# Supplementary information for protein threading using residue co-variation and deep learning

Jianwei Zhu[1,2,3], Sheng Wang[3], Dongbo Bu[1,2,*] and Jinbo Xu[3,*]

[1]Key lab of intelligent information process, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, [2]University of Chinese Academy of Sciences, Beijing, China, [3]Toyota Technological Institute at Chicago, Chicago, Illinois, United States of America.

*To whom correspondence should be addressed.

## 1 *P*-values

**Table 1 The *p*-value between DeepThreader and the other methods when the alignment accuracy on Test500 is evaluated.** The *p*-values are calculated by three different quality measures: TM-score, GDT and uGDT. Test protein pairs are split into groups based upon its TM-score.

| Group | Count | DeepThreader vs HHpred | | | DeepThreader vs CNFpred | | | DeepThreader vs EigenTHREADER | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | TM-score | GDT | uGDT | TM-score | GDT | uGDT | TM-score | GDT | uGDT |
| **(0.00, 1.00]** | **1000** | **2.7e-11** | **4.5e-08** | **7.2e-05** | **3.6e-04** | **3.9e-03** | **2.1e-02** | **1.2e-34** | **2.9e-37** | **1.1e-26** |
| (0.00, 0.25] | 16 | 4.0e-02 | 0.174 | 0.267 | 0.419 | 0.466 | 0.569 | 0.307 | 9.1e-03 | 3.6e-02 |
| (0.25, 0.40] | 166 | 6.1e-13 | 6.1e-06 | 3.9e-04 | 6.7e-04 | 1.7e-02 | 0.096 | 6.8e-10 | 1.4e-09 | 9.7e-09 |
| **(0.40, 0.65]** | **398** | **2.2e-20** | **5.3e-15** | **7.0e-07** | **6.5e-08** | **2.2e-05** | **1.8e-03** | **2.1e-41** | **1.3e-38** | **1.8e-24** |
| (0.65, 0.80] | 273 | 5.7e-07 | 2.6e-05 | 3.6e-02 | 6.2e-04 | 7.6e-03 | 0.176 | 6.0e-39 | 4.5e-39 | 2.8e-15 |
| (0.80, 1.00] | 147 | 0.082 | 0.073 | 0.61 | 0.415 | 0.536 | 0.831 | 1.9e-14 | 2.2e-16 | 3.1e-04 |

**Table 2 The *p*-value between DeepThreader and the other methods when the threading performance on Test500 is evaluated.** "TM-score<x" means that when doing threading we exclude all the templates whose structure similarity (measured by TM-score) with the test protein is larger than x. The *p*-values are calculated on the 3D models built from the first-ranked templates.

| | TM-score<0.50 | | | TM-score<0.55 | | | TM-score<0.60 | | | TM-score<0.65 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | TM-score | GDT | uGDT | TM-score | GDT | uGDT | TM-score | GDT | uGDT | TM-score | GDT | uGDT |
| vs HHpred | 2.6e-50 | 3.8e-37 | 6.8e-18 | 5.2e-36 | 1.0e-27 | 8.9e-13 | 2.5e-24 | 9.9e-20 | 1.9e-08 | 8.6e-18 | 9.5e-15 | 2.9e-06 |
| vs CNFpred | 3.8e-24 | 8.6e-19 | 7.5e-11 | 1.2e-18 | 5.8e-15 | 7.2e-08 | 6.9e-12 | 9.6e-10 | 4.0e-05 | 5.3e-08 | 6.8e-07 | 1.2e-03 |
| vs CNFpred[DL] | 8.6e-08 | 4.1e-05 | 4.6e-03 | 3.1e-05 | 7.0e-04 | 2.4e-02 | 1.9e-03 | 1.5e-02 | 0.102 | 1.1e-02 | 4.1e-02 | 0.184 |
| vs EigenTH | 4.9e-103 | 4.0e-86 | 1.6e-77 | 3.5e-107 | 2.9e-95 | 2.2e-77 | 5.8e-120 | 5.9e-108 | 9.8e-84 | 3.2e-119 | 8.6e-109 | 2.0e-86 |

| | TM-score<0.70 | | | TM-score<0.75 | | | TM-score<0.80 | | | TM-score<0.85 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | TM-score | GDT | uGDT | TM-score | GDT | uGDT | TM-score | GDT | uGDT | TM-score | GDT | uGDT |
| vs HHpred | 9.1e-12 | 7.1e-10 | 4.0e-04 | 4.3e-08 | 3.2e-07 | 4.2e-03 | 2.0e-05 | 6.6e-05 | 3.9e-02 | 7.2e-05 | 2.1e-04 | 0.079 |
| vs CNFpred | 6.7e-05 | 3.9e-04 | 2.1e-02 | 3.0e-03 | 6.0e-03 | 0.097 | 2.3e-02 | 3.4e-02 | 0.21 | 0.067 | 0.106 | 0.394 |
| vs CNFpred[DL] | 4.5e-02 | 0.116 | 0.259 | 0.181 | 0.288 | 0.421 | 0.380 | 0.467 | 0.555 | 0.416 | 0.534 | 0.686 |
| vs EigenTH | 9.5e-128 | 5.2e-118 | 5.0e-89 | 7.7e-125 | 6.1e-119 | 2.8e-87 | 1.8e-116 | 4.6e-114 | 6.5e-82 | 3.4e-122 | 2.2e-120 | 3.2e-82 |

**Table 3 The *p*-value between DeepThreader and the other methods on the 86 CASP12 domains.** Each cell shows the *p*-value calculated on the 3D models built from the first-ranked templates. The number of domains for ALL, FM, FM/TBM and FM+FM/TBM is 86, 38, 13 and 51, respectively.

| | ALL | | | FM | | | FM/TBM | | | FM+FM/TBM | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | TM-score | GDT | uGDT | TM-score | GDT | uGDT | TM-score | GDT | uGDT | TM-score | GDT | uGDT |
| **vs HHpred** | **2.20e-02** | **1.50e-02** | **0.314** | **6.50e-03** | **2.10e-02** | **3.00e-02** | **1.80e-02** | **4.20e-03** | **0.448** | **9.50e-04** | **1.70e-03** | **0.064** |
| vs CNFpred | 0.190 | 0.144 | 0.513 | 0.196 | 0.187 | 0.360 | 2.50e-02 | 2.00e-02 | 0.303 | 3.80e-02 | 4.00e-02 | 0.180 |
| vs CNFpred[DL-] | 0.530 | 0.573 | 0.832 | 0.473 | 0.594 | 0.621 | 0.308 | 0.276 | 0.765 | 0.325 | 0.405 | 0.609 |
| **vs EigenTH** | **4.40e-09** | **2.30e-11** | **1.40e-04** | **1.00e-04** | **9.10e-05** | **6.90e-05** | **4.10e-05** | **4.00e-05** | **2.50e-02** | **4.70e-07** | **8.50e-07** | **4.90e-05** |

**Table 4 The p-value between DeepThreader and the other methods on 64 CASP12 domains.** Each cell shows the *p*-value calculated on the 3D models built from the first-ranked templates. The number of domains for ALL, FM, FM/TBM and FM+FM/TBM is 64, 34, 12 and 46, respectively.

| | ALL | | | FM | | | FM/TBM | | | FM+FM/TBM | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | TM-score | GDT | uGDT | TM-score | GDT | uGDT | TM-score | GDT | uGDT | TM-score | GDT | uGDT |
| **vs HHpred** | **8.60e-03** | **6.90e-03** | **0.268** | **7.20e-03** | **2.30e-02** | **2.50e-02** | **1.50e-02** | **3.80e-03** | **0.434** | **9.10e-04** | **1.90e-03** | **0.063** |
| vs CNFpred | 0.096 | 0.073 | 0.406 | 0.173 | 0.159 | 0.306 | 2.20e-02 | 2.00e-02 | 0.292 | 3.00e-02 | 3.40e-02 | 0.153 |
| vs CNFpred[DL-] | 0.445 | 0.488 | 0.792 | 0.474 | 0.584 | 0.578 | 0.312 | 0.267 | 0.762 | 0.324 | 0.397 | 0.592 |
| **vs EigenTH** | **8.90e-09** | **5.00e-10** | **1.80e-04** | **4.30e-04** | **4.30e-04** | **2.30e-04** | **1.70e-05** | **3.10e-05** | **2.20e-02** | **8.90e-07** | **3.40e-06** | **8.50e-05** |

## 2     Evaluation of Distance Prediction
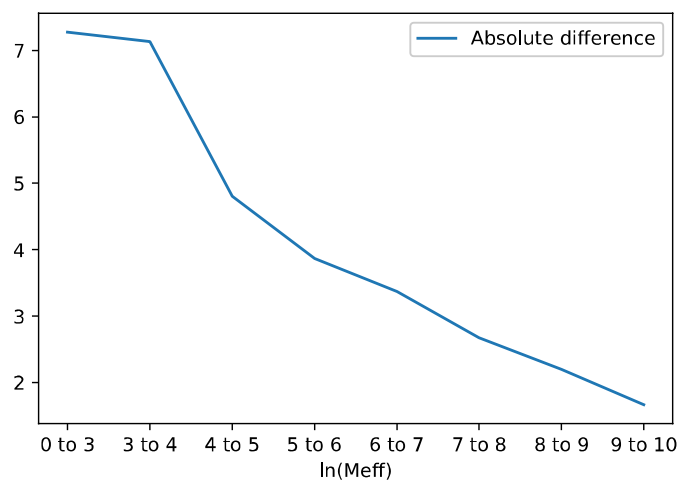


**Fig. 1** The absolute distance prediction error with respect to the number of effective sequence homologs (measured by ln*Meff*), derived from the query proteins in Test500. The absolute error is defined as the difference between predicted distance and its native value. We estimate the distance between two residues by calculating the expected distance from the predicted distance probability distribution. Only those residue pairs with predicted distance ≤15Å are considered. The error of one protein is the average prediction error of all the residue pairs with predicted distance ≤15Å. All the query proteins are split into groups by their ln*Meff* and this figure shows the average prediction error in each group.
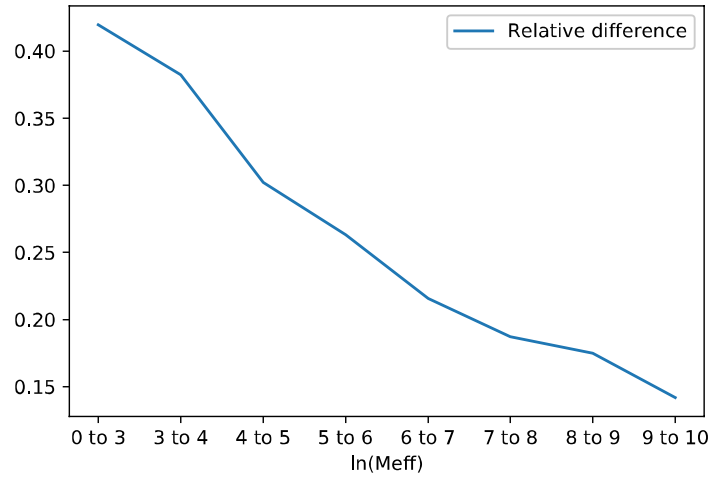
**Fig. 2** The relative distance prediction error with respect to the number of effective sequence homologs (measured by ln*Meff*), derived from the query proteins in Test500. The relative error is defined as the difference between predicted distance and its native value divided by the average of the predicted distance and its native. Only those residue pairs with predicted distance ≤15Å are considered. The error of one protein is the average prediction error of all the residue pairs with predicted distance ≤15Å. All the query proteins are split into groups by their ln*Meff* and this figure shows the average prediction error in each group.



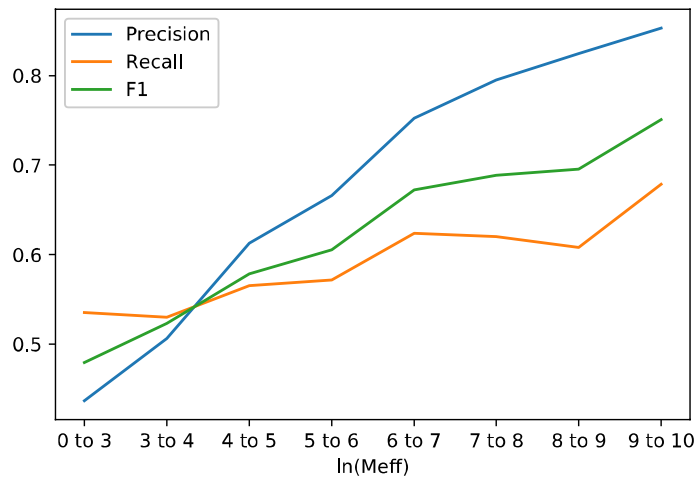**Fig. 3** The precision, recall and F1 score of predicting one pair of residues have distance ≤15Å with respect to the number of effective sequence homologs (measured by ln*Meff*), tested on Test500. Only those pairs of residues with sequence separation at least 12 are considered. The values are first averaged over all the residue pairs in a protein and then averaged over all the proteins in a group defined by ln*Meff*.