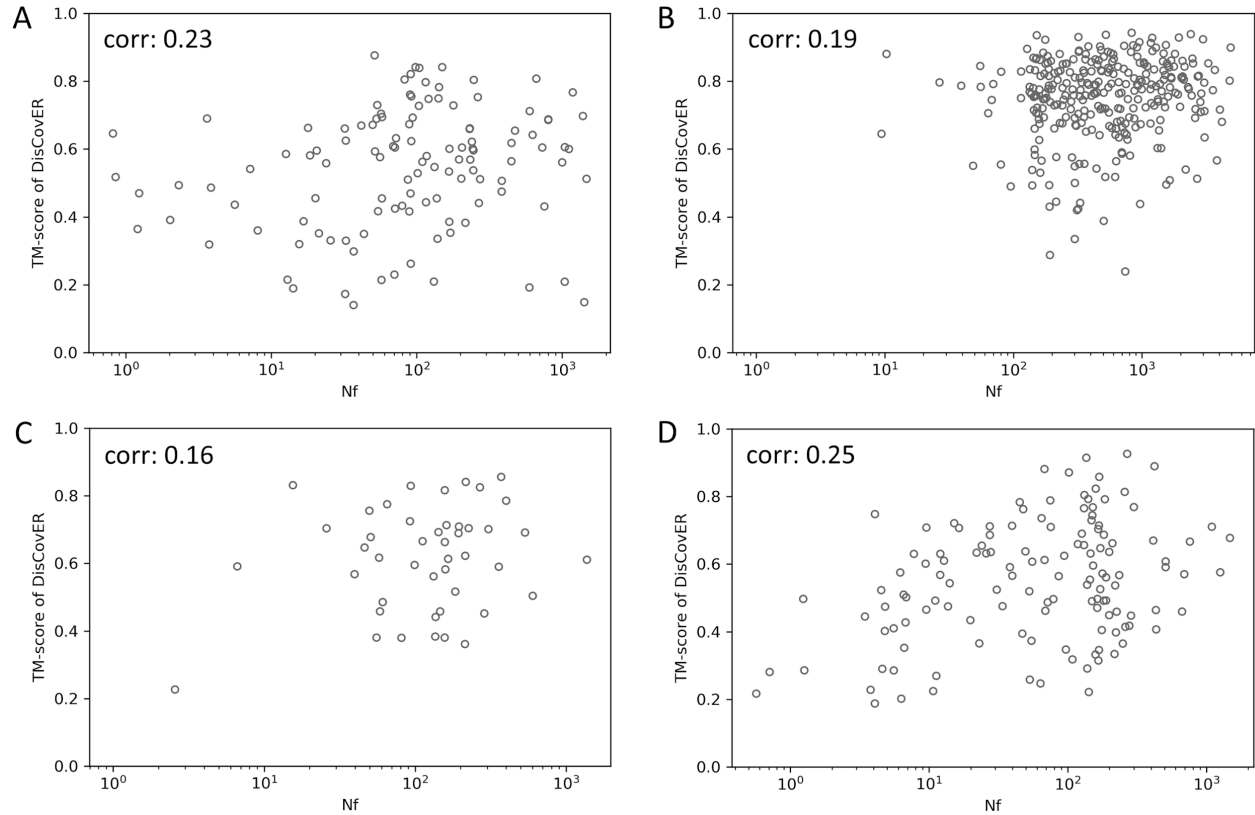
Fax: (540) 231-6075

E-mail: dbhattacharya@vt.edu

Supplementary Fig. S1. (A) TM-score of DisCovER vs DisCovER^{No orientation} on MUSTER dataset. 52.3% of the cases DisCovER outperforms DisCovER^{No orientation} whereas 26.7% of the cases DisCovER^{No orientation} outperforms DisCovER. (B) TM-score of DisCovER vs DisCovER^{No distance} on MUSTER dataset. 54.7% of the cases DisCovER outperforms DisCovER^{No distance} whereas 32.6% of the cases DisCovER^{No distance} outperforms DisCovER. DisCovER represents the full-fledged variant that incorporates both orientation and distance. DisCovER^{No orientation} represents the ablated variant that only incorporates distance but no orientation. DisCovER^{No distance} represents the ablated variant of DisCovER that only incorporates orientation but no distance.



Supplementary Fig. S2. TM-score of DisCovER vs Nf (number of sequence homologs) on (A) 117 hard targets from CAMEO having a Spearman correlation of 0.23, (B) 304 easy targets from CATHER having a Spearman correlation of 0.19, (C) 45 medium targets from CATHER having a Spearman correlation of 0.16, (D) 131 hard targets from CATHER having a Spearman correlation of 0.25.



Supplementary Table S1. Contribution of individual features to DisCovER performance on CAMEO targets 5ZER_B and 6D7Y_B. The performance is measured by TM-score. Bold values represent the best performance. Nf defines the number of effective sequence homologs.

Target (Type)	Nf	DisCovER ^{No} neighborhood, No geometry [†]	DisCovER ^{No} geometry [†]	DisCovER ^{No} distance	DisCovER ^{No} orientation	DisCovER
5ZER_B (hard)	104.3	0.782	0.797	0.798	0.802	0.826
6D7Y_B (very hard)	21.3	0.292	0.307	0.330	0.372	0.446

[†]geometry includes distance and orientation information.

Supplementary Table S2. Performance comparison of DisCovER and AlphaFold2 on representative CAMEO targets 7APJ_B and 7D8B_B. Bold values represent the best performance.

Target details			TM-score		
Target (Submission date)	Type [†]	Length	HHpred ^a	DisCovER ^b	AlphaFold2 ^c
7APJ_B (August 21, 2021)	Easy	126	0.769	0.913	0.896
7D8B_B (August 21, 2021)	Medium	161	0.487	0.847	0.873

[†] officially classified by CAMEO

^a Jobs were submitted to HHpred webserver (<https://toolkit.tuebingen.mpg.de>) using default parameters and the default template library (PDB_mmCIF70_13_Jul)

^b Template library, containing 70 670 templates, curated in 2018, significantly older than the submission date of both targets.

^c Jobs were submitted to AlphaFold2 Colab notebook (<https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb>) using default parameter settings.