SPOT-Fold: Fragment-Free Protein Structure Prediction Guided by Predicted Backbone Structure and Contact Map

Yufeng Cai,[†[a]] Xiongjun Li,[†[a]] Zhe Sun,[[a]] Yutong Lu,[[a]] Huiying Zhao,[[b]] Jack Hanson,[[c]] Kuldip Paliwal,[[c]] Thomas Litfin,[[d]] Yaoqi Zhou,[*[d]] and Yuedong Yang*[[a]]

Protein structure determination has long been one of the most challenging problems in molecular biology for the past 60 years. Here we present an *ab initio* protein tertiary-structure prediction method assisted by predicted contact maps from SPOT-Contact and predicted dihedral angles from SPIDER 3. These predicted properties were then fed to the crystallography and NMR system (CNS) for restrained structure modeling. The resulted structures are first evaluated by the potential energy calculated by CNS, followed by dDFIRE energy function for model selections. The method called SPOT-Fold has been tested on 241 CASP targets between 67 and 670 amino acid residues, 60 randomly selected globular proteins under 100 amino acids. The method has a comparable accuracy to other contact-map-based modeling techniques. © 2019 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.26132

Introduction

Protein tertiary structures are the foundations for their diverse functions. This was demonstrated by the first experimentally solved protein structure of myoglobin, which provided a deep mechanistic understanding to its function.[1] Protein structures can be experimentally determined by a number of techniques such as nuclear magnetic resonance,[2] X-ray crystallography, and cryogenic electron microscopy.[3] However, the high cost of experimental structure determination and the low cost of whole-genome sequencing make it practically necessary to model protein structures computationally.

Computational techniques for protein structure modeling are assessed in the biennial Critical Assessment of Structure Prediction (CASP) meeting,[4] in which computational scientists predict soon-to-be-solved protein structures in the summer, followed by model assessment against the structures solved experimentally afterward later in the year. The latest published CASP results[5,6] indicate that the most reliable techniques mix, match, and assemble known native structures either in whole (template-based modeling) or in part (fragment assembly).[7–12]

"*Ab initio*" structure-prediction techniques are the methods that attempt to predict protein structures without relying on known structures (template or fragment). These fragment-free techniques have the potential to model structures with unseen folds. While some successes were made for small proteins,[13–15] it remains challenging for proteins of moderate or large sizes (>100 residues). Recently, highly accurate prediction of protein contact maps can be made by integrating modern deep learning with evolutionary coupling techniques.[16–18] This led to the development of several methods for *ab initio* structure prediction restrained by predicted contacts.[19–21] The accuracy of these methods is remarkably close to those of fragment-based techniques for template-free modeling targets. Moreover, they are computationally more efficient because fewer decoys are required to capture near-native conformations.[19,21]

In this paper, we seek to further improve *ab initio* structure prediction by following the same contact-restrained approach. This work is built on our recent success in improving prediction of secondary structure (SS), backbone torsion angles, and contact maps by using two-dimensional long-short term memory bidirectional neural networks (2D-LSTM-BRNN) and residual convolution neural networks (ResNet).[16,18,22–24] We showed that the software CNS[25,26] with the input of the distance and angle restraints[27] derived from 1D backbone-structure and 2D contact predictions can produce near-native structures. The

[a] *Y. Cai, X. Li, Z. Sun, Y. Lu, Y. Yang*
School of Data and Computer Science, Sun Yat-Sen University, 132 East Circle at University City, Guangzhou 510006, China
E-mail: yangyd25@mail.sysu.edu.cn
[b] *H. Zhao*
Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510000, China
[c] *J. Hanson, K. Paliwal*
Signal Processing Laboratory, Griffith University, Brisbane, Queensland, 4122, Australia
[d] *T. Litfin, Y. Zhou*
Institute for Glycomics and School of Information and Communication Technology, Griffith University, Southport, Queensland, 4222, Australia
E-mail: yaoqi.zhou@griffith.edu.au
[†] *These authors contributed equally to this work.*
Contract Grant sponsor: Australia Research Council; Contract Grant number: DP180102060; Contract Grant sponsor: National Health and Medical Research Council; Contract Grant number: 1121629; Contract Grant sponsor: National Natural Science Foundation of China; Contract Grant numbers: 61772566, 81801132, U1611261; Contract Grant sponsor: Guangdong Introducing Innovative and Entrepreneurial Teams; Contract Grant number: 2016ZT06D211; Contract Grant sponsor: GD Frontier & Key Tech Innovation Program; Contract Grant number: 2019B020228001; Contract Grant sponsor: National Key R&D Program of China; Contract Grant number: 2018YFC0910500

performance of the method called SPOT-Fold is compared to CoinFold using 227 recent CASP targets and other methods using free modeling targets in CASP12 and CASP13.

Methodology

Test datasets

Three datasets were used to test our method of distance- and angle-restrained protein structure prediction. One dataset is based on 213 CASP targets (CASP-T213): 39 from CASP12,[6] 82 from CASP11,[28] and 92 from CASP10.[29] These targets were chosen from those with native structures released, between 67 and 670 amino acid residues after randomly removing some proteins of similar sizes to avoid biases toward a specific protein size. The whole protein sequence is employed even if it contains multiple domains. We also randomly selected 60 nonredundant small globular proteins with 80–100 amino acid residues from 1199 nonredundant test proteins collected previously (SGP60).[23] In addition, we obtained 14 free modeling targets (CASP13-FM) and 14 template-based modeling hard targets (CASP13-TBM-hard) from the most recent CASP13. These are proteins with no or marginal similarity to known structures.

Contact map prediction

Protein contact maps were predicted by SPOT-Contact.[18] The method employed an ensemble of hybrid ultra-deep ResNet and 2D-LSTM BRNNs to capture both short and long-distance interactions. A contact between two residues is defined by a distance cut off of 8 Å between two Cβ atoms of amino acid residues (Cα for Glycine).

SS, angles and torsion angles prediction

In SPOT-Fold protocol, SPIDER3[23,24] was utilized to predict SS, three backbone dihedral angles rotated about the N-Cα bond (ϕ), the Cα-C bond (ψ), and the $C\alpha_i$-$C\alpha_{i+1}$ bond (τ), and θ angle between three neighboring Cα atoms. For comparison, predicted SS was also converted to ϕ and ψ angle restraints. This was done by using the average ϕ and ψ angles in different secondary-structure states of proteins in SABmark database[30] as described by Adhikari and Cheng[19,21] before. The results with these torsion angles are labeled as SPOT-Fold(SS).

Tertiary structure construction

Here, we employed a CNS[25,31] version 1.3 modified by Adhikari and Cheng[19,21] for contact-restrained protein-structure prediction. As in CONFOLD[19,21] and CoinFold,[20] protein models were built in two stages according to predicted distance and torsion angle restraints. For distance restraints, we selected top contact pairs with the highest prediction confidence. A list of 40 sets of contact pairs was obtained for top 0.1L, 0.2L, 0.3L, … 4.0L predictions with L, the length of the target protein. All predicted contact pairs were converted to distance restraints between the two Cβ atoms (Cα for glycine) of the two amino residues. In addition to predicted contact pairs, predicted θ angles between three neighboring Cα atoms were also converted to distance restraints. On the other hand, all predicted torsion angles (ϕ, ψ, and τ) are employed as torsion angle restraints in CNS.[25,31] The distance restraining energy was calculated by the default soft-square energy function implemented in CNS.

In the first stage, 20 conformations were constructed for each of the above-mentioned 40 sets of predicted contact pairs by using the distance-geometry simulated annealing protocol in CNS. These conformations were then evaluated according to restraining energies in CNS. The top 5 of the 20 models were selected. The predicted contact pairs that were not in contact in the top 5 models were removed from the predicted contact-pair list. All 40 sets of contact pairs were updated at this stage. In the second stage, the distance-geometry simulated annealing was performed again but with the updated list of contact pairs and 20 models are built for each list. This led the final 800 structural models (20×40).

Model selection and evaluation

The final 800 structures with all heavy atoms were ranked by the dDFIRE energy function[32,33] for top 1 to top 5 models. dDFIRE is a statistical energy function based on the distance-scaled finite ideal-gas reference state that accounts for dipolar interactions. Following CASP, we evaluated our models according to the global distance test (GDT)[34] between the protein models and the experimental structures. GDT score ranges from 0 to 1, with 0 for no structural similarity and 1 for the perfect match.

Comparsion with CoinFold models

Results of CoinFold predicted contact maps and model structures were obtained from its online server at http://raptorx.uchicago.edu/ContactMap/.

Results

Effect of angle and distance-restraints on model accuracy

To illustrate the impact of different restraints on model accuracy, angles (ϕ and ψ) predicted from SS, angles (ϕ, ψ, θ, and τ) (AG), and contact maps are used separately and in combination to build protein models by the distance-geometry simulated annealing module in CNS. The results for the CASP-T213 set are shown in Table 1. It is clear that mostly random structures were generated when only angle restraints were employed. Introducing predicted contact maps yield substantial improvement (>250% increase) regardless if SS or angles were employed in conjunction with contact maps. Using SPIDER3 predicted angles provides a small (2%) but statistically significant improvement in median GDT values (the *p*-value of a two-tailed T-test of a significant level at 0.05 is 0.0051).

The small and visible improvement by using angle over that by using SS can be further illustrated by the accuracy of model structural fragments. Figure 1 shows the average of the best root-mean-squared-distance (RMSD) value between a TOP5 model and actual native structure for the CASP-T213 set as a function of the fragment size (4–12 residues). When coupled with the same contact map, AG-based restraints yielded consistently more accurate fragments than SS-based restraints

Table 1. The GDT-scores of sampled conformations in the CASP-T213 set by using restraints from predicted SS, angles (AG), contact and secondary (SPOT-Fold(SS)), and contact and angles (SPOT-Fold). The columns 2–4 are the maximum, median, and average GDT scores of the 800 models built for each target in CASP-T213. The Top 1 is the GDT score of the conformation with the lowest dDFIRE energy, and Top 5 is the maximum GDT score among 5 lowest dDFIRE energy models. These values are averaged over 213 targets.

	Max	Med	Ave	TOP1	TOP5
SS	0.1336	0.1199	0.1204	0.1219	0.1295
AG	0.1375	0.1218	0.1214	0.1266	0.1341
SPOT-Fold(SS)	0.4915	0.3184	0.3114	0.4356	0.4527
SPOT-Fold	0.4948	0.3236	0.3136	0.4364	0.4570

regardless of the specific size of a fragment. As a comparison, the result from CoinFold is also shown. Our method yields substantial improvement over CoinFold whether angles from SS or AG were used in restraints with contact maps.

It is of interest to know how much improvement is due to two separate stages. Stage 1 is construction of initial models based on raw restraints whereas stage 2 is based on updated restraints using the models generated from stage 1. As shown in Table 2, both the TOP1 and the best in TOP5 GDT scores of the CASP-T213 proteins are improved in the second stage, when predicted contact maps are combined with either with predicted SS or from predicted torsion angles. The larger Max, Med, and Ave values indicate the better overall sampling of the 800 structures in the second stage. The improvement is 5% in median GDT scores.

Model accuracy

Figure 2 shows the model accuracy for SGP60 as a function of Neff, which is a measure of the number of effective homologous sequences from Hblits.[35,36] A higher Neff indicates higher

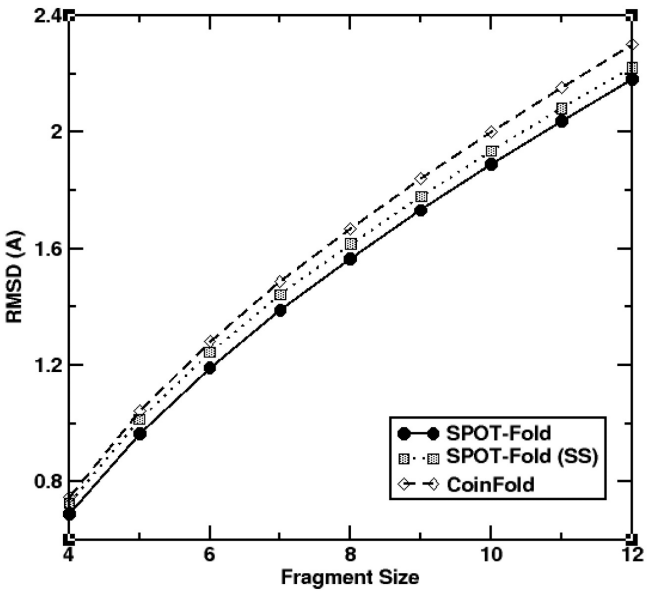

Figure 1. The average of the best root-mean-squared distance (RMSD) between a top-5 model fragment and its native structure for the CASP-T213 set as a function of the size of fragments.

Table 2. The average GDT scores of TOP1 and the best GDT scores in TOP5 models for the CASP-T213 set at two stages (1 and 2) of model constructions, given by combining predicted contact maps (SPOT-Fold) with predicted SSs or backbone angles.

	Stage	Max	Med	Ave	TOP1	TOP5
SPOT-Fold	1	0.4873	0.3073	0.3018	0.4308	0.4525
	2	0.4948	0.3236	0.3136	0.4364	0.4570
SPOT-Fold(SS)	1	0.4853	0.3031	0.3005	0.4262	0.4442
	2	0.4915	0.3184	0.3114	0.4356	0.4527

quality for evolution information and more accurate prediction for contact maps and backbone angles.[18,22,23] Results shown in Figure 3 indicate that higher Neff values lead to high-quality models.

Model quality assessment and accuracy comparison

The dDFIRE energy function[32,33] is used to evaluate the models and select the TOP1 and TOP5 predicted structures. As shown in Table 3, compared to using the restraint energy calculated from CNS, labeled as ResE, both TOP1 and TOP5 have improved when using dDFIRE energy function. For example, TOP1 GDT values of SPOT-Fold and SPOT-Fold(SS) have improved by 5%. TOP5 values of SPOT-Fold and SPOT-Fold(SS) have improved by 3%.

We submitted the proteins in the CASP-T213 set to CoinFold within the RaptorX sever.[20,37] The GDT scores of the TOP1 and the best in TOP5 structures for each protein are compared between CoinFold and our method in Figures 3A and 3B, respectively. Our method outperforms Coinfold on both the TOP1 and the best in TOP5 structures because there are substantially more proteins with higher GDT scores for our method. The average GDT scores of our TOP1 and the best in TOP 5 models are 0.436 and 0.457, respectively, which are 6–7% improvement over 0.407 and 0.433 given by CoinFold.

To further examine the accuracy of our method, we compared SPOT-Fold with CoinFold on the 28 targets from CASP13

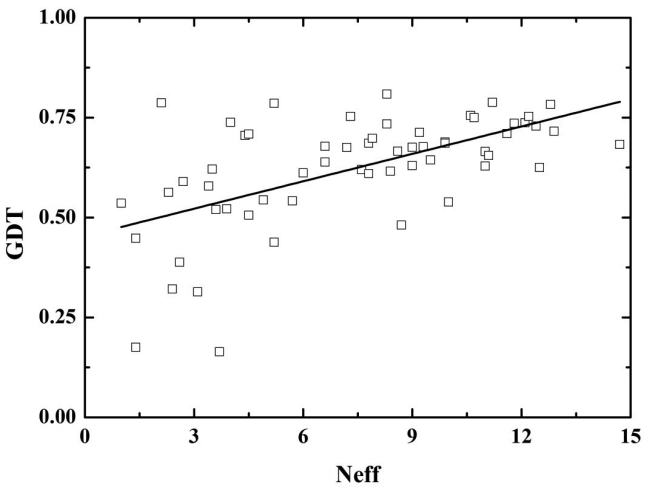

Figure 2. The TOP1 GDT scores of SGP60 sets as a function of the number of effective homologous sequences (Neff) for each protein.

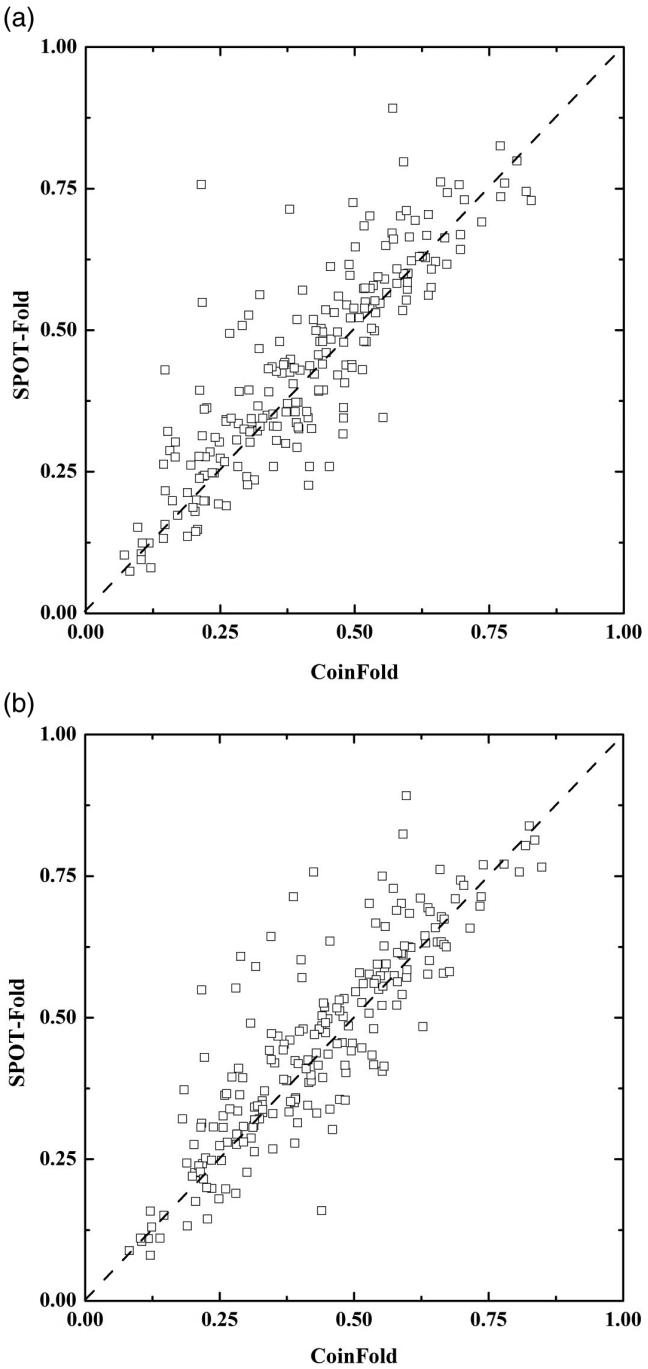

Figure 3. Comparison of (A) Top 1 and (B) Top 5 GDT scores for 213 proteins given by CoilFold and by our method (SPOT-Fold).

Table 3. : The GDT-scores of the CASP-T213 set from CoinFold, SPOT-Fold(SS) and SPOT-Fold. The TOP1 and TOP5 models are selected based on the lowest CNS restraining energy (ResE) or dDFIRE energy function in SPOT-Fold(SS) and SPOT-Fold.

		SPOT-Fold(SS)		SPOT-Fold	
	CoinFold	ResE	dDFIRE	ResE	dDFIRE
TOP1	0.4069	0.4144	0.4356	0.4155	0.4364
TOP5	0.4326	0.4474	0.4527	0.4457	0.4570

Table 4. TOP5 and TOP1 results predicted by SPOT-Fold and CoinFold methods, averaged over 28 targets from CASP13.

		Top1		Top5	
No.	Target ID	SPOT-Fold	CoinFold	SPOT-Fold	CoinFold
1	T0950	0.1564	0.2135	0.1820	0.2500
2	T0953s1	0.3576	0.3750	0.3646	0.3750
3	T0953s2	0.3105	0.2762	0.3105	0.2792
4	T0957s1	0.2341	0.2150	0.2771	0.2197
5	T0960	0.1257	0.1330	0.1257	0.1330
6	T0963	0.1168	0.1078	0.1195	0.1099
7	T0968s1	0.3276	0.5539	0.3491	0.5733
8	T0968s2	0.3596	0.4408	0.4232	0.4671
9	T0969	0.3870	0.1864	0.4216	0.2020
10	T0970	0.1366	0.1521	0.1495	0.1521
11	T0980s1	0.3087	0.3724	0.3087	0.3980
12	T0980s2	0.3468	0.3468	0.3468	0.3468
13	T1021s3	0.2336	0.2600	0.2827	0.2618
14	T1022s1	0.2444	0.2723	0.2701	0.2836
15	T0954	0.4525	0.3820	0.4525	0.3820
16	T0955	0.6524	0.5915	0.6585	0.6280
17	T0957s2	0.4045	0.5488	0.4602	0.5796
18	T0958	0.5982	0.5149	0.6250	0.5298
19	T0965	0.5667	0.5024	0.5952	0.5024
20	T0966	0.1852	0.1359	0.1893	0.1359
21	T0986s1	0.5730	0.5590	0.6292	0.5815
22	T0986s2	0.2583	0.4667	0.4617	0.4833
23	T1005	0.4224	0.2680	0.4303	0.2735
24	T1009	0.4359	0.2235	0.4648	0.2329
25	T1011	0.0867	0.3119	0.2438	0.3187
26	T01021s1	0.3607	0.3691	0.3809	0.3742
27	T01021s2	0.4499	0.2378	0.4936	0.2851
28	T01022s2	0.1835	0.1086	0.2272	0.1136
	Average	0.3413	0.3338	0.3680	0.3469

in Table 4. The average GDT scores of our TOP1 and the best in TOP 5 models are 2% and 6% higher than models from CoinFold. A further analysis (Fig. 4) indicates that SPOT-Fold improves over CoinFold significantly for CASP13-TMB-hard (No. 15–28 in Table 4) and comparable for CASP13-FM (No. 1–14 in Table 4).

We compared SPOT-Fold with other methods on the 17 FM targets from CASP13, and listed the average Top1 TMscores and GDT scores in Table 5. Our detailed results of every domains are presented in Table 6.

Performance improvement is observed for the SGP60 set. As shown in Figure 5, our method has 45 (75%) proteins with higher GDT scores for the TOP1 model than CoilFold. The average GDT score of TOP1 models from the CoinFold server is 0.5456, compared to 0.6246 by our method. This is a 14.5% improvement.

Illustrative examples

Figure 6 illustrates the model predicted (TOP1) for a target in comparison with its corresponding experimental structures. 2i5fA from SGP60 with 99 amino acids has a mixed helix and sheet conformation. The GDT score of the TOP1 models is 0.7882.

Conclusions

In this paper, we presented a new method SPOT-Fold for *ab initio* protein structure prediction. The model prediction is

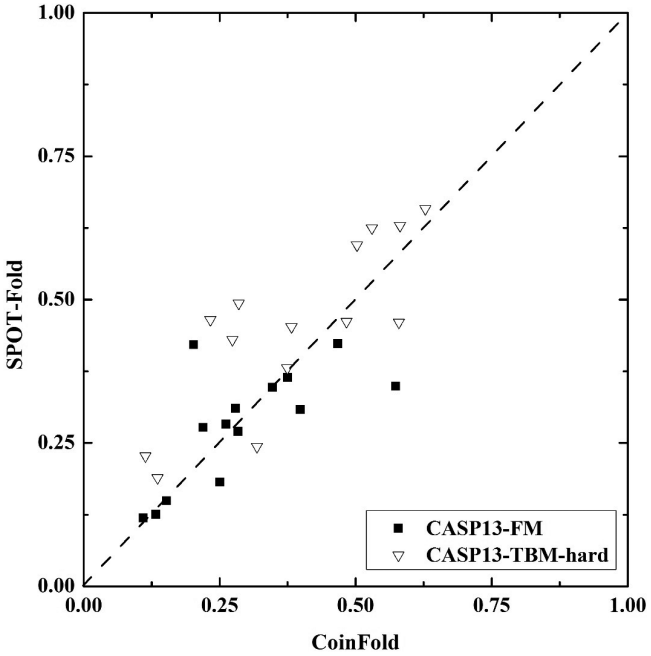

Figure 4. GDT scores of 28 targets from CASP13. TOP5 models from SPOT-Fold against TOP5 models from CoinFold.

based on the distance-geometry simulated annealing module in CNS and the two-stage protocol of CONFOLD.[19,21] The distance restraints were obtained from the contact maps predicted by SPOT-Contact and angle restraints from dihedral angles predicted by SPIDER3. The method performance was evaluated by using CASP-T213, CASP13-FM, CASP13-TBM-hard and SGP60 sets according to the commonly used GDT scores. SPOT-Fold models outperformed models from CoinFold online server of the similar approach for free modeling. The method is computationally efficient. For the smallest 67 and largest 670 amino-acid proteins in our dataset, CPU times are 38 and 248 h, respectively, to finish all the modeling by using a single core of an Intel Xeon 2.2GHz processor.

Table 5. Average TMscores and GDT scores of TOP1 results modeled by SPOT-Fold are compared to other methods, for 17 FM domains defined by CASP13.

Group	Rank	Targets number	Averge TMscore	Averge GDT score
RaptorX-Contact	1	17	0.5057059	0.46511765
RaptorX-DeepModeller	2	17	0.5024706	0.45552941
QUARK	3	17	0.5014706	0.46341176
Zhang-Server	4	17	0.4960588	0.45658824
RaptorX-TBM	5	17	0.4275882	0.39264706
BAKER-ROSETTASERVER	6	17	0.4022353	0.37288235
SPOT-Fold	7	17	0.3786471	0.287
Zhou-SPOT-3D	8	17	0.352	0.32664706
Zhang-CEthreader	9	17	0.3473529	0.31341176
AWSEM-Suite	10	17	0.3443529	0.31752941
PconsC4	11	17	0.34	0.30973333
MULTICOM_CLUSTER	12	17	0.3266471	0.30441176
Yang-Server	13	16[a]	0.343125	0.3171875
MULTICOM-CONSTRUCT	14	17	0.3225882	0.29876471
IntFOLD5	15	17	0.2730588	0.25329412
FALCON	16	17	0.2659412	0.25664706
Seok-server	17	17	0.2628824	0.23988235

[a] Yang-Server has a 0 score on T0963-D2 target.

Table 6. TMscores and GDT scores of TOP1 models from SPOT-Fold for 17 FM domains defined by CASP13.

Target	Top1 TMscore	Top1 GDT score
T0950-D1	0.261	0.124
T0953s1-D1	0.328	0.313
T0953s2-D1	0.325	0.386
T0953s2-D2	0.637	0.536
T0953s2-D3	0.29	0.223
T0957s1-D1	0.363	0.29
T0957s2-D1	0.516	0.364
T0960-D2	0.196	0.149
T0963-D2	0.209	0.171
T0968s1-D1	0.354	0.267
T0968s2-D1	0.407	0.333
T0969-D1	0.631	0.357
T0980s1-D1	0.301	0.229
T0986s2-D1	0.337	0.215
T1021s3-D1	0.527	0.358
T1021s3-D2	0.302	0.247
T1022s1-D1	0.453	0.317

Acknowledgments

The work was supported in part by the National Key R&D Program of China (2018YFC0910500), GD Frontier & Key Tech Innovation Program (2018B010109006, 2019B020228001), the National Natural Science Foundation of China (61772566, U1611261, and 81801132), the program for Guangdong Introducing Innovative and Entrepreneurial Teams (2016ZT06D211), and Australia

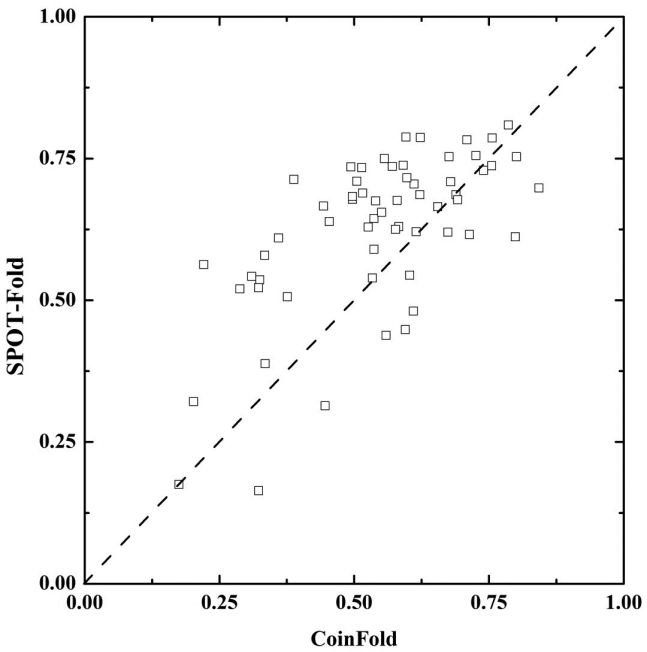

Figure 5. GDT score of SGP60 proteins. TOP1 models from SPOT-Fold against TOP1 models from CoinFold.

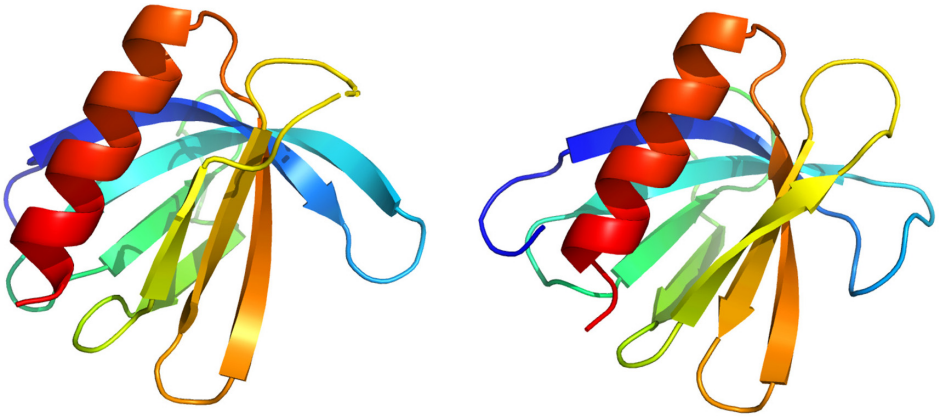

Figure 6. The crystal structure (left) and the TOP1 models from SPOT-Fold (right) of 2i5fA with GDT = 0.7882. [Color figure can be viewed at wileyonlinelibrary.com]

Research Council (DP180102060) to Y.Z. and K.P. and in part by the National Health and Medical Research Council (1121629) of Australia to Y.Z. The authors have declared that no conflict of interest exists.

Keywords: protein structure prediction · template-free modeling · molecular dynamics · contact map · deep learning

How to cite this article: Y. Cai, X. Li, Z. Sun, Y. Lu, H. Zhao, J. Hanson, K. Paliwal, T. Litfin, Y. Zhou, Y. Yang. *J. Comput. Chem* **2020**, *41*, 745–750. DOI: 10.1002/jcc.26132

[1] R. Jaenicke, *Prog. Biophys. Mol. Biol.* **1987**, *49*, 117.
[2] A. Bax, S. Grzesiek, NMR of Proteins, Springer, Berlin, **1993**, p. 33.
[3] X.-C. Bai, G. McMullan, S. H. Scheres, *Trends Biochem. Sci.* **2015**, *40*, 49.
[4] J. Moult, *Curr. Opin. Struct. Biol.* **2005**, *15*, 285.
[5] L. Hovan, V. Oleinikovas, H. Yalinca, A. Kryshtafovych, G. Saladino, F. L. Gervasio, *Proteins* **2018**, *86*, 152.
[6] A. Kryshtafovych, B. Monastyrskyy, K. Fidelis, T. Schwede, A. Tramontano, *Proteins* **2018**, *86*, 345.
[7] V. Mariani, F. Kiefer, T. Schmidt, J. Haas, T. Schwede, *Proteins* **2011**, *79*, 37.
[8] Y. J. Huang, B. Mao, J. M. Aramini, G. T. Montelione, *Proteins* **2014**, *82*, 43.
[9] Y. Zhang, *Curr. Opin. Struct. Biol.* **2009**, *19*, 145.
[10] Y. Zhou, Y. Duan, Y. Yang, E. Faraggi, H. Lei, *Theor. Chem. Accounts* **2011**, *128*, 3.
[11] Y. Yang, E. Faraggi, H. Zhao, Y. Zhou, *Bioinformatics* **2011**, *27*, 2076.
[12] K. T. Simons, C. Kooperberg, E. Huang, D. Baker, *J. Mol. Biol.* **1997**, *268*, 209.
[13] C. L. Brooks, *Acc. Chem. Res.* **2002**, *35*, 447.
[14] M. M. Seibert, A. Patriksson, B. Hess, D. Van Der Spoel, *J. Mol. Biol.* **2005**, *354*, 173.
[15] Y. Duan, P. A. Kollman, *Science* **1998**, *282*, 740.
[16] J. Xu, *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 16856.
[17] S. Wang, S. Sun, Z. Li, R. Zhang, J. Xu, *PLoS Comput. Biol.* **2017**, *13*, 1.
[18] J. Hanson, K. Paliwal, T. Litfin, Y. Yang, Y. Zhou, *Bioinformatics* **2018**, *34*, 4039.
[19] B. Adhikari, D. Bhattacharya, R. Cao, J. Cheng, *Proteins* **2015**, *83*, 1436.
[20] S. Wang, W. Li, R. Zhang, S. Liu, J. Xu, *Nucleic Acids Res.* **2016**, *44*, W361.
[21] B. Adhikari, J. Cheng, *BMC Bioinform.* **2018**, *19*, 22.
[22] R. Heffernan, A. Dehzangi, J. Lyons, K. Paliwal, A. Sharma, J. Wang, A. Sattar, Y. Zhou, Y. Yang, *Bioinformatics* **2015**, *32*, 843.
[23] R. Heffernan, K. Paliwal, J. Lyons, A. Dehzangi, A. Sharma, J. Wang, A. Sattar, Y. Yang, Y. Zhou, *Sci. Rep.* **2015**, *5*, 11476.
[24] J. Lyons, I. A. Dehzangi, R. Heffernan, A. Sharma, K. Paliwal, A. Sattar, Y. Zhou, Y. Yang, *J. Comput. Chem.* **2014**, *35*, 2040.
[25] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, *Acta Crystallogr. Sect. D* **1998**, *54*, 905.
[26] A. T. Brunger, *Nat. Protoc.* **2007**, *2*, 2728.
[27] Y. Yang, Y. Zhou, *J. Comput. Chem.* **2016**, *37*, 976.
[28] L. N. Kinch, W. Li, R. D. Schaeffer, R. L. Dunbrack, B. Monastyrskyy, A. Kryshtafovych, N. V. Grishin, *Proteins* **2016**, *84*, 20.
[29] A. Kryshtafovych, K. Fidelis, J. Moult, *Proteins* **2014**, *82*, 164.
[30] I. Van Walle, I. Lasters, L. Wyns, *Bioinformatics* **2004**, *21*, 1267.
[31] M. Nilges, A. M. Gronenborn, A. T. Brünger, G. M. Clore, *Protein Eng. Des. Select.* **1988**, *2*, 27.
[32] Y. Yang, Y. Zhou, *Proteins* **2008**, *72*, 793.
[33] Y. Yang, Y. Zhou, *Protein Sci.* **2008**, *17*, 1212.
[34] A. Zemla, Č. Venclovas, J. Moult, K. Fidelis, *Proteins* **1999**, *37*, 22.
[35] J. A. Casbon, M. A. Saqi, *BMC Bioinform.* **2004**, *5*, 200.
[36] R. I. Sadreyev, N. V. Grishin, *Bioinformatics* **2004**, *20*, 818.
[37] S. Wang, S. Fei, Z. Wang, Y. Li, J. Xu, F. Zhao, X. Gao, *Bioinformatics* **2018**, *35*, 691.

Received: 22 March 2019
Revised: 7 October 2019
Accepted: 1 December 2019
Published online on 17 December 2019