THE PROTEIN SOCIETY

OPUS-SSF: A side-chain-inclusive scoring function for ranking protein structural models

Gang Xu,[1] Tianqi Ma,[2,3] Qinghua Wang,[4] and Jianpeng Ma[1,2,3,4]*

[1]School of Life Sciences, Tsinghua University, Beijing 100084, People's Republic of China
[2]Applied Physics Program, Rice University, Houston, Texas 77005
[3]Department of Bioengineering, Rice University, Houston, Texas 77005
[4]Verna and Marrs Mclean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030

Received 26 December 2018; Accepted 27 March 2019
DOI: 10.1002/pro.3608
Published online 11 April 2019 proteinscience.org

Abstract: We introduce a side-chain-inclusive scoring function, named OPUS-SSF, for ranking protein structural models. The method builds a scoring function based on the native distributions of the coordinate components of certain anchoring points in a local molecular system for peptide segments of 5, 7, 9, and 11 residues in length. Differing from our previous OPUS-CSF [Xu et al., Protein Sci. 2018; 27: 286–292], which exclusively uses main chain information, OPUS-SSF employs anchoring points on side chains so that the effect of side chains is taken into account. The performance of OPUS-SSF was tested on 15 decoy sets containing totally 603 proteins, and 571 of them had their native structures recognized from their decoys. Similar to OPUS-CSF, OPUS-SSF does not employ the Boltzmann formula in constructing scoring functions. The results indicate that OPUS-SSF has achieved a significant improvement on decoy recognition and it should be a very useful tool for protein structural prediction and modeling.

Keywords: protein structure modeling; protein folding; coarse-graining; scoring function; decoy recognition

Introduction

Designing a good empirical function that can evaluate protein structural models is very important in protein structure prediction. Commonly, empirical potential functions can be divided into two categories: physics-based potentials[1–9] and knowledge-based potentials.[10–42] Both kinds of potentials can be in all-atom form, coarse-grained form, or mixed all-atom and coarse-grained form. For protein structural modeling, physics-based potentials are often outperformed by knowledge-based potentials.

In this article, we followed the basic idea of our recent work OPUS-CSF,[43] that is, building the scoring function based on the distribution of coordinate components of certain anchoring points extracted from short peptide segments in the native structure database, rather than on the traditional Boltzmann formula. An anchoring point is an actual atom position or position computed based on a few atoms on side chain. Similar to OPUS-CSF, we generated configurational native distribution (CND) lookup tables of small peptide segments of 5, 7, 9, and 11 residues in length by scanning through the entire Protein Data Bank (PDB). Unlike OPUS-CSF that records the distribution of main-chain C atom coordinates, OPUS-SSF

Gang Xu and Tianqi Ma contributed equally to this work.

Grant sponsor: Gillson Longenbaugh Foundation; Grant sponsor: Welch Foundation Q-1512 Q-1826.

*Correspondence to: Jianpeng Ma, Baylor College of Medicine, One Baylor Plaza, BCM-125, Houston, TX 77030. E-mail: jpma@bcm.edu

combines the side-chain and main-chain information to improve the performance of the scoring function. In this case, the local molecular coordinate system is built on the main-chain of the central residue (same as OPUS-CSF), and the recorded anchoring points are based on the side-chains of specific residues. Therefore, the main-chain and side-chain information are both taken into consideration in OPUS-SSF.

The performance of OPUS-SSF was tested on 15 decoy sets and the results showed that OPUS-SSF significantly outperforms OPUS-CSF both in terms of native structure recognition and *Z*-scores. OPUS-SSF recognized 571 out of 603 native structures from their decoys, while OPUS-CSF recognized 491. The average *Z*-score of OPUS-SSF was −5.46, while that of OPUS-CSF was −3.32. In terms of Pearson's correlation coefficients, OPUS-SSF and OPUS-CSF have similar values comparing with a generalized orientation-dependent, all-atom statistical potential (GOAP).[40] Note that GOAP potential needs all-atom coordinates, while OPUS-SSF and OPUS-CSF are highly coarse-grained.

Both OPUS-CSF and OPUS-SSF do not employ the Boltzmann formula in constructing scoring functions, which is a very different feature from most methods in literature. The performance of OPUS-SSF is significantly better than other methods, which indicates the effectiveness of the non-Boltzmann approach. Furthermore, the local molecular coordinate system in peptide segments makes it easier and quicker to describe relative configuration of the peptide than Boltzmann approach. We believe that OPUS-SSF would be a very helpful tool in modeling protein structures.

Results

Beside the 11 commonly used decoy sets used in GOAP[40] and OPUS-CSF,[43] including decoy sets of 4state_reduced,[44] fisa,[45] fisa_casp3,[45] hg_structal, ig_structal and ig_structal_hires (R. Samudrala, E. Huang, and M. Levitt, unpublished), I-TASSER,[39] lattice_ssfit,[46,47] lmds,[48] MOULDER,[49] and ROSETTA,[50] we also tested OPUS-SSF on casp-good set[39] (that contains 143 protein targets generated during CASP5-CASP8) and I-TASSER9, I-TASSER10, I-TASSER11 sets (decoy sets generated by I-TASSER server from CASP9,[51] CASP10,[52] CASP11,[53] downloaded from the website of Dr. Yang Zhang's group).

The performance of OPUS-SSF tested on 15 decoy sets are shown in Table I. OPUS-SSF recognized 571 native structures of totally 603 proteins in all decoy sets and the *Z*-score was −5.46. OPUS-CSF[43] recognized 491 native structures on the same 15 decoy sets with *Z*-score of −3.32. Also, like in OPUS-CSF,[43] a cutoff value 15 was added in OPUS-SSF on the sum of *Z*-scores of an anchoring point. For OPUS-SSF without the cutoff (see Methods), it recognized 546 out of 603 native structures from the decoys with an average *Z*-score of −3.89. Thus, the performance of OPUS-SSF without cutoff was worse than that of OPUS-CSF in native structure recognition, which supports the inclusion of the cutoff in SSF score calculation.

Table I. *The Performance of OPUS-CSF and OPUS-SSF on 15 Decoys Sets*

	# of Proteins	OPUS-CSF	OPUS-SSF
4state_reduced	7	7 (−3.31)	7 (−5.00)
fisa	4	2 (−2.55)	2 (−3.84)
fisa_casp3	5	4 (−6.72)	4 (−13.60)
hg_structal	29	23 (−2.06)	23 (−2.85)
ig_structal	61	56 (−2.14)	57 (−3.14)
ig_structal_hires	20	20 (−2.08)	20 (−2.84)
I-TASSER	56	56 (−6.39)	56 (−11.24)
lattice_ssfit	8	8 (−11.79)	8 (−18.92)
lmds	10	8 (−6.80)	9 (−9.36)
MOULDER	20	20 (−3.16)	20 (−6.16)
ROSETTA	58	53 (−4.53)	54 (−5.69)
casp_good	143	129 (−1.72)	135 (−2.27)
CASP9	85	46 (−2.80)	84 (−5.91)
CASP10	43	29 (−4.73)	43 (−8.11)
CASP11	54	30 (−3.13)	49 (−6.33)
Total	603	491 (−3.32)	571 (−5.46)

The numbers of protein targets in the decoy set, with their native structures successfully recognized by OPUS-CSF and OPUS-SSF are listed in the table. The numbers in parentheses are the average *Z*-scores of the native structures. The bigger the absolute value of *Z*-score, the better. Out of totally 603 protein targets in 15 decoy sets, OPUS-SSF recognized 571 native structures with an average *Z*-score of −5.46, both values are better than that of OPUS-CSF.

The average RMSD values of recognized structures are shown in Table II. The result of using backbone information alone (CSF case) is worse than that of using side-chain information and backbone information together (SSF case), the values are 1.661 and 0.334, respectively. This result demonstrates that OPUS-SSF can recognize the native structure from decoys that are very close to the native structure.

The Pearson's correlation coefficients between SSF score and TM-score[54] in all decoy sets were also calculated. The results are shown in Table III. The result of OPUS-SSF (0.52) was worse than that of OPUS-CSF (0.56). Side-chain configuration has larger variation than main-chain configuration that increases the uncertainty in modeling. A relatively good result of OPUS-SSF indicates that it captured the key point of the side-chain conformation. The correlation coefficients of OPUS-SSF were better than that of OPUS-SSF with no cutoff (data not shown).

Discussion

In the Results section, we briefly discussed the construction of CND lookup tables. It is worth noting that the number of segments was smaller in OPUS-SSF than that in OPUS-CSF[43] although OPUS-SSF was developed later than OPUS-CSF. Besides, the ratio between the number of 5-residue segments that appear more than five times to the total number of 5-residue segments is also smaller in OPUS-SSF. A reasonable guess is that lots of structures in PDB do not have a complete side-chain, therefore a lot of structures were excluded.

Table II. *Average RMSD Values of OPUS-CSF and OPUS-SSF on 15 Decoys Sets*

	OPUS-CSF	OPUS-SSF
4state_reduced	0	0
fisa	2.184	1.700
fisa_casp3	1.223	1.112
hg_structal	0.305	0.301
ig_structal	0.146	0.127
ig_structal_hires	0	0
I-TASSER	0	0
lattice_ssfit	0	0
lmds	0.756	0.379
MOULDER	0	0
ROSETTA	0.604	0.496
casp_good	0.973	0.264
CASP9	3.911	0.167
CASP10	2.038	0
CASP11	3.387	0.941
Average	1.661	0.334

The numbers for each decoy set are the average RMSD values of recognized structures in that decoy set by two scoring functions. The numbers in the last row are the weighted average numbers of all decoy sets. The value "0" indicates that all native structures were successfully found in that decoy set. OPUS-SSF outperformed OPUS-CSF in every decoy set.

However, the ratio between the number of 7-, 9-, and 11-residue segments that appear more than five times to the total number of 7-, 9-, and 11-residue segments increased. This means that the incomplete structures may also have uncommon sequences, excluding them in training was the right approach.

The performance of OPUS-SSF without cutoff value was also examined. It was 546 out of 603, while OPUS-SSF with cutoff was 571 out of 603 (Table I). So, the performance without cutoff was worse than that with cutoff. The final version of OPUS-SSF used 15 as the cutoff value in calculating the SSF score.

Table III. *Average Pearson's Correlation Coefficients of OPUS-CSF and OPUS-SSF Scores with TM-Scores*

	OPUS-CSF	OPUS-SSF
4state_reduced	−0.67	−0.74
fisa	−0.55	−0.63
fisa_casp3	−0.33	−0.36
hg_structal	−0.80	−0.80
ig_structal	−0.88	−0.87
ig_structal_hires	−0.90	−0.88
I-TASSER	−0.45	−0.48
lattice_ssfit	−0.15	−0.08
lmds	−0.34	−0.32
MOULDER	−0.86	−0.86
ROSETTA	−0.39	−0.38
casp_good	−0.65	−0.60
CASP9	−0.31	−0.50
CASP10	−0.25	−0.46
CASP11	−0.14	−0.29
Total	−0.52	−0.56

The correlation coefficient of a decoy set is the average coefficient of all targets in that decoy set. The native structures were excluded from the calculation. OPUS-SSF has better result than OPUS-CSF.

In constructing the CND lookup table, only the sequence and coordinates information were included, no other information such as secondary structural elements were used.

OPUS-SSF is a fast and accurate modeling method. It is highly course-grained and does not require interatomic information. In early stage of protein modeling, a fast and accurate scoring function is very important. OPUS-SSF seems to be promising in this regard.

Methods

The procedure of OPUS-SSF is similar to that of OPUS-CSF.[43] For the collections of small peptide segments with specific sequences (with length of 5, 7, 9, and 11 residues), CNDs were constructed. The distributions were constructed through analyzing all structures in the PDB, except the ones Do not match our selection criteria. The sequences that appeared less than five times in PDB were discarded. We analyzed 130,054 PDB structures up to June 7, 2017 via ftp://ftp.wwpdb.org/pub/pdb/data/structures/divided/pdb. Finally, the information extracted from CNDs was stored in CND lookup tables.

The details of the procedure of OPUS-SSF are very similar to those in OPUS-CSF.[43] In OPUS-SSF, we included the side-chain conformation, not just the main-chain atoms only as in OPUS-CSF. Based on rigid body representation of side-chain chemical structures in OPUS-DOSP,[55] the representations of side-chain configurations were further simplified to 1–3 anchoring points on side-chains, the details of which are shown in Table IV.

For each segment, a local molecular coordinate system is constructed based on the central residue of the segment via the coordinates of main-chain C atom, Ca atom, and main-chain O atom. The definition of the coordinate system is identical to that in OPUS-CSF.[43] The Ca atom is set as the origin, the line connecting Ca and C atoms is defined as *X*-axis, the parallel component of C-O vector that is perpendicular to the *X*-axis in the Ca-C-O plane is defined as *Y*-axis, and the *Z*-axis is defined by the rule of a right-handed coordinate system. More details can be found in OPUS-CSF paper.[43]

The segments of different length are marked as 5(1, 3, 5), 7(2, 4, 6), 9(1, 3, 5, 7, 9), and 11(2, 4, 6, 8, 10). In the form of 5(1, 3, 5), for example, the first number 5 is the segment length, 1, 5 in the parenthesis are the residue indices for which we record coordinate component distributions of the anchoring points in local coordinate system, 3 is the residue on which the local coordinate system is constructed.

For a 5-residue segment with a specific sequence, for example, we recorded the coordinate components of the anchoring points on the side chains of the first and fifth residues in the local coordinate system. These coordinate components were treated as independent variables. By scanning through the entire PDB, we generated distributions of these independent variables using the recorded

Table IV. *Definition of Anchoring Points on 20 Amino Acid Side Chains*

	Anchoring point 1	Anchoring point 2	Anchoring point 3
GLY	C	–	–
ALA	CB	–	–
SER	(CB + OG)/2	–	–
CYS	(CB + SG)/2	–	–
VAL	(CB + CG1 + CG2)/3	–	–
ILE	(CB + CG1 + CG2)/3	CD	–
LEU	CB	(CG + CD1 + CD2)/3	–
THR	(CB + OG1 + CG2)/3	–	–
ARG	CB	(CG + CD)/2	(NE + CZ + NH1 + NH2)/4
LYS	CB	(CG + CD)/2	(CE + NZ)/2
ASP	CB	(CG + OD1 + OD2)/3	–
GLU	(CB + CG)/2	(CD + OE1 + OE2)/3	–
ASN	CB	(CG + OD1 + ND2)/3	–
GLN	(CB + CG)/2	(CD + OE1 + NE2)/3	–
MET	(CB + CG)/2	(SD + CE)/2	–
HIS	CB	(CG + CE1 + NE2)/3	–
PRO	(CB + CG + CD)/3	–	–
PHE	CB	(CG + CE1 + CE2)/3	–
TYR	CB	(CG + CE1 + CE2)/3	OH
TRP	CB	(NE1 + CZ2 + CZ3)/3	–

coordinates, called CNDs of 5-residue segments. We then calculated the means and standard deviations of the distributions and they were kept in the CND lookup table.

For a test structure, we first split the structure by all possible segments (5, 7, 9, and 11 residues in length). Then for every segment existing in CND lookup table, we calculated the absolute values of Z-score of each independent coordinate variable based on distributions found in the CND lookup table. The final step was adding up all the absolute values of Z-score of all independent variables for all segments and the total sum was called SSF score. In the process, if the sum of absolute values of Z-scores of three coordinate variables of one anchoring point is greater than 15, we set it to 15, that is, a cutoff value of 15. This value comes from the assumption that all these coordinate variables are Gaussian. If the absolute value of Z-score is greater than 5, then this data point is extremely rare, and we want to ignore its influence. We have three coordinate variables and the cutoff value is therefore 15. If a residue has more than one anchoring point, the total sum of all anchoring points on that residue will be divided by the number of anchoring points for normalization purpose. The polypeptide structure with smallest SSF score was assumed to be the closest one to the native structure. No weighting function was included for different segment length in calculating SSF scores.

Accessibility of OPUS-SSF

The software is publicly accessible from ma-lab.rice.edu, or by contacting Jianpeng Ma at jpma@bcm.edu.

Acknowledgments

J.M. acknowledges the support from the Welch Foundation (Q-1512). Q.W. wishes to thank the support from the Welch Foundation (Q-1826), and Gillson-Longenbaugh Foundation.

References

1. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102: 3586–3616.
2. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4:187–217.
3. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) An all atom force field for simulations of proteins and nucleic acids. J Comput Chem 7:230–252.
4. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688.
5. Arnautova YA, Jagielska A, Scheraga HA (2006) A new force field (ECEPP-05) for peptides, proteins, and organic molecules. J Phys Chem B 110:5025–5044.
6. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. J Phys Chem B 111: 7812–7824.
7. Liwo A, Ołdziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA (1997) A united-residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. J Comput Chem 18:849–873.
8. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Ołdziej S, Scheraga HA (1997) A united-residue force field for off-lattice protein-structure simulations. II. Parameterization of short-range interactions and determination of weights of energy terms by Z-score optimization. J Comput Chem 18: 874–887.
9. Chebaro Y, Pasquali S, Derreumaux P (2012) The coarse-grained OPEP force field for non-amyloid and amyloid proteins. J Phys Chem B 116:8741–8752.

10. Skolnick J (2006) In quest of an empirical potential for protein structure prediction. Curr Opin Struct Biol 16: 166–171.
11. Sippl MJ (1995) Knowledge-based potentials for proteins. Curr Opin Struct Biol 5:229–235.
12. Jernigan RL, Bahar I (1996) Structure-derived potentials and protein simulations. Curr Opin Struct Biol 6: 195–209.
13. Moult J (1997) Comparison of database potentials and molecular mechanics force fields. Curr Opin Struct Biol 7:194–199.
14. Lazaridis T, Karplus M (2000) Effective energy functions for protein structure prediction. Curr Opin Struct Biol 10:139–145.
15. Gohlke H, Klebe G (2001) Statistical potentials and scoring functions applied to protein–ligand binding. Curr Opin Struct Biol 11:231–235.
16. Russ WP, Ranganathan R (2002) Knowledge-based potential functions in protein design. Curr Opin Struct Biol 12:447–452.
17. Buchete N, Straub J, Thirumalai D (2004) Development of novel statistical potentials for protein fold recognition. Curr Opin Struct Biol 14:225–232.
18. Poole AM, Ranganathan R (2006) Knowledge-based potentials in protein design. Curr Opin Struct Biol 16: 508–513.
19. Zhou Y, Zhou H, Zhang C, Liu S (2006) What is a desirable statistical energy functions for proteins and how can it be obtained? Cell Biochem Biophys 46:165–174.
20. Ma J (2009) Explicit orientation dependence in empirical potentials and its significance to side-chain modeling. Acc Chem Res 42:1087–1096.
21. Gilis D, Biot C, Buisine E, Dehouck Y, Rooman M (2006) Development of novel statistical potentials describing cation-π interactions in proteins and comparison with semiempirical and quantum chemistry approaches. J Chem Inf Model 46:884–893.
22. Hendlich M, Lackner P, Weitckus S, Floeckner H, Froschauer R, Gottsbacher K, Casari G, Sippl MJ (1990) Identification of native protein folds amongst a large number of incorrect models: the calculation of low energy conformations from potentials of mean force. J Mol Biol 216:167–180.
23. Hoppe C, Schomburg D (2005) Prediction of protein thermostability with a direction-and distance-dependent knowledge-based potential. Protein Sci 14:2682–2692.
24. Jones DT, Taylort W, Thornton JM (1992) A new approach to protein fold recognition. Nature 358:86–89.
25. Kolíński A, Bujnicki JM (2005) Generalized protein structure prediction based on combination of fold-recognition with de novo folding and evaluation of models. Proteins 61:84–90.
26. Miyazawa S, Jernigan RL (1985) Estimation of effective interresidue contact energies from protein crystal structures - quasi-chemical approximation. Macromolecules 18:534–552.
27. Sippl MJ (1990) Calculation of conformational ensembles from potentials of mena force: an approach to the knowledge-based prediction of local structures in globular proteins. J Mol Biol 213:859–883.
28. Skolnick J, Kolinski A, Ortiz A (2000) Derivation of protein-specific pair potentials based on weak sequence fragment similarity. Proteins 38:3–16.
29. Tobi D, Elber R (2000) Distance-dependent, pair potential for protein folding: results from linear optimization. Proteins 41:40–46.
30. Wu Y, Lu M, Chen M, Li J, Ma J (2007) OPUS-Ca: a knowledge-based potential function requiring only Cα positions. Protein Sci 16:1449–1463.
31. Zhang Y, Kolinski A, Skolnick J (2003) TOUCHSTONE II: a new approach to ab initio protein structure prediction. Biophys J 85:1145–1164.
32. DeBolt SE, Skolnick J (1996) Evaluation of atomic level mean force potentials via inverse folding and inverse refinement of protein structures: atomic burial position and pairwise non-bonded interactions. Protein Eng 9: 637–655.
33. Lu H, Skolnick J (2001) A distance-dependent atomic knowledge-based potential for improved protein structure selection. Proteins 44:223–232.
34. Lu M, Dousis AD, Ma J (2008) OPUS-PSP: an orientation-dependent statistical all-atom potential derived from side-chain packing. J Mol Biol 376:288–301.
35. Samudrala R, Moult J (1998) An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. J Mol Biol 275: 895–916.
36. Shen M, Sali A (2006) Statistical potential for assessment and prediction of protein structures. Protein Sci 15:2507–2524.
37. Yang Y, Zhou Y (2008) Specific interactions for ab initio folding of protein terminal regions with secondary structures. Proteins 72:793–803.
38. Zhang C, Vasmatzis G, Cornette JL, DeLisi C (1997) Determination of atomic desolvation energies from the structures of crystallized proteins. J Mol Biol 267:707–726.
39. Zhang J, Zhang Y (2010) A novel side-chain orientation dependent potential derived from random-walk reference state for protein fold selection and structure prediction. PLoS One 5:e15386.
40. Zhou H, Skolnick J (2011) GOAP: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction. Biophys J 101:2043–2052.
41. Zhou H, Zhou Y (2002) Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. Protein Sci 11:2714–2726.
42. Singh J, Thornton JM. Atlas of protein side-chain interactions. Oxford, UK: IRL Press at Oxford University Press, 1992.
43. Xu G, Ma T, Zang T, Wang Q, Ma J (2018) OPUS-CSF: a C-atom-based scoring function for ranking protein structural models. Protein Sci 27:286–292.
44. Park B, Levitt M (1996) Energy functions that discriminate X-ray and near-native folds from well-constructed decoys. J Mol Biol 258:367–392.
45. Simons KT, Kooperberg C, Huang E, Baker D (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 268:209–225.
46. Samudrala R, Xia Y, Levitt M, Huang E (1999) A combined approach for ab initio construction of low resolution protein tertiary structures from sequence. Pac Symp Biocomput 4:505–516.
47. Xia Y, Huang ES, Levitt M, Samudrala R (2000) Ab initio construction of protein tertiary structures using a hierarchical approach. J Mol Biol 300:171–185.
48. Keasar C, Levitt M (2003) A novel approach to decoy set generation: designing a physical energy function having local minima with native structure characteristics. J Mol Biol 329:159–174.
49. John B, Sali A (2003) Comparative protein structure modeling by iterative alignment, model building and model assessment. Nucleic Acids Res 31:3982–3992.
50. Tsai J, Bonneau R, Morozov AV, Kuhlman B, Rohl CA, Baker D (2003) An improved protein decoy set for testing energy functions for protein structure prediction. Proteins 53:76–87.

51. Xu D, Zhang J, Roy A, Zhang Y (2011) Automated protein structure modeling in CASP9 by I-TASSER pipeline combined with QUARK-based ab initio folding and FG-MD-based structure refinement. Proteins 79:147–160.
52. Zhang Y (2014) Interplay of I-TASSER and QUARK for template-based and ab initio protein structure prediction in CASP10. Proteins 82:175–187.
53. Zhang W, Yang J, He B, Walker SE, Zhang H, Govindarajoo B, Virtanen J, Xue Z, Shen HB, Zhang Y (2016) Integration of QUARK and I-TASSER for ab initio protein structure prediction in CASP11. Proteins 84:76–86.
54. Zhang Y, Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. Proteins 57:702–710.
55. Xu G, Ma T, Zang T, Sun W, Wang Q, Ma J (2017) OPUS-DOSP: a distance-and orientation-dependent all-atom potential derived from side-chain packing. J Mol Biol 429:3113–3120.