

A Novel Prediction Method of Protein Structural Classes Based on Protein Super-Secondary Structure

Longlong Liu, Jing Cui*, Jie Zhou

Department of Mathematical Sciences, Ocean University of China, Qingdao, China

Email: *cuijing186@hotmail.com, xinxijishu@ouc.edu.cn

How to cite this paper: Liu, L.L., Cui, J. and Zhou, J. (2016) A Novel Prediction Method of Protein Structural Classes Based on Protein Super-Secondary Structure. *Journal of Computer and Communications*, 4, 54-62.
<http://dx.doi.org/10.4236/jcc.2016.415005>

Received: September 26, 2016

Accepted: November 25, 2016

Published: November 28, 2016

Abstract

At present, the feature extraction of protein sequences is the most basic issue to predict protein structural classes and is also the key problem to decide the quality of prediction. In order to predict protein structural classes accurately, we construct a 14-dimensional feature vector based on protein secondary and super-secondary structure information to reflect the content and spatial ordering of the given protein sequences. Among the vector, seven features about α -helix bundle, hairpin β motifs, Rossman folds, $\alpha\beta$ -plaits and other super-secondary structure information are first proposed in our paper. Experiments show that our method improves overall accuracy of lower similarity datasets 1189 and 640 by 0.9% - 3.8% and 0.5% - 4.2% respectively compared with other methods and has a competitive advantage for predicting proteins in α / β and $\alpha + \beta$ classes.

Keywords

Proteins, Super-Secondary Structure, Low Similarity, New Features

1. Introduction

Modern molecular biological studies indicate that the function of a protein is determined by its spatial structure. Therefore, it's very important to predict the structural classes of the newly discovered protein accurately [1]. The types and order of 20 kinds of amino acids are the basic information in a protein sequence, then a large number of initial prediction methods use features based on amino acid composition (AAC) and the position information, which are the easiest and most intuitive methods [2] [3] [4] [5]. However, these methods have advantage to predict protein sequences with high

degree of similarity and have disadvantage to predict protein sequences with similarity less than 40%. Due to the low-similarity protein sequences always have high-similarity secondary structural contents and spatial arrangements, so a large number of researchers tried to extract features from the secondary structure of proteins predicted by PSI-PRED[6]. Under the guidance of this idea, SPRED model [2] and MODAS model [7] were constructed. Currently novel computational prediction methods [8] [9] build feature vectors by using the protein secondary structure information and protein sequences are predicted into four classes (All- α , All- β , α/β , $\alpha+\beta$) using SVM (support vector machine) classifier. Overall prediction accuracy on several datasets of these methods reach to 80% - 90%, but the prediction of α/β and $\alpha+\beta$ classes is not ideal, especially for $\alpha+\beta$ class with accuracy just about 70%.

In order to improve the prediction accuracy of α/β and $\alpha+\beta$ classes, we will extract seven different features reflecting general contents and spatial arrangements of the secondary structural elements from super-secondary structure of a given protein sequence. We use SVM to predict protein structural classes after features are extracted. Finally, our paper evaluated our model objectively.

2. Materials and Methods

2.1. Materials

Currently proposed methods widely use low-similarity benchmark datasets named 1189, FC699, 640 and 25PDB. The sequence similarity of 25PDB [10], 1189 [11], FC699 [1] and 640[12] are lower than 25%, 40%, 40% and 25% respectively. In order to compare with other methods, we choose 25PDB dataset as training set for SVM classifier, while the other datasets 1189, FC699 and 640 are test sets.

2.2. Feature Vector

Nowadays, a lot of methods are used to predict amino acids sequences into secondary structural sequences (SSS) constructed by three secondary structural elements, α -helix (H), β -strand (E), and random coil (C). Firstly, we obtain corresponding secondary structural sequences by PSI-PRED (version 2.6) [6]. It's difficult to distinguish α/β and $\alpha+\beta$ classes for both of them contain α -helices and β -strands. α -helices and β -strands are usually separated in α/β class, while they are usually interspersed in $\alpha+\beta$ class. In order to better represent the distribution of α -helices and β -strands, every segment H, E and C in secondary structure sequence (SSS) is replaced by α , β and ζ respectively, the new sequence is secondary structure segment sequence (SS), all element ζ are removed from SS to form a new sequence represented by SSW[13].

Our novel method mapped each protein sequence into a 14-dimensional vector that can be defined as

$$P = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7, \varphi_8, \varphi_9, \varphi_{10}, \varphi_{11}, \varphi_{12}, \varphi_{13}, \varphi_{14}\}^T \quad (1)$$

where T is a transpose symbol, φ_i ($i = 1, 2, \dots, 14$) is one of the features in feature

vector P . The elements $\{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_9, \varphi_{10}\}$ of Equation (1) are based on super-secondary structure information and are first proposed in this paper. What's more, φ_{3-5} are proposed specially to make a distinction between α/β and $\alpha+\beta$ classes. A number of secondary structure elements interact with each other and form a regular combination of secondary structure, which acts as a structural member of tertiary structure in much protein and is known as super-secondary structure (motif). Some super-secondary structures are related to specific functions and there are three basic forms of combination: $\alpha\alpha$, $\beta\alpha\beta$ and $\beta\beta$. In order to tap the structural characteristics of each class, we extract several typical features of folds and combinations.

1) The easiest $\beta\beta$ structure is hairpin β motif which is connected by a short loop. Multiple hairpin β motifs together will form a stable and widespread β -turns, so extracting the number of hairpin β motifs (defined as $Con_{\beta\beta}$) are very meaningful. Super-secondary structure $\alpha\alpha$ is a α helix bundle and is often formed by two intertwined spiral parallel or ant parallel α -helices. Structure $\beta\alpha\beta\alpha\beta$ named Rossman folds is one of the most special structures of α/β class. So we extract the number of super-secondary structure $\alpha\alpha$ ($Con_{\alpha\alpha}$) and $\beta\alpha\beta\alpha\beta$ ($Con_{\beta\alpha\beta\alpha\beta}$) as features. Then the corresponding features can be defined as $\varphi_1 = Con_{\beta\beta} / N_2$, $\varphi_2 = Con_{\alpha\alpha} / N_3$, $\varphi_3 = Con_{\beta\alpha\beta\alpha\beta} / N_3$.

2) In proteins of α/β class, β -strands and α -helices are parallel, α -helix and β -sheets appear alternatively, such as $\alpha-\beta-\alpha-\beta-\alpha\dots$. Proteins of α/β class are two types. Such as " $\alpha\beta$ -plaits" in which α -helices and β -strands appear alternatively, this characteristic is same to proteins in α/β class. In $\alpha\beta$ -plaits, the alternative form of α -helices and β -strands is $\dots\alpha-\beta-\beta-\alpha-\beta\dots$ or $\dots\alpha-\beta-\beta-\beta-\beta-\alpha-\beta-\beta\dots$. In another type α -helices and β -strands are separated. Therefore the number of occurrences of α segments (seg_α) and β (seg_β) segments is an obvious characteristic to distinguish proteins in α/β and $\alpha+\beta$ classes.

Features mentioned above can be represented as:

$$\varphi_4 = Maxseg_\alpha / N_3, \quad \varphi_5 = Maxseg_\beta / N_3$$

where $Maxseg_\alpha$ and $Maxseg_\beta$ are the maximal lengths of α segments and β segments in SSW.

3) In proteins of the α/β class, α -helices and β -strands alternate more frequently than in proteins of the $\alpha+\beta$ class. Based on this characteristic, we can design a feature $\varphi_6 = Altn / N_1$, where $Altn$ is the alternating frequency of α -helices and β -strands in SSS.

4) Because the length of the secondary structural segments will affect the assignment of the structural class, we define new features

$$\varphi_7 = Maxseg_H / N_1 \quad \text{and} \quad \varphi_8 = Maxseg_E / N_1$$

where $Maxseg_H$ and $Maxseg_E$ are the maximal lengths of α -helix and β -strand segments in SSS.

5) Position information of SS is also the deciding factor. Herein, the position of a segment is defined as a starting position of the segment. The corresponding features can be defined as

$$\varphi_9 = \sum_{i=1}^{Con_{seg_H}} P_{i\ seg_H} / N_1, \quad \varphi_{10} = \sum_{i=1}^{Con_{seg_E}} P_{i\ seg_E} / N_1$$

where $P_{i\ seg_H}$ ($P_{i\ seg_E}$) is the starting order of the α -helix (β -strand) segment in SSS, Con_{seg_H} and Con_{seg_E} are the occurrences of α -helix and β -strand segments in SSS.

6) To reflect the position information of protein secondary structure, two features $\varphi_{11}, \varphi_{12}$ can be defined as

$$\varphi_{11} = \sum_{j=1}^{ConH} P_{Hj} / (N_1(N_1 - 1)), \quad \varphi_{12} = \sum_{j=1}^{ConE} P_{Ej} / (N_1(N_1 - 1))$$

where P_{Hj} and P_{Ej} are the j -th order of H and E in SSS, $ConH$ and $ConE$ are the number of H and E in protein secondary structure sequence (SSS).

7) The probability of content C can be ignored due to the sum of the three probabilities of H, E and C is 1 [1]. Hence, the two features are expressed as

$$\varphi_{13} = P(H), \quad \varphi_{14} = P(E)$$

where $P(H) = ConH / N_1$, $P(E) = ConE / N_1$.

2.3. Classification Algorithm Construction

Protein secondary structure prediction is a multiclass classification problem. With high prediction accuracy, support vector machine (SVM) has been widely used for protein secondary structure classification [4] [5]. Here we use of the "one to one" multiclass classification method that construct a multiclass classifier by combining six binary classifiers. We choose Gaussian radial basis function (RBF) as the kernel function for SVM [14]. Using a grid search on the training set (25PDB) by tenfold cross-validation, we can find out the penalty parameter C and kernel parameter γ , the final parameters are $C = 80$, $\gamma = 0.8$.

2.4. Performance Measures

In this paper, we use an independent testing dataset cross-validation. There are many indicators to evaluate model's performance, sensitivity (Sens), specificity (Spec), Matthew's correlation coefficient (MCC) and overall accuracy (OA) are widely used in protein structure prediction [15]. The total number of proteins, classes and proteins in k -th class are denoted by N , k and N_k respectively, so $N = \sum_{k=1}^{\mu} N_k$. Usually, four parameters are used by studies for examining a predictor's effectiveness:

The number of proteins which is correctly predicted as k th class and non- k th class are denoted by TP_k and TN_k . The number of proteins which is incorrectly predicted as k th class and non- k th class are denoted by FP_k and FN_k . Where $TP_k + FN_k = N_k$, $TN_k + FP_k = N - N_k$. Using these parameters, we can obtain Equation (2):

$$\begin{cases} Sens_k = TP_k / (TP_k + FN_k) \\ Spec_k = TN_k / (TN_k + FP_k) \\ MCC_k = (TP_k \times TN_k - FP_k \times FN_k) / \sqrt{(TP_k + FP_k)(TP_k + FN_k)(TN_k + FP_k)(TN_k + FN_k)} \\ OA = \sum_{k=1}^{\mu} TP_k / N \end{cases} \quad (2)$$

3. Results and Discussion

3.1. Structural Class Prediction Accuracies

In our experiment, we use 25PDB dataset as a training set and other three datasets as testing sets. We not only report the values of Sens, Spec, MCC and overall accuracy (OA) of every structural class of testing set, but also report the average of Sens, Spec, MCC and overall accuracy (OA). The detail results can be seen in **Table 1**. The overall accuracy is more than 84% for each test set and it reaches 90% for FC699 dataset. What's more, the average overall accuracy of 3 test sets is up to 86.6%. The Sens and

Table 1. The prediction quality of our method on test datasets.

Dataset	Class	Sens (%)	Spec (%)	MCC (%)
FC699	All- α	96.9	98.9	94.6
	All- β	89.2	98.3	89.4
	α / β	89.4	98.8	89.3
	$\alpha + \beta$	89.0	92.5	66.7
	OA	90.4	-	-
1189	All- α	91.9	96.9	87.5
	All- β	87.8	96.9	85.7
	α / β	80.2	95.9	78.7
	$\alpha + \beta$	79.3	89.8	65.7
	OA	84.4	-	-
640	All- α	92.0	98.2	90.7
	All- β	86.4	95.7	82.0
	α / β	88.1	95.0	82.9
	$\alpha + \beta$	74.9	90.8	65.7
	OA	85.0	-	-
Average	All- α	93.6	98.0	91.0
	All- β	87.8	97.0	85.7
	α / β	85.9	96.6	83.6
	$\alpha + \beta$	81.0	91.0	66.0
	OA	86.6	-	-

MCC values of $\alpha + \beta$ class are the lowest. This is not strange because $\alpha + \beta$ class is complicated and it not only contains α and β classes but also contains $\alpha\beta$ -plaits structure, so the prediction accuracy of $\alpha + \beta$ class is lower.

3.2. Feature Vector Analysis

To better verify the effect of the new proposed seven features, we do the following experiment with FC699, 1189 and 640 datasets. The comparison of obtained accuracies between our method including 14 features and our method including 7 features can be seen in **Table 2**. After added new features, the average overall accuracy increases by 2.6% up to 86.6%. For FC699 dataset, the overall accuracy and accuracies of All- α , All- β and α / β classes are improved by 2.7%, 1.5%, 4.4% and 2.7% respectively. For 1189 and 640 datasets, the overall accuracy increases by 2.6% and 2.3%, respectively. However, the results of α / β and $\alpha + \beta$ classes are not obvious because of the interference of other classes [13].

To further validate effect of super-secondary structure features, we do experiment just on proteins in α / β and $\alpha + \beta$ classes with a 14-dimensional feature vectors. The 25PDBS, FC699S, 1189S and 640S sets are the subsets formed by removing all the proteins in the All- α and All- β classes from 25PDB, FC699, 1189 and 640 datasets respectively. Hence, we use the 25PDBS to train SVM classifier and other subsets to test. The parameters C and γ ($C = 60$, $\gamma = 0.9$) are selected by tenfold cross-validation on 25PDBS with a grid search method. The corresponding experimental results are shown in **Table 3**. In **Table 3**, the overall accuracies of all datasets predicted by our method are higher than 80%. The overall accuracies and accuracies of $\alpha + \beta$ class predicted by our method are the highest compared with other competitive methods. The prediction accuracies of all structural classes are higher than 90% on FC699S subset. The accuracies of $\alpha + \beta$ class and the overall accuracy are increased by 9.2% - 27.4% and 2.4% - 7.3% on 1189S respectively. For 640S, the accuracies of $\alpha + \beta$ class and the overall accuracy are improved by 2.9% - 11.7% and 0.6% - 5.2%

Table 2. Comparison of the accuracies between the method including 14 features and one including only 7 features.

Dataset	Features	Accuracy (%)				
		All- α	All- β	α / β	$\alpha + \beta$	Overall
FC699	All features included	96.9	89.2	89.4	89.0	90.4
	Novel features excluded	95.4	84.8	86.7	89.0	87.7
1189	All features included	91.9	87.8	80.2	79.3	84.4
	Novel features excluded	91.0	85.7	77.5	74.3	81.8
640	All features included	92.0	86.4	88.1	74.9	85.0
	Novel features excluded	89.1	84.4	84.8	73.7	82.7
Average	All features included	93.6	87.8	85.9	81.0	86.6
	Novel features excluded	91.9	85.0	83.0	79.0	84.0

Table 3. The accuracy of differentiating between the α / β and $\alpha + \beta$ class.

Dataset	Method	Reference	Accuracy (%)		
			α / β	$\alpha + \beta$	Overall
FC699S	Kong <i>et al.</i>	Kong <i>et al.</i> (2013)	98.4	79.3	90.5
	Our method	This paper	90.7	96.3	91.7
1189S	SCPRED	Kurgan <i>et al.</i> (2008b)	88.6	63.1	77.9
	PKS-PPSC	Yang <i>et al.</i> (2010)	83.8	81.3	82.8
	Ding <i>et al.</i>	Ding <i>et al.</i> (2012)	85.9	73.9	80.9
	Kong <i>et al.</i>	Kong <i>et al.</i> (2013)	84.7	77.2	81.6
	Our method	This paper	81.4	90.5	85.2
	SCPRED	Kurgan <i>et al.</i> (2008b)	89.3	77.2	83.3
640S	PKS-PPSC	Yang <i>et al.</i> (2010)	88.1	83.6	85.9
	Ding <i>et al.</i>	Ding <i>et al.</i> (2012)	89.3	83.6	86.5
	Kong <i>et al.</i>	Kong <i>et al.</i> (2013)	89.8	86.0	87.9
	Our method	This paper	88.1	88.9	88.5
	SCPRED	Kurgan <i>et al.</i> (2008b)	89.3	77.2	83.3

respectively, however, the accuracy of α / β class is consistent with the result produced by PKS-PPSC model. Experimental results show that our method for predicting proteins in α / β and $\alpha + \beta$ classes is very effective.

3.3. Comparison with Other Prediction Method

It's known to all, SCPRED [2] and MODAS [7] are famous in predicting protein secondary structure and are often used as baseline for comparison. From **Table 4** we can see, our method improves the overall accuracies by 0.9% - 3.8% and 0.5% - 4.2% on 1189 and 640 datasets compared with other competing prediction methods including SCPRED and MODAS. And only for FC699 dataset, the overall accuracy is lower than Kong *et al.* and is not the highest, but it is increased by 0.8% - 2.9% compared with the rest methods. Compared with model SCPRED and the method of Liu and Jia, the overall accuracy predicted by our method is improved by 2.9% and 0.8% on FC699 dataset, respectively, besides the accuracy of $\alpha + \beta$ class is increased by 4.8% and 19.5%. Our method obtains the highest accuracies for the All- β and $\alpha + \beta$ classes which reach to 87.8% and 79.3% on 1189 dataset and the overall accuracy is the highest than other exiting methods. For 640 dataset, the overall accuracy and the accuracy of All- β class are the highest.

Therefore our method extracting features based on super-secondary structure has the ability to reflect the realistic characteristics of proteins more accurately. Specially, our method improved the accuracies of α / β and $\alpha + \beta$ classes greatly. According **Table 4**, we find our method is not always the best. The reason is that some methods not only extract features based on protein secondary structure but also combine other information. In contrast, our method is aimed to predict secondary structural classes by extracting features more effectively just on the basis of secondary structure.

Table 4. Performance comparison of difference methods on 3 test datasets.

Dataset	Method	Reference	Accuracy (%)				Overall
			All- α	All- β	α / β	$\alpha + \beta$	
FC699	SCPRED	Kurgan <i>et al.</i> (2008b)	-	-	-	-	87.5
	Liu and Jia	Liu and Jia(2010)	97.7	88.0	89.1	84.2	89.6
	Kong <i>et al.</i>	Kong <i>et al.</i> (2013)	96.2	90.7	96.3	69.5	92.0
	Our method	This paper	96.9	89.2	89.4	89.0	90.4
1189	SCPRED	Kurgan <i>et al.</i> (2008b)	89.1	86.7	89.6	53.8	80.6
	MODAS	Mizianty <i>et al.</i> (2009)	92.3	87.1	87.9	65.4	83.5
	RKS-PPSC	Yang <i>et al.</i> (2010)	89.2	86.7	82.6	65.6	81.3
	Zhang <i>et al.</i>	Zhang <i>et al.</i> (2011)	92.4	87.4	82.0	71.0	83.2
	Ding <i>et al.</i>	Ding <i>et al.</i> (2012)	93.7	84.0	83.5	66.4	82.0
	Kong <i>et al.</i>	Kong <i>et al.</i> (2013)	91.9	84.4	85.3	72.2	83.5
	Our method	This paper	91.9	87.8	80.2	79.3	84.4
640	SCPRED	Kurgan <i>et al.</i> (2008b)	90.6	81.8	85.9	66.7	80.8
	RKS-PPSC	Yang <i>et al.</i> (2010)	89.1	85.1	88.1	71.4	83.1
	Ding <i>et al.</i>	Ding <i>et al.</i> (2012)	94.9	76.6	89.3	74.3	83.4
	Kong <i>et al.</i>	Kong <i>et al.</i> (2013)	94.2	80.5	87.6	77.2	84.5
	Our method	This paper	92.0	86.4	88.1	74.9	85.0

4. Conclusion

In this paper, a novel method is proposed based on protein super-secondary structure information. Seven new features related to α -helix bundle, hairpin β motifs, Rossmann folds, $\alpha\beta$ -plaits and other information are very useful to predict protein secondary structural classes. We adopt advanced SVM classifier which use little computational time and space, is accurate and is very suitable for large-scale protein sequence databases. Finally, experimental results show that this new prediction method not only improve the overall prediction accuracy but also improve the accuracies of all structural classes, especially, the accuracies of α / β and $\alpha + \beta$ classes are improved greatly. Hence, the new extracted features can reflect the characteristics of different structural classes more accurately and our method is more effective than previous methods.

Acknowledgements

The authors would like to thank all of the researchers who made publicly available data used in this study and thank the National Natural Science Foundation of China (No: 61303145) for the support to this work.

References

- [1] Kurgan, L.A., Zhang, T., Zhang, H., Shen, S. and Ruan, J. (2008) Secondary Structure-Based Assignment of the Protein Structural Classes. *Amino Acids*, **35**, 551-564.

- <http://dx.doi.org/10.1007/s00726-008-0080-3>
- [2] Kurgan, L., Cios, K. and Chen, K. (2008) Scpred: Accurate Prediction of Protein Structural Class for Sequences of Twilight-Zone Similarity with Predicting Sequences. *BMC Bioinformatics*, **9**, 815-818. <http://dx.doi.org/10.1186/1471-2105-9-226>
- [3] Olyaei, M.H., Yaghoubi, A. and Yaghoobi, M. (2016) Predicting Protein Structural Classes Based on Complex Networks and Recurrence Analysis. *Journal of Theoretical Biology*, **404**, 375-382. <http://dx.doi.org/10.1016/j.jtbi.2016.06.018>
- [4] Zhang, S. (2015) Accurate Prediction of Protein Structural Classes by Incorporating PSSS and PSSM into Chou's General PseAAC. *Chemometrics & Intelligent Laboratory Systems*, **142**, 28-35. <http://dx.doi.org/10.1016/j.chemolab.2015.01.004>
- [5] Ding, S., Yan, L., Shi, Z. and Yan, S. (2014) A Protein Structural Classes Prediction Method Based on Predicted Secondary Structure and Psi-Blast Profile. *Biochimie*, **97**, 60-65. <http://dx.doi.org/10.1016/j.biochi.2013.09.013>
- [6] Jones, D.T. (1999) Protein Secondary Structure Prediction Based on Position-Specific Scoring Matrices. *Journal of Molecular Biology*, **292**, 195-202. <http://dx.doi.org/10.1006/jmbi.1999.3091>
- [7] Mizianty, M.J. and Lukasz, K. (2009) Modular Prediction of Protein Structural Classes from Sequences of Twilight-Zone Identity with Predicting Sequences. *BMC Bioinformatics*, **10**, 1-24. <http://dx.doi.org/10.1186/1471-2105-10-414>
- [8] Liu, T. and Jia, C. (2010) A High-Accuracy Protein Structural Class Prediction Algorithm Using Predicted Secondary Structural Information. *Journal of Theoretical Biology*, **267**, 272-275. <http://dx.doi.org/10.1016/j.jtbi.2010.09.007>
- [9] Zhang, S., Ding, S. and Wang, T. (2011) High-Accuracy Prediction of Protein Structural Class for Low-Similarity Sequences Based on Predicted Secondary Structure. *Biochimie*, **93**, 710-714. <http://dx.doi.org/10.1016/j.biochi.2011.01.001>
- [10] Kurgan, L.A. and Homaeian, L. (2006) Prediction of Structural Classes for Protein Sequences and Domains—Impact of Prediction Algorithms, Sequence Representation and Homology, and Test Procedures on Accuracy. *Pattern Recognition*, **39**, 2323-2343. <http://dx.doi.org/10.1016/j.patcog.2006.02.014>
- [11] Wang, Z. and Zheng, Y. (2000) How Good Is Prediction of Protein Structural Class by the Component-Coupled Method? *Proteins Structure Function & Bioinformatics*, **38**, 165-175. [http://dx.doi.org/10.1002/\(SICI\)1097-0134\(20000201\)38:2<165::AID-PROT5>3.0.CO;2-V](http://dx.doi.org/10.1002/(SICI)1097-0134(20000201)38:2<165::AID-PROT5>3.0.CO;2-V)
- [12] Yang, J.Y., Peng, Z.L. and Xin, C. (2010) Prediction of Protein Structural Classes for Low-Homology Sequences Based on Predicted Secondary Structure. *BMC Bioinformatics*, **11**, 1-10. <http://dx.doi.org/10.1186/1471-2105-11-s1-s9>
- [13] Liang, K., Zhang, L. and Lv, J. (2014) Accurate Prediction of Protein Structural Classes by Incorporating Predicted Secondary Structure Information into the General Form of Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, **344**, 12-18. <http://dx.doi.org/10.1016/j.jtbi.2013.11.021>
- [14] Zheng, Y., Bailey, T.L. and Teasdale, R.D. (2005) Prediction of Protein b-Factor Profiles. *Proteins Structure Function & Bioinformatics*, **58**, 905-912. <http://dx.doi.org/10.1002/prot.20375>
- [15] Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A. and Nielsen, H. (2000) Assessing the Accuracy of Prediction Algorithms for Classification: An Overview. *Bioinformatics*, **16**, 412-424. <http://dx.doi.org/10.1093/bioinformatics/16.5.412>



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact jcc@scirp.org