

Real-value prediction of backbone torsion angles

Bin Xue,^{1,2} Ofer Dor,³ Eshel Faraggi,^{1,2} and Yaoqi Zhou^{1,2*}

¹ Indiana University School of Informatics, Indiana University-Purdue University, Indianapolis, Indiana 46202

² Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana 46202

³ Faculty of Engineering, Tel Aviv University, Tel Aviv 69978, Israel

ABSTRACT

The backbone structure of a protein is largely determined by the ϕ and ψ torsion angles. Thus, knowing these angles, even if approximately, will be very useful for protein-structure prediction. However, in a previous work, a sequence-based, real-value prediction of ψ angle could only achieve a mean absolute error of 54° (83° , 35° , 33° for coil, strand, and helix residues, respectively) between predicted and actual angles. Moreover, a real-value prediction of ϕ angle is not yet available. This article employs a neural-network based approach to improve ψ prediction by taking advantage of angle periodicity and apply the new method to the prediction to ϕ angles. The 10-fold-cross-validated mean absolute error for the new method is 38° (58° , 33° , 22° for coil, strand, and helix, respectively) for ψ and 25° (35° , 22° , 16° for coil, strand, and helix, respectively) for ϕ . The accuracy of real-value prediction is comparable to or more accurate than the predictions based on multistate classification of the ϕ - ψ map. More accurate prediction of real-value angles will likely be useful for improving the accuracy of fold recognition and ab initio protein-structure prediction. The Real-SPINE 2.0 server is available on the website <http://sparks.informatics.iupui.edu>.

Proteins 2008; 72:427–433.
© 2008 Wiley-Liss, Inc.

Key words: artificial neural networks; dihedral angle; secondary structure.

INTRODUCTION

The polypeptide backbone of a protein is a linked sequence of rigid planar peptide groups. The rotational angle about the C—N bond, ω , is fixed at 180° for the common trans and 0° for the rare cis conformation. Thus, protein backbones can be described well by two rotation angles (torsion angles) about the C_α —N bond (ϕ) and the C_α —C bond (ψ). By convention, they vary from -180° to 180° . However, not all angles can be sampled by the backbone because of internal steric constraints as discovered by Ramachandran and Sasisekharan.¹ As a result, sampling in torsional space is one of the most commonly used methods for efficient exploration of protein conformational space [for example, see Refs. 2–4]. This is also the motivation behind the development of simple models based on a few torsion-angle states.^{5–8}

Several methods have been developed for direct prediction of ϕ/ψ angles. The majority of them predict discrete dihedral-angle states based on local (fragment) structural patterns using either machine-learning techniques or classification schemes.^{9–16} These studies centered on predicting more accurate local structures of short peptide segments than simple three states usually predicted by secondary-structure prediction methods. The predicted angles have been used to improve fold recognition¹² sequence alignment,¹⁷ and the accuracy of secondary structure prediction.^{15,18}

However, backbone angles are continuously varying variables because proteins can move freely in a three-dimensional space. Thus, we prefer a real-value prediction rather than arbitrarily dividing the angles into a few states. Wood and Hirst¹⁸ made the first real-value prediction of ψ angles for the purpose of improving the accuracy of secondary-structure prediction. The correlation coefficient between predicted and actual values of the angles, however, was only 0.47. We developed a method called Real-SPINE that predicts the real values of structural properties of proteins using integrated neural networks.¹⁹ The method substantially improved the correlation coefficient to 0.62 (10-fold cross-validated) through large-scale learning with a slow learning rate and overfitting protection.

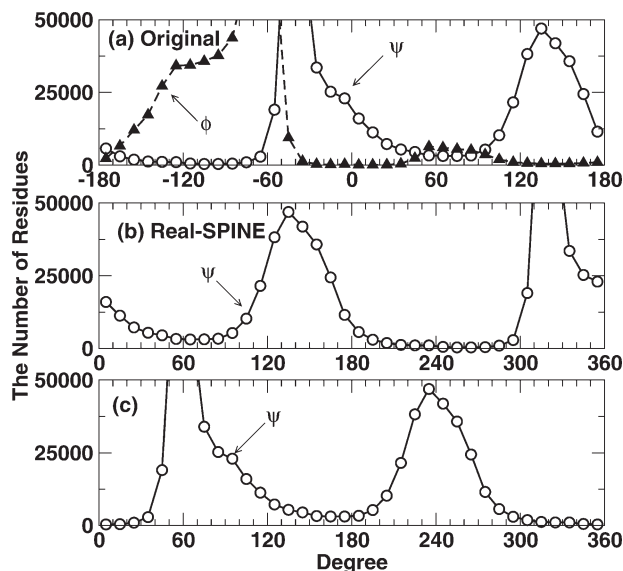
However, Real-SPINE makes a relatively poor prediction for the ψ angle around 0° .¹⁹ In fact, its accuracy between -36° and 36° is even lower than the accuracy by random prediction. We attribute the poor performance near 0° to the angle periodicity because a prediction of 360° should be perfect for an actual angle of 0° , rather than an error of 360° . Another reason for the poor

Grant sponsor: NIH; Grant numbers: GM966049, GM068530.

*Correspondence to: Dr. Yaoqi Zhou, Indiana University School of Informatics, Indiana University-Purdue University, Indianapolis, Indiana 46202. E-mail: yqzhou@iupui.edu.

Received 2 October 2007; Revised 12 November 2007; Accepted 3 December 2007

Published online 23 January 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21940

**Figure 1**

The distribution of ψ and ϕ angles from a dataset of 2640 proteins. (a) Original ψ angle from -180° to 180° . A plot of ϕ angle is also shown (in triangle) (b) Transformation of ψ angles to 0° – 360° in real-SPINE 1.0. (c) The proposed transformation (simple shift) of ψ angles.

accuracy near 0° is that the angle transformation of Real-SPINE positioned a considerable amount of angles near the edges of the neural network prediction range, where it is very difficult for the neural network to predict (due to the shape of the sigmoidal function).

The purpose of this article is to further improve the accuracy of predicted ψ angles by reducing the impact of angle periodicity. Moreover, we will establish the first sequence-based method that predicts the real values of both ϕ and ψ angles. The knowledge of both angles will be useful for aiding protein-structure prediction.

RESULTS

To improve Real-SPINE, we first analyze the distribution of ψ angles obtained from the structural database of 2640 proteins [Fig. 1(a)]. There are two peaks contributed mostly from α helical and β -strand conformations, respectively. In Real-SPINE,¹⁹ the ψ dihedral angles are converted to angles between 0° to 360° by keeping the angles between 0° to 180° unchanged and adding 360 for angles between -180° and 0° [Fig. 1(b)]. This transformation, however, leads to significant population at 0° or 360° .

To minimize the impact of the angle periodicity at 0° (or 360°), we make a simple angle transformation so that the probability of the transformed angle is close to zero at 0° or 360° . We observe the lowest number of residues with $\psi \sim -100^\circ$ [Fig. 1(a)]. Thus, we can convert ψ angles by adding 100 to the angles between -100° to 180° and adding 460 to the angles between -180° and -100° . This simple shift operation makes the occurrence of 0° the lowest probability for the shifted angles [Fig. 1(c)]. The simple shift makes the ψ angle more learnable because of less confusion at 0° and 360° .

The distribution of ϕ angles also has two peaks [the major one at around -60° and the minor one at around 60° , Fig. 1(a)]. Because the lowest number of residues is at $\psi \sim 10^\circ$, the ϕ dihedral angles are converted to angles between 0° to 360° by adding -10 to the angles between 10° to 180° and adding 350 for angles between -180° and 10° . All shifted angles are normalized by 360 so that the output values range from 0 to 1. The clustering of ϕ angles around -60° does not mean that an accurate prediction is assured because the accuracy of prediction is measured by real values rather than two-state or one-state approximation although it will make ϕ relatively easier to predict than ψ , as we shall see below.

Table I compares the accuracy of predicted ψ angles between Real-SPINE¹⁹ and this work based on shifted angles when both use two independent predictors (A and

Table I

The Accuracy of the ψ/ϕ Angle Prediction by 10-Fold Cross Validation Using a Dataset of 2640 Proteins

Method	ψ			ϕ
	RS ^a (AB) ^b	This work (AB) ^b	This work (ABCDE) ^b	This work (ABCDE) ^b
Q_{10} ^c	41.2%	45.8% \pm 1.1%	46.6% \pm 0.2%	45.0% \pm 1.1%
$Q_{10\%}$ ^d	—	64.5% \pm 1.2%	64.7% \pm 0.2%	80.1% \pm 0.2%
Corr. coeff. ^e	0.619	0.740 \pm 0.008	0.745 \pm 0.002	0.707 \pm 0.003
Abs. Error ^f	54.0°	38.9°	38.2°	24.8°
–Helix ^g	33.1°	22.4°	22.1°	15.9°
–Strand ^g	34.9°	33.5°	33.1°	22.0°
–Coil ^g	82.8°	57.4°	56.9°	35.4°

^aReal-SPINE for ψ only.

^bA consensus of two (AB) or five independent predictors (ABCDE). The error bar is estimated from one standard deviation from two or five independent predictions.

^c Q_{10} : fraction of residues with correctly predicted states. Angles are divided into 10 states with 36° per bin.

^d $Q_{10\%}$: fraction of residues whose angles are predicted within 36° from the true value.

^eCorrelation coefficient between predicted and actual angles.

^fMean-absolute error (MAE) between predicted and actual angles.

^gMAE for helical, strand, and coil residues, respectively.

B) for a consensus prediction. Marked improvement is observed for all the parameters used to measure the accuracy of the method. The correlation coefficient increases from 0.62 to 0.75, the fraction of residues within their correct 36° bins (Q_{10}) increases from 41 to 46%. Meanwhile, the mean absolute error reduces from 54° to 39° . More specifically, the error reduces from 33° to 22° for helix, 35° to 33° for strand, and 83° to 57° for coil residues. The largest improvement is for helix and coil residues because ψ angles of some helical and coil residues are close to 0° .¹ The accuracy of angle prediction increases slightly by expanding a two-predictor-based consensus to a five-predictor-based consensus (1% in Q_{10} , 1° in mean absolute error, and 0.005 in correlation coefficient).

Table I also reports the accuracy for ϕ prediction. The Q_{10} score is similar to the accuracy for ψ prediction whereas the mean absolute error is smaller. A smaller MAE for ϕ prediction is because ϕ angles are dominated by one peak [Fig. 1(a)].

The distribution of Q -score for ψ in 10-angle bins is shown in Figure 2. Interestingly, the new method has a higher Q -score than Real-SPINE only in five angle bins from -36° to 144° . These five angle bins are mostly located at the regions with significant angle populations [-72° to 36° and 108° to 144° , Fig. 1(a)]. Thus, the new method improves over Real-SPINE by focusing on highly populated regions for ψ angles. The highest Q -score is close to 80% for angles between -72° and -36° (helical regions) and followed by about 55% for angles between 108° and 144° (strand regions). Q -scores from this work are not always higher than the accuracy from random prediction (from -180° to -72° , from -36° to 0° and from 144° to 180°).

Figure 3 shows the Q -score distribution for ϕ . Unlike ψ prediction, the accuracy of predicted ϕ angles is better

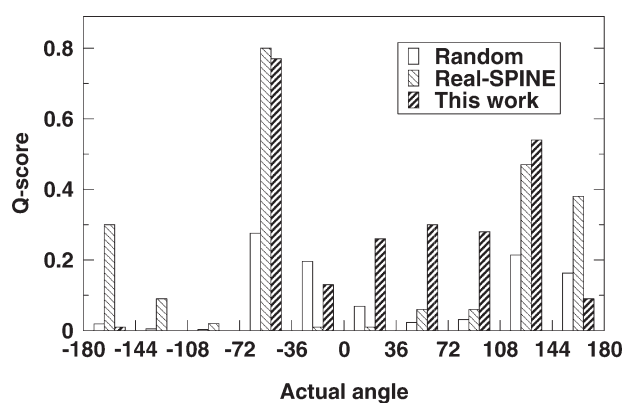


Figure 2

Q-score of ψ backbone-angle prediction as a function of ψ angles from Real-SPINE and this work. The accuracy from random prediction is also shown.

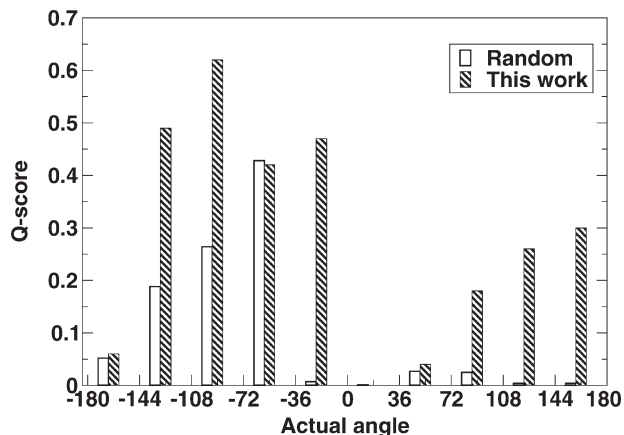


Figure 3

Q-score of ϕ backbone-angle prediction as a function of ϕ angles.

than that of random prediction in eight angle bins. The accuracy of prediction for the ϕ angles between -72° and -36° is slightly less than that of the random prediction. This poor accuracy at this highly populated region highlights the need for further improvement.

We further analyze Q_{10} as a function of residue solvent accessibility. Residue solvent accessibility (RSA) is the solvent-accessible surface areas of a residue in a protein normalized by the accessible surface area of the residue in its “unfolded” state.^{19,20} Figure 4 plots the fraction of residues at a given solvent-accessibility bin with correctly classified angle states for ψ angles. Random prediction, Real-SPINE, and this work all give the highest Q -score

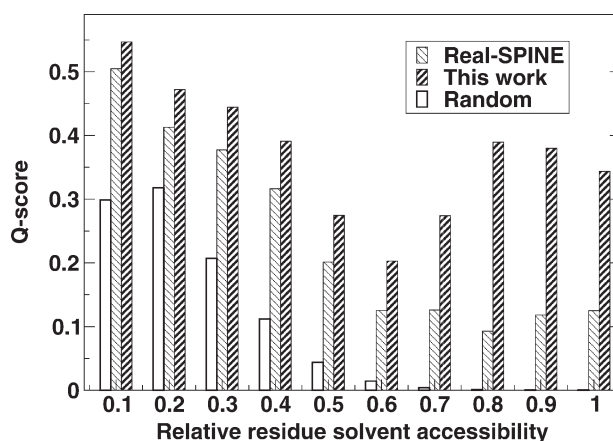


Figure 4

The fraction of residues with correctly classified states for ψ torsion angles as a function of residue solvent accessibility (1 for fully exposed and 0 for fully buried).

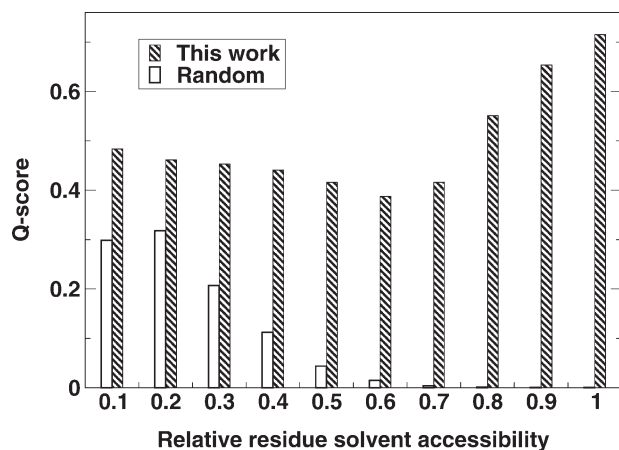


Figure 5

The fraction of residues with correctly classified states for ϕ torsion angles as a function of residue solvent accessibility (1 for fully exposed and 0 for fully buried).

for the residues buried most (solvent accessibility close to zero). Interestingly, Q-score is the lowest ($\sim 20\%$) at 60% accessibility for this work but is essentially unchanged ($\sim 10\%$) for exposed residues with $>60\%$ accessibility in Real-SPINE. Figure 5 further shows that ϕ angles of exposed residues are even more accurately predicted than those of buried residues. Obviously, there are two competing factors for the accuracy of prediction. Torsion angles of locally interacting residues (exposed residues) are easier to predict and buried residues having the highest populations are also accurately predicted with more learning for the neural network.

It is of interest to know the accuracy of prediction for individual residues. Table II compares the mean absolute errors for each residue given by Real-SPINE and by this work. Results for ϕ angles are also shown in the table. It is clear that the prediction error obtained in this work is reduced for every residue from the error in Real-SPINE. The residue with the largest error is Glycine (G) for both ϕ and ψ angles. This corresponds well to the fact that Glycine is the residue with the least restriction in the Ramachandran diagram.¹ Residue Proline (P), on the other hand, has the second largest error in ψ but the smallest error in ϕ . Proline has special restrictions in ϕ - ψ space that arise from its five-member ring. Its ϕ angle is restricted at around -60° and thus easier to predict than ϕ angles of other residues. The ψ angle of Proline is difficult to predict because it is not favored in either helix or sheet regions.²¹

Figure 6 shows the fraction of proteins with more than a given fraction of correctly predicted angles (ψ , ϕ , or both). Here, a correct prediction is defined as 36° or less from the actual angle. There are 93.1% of proteins with

Table II

Mean Absolute Error by 20 Types of Amino Acids (Normalized by 360)

AA type	ψ			ϕ
	RS ^a (AB) ^b	This work (AB) ^b	This work (ABCDE) ^b	This work (ABCDE) ^b
A	0.123	0.093	0.092	0.060
R	0.140	0.100	0.099	0.066
N	0.203	0.116	0.115	0.106
D	0.191	0.116	0.115	0.080
C	0.151	0.111	0.110	0.071
Q	0.145	0.098	0.097	0.064
E	0.136	0.094	0.093	0.060
G	0.228	0.169	0.167	0.159
H	0.177	0.114	0.113	0.080
I	0.103	0.080	0.078	0.047
L	0.124	0.086	0.084	0.047
K	0.148	0.103	0.102	0.066
M	0.127	0.092	0.091	0.057
F	0.139	0.100	0.098	0.065
P	0.168	0.140	0.139	0.037
S	0.169	0.131	0.129	0.075
T	0.160	0.123	0.121	0.059
W	0.143	0.105	0.104	0.067
Y	0.144	0.104	0.103	0.067
V	0.106	0.084	0.083	0.050

^aReal-SPINE for ψ only.

^bA consensus of two (AB) or five independent predictors (ABCDE).

more than 80% correctly predicted ϕ angles. In contrast, only 35.2% proteins are predicted with more than 80% correctly predicted ψ angles. More accurate prediction of ϕ angle results from the simpler distribution of ϕ angle that is dominated by a single peak (Fig. 1). Nevertheless, 73.7% proteins have 60% or more correctly predicted angles for both ϕ and ψ . We also found 78.5% (81.8%)

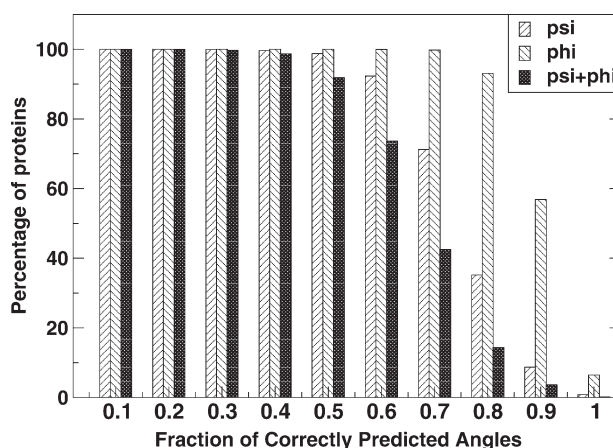


Figure 6

Percentage of proteins with more than a fraction of correctly predicted dihedral angles (ϕ , ψ , or both for $|x^{\text{pred}} - x^{\text{actual}}| \leq 36^\circ$). There is 73.7% proteins with 60% or more correctly predicted angles (for both ϕ and ψ).

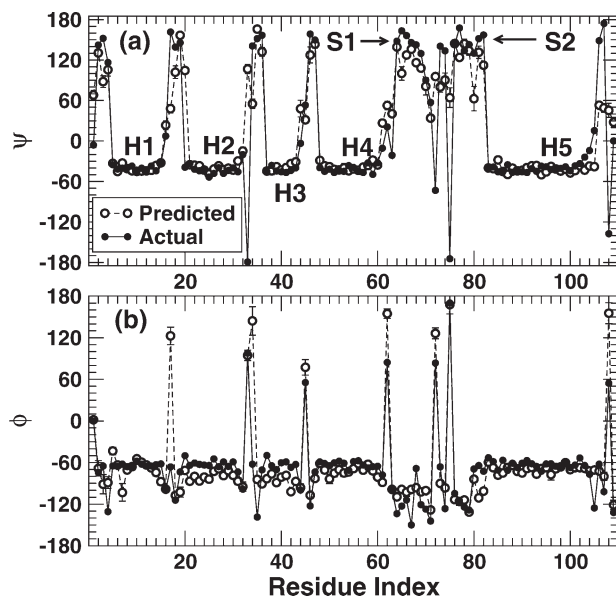


Figure 7

The comparison between predicted and actual ψ (a) and ϕ angles (b) for 1sfx (pdb id). The error bars in the figure are from one standard deviation from five independent predictions.

of proteins with a correlation coefficient of 0.6 or above between predicted and measured ψ (ϕ) angles.

We further investigate whether or not the residues with incorrectly predicted angles in a protein are isolated or clustered in a segment. If a correct prediction is defined by within 36° of the actual angle, the populations of the incorrect predictions associated with 1, 2, 3, 4, 5, 6 neighboring residues are 6.1, 5.8, 5.2, 4.5, 3.3, and 2.6% for ψ and 9.9, 5.6, 2.6, 1.1, 0.5, and 0.2% for ϕ , respectively. Thus, the most frequent error is a residue with an incorrectly predicted angle surrounded by two residues with correctly predicted angles.

As an example, Figure 7 compares predicted and actual ϕ and ψ angles for chain A of 1sfx (pdb id). The correlation between predicted and actual values for this protein is 0.88 for ψ and 0.80 for ϕ . All five helices in 1sfx are accurately predicted well. The two-strand regions are also predicted reasonably well. As shown in the figure, the error bar estimated from five independent predictors is usually smaller than the actual error for the angle. Thus, it is not a reliable predictor for the error associated with predicted angles.

It is tempting to know if there is any similarity between native structures and the structures built from predicted ϕ and ψ angles. We can assess the structural similarity by MaxSub score.²² MaxSub score between the predicted (model) structure and the native structure is a measure of similarity between 0.0 (no similarity) and 1.0

(perfect similarity). The value is calculated by searching the largest subset of well-superimposed residues ($\leq 3.5\text{\AA}$). A nonzero MaxSub score is used as an indication of detecting structural similarity.²³ The MaxSub score for the model built for Chain A of 1sfx is 0.09. We built the models based on predicted backbone angles for all 2640 proteins. We found that there are 2105 (80%) proteins with nonzero MaxSub score and 269 (10%) proteins with MaxSub >0.1 . The largest MaxSub score, however, is only 0.25. Thus, it is not yet reliable to predict the three-dimensional structure directly based on predicted torsion angles. This is expected because a small change in torsion angles may lead to a large change in structures. In fact, native backbone torsion angles with standard bond lengths and angles do not necessarily yield native structures correctly. Nevertheless, the constructed models likely reduce the conformational space of searching native structures significantly because of prefolded secondary structural elements.

DISCUSSION

A purely sequence-based method is improved in this study to predict backbone ψ torsion angles of proteins. We proposed a simple angle transformation to reduce the effect of angle periodicity. The 10-fold cross-validated correlation coefficient between predicted and actual ψ angles (0.75) is significantly improved over 0.62 obtained by Real-SPINE. Meanwhile, the mean absolute error reduces by 16° from 54° to 38° and Q_{10} increases from 41 to 47%. We also predict ϕ angles. The 10-fold cross-validated correlation coefficient between predicted and actual ϕ angles is 0.71, the mean absolute error is 25° , and Q_{10} is 45%.

Errors associated with predicted angles are analyzed in details. The largest prediction errors are associated with coil residues. This is followed by strand and helical residues. Among 20 amino acid types, the ψ angles of Gly and Pro are the most difficult to predict. In the less populated region of ψ angle, the accuracy of prediction is worse than that from random prediction. Moreover, the accuracy for predicting locally interacting residues (fully exposed) is only $\sim 35\%$ for ψ and 70% for ϕ (Q -score). This can be compared with a success rate of 87% in secondary structure prediction for exposed residues by SPINE.²⁹ There is no correlation between the standard deviation in five predictors and the actual error of predicted angles. Thus, how to predict the error of predicted angles is yet to be solved.

It is of interest to compare the accuracy of real-value prediction with that of multiclass prediction. Kuang *et al.*¹⁴ divided ψ - ϕ map into four states. They have achieved a 10-fold-cross-validated success rate of 77% for identifying correct states (Q_4). This was compared with a success rate of 74% given by HMMSTR.^{10,14} The corresponding Q_4 accuracy in this work is 81%. Another assessment is MDA score, the percentage of all residues

in correctly predicted eight-residue segments (dihedral angles within 120° from true values).¹⁰ HMMSTR achieved 54% for prediction and 59% for training.¹⁰ Our 10-fold cross-validated MDA score is 55.8%. More recently, Zimmermann and Hansmann divided ψ - ϕ map into three states (alpha, beta, outlier) according to grids with 1° spacing. They trained SVM classifier with 499 proteins and tested only on one set of 97 proteins. Their accuracy can be converted to a Q_3 score of 82%. The corresponding 10-fold cross-validated Q_3 score of this work (i.e. 10 test sets of 264 proteins each) is also 82%. Thus, the real-value prediction is at least as accurate as multi-state prediction when the real-value prediction is converted to multistate prediction.

More accurately predicted angles will likely be more useful for improving fold recognition and conformational sampling of protein structures. Eighty percent of proteins built with predicted torsion angles have a positive Max-Sub score when compared with corresponding native structures. This indicates the prevalence of some native structures in models built with predicted torsion angles. Our initial studies indicate that fold-recognition alignment can be improved with predicted torsion angles as well as with predicted secondary structures²⁴ and predicted solvent accessibility surface areas.²⁵ These predicted angles can also be used as restraints with an appropriate energy function for *ab initio* folding and/or conformational sampling.^{26,27} In fact, the constructed model based on predicted backbone angles can be used as a starting structure for *ab initio* structure prediction. Work in this area is in progress.

METHODS

The dataset (2640 nonredundant high-resolution proteins), the input features (a window of 21 residues with sequence profiles, seven representative amino-acid properties and predicted secondary structures by SPINE²⁹), and the algorithms (consensus over independent neural-network predictors with one hidden layer) are exactly the same as used in Real-SPINE¹⁹ where readers can find more detailed information. The key new features in this article are (1) the prediction of ϕ angles in addition to ψ angles, (2) shifted angles for training and testing, (3) ϕ prediction is made without using predicted secondary structures because their usage does not improve the accuracy of prediction in the current approach, (4) a consensus over five independent predictors rather than two independent predictors, and (5) 100 hidden units rather than 200 hidden units are used because 200 hidden units do not make significant improvement over 100 hidden units. As in Real-SPINE,¹⁹ 5% of randomly selected data is used for independent test to avoid overfitting. ψ and ϕ dihedral angles of a given structure are calculated by the DSSP program.²⁸

A 10-fold cross validation is used to assess the accuracy of prediction. The training set is randomly divided into 10 parts, nine of which are for training and the rest for testing. The process is repeated 10 times. The final result is assessed by multiple parameters. They are the mean absolute error (MAE) and accuracy according to Q-score. MAE is the absolute difference between predicted and actual values of a normalized structural property that is averaged over all predicted residues. Q-score is used when structural properties are simplified into several classes. Q-score is the number of correctly classified residues in total number of residues in that class of structural properties. For example, Q_{10} is a measure of success when torsion angles are equally divided into 10 bins (36° per bin). Q_{10} is the fraction of residues whose angles are in correctly classified states. Also introduced is $Q_{10\%}$, the fraction of residues whose angles are within 36° from the actual value. We further report Pearson's correlation coefficient (PCC) between predicted and actual values of angles in all 10 tests to compare with previous work although it is not suitable for circular data of angles.

REFERENCES

1. Ramachandran GN, Sasisekharan V. Conformation of polypeptides and proteins. *Adv Protein Chem* 1968;23:283–437.
2. Abagyan RA, Totrov MM, Kuznetsov DA. ICM: a new method for structure modeling and design: applications to docking and structure prediction from the distorted native conformation. *J Comp Chem* 1994;15:488–506.
3. Rice LM, Brunger AT. Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins* 1994;19:277–290.
4. Evans JS, Mathiowetz AM, Chan SI, Goddard WA, III. De novo prediction of polypeptide conformations using dihedral probability grid Monte Carlo methodology. *Protein Sci* 1995;4:1203–1216.
5. Rooman MJ, Kocher J-PA, Wodak SJ. Prediction of protein backbone conformation based on seven structure assignments: Influence of local interactions. *J Mol Biol* 1991;221:961–979.
6. Park BH, Levitt M. The complexity and accuracy of discrete state models of protein structure. *J Mol Biol* 1995;249:493–507.
7. de laCruz XF, Mahoney MW, Lee B. Discrete representations of the protein c_α chain. *Fold Des* 1997;2:223–234.
8. Yang Y, Liu H. Genetic algorithms for protein conformation sampling and optimization in a discrete backbone dihedral angle space. *J Comput Chem* 2006;27:1593–1602.
9. Kang HS, Kurochkina NA, Lee B. Estimation and use of protein backbone angle probabilities. *J Mol Biol* 1993;229:448–460.
10. Bystroff C, Thorsson V, Baker D. HMMSTR: a hidden Markov model for local sequence-structure correlations in proteins. *J Mol Biol* 2000;301:173–190.
11. deBevern AG, Etchebest C, Hazout S. Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. *Proteins* 2000;41:271–287.
12. Karchin R, Cline M, Mandel-Gutfreund Y, Karplus K. Hidden Markov models that use predicted local structure for fold recognition: alphabets of backbone geometry. *Proteins* 2003;51:504–514.
13. deBevern AG, Benros C, Gautier R, Valadie H, Hazout S, Etchebest C. Local backbone structure prediction of proteins. *In Silico Biol* 2004;4:31.
14. Kuang R, Leslie CS, Yang A-S. Protein backbone angle prediction with machine learning approaches. *Bioinformatics* 2004;20:1612–1621.

15. Mooney C, Vullo A, Pollastri G. Protein structural motif prediction in multidimensional ϕ - ψ space leads to improved secondary structure prediction. *J Comput Biol* 2006;13:1489–1502.
16. Zimmermann O, Hansmann UHE. Support vector machines for prediction of dihedral angle regions. *Bioinformatics* 2006;22:3009–3015.
17. Huang YM, Bystroff C. Improved pairwise alignments of proteins in the twilight zone using local structure predictions. *Bioinformatics* 2006;22:413–422.
18. Wood MJ, Hirst JD. Protein secondary structure prediction with dihedral angles. *Proteins* 2005;59:476–481.
19. Dor O, Zhou Y. Real-SPINE: An integrated system of neural networks for real-value prediction of protein structural properties. *Proteins* 2007;68:76–81.
20. Ahmad S, Gromiha MM, Sarai A. Real value prediction of solvent accessibility from amino acid sequence. *Proteins* 2003;50:629–635.
21. Chou PY, Fasman GD. Empirical predictions of protein conformation. *Annu Rev Biochem* 1978;47:251–276.
22. Siew N, Elofsson A, Rychlewski L, Fischer D. Maxsub: an automated measure for the assessment of protein structure prediction quality. *Bioinformatics* 2000;16:776–785.
23. Rychlewski L, Fischer D. LiveBench-8: the large-scale, continuous assessment of automated protein structure prediction. *Protein Sci* 2005;14:240–245.
24. Zhou H, Zhou Y. Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. *Proteins* 2005;58:321–328.
25. Liu S, Zhang C, Liang S, Zhou Y. Fold recognition by concurrent use of solvent accessibility and residue depth. *Proteins* 2007;68:636–645.
26. Rose GD, Fleming PJ, Banavar JR, Maritan A. A backbone-based theory of protein folding. *Proc Natl Acad Sci USA* 2006;103:16623–16633.
27. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M. Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci USA* 2007;104:9615–9620.
28. Kabsch W, Sander C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;22:2577–2637.
29. Dor O, Zhou Y. Achieving 80% ten-fold cross-validated accuracy for secondary structure prediction by large-scale training. *Proteins* 2007;66:838–845.