Chapter 6

SPIDER2: A Package to Predict Secondary Structure, Accessible Surface Area, and Main-Chain Torsional Angles by Deep Neural Networks

Yuedong Yang, Rhys Heffernan, Kuldip Paliwal, James Lyons, Abdollah Dehzangi, Alok Sharma, Jihua Wang, Abdul Sattar, and Yaoqi Zhou

Abstract

Predicting one-dimensional structure properties has played an important role to improve prediction of protein three-dimensional structures and functions. The most commonly predicted properties are secondary structure and accessible surface area (ASA) representing local and nonlocal structural characteristics, respectively. Secondary structure prediction is further complemented by prediction of continuous main-chain torsional angles. Here we describe a newly developed method SPIDER2 that utilizes three iterations of deep learning neural networks to improve the prediction accuracy of several structural properties simultaneously. For an independent test set of 1199 proteins SPIDER2 achieves 82 % accuracy for secondary structure prediction, 0.76 for the correlation coefficient between predicted and actual solvent accessible surface area, 19° and 30° for mean absolute errors of backbone φ and ψ angles, respectively, and 8° and 32° for mean absolute errors of Cα-based θ and τ angles, respectively. The method provides state-of-the-art, all-in-one accurate prediction of local structure and solvent accessible surface area. The method is implemented, as a webserver along with a standalone package that are available in our website: http://sparks-lab.org.

Key words Secondary structure prediction, Solvent accessible surface area, Backbone torsion angles, Deep neural networks, C alpha-based angles

1 Introduction

With the rapid development of DNA sequencing techniques, there is a continuously increasing gap between the number of sequences available from genomic analysis and the number of structures and functions determined or annotated by expensive experimental techniques. It is highly desirable to develop theoretical methods to predict protein structures and functions from their one-dimensional

Yaoqi Zhou et al. (eds.), *Prediction of Protein Secondary Structure*, Methods in Molecular Biology, vol. 1484,
DOI 10.1007/978-1-4939-6406-2_6, © Springer Science+Business Media New York 2017

sequences. However, methods for highly accurate prediction of protein three-dimensional structures (except homology modeling) are not yet available. This has significantly limited the ability to annotate protein functions based on their three-dimensional structures. As a result, predicted one-dimensional structural properties of proteins have often been utilized for predicting protein functions [1–4], their binding sites to other molecules [5–7], and other studies [8–11]. They have also been widely employed to improve protein structure prediction methods: both ab initio [12–14] and template-based techniques [15–18]. Thus any improvement in predicted one-dimensional structural properties will benefit protein structure and function modeling.

The most commonly predicted one-dimensional structural property of a protein is three-state secondary structure (helix, sheet, and coil). Secondary structure prediction accuracy without using homologous sequences in training has gradually been improved to above 81 % in recent years [19, 20], due to improved machine-learning algorithms, better features, and available larger training datasets.

An alternative to secondary structures is angle-based representation of backbone structure. Angle-based description such as torsion angles φ and ψ offers a continuous representation of local conformation [12], rather than discontinuous and somewhat arbitrary definition of three secondary-structure states. The advantage of angle-based representation leads to methods for predicting torsional angles φ and ψ [12, 21], and Cα-based angles [an angle between $C\alpha_{i-1} - C\alpha_i - C\alpha_{i+1}$ (θ) and a dihedral angle rotated about the $C\alpha_{i-1} - C\alpha^i$ bond (τ)] [22].

Another important one-dimensional structure property is solvent Accessible Surface Area (ASA) that measures exposure of amino acid residues of proteins to solvent, which is important for understanding and predicting protein structure, function, and interactions [23–26]. Earlier multistate prediction [23, 27, 28] has been gradually moved to continuous real value prediction [29–33].

In a recent study, we have developed SPIDER2, an iterative deep-learning neutral network, to predict all above-mentioned structural properties at the same time [34]. The iterative and cross-learning method achieved 82 % accuracy for secondary structure prediction, 0.76 for the correlation coefficient between predicted and actual solvent accessible surface area, 19° and 30° for mean absolute errors of backbone φ and ψ angles, respectively, and 8° and 32° for mean absolute errors of Cα-based θ and τ angles, respectively, for an independent test dataset of 1199 proteins. The resulting method provides state-of-the-art, all-in-one accurate prediction of local structure and solvent accessible surface area.

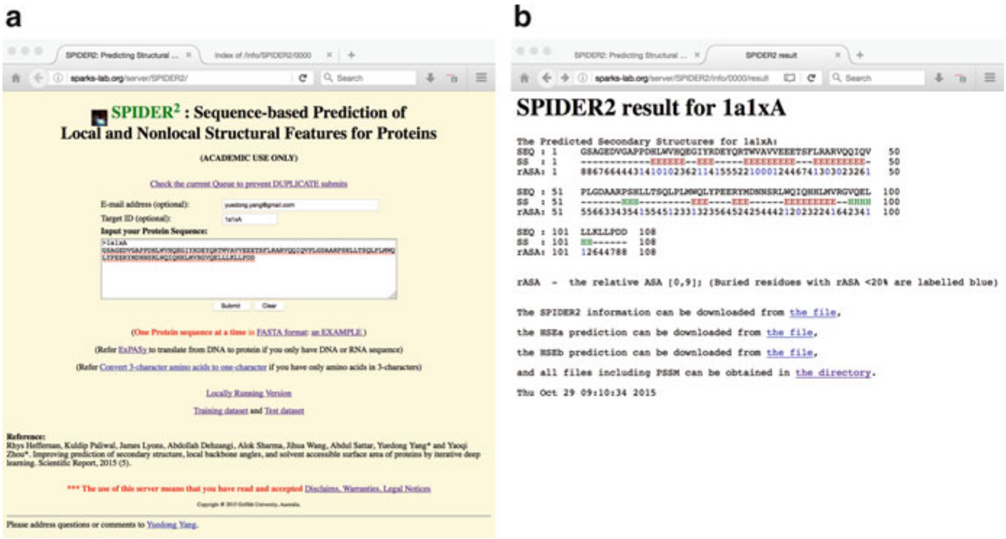
2 Algorithm

SPIDER2 server version was trained on a dataset of 5789 nonredundant (25 % cutoff), high resolution (<2.0 Å) structure by employing a three consecutive deep neural networks trained iteratively. In each iteration, we employed a deep neural network (DNN) consisting of three hidden layers with 150 hidden nodes in each layer. The weights were initialized by stacked sparse autoencoder [35] and then refined by standard back-propagation through fine-tuned supervised training [36, 37]. The learning rates for backward propagation were 1, 0.5, 0.2, and 0.05, respectively, with 30 epochs at each learning rate. The input layer for the DNN in the first iterative learning consists of 459 features (27 features per residue for a sliding window of 17 residues centered at the query residue). These 27 features include seven representative physical chemical properties parameters (steric parameter (graph shape index), hydrophobicity, volume, polarizability, isoelectric point, helix probability, and sheet probability properties of the amino acids), and 20 substitution probabilities obtained from 3 iterations searching by PSIBLAST [38]. All input features are normalized to the range of 0 to 1. For residues near the ends of a protein, the features of the amino acid residue at the current end of the protein were duplicated so that a full window could be used. Predicted outputs are 12 values of predicted probabilities for three secondary structure states, relative ASA, and sine and cosine of four angles θ, τ, ϕ, and ψ. The input layers for the DNN in the second and third iterative learning are 12 predicted values in the previous iteration plus 27 above-employed features per residue, that is, 663 features [=(12 + 27) × 17].

3 Web Server

The simplest way to use SPIDER2 is to submit a query sequence to our server at http://sparks-lab.org/yueyang/server/SPIDER2.

1. As shown in Fig. 1a, your protein sequence can be entered (or copy-pasted) in the FASTA format into the text area. Only one protein sequence is allowed each time. The sequence must contain 20 standard amino acids only. The first comment line in the FASTA format (“>” followed by the protein name) is employed to identify the name of the query protein. Without this line, the protein name will be set as “unknown” by default. The email address and target name in the webpage are optional. If you have a DNA/RNA sequence, you need first to convert them into a protein sequence (*see* **Note 1**).


a
b
SPIDER2: Predicting Structural ...
Index of /info/SPIDER2/0000
sparks-lab.org/server/SPIDER2/
Search
SPIDER2 result
sparks-lab.org/server/SPIDER2/info/0000/result

SPIDER² : Sequence-based Prediction of
Local and Nonlocal Structural Features for Proteins

(ACADEMIC USE ONLY)

Check the current Queue to prevent DUPLICATE submits

E-mail address (optional): yuedong.yang@gmail.com
Target ID (optional): 1a1xA
Input your Protein Sequence:

```
>1a1xA
GSAGEDVGAPPDHLWVHQEGIYRDEYQRTWVAVVEEETSFLRARVQQIQVPLGDAARPSHLLTSQLPLMWQ
LYPEERYMDNNSRLWQIQHHLMVRGVQELLLKLLPDD
```

Submit Clear

(One Protein sequence at a time in FASTA format: an EXAMPLE.)

(Refer ExPASy to translate from DNA to protein if you only have DNA or RNA sequence)

(Refer Convert 3-character amino acids to one-character if you have only amino acids in 3-characters)

Locally Running Version

Training dataset and Test dataset

Reference:
Rhys Heffernan, Kuldip Paliwal, James Lyons, Abdollah Dehzangi, Alok Sharma, Jihua Wang, Abdul Sattar, Yuedong Yang* and Yaoqi Zhou*. Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. Scientific Report, 2015 (5).

*** The use of this server means that you have read and accepted Disclaims, Warranties, Legal Notices

Copyright © 2015 Griffith University, Australia

Please address questions or comments to Yuedong Yang.

SPIDER2 result for 1a1xA

```
The Predicted Secondary Structures for 1a1xA:
SEQ : 1     GSAGEDVGAPPDHLWVHQEGIYRDEYQRTWVAVVEEETSFLRARVQQIQV   50
SS  : 1     ------------EEEEEE--EEE-----EEEEEEEEE---EEEEEEEEE-   50
rASA: 1     886766644431410102362114155522100012446741303023261  50

SEQ : 51    PLGDAARPSHLLTSQLPLMWQLYPEERYMDNNSRLWQIQHHLMVRGVQEL  100
SS  : 51    -------HHH---------EEE----EEE------EEEEEEEEE--HHHH  100
rASA: 51    55663343541554512331323564524254444212023224164234 1  100

SEQ : 101   LLKLLPDD  108
SS  : 101   HH------  108
rASA: 101   12644788  108

rASA  -  the relative ASA [0,9]; (Buried residues with rASA <20% are labelled blue)

The SPIDER2 information can be downloaded from the file,

the HSEa prediction can be downloaded from the file,

the HSEb prediction can be downloaded from the file,

and all files including PSSM can be obtained in the directory.

Thu Oct 29 09:10:34 2015
```

Fig. 1 The webserver input (**a**) and output screenshots (**b**) for example sequence"1a1xA.seq" by SPIDER2

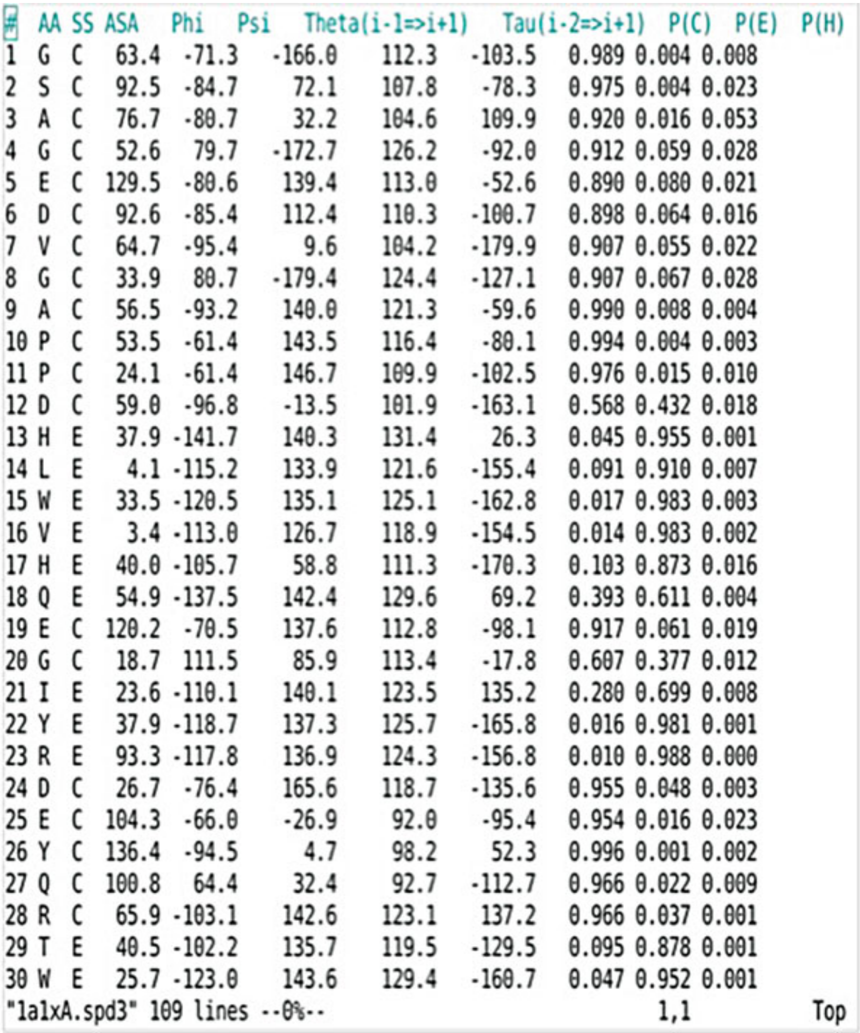
2. By clicking the "submit" button, the job will be sent to a queue, and the webpage will be directed to a new page, where the "Click the link" points to a to-be-available result file. This webpage will be automatically refreshed every 60 s until the job is completed and the result is displayed on the web page.
3. Each prediction is usually completed within 10 min, but may take up to a few hours depending on how busy the server is and how long the protein chain is. If an email address is provided in submission, the link to the result webpage will be sent to the mailbox as soon as the prediction is finished. All prediction results are kept in the server for 1 month and automatically deleted afterwards.
4. If the users have their own Position Specific Substitution Matrix (PSSM) file for their query protein sequence, SPIDER2 prediction can be made by submitting the PSSM file to the server. Using an external PSSM file can skip the most time-consuming step of generating the evolution profile by PSI-BLAST, and the executive time reduce to a few seconds.
5. To save computing resources, please do not submit query sequences more than once. The status of your job can be found by clicking the link "Check the current Queue to prevent DUPLICATE submission" on the server webpage.
6. Figure 1b shows an example for the output webpage. Aligned lines started with "SEQ," "SS," and "rASA" represent query sequence, predicted secondary structure, and predicted relative accessible surface area, respectively. For SS, predicted coil, helix and sheet residues are represented by "–," red "E," and green "H," respectively. For rASA, the relative ASA is represented by 0–9 with "0" for up to 10% of its surface exposed and "9" for above 90% exposed. The residues of rASA less

than 20 % (buried residues) are labeled in blue. Here, rASA is normalized by a residue-specific reference value (the ASA in the fully exposed state of a residue when connected by an ALA in each side). This output page does not contain predicted secondary structure probability, predicted angles, and actual real values of ASA. The complete prediction file "pro1.spd3" (*see* Subheading 4, **step 6** for explanation of the file) together with other intermediate files such as PSSM can be downloaded following the link in this output webpage.

4 Standalone Software

SPIDER2 is also available as a standalone software package. The program was designed to run in a Linux environment with python 2.7 and numpy version 1.4 or above. The input is a protein sequence in FASTA format, and outputs include predicted secondary structure, accessible surface area, main-chain torsional angles (phi/psi and theta/tau). The program can be installed in following steps.

1. Download the software package from our homepage with a shortcut link: http://sparks-lab.org/pmwiki/download/index.php?Download=yueyang/SPIDER2_local.tgz after entering your name and email address. This information will be used only for notification of future updates. You can fill in "none" if you prefer not to leave your information.
2. Unzip the package by command "tar zxvf SPIDER2_local.tgz" which creats the directory "SPIDER2_local" containing a "Readme" file and three subdirectories "dat," "ex," and "misc." The "dat" directory contains three npz files of trained parameters for three iterative neural networks, respectively, and the "misc" directory contains the program and auxiliary script files.
3. If BLAST or BLAST+ package is not installed in your computer, the software can be obtained from NCBI website. This program further requires correctly formatted nonredundant protein sequence databases, which can be downloaded from NCBI ftp://ftp.ncbi.nlm.nih.gov/blast/db (all files starting with "nr"). Until Oct 2015, the NR database contains a total of 40 files in 22GB before uncompressing. Alternatively, you can utilize a database by removing highly homology sequences, e.g., Uniref90 (*see* **Note 2**). This will speed up the calculation without making significant changes in prediction accuracy. This step can be skipped if you have prepared PSSM files (*see* **Note 3**).
4. SPIDER2 is called by the command "run_local.sh," followed by all sequence files in FASTA format. Here, one input file can contain a protein sequence only (*see* **Note 4**).



```
#  AA SS ASA   Phi    Psi    Theta(i-1=>i+1)  Tau(i-2=>i+1)  P(C)  P(E)  P(H)
1  G  C   63.4  -71.3  -166.0   112.3  -103.5   0.989 0.004 0.008
2  S  C   92.5  -84.7    72.1   107.8   -78.3   0.975 0.004 0.023
3  A  C   76.7  -80.7    32.2   104.6   109.9   0.920 0.016 0.053
4  G  C   52.6   79.7  -172.7   126.2   -92.0   0.912 0.059 0.028
5  E  C  129.5  -80.6   139.4   113.0   -52.6   0.890 0.080 0.021
6  D  C   92.6  -85.4   112.4   110.3  -100.7   0.898 0.064 0.016
7  V  C   64.7  -95.4     9.6   104.2  -179.9   0.907 0.055 0.022
8  G  C   33.9   80.7  -179.4   124.4  -127.1   0.907 0.067 0.028
9  A  C   56.5  -93.2   140.0   121.3   -59.6   0.990 0.008 0.004
10 P  C   53.5  -61.4   143.5   116.4   -80.1   0.994 0.004 0.003
11 P  C   24.1  -61.4   146.7   109.9  -102.5   0.976 0.015 0.010
12 D  C   59.0  -96.8   -13.5   101.9  -163.1   0.568 0.432 0.018
13 H  E   37.9 -141.7   140.3   131.4    26.3   0.045 0.955 0.001
14 L  E    4.1 -115.2   133.9   121.6  -155.4   0.091 0.910 0.007
15 W  E   33.5 -120.5   135.1   125.1  -162.8   0.017 0.983 0.003
16 V  E    3.4 -113.0   126.7   118.9  -154.5   0.014 0.983 0.002
17 H  E   40.0 -105.7    58.8   111.3  -170.3   0.103 0.873 0.016
18 Q  E   54.9 -137.5   142.4   129.6    69.2   0.393 0.611 0.004
19 E  C  120.2  -70.5   137.6   112.8   -98.1   0.917 0.061 0.019
20 G  C   18.7  111.5    85.9   113.4   -17.8   0.607 0.377 0.012
21 I  E   23.6 -110.1   140.1   123.5   135.2   0.280 0.699 0.008
22 Y  E   37.9 -118.7   137.3   125.7  -165.8   0.016 0.981 0.001
23 R  E   93.3 -117.8   136.9   124.3  -156.8   0.010 0.988 0.000
24 D  C   26.7  -76.4   165.6   118.7  -135.6   0.955 0.048 0.003
25 E  C  104.3  -66.0   -26.9    92.0   -95.4   0.954 0.016 0.023
26 Y  C  136.4  -94.5     4.7    98.2    52.3   0.996 0.001 0.002
27 Q  C  100.8   64.4    32.4    92.7  -112.7   0.966 0.022 0.009
28 R  C   65.9 -103.1   142.6   123.1   137.2   0.966 0.037 0.001
29 T  E   40.5 -102.2   135.7   119.5  -129.5   0.095 0.878 0.001
30 W  E   25.7 -123.0   143.6   129.4  -160.7   0.047 0.952 0.001
"1a1xA.spd3" 109 lines --0%--                     1,1          Top
```

Fig. 2 The partial prediction results by SPIDER2 for the example sequence"1a1xA.seq"

5. Results will be saved in an output file with extension "spd3." An example of output is shown in Fig. 2. The output file contains 11 columns that represent the residue index, residue type, predicted secondary structure type, ASA, φ, ψ, θ, τ, and probabilities as coil (C), sheet (E), and helix (H). The predicted secondary structure is the secondary structure type with the highest probability. The θ angle at residue index i is the angle between $C\alpha_{i-1} - C\alpha_i - C\alpha_{i+1}$, and τ is the dihedral angle formed by $C\alpha_{i-2} - C\alpha_{i-1} - C\alpha_i - C\alpha_{i+1}$. Three torsional angles φ, ψ, and τ range from −180 to 180°, and angle θ mostly ranges between 70 and 180°.

6. In addition, the package includes one program "pred_nopssm.py" that makes prediction without using the PSSM from PSI-BLAST. Instead, the profile is replaced by the BLOSUM62 substitution matrix. This replacement allows a fast calculation

at a lower accuracy (For example, secondary structure accuracy at 68.9%, compared to 81.8% by using PSI-BLAST profile). This may be useful for large-scale calculations in genome level. However, it should be noted that all parameters were not optimized for the evolution-profile free prediction, and the development of a specific predictor by using sequence only is in progress.

5 Notes

1. The query sequence must be a protein sequence in the FASTA format. The gene in the DNA/RNA sequence has to be converted to the sequence of amino acids first. This conversion can be made by using http://web.expasy.org/translate or any other tools. Nonstandard amino acids (e.g., X) must be removed, prior to the use of SPIDER2.
2. The package employs PSI-BLAST to generate PSSM generated by scanning NR database. Alternatively, you can employ the sequence database uniref90 that can be downloaded from ftp://ftp.uniprot.org/pub/databases/uniprot/uniref/uniref90/uniref90.fasta.gz. This database can be converted to BLAST-readable format by the command "gunzip -c uniref90.fasta.gz | ~/aspen/software/ncbi-blast-2.2.30+/bin/makeblastdb -in - -dbtype prot -parse_seqids -out uniref90 -title uniref90." This operation skips the step of unzipping the large database.
3. For users with their own PSSM files, they can obtain predictions by utilizing the script "pred_pssm.py" followed by PSSM file names. This command will skip running PSI-BLAST and prediction can be finished in a few seconds.
4. If your sequence file contains more than one protein sequence, you can use the script file "splitseq.py" to split your sequence files to many files, and each file will be named according to protein names in the FASTA file.

Acknowledgements

This work was supported in part by National Health and Medical Research Council (1059775) of Australia and Australian Research Council's Linkage Infrastructure, Equipment and Facilities funding scheme (project number LE150100161), the Taishan Scholars Program of Shandong province of China, National Natural Science Foundation of China (61540025) to Y.Z. and National Natural Science Foundation of China (61271378) to Y.Y. and J.W. We also gratefully acknowledge the support of the Griffith University eResearch Services Team and the use of the High Performance Computing Cluster “Gowonda” to complete this research. This research/project has also been undertaken with the aid of the research cloud resources provided by the Queensland Cyber Infrastructure Foundation (QCIF).

References

1. Lou W, Wang X, Chen F, Chen Y, Jiang B, Zhang H (2014) Sequence based prediction of DNA-binding proteins based on hybrid feature selection using random forest and Gaussian naïve Bayes. PLoS One 9(1)
2. Zhao H, Yang Y, Zhou Y (2011) Highly accurate and high-resolution function prediction of RNA binding proteins by fold recognition and binding affinity prediction. RNA Biol 8(6):988–996. doi:10.4161/rna.8.6.17813
3. Zhao H, Yang Y, von Itzstein M, Zhou Y (2014) Carbohydrate-binding protein identification by coupling structural similarity searching with binding affinity prediction. J Comput Chem 35(30):2177–2183
4. Zhao H, Wang J, Zhou Y, Yang Y (2014) Predicting DNA-binding proteins and binding residues by complex structure prediction and application to human proteome. PLoS One 9(5):e96694
5. Zhang T, Zhang H, Chen K, Ruan J, Shen S, Kurgan L (2010) Analysis and prediction of RNA-binding residues using sequence, evolutionary conservation, and predicted secondary structure and solvent accessibility. Curr Protein Peptide Sci 11(7):609–628
6. Zhang Z, Li Y, Lin B, Schroeder M, Huang B (2011) Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. Bioinformatics 27(15):2083–2088
7. Bradford JR, Westhead DR (2005) Improved prediction of protein–protein binding sites using a support vector machines approach. Bioinformatics 21(8):1487–1494
8. Folkman L, Yang Y, Li Z, Stantic B, Sattar A, Mort M, Cooper DN, Liu Y, Zhou Y (2015) DDIG-in: detecting disease-causing genetic variations due to frameshifting indels and nonsense mutations employing sequence and structural properties at nucleotide and protein levels. Bioinformatics 31(10):1599–1606
9. Zheng W, Zhang C, Hanlon M, Ruan J, Gao J (2014) An ensemble method for prediction of conformational B-cell epitopes from antigen sequences. Comput Biol Chem 49:51–58
10. Zhao H, Yang Y, Lin H, Zhang X, Mort M, Cooper DN, Liu Y, Zhou Y (2013) DDIG-in: discriminating between disease-associated and neutral non-frameshifting micro-indels, Genome Biology, 14, R43
11. Lyons J, Dehzangi A, Heffernan R, Yang Y, Zhou Y, Sharma A, Paliwal K (2015) Advancing the accuracy of protein fold recognition by utilizing profiles from Hidden Markov models, IEEE Transactions on NanoBioscience, 14, 761–772
12. Faraggi E, Yang Y, Zhang S, Zhou Y (2009) Predicting continuous local structure and the effect of its substitution for secondary structure in fragment-free protein structure prediction. Structure 17(11):1515–1527. doi:10.1016/j.str.2009.09.006
13. Bradley P, Chivian D, Meiler J, Misura KM, Rohl CA, Schief WR, Wedemeyer WJ, Schueler-Furman O, Murphy P, Schonbrun J, Strauss CE, Baker D (2003) Rosetta predictions in CASP5: successes, failures, and prospects for complete automation. Proteins 53(Suppl 6):457–468. doi:10.1002/prot.10552
14. Handl J, Knowles J, Vernon R, Baker D, Lovell SC (2012) The dual role of fragments in fragment-assembly methods for de novo protein structure prediction. Proteins 80(2):490–504
15. Yang Y, Faraggi E, Zhao H, Zhou Y (2011) Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. Bioinformatics 27(15):2076–2082. doi:10.1093/bioinformatics/btr350
16. Zhang Y (2009) I-TASSER: fully automated protein structure prediction in CASP8. Proteins 77(S9):100–113
17. Remmert M, Biegert A, Hauser A, Söding J (2011) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat Meth 9(2):173–175
18. Cheng J, Wang Z, Tegge AN, Eickholt J (2009) Prediction of global and local quality of CASP8 models by MULTICOM series. Proteins 77(S9):181–184
19. Faraggi E, Zhang T, Yang Y, Kurgan L, Zhou Y (2011) SPINE X: improving protein second-

ary structure prediction by multi-step learning coupled with prediction of solvent accessible surface area and backbone torsion angles. J Comput Chem 33:259–263

20. Yaseen A, Li YH (2014) Context-based features enhance protein secondary structure prediction accuracy. J Chem Inf Model 54(3):992–1002. doi:10.1021/Ci400647u
21. Wu S, Zhang Y (2008) ANGLOR: a composite machine-learning algorithm for protein backbone torsion angle prediction. PLoS One 3(10):e3400
22. Lyons J, Dehzangi A, Heffernan R, Sharma A, Paliwal K, Sattar A, Zhou Y, Yang Y (2014) Predicting backbone Calpha angles and dihedrals from protein sequences by stacked sparse autoencoder deep neural network. J Comput Chem 35(28):2040–2046. doi:10.1002/jcc.23718
23. Rost B, Sander C (1994) Conservation and prediction of solvent accessibility in protein families. Proteins 20(3):216–226
24. Gilis D, Rooman M (1997) Predicting protein stability changes upon mutation using database-derived potentials: solvent accessibility determines the importance of local versus non-local interactions along the sequence. J Mol Biol 272(2):276–290
25. Tuncbag N, Gursoy A, Keskin O (2009) Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. Bioinformatics 25(12):1513–1520
26. Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. J Mol Biol 55(3):379–400
27. Holbrook SR, Muskal SM, Kim SH (1990) Predicting surface exposure of amino acids from protein sequence. Protein Eng 3(8):659–665
28. Pollastri G, Baldi P, Fariselli P, Casadio R (2002) Prediction of coordination number and relative solvent accessibility in proteins. Proteins 47(2):142–153
29. Dor O, Zhou Y (2007) Real-SPINE: an integrated system of neural networks for real-value prediction of protein structural properties. Proteins 68(1):76–81
30. Garg A, Kaur H, Raghava GP (2005) Real value prediction of solvent accessibility in proteins using multiple sequence alignment and secondary structure. Proteins 61(2):318–324. doi:10.1002/prot.20630
31. Yuan Z, Huang B (2004) Prediction of protein accessible surface areas by support vector regression. Proteins 57(3):558–564. doi:10.1002/prot.20234
32. Ahmad S, Gromiha MM, Sarai A (2003) Real value prediction of solvent accessibility from amino acid sequence. Proteins 50(4):629–635. doi:10.1002/prot.10328
33. Adamczak R, Porollo A, Meller J (2004) Accurate prediction of solvent accessibility using neural networks-based regression. Proteins 56(4):753–767. doi:10.1002/prot.20176
34. Heffernan R, Paliwal K, Lyons J, Dehzangi A, Sharma A, Wang J, Sattar A, Yang Y, Zhou Y (2015) Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. Sci Rep 5:11476
35. Bengio Y, Lamblin P, Popovici D, Larochelle H (2007) Greedy layer-wise training of deep networks. Adv Neural Inform Process Syst 19:153
36. Hinton GE (2007) Learning multiple a layers of representation. Trends Cogn Sci 11(10):428–434. doi:10.1016/J.Tics.2007.09.004
37. Bengio Y (2009) Learning deep architectures for AI. Found Trends Mach Learn 2(1):1–127
38. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17):3389–3402