

Using Fuzzy Support Vector Machine Network to Predict Low Homology Protein Structural Classes

Tongliang Zhang¹, Rong Wei³, and Yongsheng Ding^{1,2}

¹ College of Information Sciences and Technology
tl.zhang@mail.dhu.edu.cn

² Engineering Research Center of Digitized Textile & Fashion Technology,
Ministry of Education
ysding@dhu.edu.cn

Donghua University, Shanghai 201620, P.R. China

³ College of Sciences, Hebei Polytechnic University,
Hebei Tangshan 063009, P.R. China
wr@heut.edu.cn

Abstract. Prediction of protein structural classes for low homology proteins is a challenging research task in bioinformatics. A dual-layer fuzzy support vector machine (FSVM) network approach is proposed to predict protein structural classes. A protein sample can be represented by nine representation feature vectors: pair couple amino acid (210-D) and eight pseudo amino acid composition vectors (PseAAC). Eight physicochemical properties of amino acids extracted from AAIndex databank are used to calculate low frequencies of power spectrum density of sequence-order correlation in protein sequence. In the first layer of FSVM network, nine FSVM classifiers are established, which are trained by different protein feature vectors, respectively. The outputs of the first layer are reclassified by FSVM classifier in 2nd layer of the network. The performance of proposed method is validated by low homology (average 25%) dataset covering 1673 proteins. The promising results indicate that the new method may become a useful tool for predicting not only the structural classification of proteins but also their other attributes.

1 Introduction

In structural classification of proteins databank (SCOP) [1-3], proteins are classified into seven structural classes: *all- α* , *all- β* , *$\alpha+\beta$* , *α/β* , multi-domain, small protein, and peptide. More than 80% proteins are deposited into the former four classes. Many efforts were focused on the four structural classes, i.e., *all- α* , *all- β* , *$\alpha+\beta$* , and *α/β* .

Numerous prediction methods for protein structural classes have been proposed based on the primary amino acid sequence [4-15], since the work of Klein and Delisi [4]. During the twenty years, the performance of these methods are increasing with the combination of new pattern recognize algorithms and effective protein sequence representation. Perfect accuracy rates (about 95%) have been achieved in some prediction methods. However, these methods were often tested on small datasets, and characterized by different homology of sequences. Kurgan and Homaeian [22] indicated that sequence homology in dataset have a significant impact on the predictive

accuracy. The best achieved prediction accuracy for low homology datasets is about 57%. Wang and Yuan [17] have stated that the prediction method should aim only at proteins with lower 30% homology. So, it is crucial to develop the prediction methods or algorithms for structural classes of protein with lower homology.

Several studies have testified that the performance of ensemble machine learning approaches is superior to individual learning algorithm [7, 18-21]. Recently, the methods of ensembles have been used in this area. Kedarisetti et al. [16] established an ensembles method with heterogeneous classifiers validating on the datasets of varying homology. Chen et al. [7] developed support vector machine fusion network algorithm.

Compared with amino acid composition frequently used in prediction methods of protein structural classes, pseudo amino acid composition (PseAAC) as introduced by Chou [23] can incorporate more information and remarkable enhance prediction performance in various attribute of protein. In this study, a sample of proteins is represented by nine kinds of feature vectors, including pair couple amino acid composition (PcAA) (210-D) and eight PseAAC feature vector. Eight physicochemical properties extracted from AAIndex database [24] are used to calculate sequence-order correlation that introduced by Chou [25]. Low frequencies of power spectrum density of different sequence-order effect are used to construct PseAAC. A dual-layer fuzzy support vector machine (FSVM) network is used as prediction engine. The low homology dataset 25PDB, constructed by Kurgan et al. [22], is applied to verify the new method. Promising results obtained on self-consistent and jackknife cross-validating test methods show that it is effective and practical.

2 Methods

2.1 Protein Sequence Representation

Pair-coupled amino acid composition (PcAA) attempts to extract the information of local order of amino acids in sequence. This concept has been used in protein secondary structure content prediction [7, 26, 27] and other attributes of protein prediction [25]. The PcAA is formulated as follows:

$$\mathfrak{R}_{210} = \left\{ \begin{array}{l} f(AA), f(AC), f(AD), \dots, f(AY) \\ f(CC), f(CD), \dots, f(CY) \\ f(DD), \dots, f(DY) \\ \dots \\ f(YY) \end{array} \right\} \quad (1)$$

where, $f(AC)$ is the sum of AC pair occurrence frequency and CA pair occurrence frequency in protein sequence. Thus, the pair-couple AA is a 210-D feature vector.

$$\vec{S}_1 = [x_1, x_2, \dots, x_{210}] \quad (2)$$

where, $x_1 = f(AA), x_2 = f(AC), \dots, x_{210} = f(YY)$. The 210-D (dimensional) vector is normalized to meet the condition that the sum is 1.

$$\sum_{i=1}^{210} x_i = 1 \quad (3)$$

Compared with conventional protein composition (20-D), the concept of PseAAC as originally introduced by Chou [23], which is defined in a $(20 + \lambda)$ -D features space, will contain much more sequence-order information. A protein sample can be represented by $(20 + \lambda)$ -D vectors, where the λ is the number of additional properties of sequence.

$$x_i = \begin{cases} \frac{f_i}{\sum_{j=1}^{20} f_j + w \sum_{j=1}^{\lambda} p_j} & (1 \leq i \leq 20) \\ \frac{w p_i}{\sum_{j=1}^{20} f_j + w \sum_{j=1}^{\lambda} p_j} & (21 \leq i \leq 20 + \lambda) \end{cases} \quad (4)$$

where, the f_i is the normalized occurrence frequency of the 20 amino acid in the protein, that is amino acid composition (AA), and the P_i is the additional properties of protein sequence. w is weight factor of additional characteristics. In this study, the properties of protein sequence are the low frequencies of power spectrum density of sequence-order effect that introduced by Chou [25]. Protein sequence with N residues can be written as $R = R_1 R_2 \dots R_N$. Protein sequence order effect can be reflected through Eq. (5). It is actually the same as Eq. (2) of Chou [25].

$$\tau_m = \frac{1}{L-m} \sum_{i=1}^{L-m} J_{i,i+m} \quad (5)$$

where, $J_{i,i+m}$ is the correlation factor of residue i and $i+m$. m is the distance of two residues have correlation in sequence. In this study, $J_{i,i+m}$ is defined as the product of physicochemical properties of two residues, see Eq. (6), which is actually the same as Eq. (3) of Chou[25].

$$J_{i,i+\lambda} = h(R_i)h(R_{i+m}) \quad (6)$$

Before substituting the properties into Eq. (6), they are normalized according to the Eq. (7)

$$h(R_i) = \frac{h_0(R_i) - \bar{h}_0}{SD(h_0)} \quad (7)$$

In Eq. (7), $h_0(R_i)$ ($i=1,2,\dots,20$) are the original physicochemical properties of 20 amino acids. \bar{h}_0 denotes that the average property values of 20 amino acids. $SD(h_0)$ presents the standard deviation.

We use standard function 'PWELCH' in Matlab 7.0 environment to calculate power spectral density of the sequence-order effect. Based on the theory of digital signal processing, the high-frequency components are more noisies, and hence only the

low-frequency components are more important. Similarly, low frequencies of Fourier transform of protein sequence-order correlation have been used in prediction of membrane protein types [28, 29]. In our previous work [30], low frequencies of energy spectrum density of protein sequence-order correlation are used to construct PseAAC. This is just like the case of protein internal motions where the low-frequency components are functionally more important [31, 32].

In some prior works, hydrophobicity scale of amino acid was usually used to calculate sequence-order effect [7, 28-30, 33-35]. Except for hydrophobicity, some other physicochemical properties of amino acids are also important in the fold process, such as volume, polarity, average accessible surface area, and so on. Here, eight physicochemical properties extracted from AAIndex database [24] are used to compute the sequence-order correlation of protein sequence through Eq. (6) and (7), respectively. The physicochemical properties are listed in Table 1. Eight PseAAC vectors are obtained and are named as \vec{S}_i ($i = 2, 3, \dots, 9$).

Table 1. The eight physicochemical properties used in this work

No	AA Index	Description
1	PRAM900101	Hydrophobicity
2	COSI940101	Electron-ion interaction potential values
3	RADA880108	Mean polarity
4	PONJ960101	Average volumes of residues
5	KUHL950101	Hydrophilicity scale
6	JANJ790102	Transfer free energy
7	JANJ780101	Average accessible surface area
8	FAUJ880103	Normalized van der Waals volume

2.2 Fuzzy Support Vector Machine

Support vector machine is a typical binary-class classifier based on the statistic learning theory[36]. The task of protein structural classes' prediction is a four classes classification problem. There are many multi-classes SVMs methods to solve the problem, such as one against one, one against others, DAG, etc. However, there are some unclassifiable points still existing in these multi-class SVMs methods. FSVM algorithm as introduced by Abe [37] has capability to solve unclassifiable points effectively.

Compared with conventional SVM algorithm, membership function is defined in FSVM algorithm. When solving k classes classification task, $k(k-1)/2$ SVM classifiers have to be established with one against one method. Toward to the SVM classifier between class i and class j , the decision function of input vector x is

$$D_{ij} = w_{ij} \cdot x + b_{ij} \quad (8)$$

where w_{ij} is the m-D vector, b_{ij} is a scalar, and $D_{ij} = -D_{ji}$

For the input vector x , we assemble

$$D_i = \sum_{j \neq i, j=1}^n \text{sign}(D_{ij}(x)) \tag{9}$$

where,

$$\text{sign}(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{10}$$

For optimal separating hyperplane $D_{ij} = 0$ ($i \neq j$), the membership function m_{ij} is defined as below:

$$m_{ij} = \begin{cases} 1 & D_{ij}(x) \geq 1 \\ D_{ij}(x) & \text{otherwise} \end{cases} \tag{11}$$

We define the class i membership function of x using the minimum operator:

$$m_i(x) = \min_{j=1, \dots, n} m_{ij}(x) \tag{12}$$

The shape of the membership function is a truncated polyhedral pyramid. An unknown protein sequence is classified into the class with maximum membership value.

$$\arg \max_{i=1, \dots, n} m_i(x) \tag{13}$$

In this study, six binary-class SVMs have been developed through the method of one against one for solving protein structural classes prediction (four classes). LS-SVMLab1.5 toolbox [38] in MatLab environment is selected as binary-class SVM classifier which is capable of searching the fittest parameters in SVM automatically. The Radial Basis Function (RBF) kernel is used in SVM. The membership function is calculated according to the Eqs (8)-(12). The output of each FSVM classifier in the first layer is not a rigid class label but a 4-D vector. The vector indicates that the membership values of protein sample belong to four structural classes.

2.3 FSVM Network

The SVM fusion uses the FSVM classifier to reclassify the outputs from all sub-classifiers. The protein sample is predicted by FSVM 1 to FSVM 8, and the output

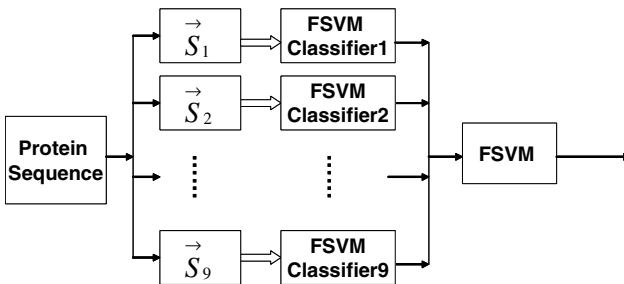


Fig. 1. The work procedure of FSVM network for prediction of protein structural classes

of the k -th FSVM ($1 \leq k \leq 9$) is represented as 4-D vector. Then the input feature vector is defined as $V = [V_1, V_2, \dots, V_9]$ to the FSVM fusion classifier for final decision. The process of the new predictive method is illustrated in Fig. 1.

2.4 Dataset and Measure Methods

The dataset constructed by Kurgan and Homaeian [22] is used to validate the performance of the new method, which includes proteins scanned with high resolution and with low on average 25% homology, named as 25PDB. The dataset contains 1673 proteins classified into four structural classes: 443 *all- α* , 443 *all- β* , 346 *α/β* , and 441 *$\alpha+\beta$* .

Three indexes are applied to evaluate the prediction accuracy, that is, sensitivity (S_n), specificity (S_p), and Mathew's correlation coefficient (MCC).

$$S_n = \frac{TP}{TP + FN} \quad (14)$$

$$S_p = \frac{TP}{TP + FP} \quad (15)$$

$$MCC = \frac{TP \times TN - FP \times FN}{[(TP + FN)(TP + FP)(TN + FN)(TN + FP)]^{1/2}} \quad (16)$$

where, TP (true positives) is the protein number of right prediction in a structure class, FN (false negatives) is the protein number of wrong prediction in a structure class, and FP (false positives) is the number of the proteins in other classes to be predicted in this class. TN (true negatives) is the number of proteins observed in other classes that are not predicted in this class. S_n represents the accuracy, and S_p represents the reliability in procedure of prediction. The MCC is a single parameter characterizing the matching extent between the observed and predicted structural classes. Self-consistency and jackknife test methods are used to test the performance of the new approach. The jackknife test is thought the most rigorous and objective one [see [39] for a comprehensive review in this regard], and hence has been used by more and more investigators in examining the power of various prediction methods.

3 Results and Discussion

In the first layer of the FSVM network, nine FSVM classifiers trained by different representation methods of protein are established. The three indexes values of each FSVM classier are tested under self-consistency and jackknife cross-validating test methods. The number of addition characteristics λ and factor parameter w in Eq. (4) are important in determining the performance of PseAAC. Different parameters $\{\lambda, w\}$ in Eq. (4) are tested, and the PseAAC feature vector is input into FSVM classifier. The parameters $\{\lambda, w\}$ are determined when the accuracy of jackknife test method is the highest. The jackknife test results of nine FSVM classifier trained by different PseAAC feature vectors are listed in Table 2. The highest accuracy is 55.9%, and the lowest is

Table 2. Performance of each FSVM classifier

Representation	\bar{S}_1	\bar{S}_2	\bar{S}_3	\bar{S}_4	\bar{S}_5	\bar{S}_6	\bar{S}_7	\bar{S}_8	\bar{S}_9
λ	-	11	11	7	9	15	11	20	3
w	-	0.15	0.05	0.02	0.1	0.1	0.1	0.15	0.2
Accuracy (%)	32.7	53.2	55.6	53.4	47.1	55.9	50.6	43.9	53.4

32.7%. The outputs of all FSVM classifiers are combined into a vector, and used as input of the FSVM classifier in the second layer of the network.

The prediction results of four structural classes on self-consistency and jackknife test methods are listed in Table 3. The three indexes are calculated respectively. The overall accuracy of self-consistency is 91.87%, and that of jackknife test is 62.9%. Class *all- β* achieves the highest accuracy (79.9%) in four structural classes, and that of class *all- α* is more than 70%. However, accuracy of class *α/β* is 36.7%, and that of class *$\alpha+\beta$* is 56%. The *Sp* value of class *all- α* is 53.9%. It demonstrate that some proteins in other structural class are incorrect classified into class *all- α* . According to the concept of Levitt and Chothia [30], it is true that the class *$\alpha+\beta$* and *α/β* are more complex than class *all- α* and *all- β* . It might be the reason to explain the phenomenon that accuracies of class *all- α* and class *all- β* are higher than that of class *$\alpha+\beta$* and class *α/β* .

Table 3. Results on self-consistency and jackknife test methods

Structural classes	Self-consistency test			Jackknife test		
	S_n	S_p	MCC	S_n	S_p	MCC
<i>all-α</i>	95.7%	85.3%	0.86	73.1%	53.9%	0.59
<i>all-β</i>	87.8%	95.3%	0.88	79.9%	75.5%	0.53
<i>α/β</i>	87.6%	94.1%	0.88	36.7%	52.0%	0.32
<i>$\alpha+\beta$</i>	95.4%	94.4%	0.93	56.0%	68.8%	0.41
Total	1537/1673=91.87%			1052/1673=62.9%		

The results of proposed method are compared with that of others using the same dataset, 25PDB. The accuracy rates of self-consistency and jackknife test methods together with the protein representation are deposited in Table 4. In same measure method and validation dataset, proposed method obtains 4.2% and 3% improvement on jackknife test when compared with the results of Logistic regression method [22] and StackingC ensemble [16], respectively. Meanwhile, the highest accuracy is achieved on self-consistency test method compared with other two methods. StackingC ensemble algorithm uses four heterogeneous classifiers. Protein sequence is initially represented using comprehensive set of 122 features which is reduced to 34 features through application of several feature selection algorithms [16]. The proposed method use different protein representation and same classifier. FSVM classifier has capability to resolve unclassifiable region effectively. FSVM is a strong classifier with excellent classification performance. It is widely accepted that ensemble method can enhance classification performance than single classifier [31]. Multiple FSVM classifiers are combined into FSVM network where various physicochemical properties of amino

Table 4. Comparison with other prediction methods on same dataset

Classification algorithm	Representation	Accuracy rates (%)	
		Self-consistency	Jackknife
Logistic regression[22]	66 feature	62.2	57.1
StackingC ensemble[16]	34 feature	87.6	59.9
This paper	Multi-feature	91.87	62.9

acid are taken into account. The promising prediction results illuminate that the FSVM network incorporating various physicochemical properties of amino acid is effective and practical. It might become potential tool for prediction of protein structural classes and other attributes of protein.

4 Conclusions

A dual-layer FSVM network is established to predict protein structural classes. Protein sample is represented by nine kinds of feature vectors including pair-couple amino acid composition (210-D) and eight PseAAC vectors. Eight physicochemical properties extracted from AAIndex databank are used to calculate sequence-order correlation in protein sequence. Low-frequencies of power spectrum density of different sequence-order correlations are used to construct PseAAC vectors. In the first layer of FSVM network, nine FSVM classifiers are obtained which are trained by different feature representation methods, respectively. The low homology dataset (average about 25% homology) is applied to verify the proposed method. The results of jackknife test in different structural classes illuminate that accuracies of class all- α and class all- β are higher than class $\alpha+\beta$ and class α/β . The phenomenon meets the truth of more complexity existing in class $\alpha+\beta$ and class α/β . Compared with other two methods tested on same dataset, the proposed methods obtain the highest accuracy. Promising results show the new method is effective and practical. It might become potential tool for protein structural class and other attribute of proteins.

Acknowledgments

This work was supported in part by Program for New Century Excellent Talents in University from Ministry of Education of China (No. NCET-04-415), the Cultivation Fund of the Key Scientific and Technical Innovation Project from Ministry of Education of China (No. 706024), International Science Cooperation Foundation of Shanghai (No. 061307041), and Specialized Research Fund for the Doctoral Program of Higher Education from Ministry of Education of China (No. 20060255006).

References

1. Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C.: J. Mol. Biol. 247, 536–540 (1995)
2. Lo Conte, L., Brenner, S.E., Hubbard, T.J.P., Chothia, C., Murzin, A.: Nucl. Acid Res. 30(1), 264–267 (2002)

3. Andreeva, A., Howorth, D., Brenner, S.E., Hubbard, T.J.P., Chothia, C., Murzin, A.: Nucl. Acid Res. 32, D226–D229 (2004)
4. Klein, P., Delisi, C.: Prediction of Protein Structural Class from the Amino Acid Sequence. *Biopolymers* 25, 1659–1672 (1986)
5. Cai, Y.D., Liu, X.J., Xu, X., Zhou, G.P.: Support Vector Machines for Predicting Protein Structural Class. *BMC Bioinformatics* 2, 3–7 (2001)
6. Cao, Y., Liu, S., Zhang, L., Qin, J., Wang, J., Tang, K.: Prediction of Protein Structural Class with Rough Sets. *BMC Bioinformatics* 7, 20 (2006)
7. Chen, C., Zhou, X., Tian, Y., Zou, X., Cai, P.: Predicting Protein Structural Class with Pseudo Amino Acid Composition and Support Vector Machine Fusion Network. *Anal. Biochem.* 357, 116–121 (2006)
8. Chou, K.C., Cai, Y.D.: Predicting Protein Structural Class by Functional Domain Composition. *Biochemical and Biophysical Research Communications (Corrigendum: ibid., 2005, Vol.329, 1362)* 321, 1007–1009 (2004)
9. Du, Q.S., Jiang, Z.Q., He, W.Z., Li, D.P., Chou, K.C.: Amino Acid Principal Component Analysis (AAPCA) and Its Applications in Protein Structural Class Prediction. *Journal of Biomolecular Structure and Dynamics* 23, 635–640 (2006)
10. Feng, K.Y., Cai, Y.D., Chou, K.C.: Boosting Classifier for Predicting Protein Domain Structural Class. *Biochemical and Biophysical Research Communications* 334, 213–217 (2005)
11. Luo, R.Y., Feng, Z.P., Liu, J.K.: Prediction of Protein Structural Class by Amino Acid and Polypeptide Composition. *Eur. J. Biochem.* 269, 4219–4225 (2002)
12. Niu, B., Cai, Y.D., Lu, W.C., Zheng, G.Y., Chou, K.C.: Predicting Protein Structural Class with AdaBoost learner. *Protein & Peptide Letters* 13, 489–492 (2006)
13. Shen, H.B., Yang, J., Liu, X.J., Chou, K.C.: Using Supervised Fuzzy Clustering to Predict Protein Structural Classes. *Biochemical and Biophysical Research Communications* 334, 577–581 (2005)
14. Sun, X.D., Huang, R.B.: Prediction of Protein Structural Classes Using Support Vector Machines. *Amino Acids* 30, 469–475 (2006)
15. Xiao, X., Shao, S.H., Huang, Z.D., Chou, K.C.: Using Pseudo Amino Acid Composition to Predict Protein Structural Classes: Approached with Complexity Measure Factor. *Journal of Computational Chemistry* 27, 478–482 (2006)
16. Kedariseti, K.D., Kurgan, L., Dick, S.: Classifier Ensemble s for Protein Structural Class Prediction with Varying Homology. *Biochemical and Biophysical Research Communications* 348, 981–988 (2006)
17. Wang, Z.X., Yuan, Z.: How Good is the Prediction of Protein Structural Class by the Component-coupled Method? *Proteins* 38, 165–175 (2000)
18. Chou, K.C., Shen, H.B.: Hum-PLoc: A Novel Ensemble Classifier for Predicting Human Protein Subcellular Localization. *Biochem. Biophys. Res. Commun.* 347, 150–157 (2006)
19. Nanni, L., Lumini, A.: MppS: An Ensemble of Support Vector Machine Based on Multiple Physicochemical Properties of Amino Acids. *Eurocomputing* 69, 1688–1690 (2006)
20. Nanni, L., Lumini, A.: Ensemblator: An Ensemble of Classifiers for Reliable Classification of Biological Data. *Pattern Recognition Letters* 28, 622–630 (2007)
21. Peng, Y.H.: A Novel Ensemble Machine Learning for Robust Microarray Data Classification. *Computers in Biology and Medicine* 36, 553–573 (2006)
22. Kurgan, L., Homaeian, L.: Prediction of Structural Classes for Protein Sequences and Domain: Impact of Prediction algorithms, Sequence Representation and Homology, and Test Procedures on Accuracy. *Pattern Recognition* 39, 2323–2343 (2006)

23. Chou, K.C.: Prediction of Protein Structural Classes and Subcellular Locations. *Curr. Protein Peptide Sci.* 1, 171–208 (2000)
24. Kawashima, S., Ogata, H., Kanehisa, M.: AAindex: Amino Acid Index Database. *Nucleic Acids Res.* 27, 368–369 (1999)
25. Chou, K.C.: Prediction of Protein Cellular Attributes Using Pseudo Amino Acid Composition. *PROTEINS: Structure, Function, and Genetics* (Erratum: *ibid.*, 2001, Vol.44, 60) 43, 246–255 (2001)
26. Cai, Y.D., Liu, X.J., Xu, X.B., Chou, K.C.: Artificial Neural Network Method for Predicting Protein Secondary Structure Content. *Computers and Chemistry* 26, 347–350 (2002)
27. Chou, K.C.: Using Pair-coupled Amino Acid composition to Predict Protein Secondary Structure Content. *J. Protein Chem.* 18, 473–480 (1999)
28. Liu, H., Wang, M., Chou, K.C.: Low-frequency Fourier Spectrum for Predicting Membrane Protein Types. *Biochem. Biophys. Res. Commun.* 336, 737–739 (2005)
29. Liu, H., Yang, J., Wang, M., Xue, L., Chou, K.C.: Using Fourier Spectrum Analysis and Pseudo Amino Acid Composition for Prediction of Membrane Protein Types. *The Protein Journal* 24, 385–389 (2005)
30. Zhang, T.L., Ding, Y.S.: Using Pseudo Amino Acid Composition and Binary-tree Support Vector Machines to Predict Protein Structural Classes. *Amino Acids* (2007) 10.1007/s00726-007-0496-1
31. Chou, K.C.: Review: Low-frequency Collective Motion in Biomacromolecules and Its Biological functions. *Biophysical Chemistry* 30, 3–48 (1988)
32. Chou, K.C.: Low-frequency Resonance and Cooperativity of Hemoglobin. *Trends in Biochemical Sciences* 14, 212–213 (1989)
33. Shen, H.B., Chou, K.C.: Ensemble Classifier for Protein fold pattern recognition. *Bioinformatics* 22, 1717–1722 (2006a)
34. Shen, H.B., Chou, K.C.: Using Ensemble Classifier to Identify Membrane Protein Types. *Amino Acids* (2006) 10.1007/s00726-006-0439-2
35. Shen, H.B., Yang, J., Chou, K.C.: Fuzzy KNN for Predicting Membrane Protein Types from Pseudo Amino Acid Composition. *Journal of Theoretical Biology* 240, 9–13 (2006)
36. Vapnik, V.N.: *The Nature of Statistical Learning Theory*. Springer, Heidelberg (1995)
37. Abe, S.: Fuzzy LP-SVM for multiClass problems. In: *ESANN 2004 proceedings- European symposium on artificial neural networks Bruges (Belgium), 28-30 April 2004 d-side public*, pp. 429–434 (2004) ISBN 2-930307-04-8
38. Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J.: *Least Squares Support Vector Machines*. World Scientific, Singapore (2002)
39. Chou, K.C., Zhang, C.T.: Review: Prediction of Protein Structural Classes. *Critical Reviews in Biochemistry and Molecular Biology* 30, 275–349 (1995)