Journal of
Computational and Theoretical Nanoscience
Vol. 11, 1–6, 2014

Apoptosis Protein Subcellular Location Prediction Based on Position-Specific Scoring Matrix

Yu-Hua Yao*, †, Zhuo-Xing Shi†, and Qi Dai

College of Life Sciences, Zhejiang Sci-Tech University, Hangzhou 310018, China

Based on Position-Specific Scoring Matrix (PSSM), average mutation probability from one particular amino acid to 20 types residues and average mutation rate of 20 types of amino acids within query sequences during evolution are extracted, the new method which combines these evolutionary information is proposed for apoptosis protein subcellular location prediction. Principal component analysis was employed to extract useful features. The proposed method is tested by the support vector machine classifier, the prediction accuracy in dataset ZD98 and CL317 are reaches 92.9% and 90.5%, respectively. The experiment results obtained by jackknife test can almost reach the highest level through comparison with other methods. In addition, it's worth to pointing out that the proposed method is better at small set predicting than others methods. All of the results confirm that the proposed novel sequence information obtained from Position-Specific Scoring Matrix are remarkable, it's heralds that the proposed method might serve as an efficient prediction model for apoptosis protein subcellular location prediction.

Keywords: Apoptosis Proteins, Subcellular Location, PSI-BLAST, Support Vector Machine, Position-Specific Scoring Matrix.

RESEARCH ARTICLE

1. INTRODUCTION

Apoptosis is a form of cell death which plays a central role in normal tissue homeostasis by regulating a balance between cell proliferation and death.[1–4] Cells undergoing apoptosis usually exhibit a characteristic morphology, including fragmentation of the cell into membrane-bound apoptotic bodies, nuclear and cytoplasm condensation and hemolytic cleavage of the DNA into small oligonucleosomal fragments.[5, 6] Unregulated excessive apoptosis may cause various degenerative and autoimmune diseases. Conversely, an inappropriately low rate of apoptosis may promote survival and accumulation of abnormal cells that can give rise to tumor formation and prolonged autoimmune stimulation such as in cancers and Graves' disease.[7] The study on apoptosis proteins can help us to understand the mechanism of apoptosis and provide many targets for therapeutic intervention.[8–10]

Zhou and Doctor[11] firstly investigated the prediction of subcellular location of apoptosis proteins. They explored amino acid composition and the covariant discrimination function to predict the four kinds of subcellular locations for 98 apoptosis proteins dataset. Through more than ten years of efforts, the prediction accuracy is improved, researchers have proposed many methods. Though the overall predictive accuracy have been improved for apoptosis proteins using existed methods, the representation of protein sequence was mainly by using the amino acid composition, or dipeptide composition. These representations will lead to ignore the sequence-order information of protein.

In recent research, evolution-based methods use the query sequence to search protein databases for extract evolutionary information and further predict the subcellular location of the query sequence. Given a query sequence, it is searched against a database of proteins using position-specific iterated BLAST (PSI-BLAST), where PSI-BLAST is a search tool hanging the double sequence alignment and the multiple sequence alignment together. Evolutionary information of protein sequence like Position Specific Scoring Matrix (PSSM) can be extracted from PSI-BLAST profile. The idea of adopting PSSM extracted from sequence profiles generated by PSI-BLAST as input information was first proposed by Jones.[12] This method has earlier been used for protein subcellular localization by Xie.[13]

With help of PSI-BLAST, we proposed a novel method to combination two aspects of evolution information for apoptosis protein subcellular location prediction which is extracted from Position Specific Scoring Matrix. Proposed novel method is tested with two benchmark datasets and

*Author to whom correspondence should be addressed.
†These two authors contributed equally to this work.

compared with the other prediction methods, the experiment results confirm that the proposed novel sequence information based on PSI-BLAST is promising.[30–33]

2. MATERIAL AND METHODS

2.1. Describe of Two Banchmark Datasets

Two benchmark datasets were adopted in this work; proteins in those datasets were extracted from SWISS-PROT (version 49.5). The ZD98 dataset consists of 98 apoptosis protein sequences, which include 43 cytoplasmic proteins, 30 plasma membrane-bound proteins, 13 mitochondrial proteins and 12 other proteins.[11] The dataset CL317 consists of 317 apoptosis protein sequences constructed by Chen and Li, which include 112 cytoplasmic proteins, 55 membrane proteins, 34 mitochondrial proteins, 17 secreted proteins, 52 nuclear proteins and 47 endoplasmic reticulum proteins.[14]

2.2. Derived Features from Position-Specific Scoring Matrix

Position-Specific Scoring Matrix (PSSM) from position-specific profiles generated by PSI-BLAST. PSSM is a commonly used representation of motifs in biological sequences.[15] This method has been used for predicting protein subcellular localization,[13] subnuclear localization[16] and predicting protein structural class.[17, 18] For a protein sequence S with L amino acid residues, PSSM is obtained according to the following equation:

$$PSSM_S = \begin{bmatrix} P_{1\to A} & P_{1\to R} & \cdots & P_{1\to j} & \cdots & P_{1\to V} \\ P_{2\to A} & P_{2\to R} & \cdots & P_{2\to j} & \cdots & P_{2\to V} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{i\to A} & P_{i\to R} & \cdots & P_{i\to j} & \cdots & P_{i\to V} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{L\to A} & P_{L\to R} & \cdots & P_{L\to j} & \cdots & P_{L\to V} \end{bmatrix} \quad (1)$$

Where $i \to j$ describes i-th amino acid residue of the protein sequence S being mutated to amino acid type j in the biology evolution process, $P_{i\to j}$ is the score of this mutation, L is the length of the sequence S. To get the $L \times 20$ scores of the $PSSM_S$ we used three iterations with E-value is 0.001 of PSI-BLAST to search a protein database for multiple sequence alignment against the protein S. $P_{i\to j}$ in $PSSM_S$ represents the 'normalized probability' that i-th amino acid residue of the protein sequence S being mutated to amino acid type j in the biology evolution process.

In this work we proposed a new method to derived features from the position-specific scoring matrix, which combines evolutionary information from both aspects, first is calculate average probabilities of mutation from one particular amino acid to 20 types residues within query sequences during the evolution process, here representation to *MAP*; second is calculate the average mutation rate of 20 types of amino acid in query sequences during the evolution process, here representation to *AMP*.

To extract the features *MAP* as follows Eqs. (2) and (3):

$$MAP_{Si} = \frac{\sum_{j=1}^{L} P_{i\to j}}{L}, \quad i = 1, 2, \ldots, 20 \quad (2)$$

$$MAP_S = [MAP_{SA}, MAP_{SR}, \ldots, MAP_{SV}] \quad (3)$$

On the basis of Eq. (2), obviously, MAP_{Si} is the mean value of the elements in the i-th column of $PSSM_S$, here MAP_S are the 20-D features vector representing the average probability of amino acid residue of the protein sequence S being mutated to 20 types of amino acid in the biology evolution process. Here we used the numerical codes $1, 2, 3, \ldots, 20$ to represent the single character of ordered 20 native amino acid types in Eq. (2).

To calculate the features *AMP* has to extract the submatrix for 20 kinds of amino acids from $PSSM_S$, for example to extract the submatrix of type of residue A in $PSSM_S$ as follows Eq. (4):

$$PSSM_{SA} = \begin{bmatrix} P_{1A\to A} & P_{1A\to R} & \cdots & P_{1A\to j} & \cdots & P_{1A\to V} \\ P_{2A\to A} & P_{2A\to R} & \cdots & P_{2A\to j} & \cdots & P_{2A\to V} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{iA\to A} & P_{iA\to R} & \cdots & P_{iA\to j} & \cdots & P_{iA\to V} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{lA\to A} & P_{lA\to R} & \cdots & P_{lA\to j} & \cdots & P_{lA\to V} \end{bmatrix} \quad (4)$$

Here l is the number of residue A in the sequence S. To calculated the feature *AMP* in the submatrix as follows Eqs. (5) and (6):

$$AMP_{Si} = \frac{\sum_{t=1}^{20} \sum_{j=1}^{l} P_{i\to t}}{l \times 20 + L}, \quad i = 1, 2, \ldots, 20 \quad (5)$$

$$AMP_S = [AMP_{SA}, AMP_{SR}, \ldots, AMP_{SV}] \quad (6)$$

On the basis of Eq. (5), here MAP_{Si} is the weighted average value of the elements in the submatrix from $PSSM_S$, MAP_S are the 20-D features vectors representing the average mutation probability of 20 kinds of amino acids residues of the protein sequence S in the biology evolution process. Here we used the numerical codes $1, 2, 3, \ldots, 20$ to represent the single character of ordered 20 native amino acid types in Eq. (5).

Through the above two-step process, finally, we get a 40-D features vector that was combination from two aspects of evolution information *MAP* and *AMP*; here representation to *PE* with Eq. (7).

$$PE_S = [MAP_S, AMP_S] \quad (7)$$

2.3. Principal Component Analysis

Feature selection is the process of identifying and removing as much irrelevant and redundant information as possible. This reduces the dimensionality of the data and may

allow learning algorithms to operate faster and more effectively. In this work we apply the dimension reduction techniques principal component analysis (PCA) to selection feature. Principal component analysis is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance, and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to the preceding components. Principal components are guaranteed to be independent only if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables. Principal component analysis (PCA) was invented by Pearson in 1901.[19]

In this work, through PCA processing the original dataset transform a new dataset that according to the contribution rate arrangement from high to low, the new dataset equal dimension with original dataset and new dataset each dimension is uncorrelated. We select the feature using a simple grid search strategy based on the jackknife test for two datasets.

2.4. Classifier and Evaluation

This work adopts Vapnik's support vector machine[20] to predict the subcellular location of apoptosis proteins. Prediction of protein subcellular location is a multi-classification problem. Therefore, we adopt the multi-class prediction method. Support vector machine using "one-against-others" strategy, given a test protein of unknown category, SVM first map the input vectors into one feature space. Then SVM finds an optimized linear division to solve two-class or multi-class problem in feature space. Finally, a prediction label is assigned to the test sample. In our study, the LIBSVM package is used to implement the SVM classifier. The radial basis function (RBF) is chosen as the kernel function. The regularization parameter c and the kernel width parameter g are optimized on the training set using a grid search strategy in the LIBSVM.

In this work, jackknife test is employed to evaluate the prediction performance of our method. Here, we used it to evaluate the performance of the proposed method. We also considered standard performance measures over class, including the accuracy for class C_i here record CA and overall accuracy here record OA, which was defined as the fraction of class C_i or all the proteins tested that are classified correctly.

$$CA_i = \frac{TP_i}{|C_i|} \tag{8}$$

$$OA = \frac{\sum_i TP_i}{\sum_i |C_i|} \tag{9}$$

Where TP_i is the number of true positives, $|C_i|$ is the number of proteins in each class C_i.

3. RESULTS AND DISCUSSION

This section should contain the discussion of the selected feature and experiment results on two benchmark datasets. First we used the MAP and AMP to create novel sequence information PE of apoptosis proteins. Then principal component analysis was employed to extract useful features. Finally, the selected features of the novel combined sequence information were fed into support vector machine to make subcellular location prediction of apoptosis proteins.

3.1. Prediction Results on Two Benchmark Datasets

Two benchmark datasets were used in this work. Dataset ZD98 consists of 98 apoptosis protein sequences, dataset CL317 consists of 317 apoptosis protein sequences. The results obtained by the proposed method were shown in Table I. In the ZD98 experiment, the overall accuracy obtained by the proposed method is 92.9%, with 95.3%, 93.3%, 84.6%, 91.7% for Cyto, Memb, Mito and Other class, respectively. In the CL317 experiment, the overall accuracy obtained by the proposed method is 90.5%, with 92.0%, 92.7%, 82.4%, 76.5%, 90.4% and 93.6% for Cyto, Memb, Mito, Secr, Nucl, Endo, respectively.

3.2. Effect of Source Database for PSI-BLAST Search

In many study are adopted NR database for PSI-BLAST search, but in this work, every protein sequence in dataset was searched against its SWISS-PROT database because we found that use SWISS-PROT database to be source database for PSI-BLAST search make the highly contributions to subcellular location prediction of apoptosis proteins. We used three iterations of PSI-BLAST with E-value 0.001 to search a database for multiple sequence alignment against the protein S to get the PSSM. Here, every protein sequence in dataset was searched against with SWISS-PROT database. SWISS-PROT database is published: May 20 2013; its number of sequences is 540052. For a better understanding of the influence of the source database for PSI-BLAST search, we also searched every protein sequence in dataset against the NR database and got

Table I. Prediction results on two datasets evaluated with Jackknife test.

	CA (%)						
	Cyto	Memb	Mito	Other			*OA* (%)
ZD98	95.3	93.3	84.6	91.7			92.9
	CA (%)						
	Cyto	Memb	Mito	Secr	Nucl	Endo	*OA* (%)
CL317	92.0	92.7	82.4	76.5	90.4	93.6	90.5

RESEARCH ARTICLE

the evolutionary profile, where NR database is published: November 5, 2012, its number of sequences is 21171493.

The comparison of performance of EP_{SW} and PE_{NR} on two datasets is illustrated in Figure 1, we can found that the PE_{SW} significantly outperforms the PE_{NR} on datasets ZD98 and CL317. These results demonstrate that the evolutionary profile gotten by searching against the SWISS-PROT have a better performance than by searching against NR database. We speculate that the reason may lie in the fact that NR database is so big and some low-homology sequences which may add the noise for apoptosis proteins subcellular location prediction. In addition, as we know that the two benchmark datasets are extract form SWISS-PROT database, as a matter of course that searching against SWISS-PROT database can obtain more targeted information for subcellular location prediction of apoptosis proteins.

There has another important point that selected the SWISS-PORT database to instead of NR database for PSI-BLAST search, which was NR its number of sequences is 21171493 it's 39.2 times larger than SWISS-PORT's 540052. In a normal two cores desktop computer, assume that a query sequence have 200 amino acid residues it's search in NR were spend 50 minutes, but search in SWISS-PORT only spend about 1 minutes, it's 50 times fast than NR. Obviously, select the SWISS-PORT database for PSI-BLAST search has higher time efficiency then NR database for researchers. Table II shows the comparison the time efficiency with SWISS-PORT and NR database for PSI-BLAST search on two datasets used the normal two cores desktop computer.

Table II. Comparison the time efficiency with SWISS-PORT and NR database for PSI-BLAST.

	Spend-time (hours)		
	SWISS-PROT	NR	SW/NR speed-up (%)
ZD98	1.6	49	30.6
CL317	5.3	158.5	29.8

3.3. Influence of Feature Selection Process

Principal component analysis (PCA) is employed to selects the useful features for predicting the apoptosis proteins subcellular location. For a better understanding of the PCA efficiency, we used the jackknife test to examine the performance of the proposed sequence information without using PCA in predicting the subcellular location on datasets ZD98 and CL317.

The prediction accuracy of the proposed sequence information with and without PCA method is listed in Table III. The experiment that the overall accuracies of the *PE* using the PCA features extraction algorithm are higher than those without using the PCA algorithm, the prediction accuracies is improved by 0.3% in CL317. Specifically, the dimensions of the PE^P are 12 and 18 for the datasets ZD98, and CL317, which are about 70% and 55% smaller than that of the *PE*, respectively.

Before the PCA process, a protein can be represented by a combination of two different 20-D vectors *MPA* and *AMP*. In fusion the above two vectors into one, we got the proposed novel sequence information denoted by *PE*, which is a 40-D vector. By using the PCA algorithm, we extracted the most important features from the 40-D vector of *PE*, as denoted by PE^P. Figure 2 shows the relations between the dimension of the PE^P and its overall accuracy for the subcellular location prediction on datasets ZD98 and CL317. Figure 2 show that the higher overall prediction accuracy is, the higher the dimension of the PE^P will be. Considering the time efficiency to extract the most important features, we select the PE^P using a simple grid search strategy based on the jackknife test for two datasets.

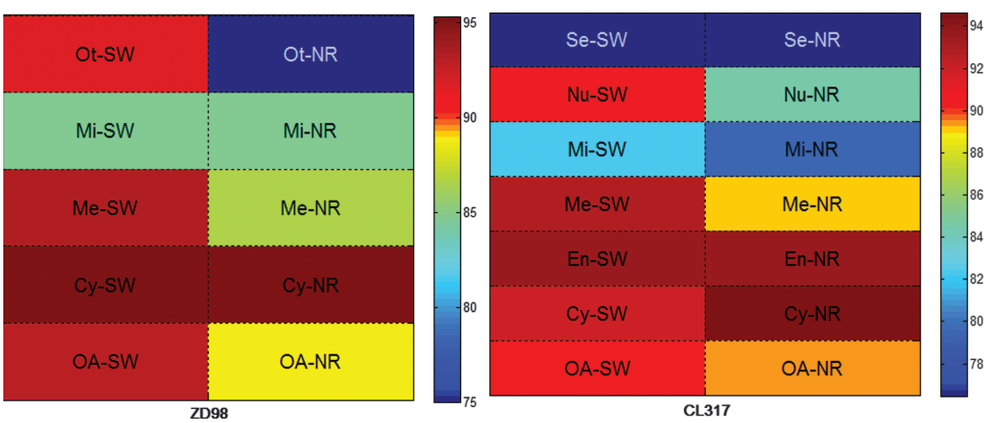

Fig. 1. Comparison of performance with EP_{SW} and PE_{NR} on two datasets, here EP_{SW} is denoted the features *PE* obtained by searching against with SWISS-PROT database; PE_{NR} is denoted the features *PE* obtained by searching against with NR database; OA-SW is denoted the overall accuracy obtained by searching against with SWISS-PROT database; OA-NR is denoted the overall accuracy *PE* obtained by searching against with NR database; Cy-SW/NR, Me-SW/NR, Mi-SW/NR, Ot-SW/NR, En-SW/NR, Nu-SW/NR, Se-SW/NR are denoted the accuracy in class Cyto, Memb, Mito, Other, Endo, Nucl, Secr obtained by searching against with SWISS-PORT/NR database, respectively.

RESEARCH ARTICLE

Table III. Overall accuracy and dimension of the *PE* and PE^P for datasets ZD98 and CL317.

	Features *PE*		Features *PEP*	
	OA	Dimension	*OA*	Dimension
ZD98	92.9	40	**92.9**	**12**
CL317	90.2	40	**90.5**	**18**

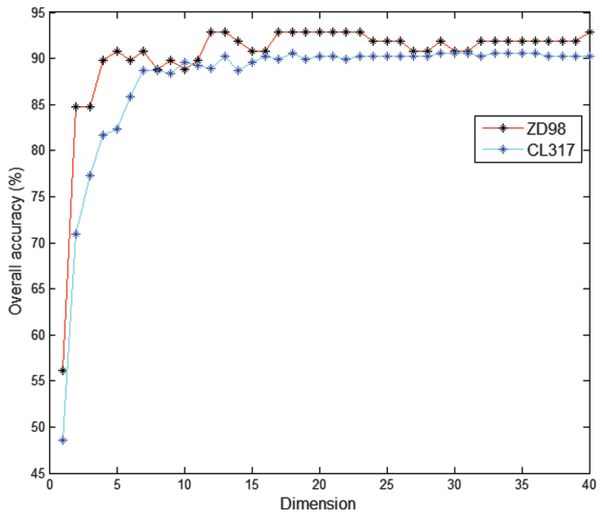

Fig. 2. The relationship between the *OA* and feature dimension in PCA process of the two datasets.

3.4. Comparison with Other Methods

In order to evaluate the efficiency of the proposed method, we compared it with the competing prediction methods on the ZD98 and CL317. We selected the accuracy of each class and overall accuracy as evaluation methods. And the experimental results are shows in Tables IV–V.

In dataset ZD98, we compared the proposed method with the others methods that include Instab_SVM,[21] Dipep_Diver,[22] AAC_CCA[23] and EBGW_SVM.[24] Table IV shows that the proposed method reached the highest overall accuracy 82.9% same with Zhang's method.[22] But the class-Other accuracy of the proposed method is 91.9%, which is 8.4% higher than Zhang's method, as we know that class-Other's only contain 12 proteins that was the small set in ZD98, as is known to all that at SVM classifier the small set is hard to predict, this result explain that the proposed method is good at the small set predicting.

Table IV. Prediction results with different models on ZD98 in Jackknife test.

	CA (%)				
Model	Cyto	Memb	Mito	Other	*OA* (%)
Instab_SVM[21]	76.8	83.3	92.5	50.0	77.6
Dipep_Diver[22]	88.4	90.0	92.3	50.0	84.7
AAC_CCA[23]	97.7	73.3	30.8	25.0	72.5
EBGW_SVM[24]	97.7	90.0	92.3	83.3	92.9
This work	95.3	93.3	84.6	**91.7**	**92.9**

Table V. Prediction results with different models on CL317 in Jackknife test.

	CA (%)						
Model	Cyto	Memb	Mito	Secr	Nucl	Endo	*OA* (%)
ID[22]	81.3	81.8	85.3	88.2	82.7	83.0	82.7
ID_SVM[25]	91.1	89.1	79.4	58.8	73.1	87.2	84.2
DF_SVM[26]	92.9	85.5	76.5	76.5	93.6	86.5	88.0
FKNN[27]	92.0	89.1	85.3	76.5	92.3	93.7	90.2
PseAAC_SVM[28]	93.8	90.9	85.3	76.5	90.4	95.7	91.1
SMAC_SVM[29]	86.4	90.7	93.8	85.7	92.1	93.8	90.0
This work	92.0	**92.7**	82.4	76.5	90.4	93.6	**90.5**

In dataset CL317, we compared the proposed method with the others methods such as ID,[22] ID_SVM,[25] DF_SVM,[26] FKNN,[27] PseAAC_SVM,[28] and SMAC_SVM.[29] Table V indicates that the proposed method only inferior to Lin's method.[25] As the same to ZD98, at the small set class-Memb's accuracy of the proposed method is 92.7%, it's 1.8% higher than Lin's method.

4. CONCLUSIONS

Apoptosis proteins play an important role in the development and homeostasis of an organism. Obtaining information and the accurate prediction of subcellular location for apoptosis proteins is very helpful for understanding the mechanism of apoptosis and their biological functions.

In this work we proposed a new method to derived features from the Position-Specific scoring matrix, which combines evolutionary information from both aspects, first is calculate average probabilities of mutation from one particular amino acid to 20 types residues within query sequences during the evolution process, here representation to *MAP*; second is calculate the average mutation rate of 20 types of amino acid in query sequences during the evolution process, here representation to *AMP*. In fusion the above information into one, we got the novel sequence information of a protein. At last, we used principal component analysis to reduce the feature space and then fed them into the support vector machine classify to predict subcellular location of apoptosis proteins.

This work made the remarkable result and have investigate the two influence factors of very meaningful its research on this area has a certain contribution. Out of the results, the first contribution can be seen from the selection of source database for PSI-BLAST search, we found that the evolutionary profile gotten by searching against the SWISS-PROT database is more useful than by searching against NR database for subcellular location prediction of apoptosis proteins. Second is the result show that the proposed method is good at the small set predicting. Through the above analysis we could summarize that the proposed

method might serve as an efficient prediction model for apoptosis protein subcellular location prediction.

Acknowledgments: We appreciate the financial support of this work that was provided by Zhejiang Provincial Natural Science Foundation of China (No. LY12F02043). This work was also supported by the National Natural Science Foundation of China (Nos. 61272312, 61170316).

References

1. K. C. Chou, T. C. Zhang, and M. G. Maggiora, *Proteins* 28, 99 (**1997**).
2. J. J. Chou, H. Li, G. S. Salvessen, J. Yuan, and G. Wagner, *Cell* 96, 615 (**1999**).
3. K. C. Chou, A. G. Tomasselli, and R. L. Heinrikson, *FEBS Lett.* 470, 249 (**2000**).
4. M. D. Jacobson, M. Weil, and M. C. Raff, *Cell* 88, 347 (**1997**).
5. J. F. Kerr, A. H. Wyllie, and A. R. Currie, *Br. J. Cancer* 26, 239 (**1972**).
6. H. Steller, *Science* 267, 1445 (**1995**).
7. M. E. Peter, A. E. Heufelder, and M. O. Hengartner, *Proc. Natl. Acad. Sci. USA* 94, 12736 (**1997**).
8. K. C. Chou, *Bioinformatics* 21, 10 (**2005**).
9. K. C. Chou, *J. Proteome. Res.* 4, 1681 (**2005**).
10. K. C. Chou, *J. Proteome. Res.* 4, 1413 (**2005**).
11. G. P. Zhou and K. Doctor, *Proteins: Struct. Funct. Genet.* 50, 44 (**2003**).
12. D. T. Jones, *J. Mol. Biol.* 292, 195 (**1999**).
13. D. Xie, A. Li, M. Wang, Z. Fan, and H. Feng, *Nucleic Acids Research* 33, W105 (**2005**).
14. Y. L. Chen and Q. Z. Li, *J. Theor. Biol.* 245, 775 (**2007**).
15. I. Ben-Gal, A. Shani, A. Gohr, J. S. A. Grau, A. Shmilovici, S. Posch, and I. Grosse, *Bioinformatics* 21, 2657 (**2005**).
16. W. L. Huang, C. W. Tung, H. L. Huang, S. F. Hwang, and S. Y. Ho, *Biosystems* 90, 573 (**2007**).
17. L. Kurgan, K. Cios, and K. Chen, *BMC Bioinformatics* 9, 22 (**2008**).
18. L. A. Kurgan, T. Zhang, H. Zhang, S. Y. Shen, and J. S. Ruan, *Amino Acids* 35, 551 (**2008**).
19. K. Pearson, *Philosophical Magazine* 2, 559 (**1901**).
20. V. Vapnik, The Nature of Statistical Learning Theory, Springer Verlag, New York (**1995**), pp. 1–8.
21. J. Huang, F. Shi, and H. B. Zhou, *China J. Bioinf.* 3, 121 (**2005**).
22. Y. L. Chen and Q. Z. Li, *Acta Sci. Nat. Univ. NeiMongol.* 25, 413 (**2004**).
23. G. P. Zhou and K. Doctor, *Proteins* 50, 44 (**2003**).
24. Z. H. Zhang and Z. H. Wang, *FEBS Lett.* 580, 6169 (**2006**).
25. Y. L. Chen and Q. Z. Li, *J. Theor. Biol.* 248, 377 (**2007**).
26. L. Zhang, B. Liao, D. Li, and W. Zhu, *J. Theor. Biol.* 259, 361 (**2009**).
27. X. Jiang, R. Wei, T. Zhang, and Q. Gu, *Protein. Pept. Lett.* 15, 392 (**2008**).
28. H. Lin, H. Wang, H. Ding, Y. L. Chen, and Q. Z. Li, *Acta Biotheor.* 57, 321 (**2009**).
29. X. Q. Yu, X. Q. Zheng, T. G. Liu, Y. C. Dou, and J. Wang, *Amino Acids* 42, 1619 (**2012**).
30. B. Liao, B. Y. Liao, X. M. Sun, and Q. G. Zeng, *Bioinformatics* 26, 2678 (**2010**).
31. T. Ono, Y. Fujimoto, and S. Tsukamoto, *Quantum Matter* 1, 4 (**2012**).
32. M. Narayanan and A. John Peter, *Quantum Matter* 1, 53 (**2012**).
33. Q. Zhao, *Rev. Theor. Sci.* 1, 83 (**2013**).
34. A. Khrennikov, *Rev. Theor. Sci.* 1, 34 (**2013**).
35. Q. Xiang, B. Liao, Q. Liu, X. Lu, and W. Zhu, *J. Comput. Theor. Nanosci.* 10, 27 (**2013**).
36. L. Huang, H. Tan, and B. Liao, *J. Comput. Theor. Nanosci.* 10, 257 (**2013**).

Received: 5 August 2013. Accepted: 26 August 2013.

RESEARCH ARTICLE