Prediction of Subcellular Localization of Apoptosis Protein Using Chou's Pseudo Amino Acid Composition

Hao Lin · Hao Wang · Hui Ding · Ying-Li Chen · Qian-Zhong Li

Abstract Apoptosis proteins play an essential role in regulating a balance between cell proliferation and death. The successful prediction of subcellular localization of apoptosis proteins directly from primary sequence is much benefited to understand programmed cell death and drug discovery. In this paper, by use of Chou's pseudo amino acid composition (PseAAC), a total of 317 apoptosis proteins are predicted by support vector machine (SVM). The jackknife cross-validation is applied to test predictive capability of proposed method. The predictive results show that overall prediction accuracy is 91.1% which is higher than previous methods. Furthermore, another dataset containing 98 apoptosis proteins is examined by proposed method. The overall predicted successful rate is 92.9%.

Keywords Apoptosis protein · Subcellular localization · Pseudo amino acid composition · Support vector machine

1 Introduction

Apoptosis is a type of cell death regulated growth, development and immune response, and clearing redundant or abnormal cells in organisms (Raff 1998; Steller 1995). It plays a key role in development and tissue homeostasis (Chou et al. 1998, 1999). The malfunctions of apoptosis will deal to a variety of formidable diseases,

H. Lin (✉) · H. Wang
Center for Bioinformatics, School of Life Science and Technology, University of Electronic Science and Technology of China, 610054 Chengdu, China
e-mail: hlin@uestc.edu.cn

H. Ding · Y.-L. Chen · Q.-Z. Li
Laboratory of Theoretical Biophysics, School of Physics Sciences and Technology, Inner Mongolia University, 010021 Hohhot, China

for example, blocking apoptosis is associated with cancer (Adams and Cory 1998; Evan and Littlewood 1998) and autoimmune disease, whereas unwanted apoptosis can possible lead to ischemic damage (Reed and Paternostro 1999) or neurodegenerative disease (Schulz et al. 1999). Because the localization of proteins in cellular is closely associated with the protein function, the study of subcellular localization of apoptosis protein is very important for elucidating functions of apoptosis protein involved in various cellular processes (Schulz et al. 1999; Suzuki et al. 2000) and drug development (Chou et al. 1997, 2000; Chou 2004).

Computational approaches, such as structural bioinformatics (Chou 2004), molecular docking (Chou et al. 2003; Li et al. 2007; Wang et al. 2008; Zheng et al. 2007), molecular packing (Chou et al. 1984, 1988), pharmacophore modeling (Sirois et al. 2004; Chou et al. 2006), Mote Carlo simulated approach (Chou 1992), diffusion-controlled reaction simulation (Chou and Zhou 1982), biomacromolecular internal collective motion simulation (Chou 1988), QSAR (Du et al. 2008), protein subcellular location prediction (Chou and Shen 2007a, 2008a) identification of membrane proteins and their types (Chou and Shen 2007b), identification of enzymes and their functional classes (Shen and Chou 2007), identification of GPCR and their types (Chou 2005), identification of proteases and their types (Chou and Shen 2008b), protein cleavage site prediction (Shen and Chou 2008b), and signal peptide prediction (Chou and Shen 2007c) and so on can timely provide very useful information and insights for both basic research and drug design and hence are widely welcome by science community. The present study is attempted to develop a computational approach for predicting the subcellular localization of apoptosis proteins in hope to stimulate the development of the relevant areas.

In the past 5 years, several algorithms such as covariant discriminant function (Zhou and Doctor 2003), support vector machine (SVM) (Huang and Shi 2005; Zhang et al. 2006; Zhou et al. 2008; Shi et al. 2008), Bayesian classifier (Bulashevska and Eils 2006), increment of diversity (ID) (Chen and Li 2007a), increment of diversity combined with support vector machine (ID_SVM) (Chen and Li 2007b) and fuzzy K-nearest neighbor (FKNN) (Jiang et al. 2008; Ding and Zhang 2008) have been proposed to predict subcellular localization of apoptosis protein based on various amino acid composition or pseudo amino acid composition. The pseudo amino acid composition (PseAAC) was firstly proposed by Chou to efficiently improve prediction quantity of protein subcellular localization (Chou 2001; Chou and Shen 2007a). PseAAC can represent a protein sequence with a discrete model yet without completely losing its sequence order information.

In this paper, based on the concept of Chou's PseAAC, SVM is applied to the latest dataset with 317 apoptosis proteins. The jackknife cross-validation is applied to examine the predictive ability of method. Moreover, another 98 apoptosis proteins built by Zhou and Doctor (2003) are examined by proposed method. The predictive results of proposed method can improve the predictive success rates, and hence the current method may play a complementary role to other existing methods for predicting protein subcellular localization of apoptosis protein.

2 Materials and Methods

2.1 Data Sets

The 317 apoptosis proteins extracted from Swiss-Prot 49.0 can be classified into six subcellular locations: 112 cytoplasmic proteins, 55 membrane proteins, 34 mitochondrial proteins, 17 secreted proteins, 52 nuclear proteins and 47 endoplasmic reticulum proteins. The distribution of the sequence identity percentage is 40.1% with ≤40% sequence identity, 15.5% with sequence identity from 41% to 80%, 18.9% with sequence identity from 81% to 90% and 25.6% with ≥91% sequence identity (Chen and Li 2007a, b).

In addition, the 98 apoptosis proteins containing 43 cytoplasmic proteins, 30 plasma membrane-bound proteins, 13 mitochondrial proteins and 12 other proteins (Zhou and Doctor 2003) are also used to estimate the effectiveness of the method.

2.2 Pseudo Amino Acid Composition

The appropriate parameter is one of the most important aspects for prediction issues. The essence of PseAAC includes not only the main feature of amino acid composition, but also the sequence order correlation (Chou 2001; Chou and Shen 2007a; Shen and Chou 2008a). Consider a protein (X) chain with length L amino acid residues:

$$R_1R_2R_3...R_L \tag{1}$$

Then a protein may be denoted as a $(20 + \lambda)$ dimension vector defined by $20 + \lambda$ discrete numbers; i.e.

$$X = [x_1...x_{20}x_{20+1}...x_{20+\lambda}]^T \tag{2}$$

$$\text{here } x_u = \begin{cases} \dfrac{f_u}{\sum\limits_{i=1}^{20} f_i + \omega \sum\limits_{j=1}^{\lambda} \theta_j}, & (1 \leq u \leq 20) \\ \dfrac{\omega \theta_{u-20}}{\sum\limits_{i=1}^{20} f_i + \omega \sum\limits_{j=1}^{\lambda} \theta_j}, & (21 \leq u \leq 20 + \lambda) \end{cases} \tag{3}$$

In Eq. 3, the f_i is the normalized frequency of the 20 amino acids in protein X, ω is the weight factor for sequence order effect. θ_j is the j-tier sequence correlation factor computed by the following formula:

$$\theta_j = \frac{1}{L-j} \sum_{i=1}^{L-j} \Theta(R_i, R_{i+j}), \quad (j < L) \tag{4}$$

where $\Theta(R_i, R_{i+j})$ is the correlation function and can be given by

$$\Theta(R_i, R_{i+j}) = \frac{1}{k} \sum_{l=1}^{k} \left[H_l(R_{i+j}) - H_l(R_i) \right]^2 \tag{5}$$

In Eq. 5, k is the number of factors. $H_l(R_i)$ is any one of the physico-chemical characteristics values of the amino acid R_i. These physico-chemical characteristics mainly include hydrophobicity, hydrophilicity, side chain mass, pK of the α-COOH group, pK of the α-NH$_3^+$ group and pI at 25°C. The hydrophobicity, hydrophilicity and side chain mass are used for the current study. The physico-chemical characteristics values must convert to standard type by the following equation:

$$H_l(R_i) = \frac{H_l^0(i) - \sum_{i=1}^{20}\left(H_l^0(i)/20\right)}{\sqrt{\dfrac{\sum_{i=1}^{20}\left[H_l^0(i) - \sum_{i=1}^{20}\left(H_l^0(i)/20\right)\right]^2}{20}}} \tag{6}$$

where $H_l^0(i)$ is the original physico-chemical characteristics values of the i-th amino acid. We use the numerical indices 1,2,3,…,20 to represent the 20 native amino acids according to the alphabetical order of their single-letter codes: A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W and Y. The data calculated by standard conversion will have a zero mean value and will remain unchanged if going through the same conversion procedure again.

2.3 Support Vector Machine

SVM is a kind of machine learning method based on statistical learning theory (Vapnik 1998). As a supervised machine learning technology, it has been successfully used in wide fields of bioinformatics by transforming the input vector into a high-dimension Hilbert space and to seek a separating hyperplane in this space. Now, we briefly explain the basic idea of the SVM. For a two-class classification problem, a series of training vectors $\vec{X}_i \in R^d$ ($i = 1, 2, …, N$) with corresponding labels $y_i \in \{+1, -1\}$ ($i = 1, 2, …, N$). Here, ± 1 and -1, respectively indicate the two classes. SVM maps the input vectors $\vec{X}_i \in R^d$ into a high dimensional feature space for constructing an optimal separating hyperplane with the largest distance between two classes, measured along a line perpendicular to this hyperplane. The decision function implemented by SVM can be written as:

$$f(\vec{X}) = \mathrm{sgn}\left(\sum_{i=1}^{N} y_i\alpha_i \cdot K(\vec{X}, \vec{X}_i) + b\right) \tag{7}$$

where $K(\vec{X}, \vec{X}_i)$ is a kernel function which defines an inner product in a high dimensional feature space. Three kinds of kernel functions may be defined as: Polynomial function:

$$K(\vec{X}_i, \vec{X}_j) = \left(\vec{X}_i \cdot \vec{X}_j + 1\right)^d \tag{8}$$

Radial basis function (RBF):

$$K\left(\vec{X_i}, \vec{X_j}\right) = \exp\left(-\gamma || \vec{X_i} - \vec{X_j} ||^2\right) \tag{9}$$

Sigmoid function:

$$K\left(\vec{X_i}, \vec{X_j}\right) = \tanh\left[b\left(\vec{X_i} \cdot \vec{X_j}\right) + c\right]. \tag{10}$$

The coefficients α_i can be solved by the following convex Quadratic Programming (QP) problem: Maximize

$$\sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_j \cdot y_iy_j \cdot K\left(\vec{X_i}, \vec{X_j}\right) \quad \text{subject to} \quad 0 \leq \alpha_i \leq C \tag{11}$$

here $\sum_{i=1}^{N} \alpha_iy_i = 0$, $i = 1, 2, \ldots, N$. The regularization parameter C can control the trade off between margin and misclassification error. These $\vec{X_i}$ are called Support Vectors only if the corresponding $\alpha_i > 0$.

In general, One-Versus-Rest (OVR) and One-Versus-One (OVO) are the most commonly used approach for solving multi-class problems by reducing a single multi-class problem into multiple binary problems. This paper used the OVO strategy. The software used to implement SVM is LibSVM2.83 written by Lin's lab and can be freely downloaded from: http://www.csie.ntu.edu.tw/~cjlin/libsvm (Chang and Lin 2001). Here, the RBF is used for all our calculations. The regularization parameter C and the kernel parameter γ of the RBF must be determined in advance.

2.4 The Criteria Definitions

The predictive capability of the algorithm is estimated by four parameters: sensitivity (S_n), specificity (S_p) and correlation coefficient (CC) defined as follows (Chen and Li 2007a, b):

$$S_n = TP/(TP + FN) \tag{12}$$

$$S_p = TP/(TP + FP) \tag{13}$$

$$CC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP) \times (TN + FN) \times (TP + FN) \times (TN + FP)}} \tag{14}$$

here TP denotes the numbers of the correctly recognized positives, FN denotes the numbers of the positives recognized as negatives, FP denotes the numbers of the negatives recognized as positives, TN denotes the numbers of correctly recognized negatives.

3 Results and Discussion

In statistical prediction, the following three cross-validation tests are often used to examine the power of a predictor: independent dataset test, sub-sampling (such fivefold or tenfold sub-sampling) test, and jackknife test. Of these three examine

method, the jackknife test is deemed the most objective and rigorous one (Chou and Zhang 1995) that can always yield a unique outcome as demonstrated by a penetrating analysis in a recent comprehensive review (Chou and Shen 2007a) and has been widely and increasingly adopted by investigators to test the power of various prediction methods (Lin and Li, 2007a, b; Lin 2008; Li and Li 2008a, b; Jia et al. 2008; Jin et al. 2008; Zhang and Fang 2008; Munteanu et al. 2008; Niu et al. 2008; Lin et al. 2008; Gao et al. 2008). For the jackknife cross-validation, each proteins in the dataset is in turn singled out as an independent test sample and all the rule parameters are calculated based on the remaining proteins without including the one being identified. Therefore, we also use the jackknife cross-validation to examine proposed method.

The weight factor w and correlation factor λ in the Chou's PseAAC are two kind important parameters. Usually, the larger the λ, the more information the representation bears. However, if the PseAAC contains too many components, it would reduce the cluster-tolerant capacity (Chou 1999) so as to lower down the jackknife success rate. We examine a great deal of parameters of PseAAC (ω and λ) and SVM (C and r) by using jackknife cross-validation. For the current study, we found that, when $w = 0.1$, $\lambda = 3$, $C = 1,000$ and $r = 0.04$, the predicted successful rate is the highest. The results of 317 apoptosis proteins are listed in Table 1. The results show that the sensitivity, specificity and CC of endoplasmic reticulum proteins are 95.7, 95.7 and 94.9%, respectively, which is higher than other subcellular location.

The compared results with other methods are shown in Table 2. Table 2 exhibits that the sensitivities of SVM combined with PseAAC are higher than other methods

Table 1 The predictive results of jackknife cross-validation for 317 apoptosis proteins

	Sn	Sp	CC
Cyto	0.938	0.921	0.890
Memb	0.909	0.893	0.880
Mito	0.853	0.935	0.881
Secr	0.765	0.813	0.777
Nucl	0.904	0.887	0.874
Endo	0.957	0.957	0.949
Overall prediction rate	0.911		

Table 2 The predictive results of different methods by the jackknife test for 317 apoptosis proteins

Method	$Sn \times 100\%$						
	Cyto	Memb	Mito	Secr	Nucl	Endo	Overall
ID[a]	81.3	81.8	85.3	88.2	82.7	83.0	82.7
ID_SVM[b]	91.1	89.1	79.4	58.8	73.1	87.2	84.2
FKNN[c]	92.0	89.1	85.3	76.5	92.3	93.7	90.2
FKNN[d]	93.8	92.7	82.4	76.5	90.4	93.6	90.9
SVM + PseAAC (This paper)	93.8	90.9	85.3	76.5	90.4	95.7	91.1

[a] Comes from Chen and Li (2007a). [b] Comes from Chen and Li (2007b). [c] Comes from Jiang et al. (2008). [d] Comes from Ding and Zhang (2008)

Table 3 The predictive results of different methods by the jackknife test for 98 apoptosis proteins

Method	$Sn \times 100\%$				
	Cyto	Memb	Mito	Others	Overall
Covariant[a]	97.7	73.3	30.8	25.0	72.5
SVM + 20 sqrt-amino acid composition[b]	86.0	90.0	100.0	100.0	90.8
EBGW_SVM[c]	97.7	90.0	92.3	83.3	92.9
HensBC-approach[d]	95.3	90.0	92.3	66.7	89.8
Dual-layer SVM[e]	95.4	96.7	92.3	91.7	94.9
ID[f]	90.7	90.0	92.3	91.7	90.8
ID_SVM[g]	95.3	93.3	84.6	58.3	88.8
Hilber Huang_SVM[h]	95.3	96.7	96.7	75.7	92.9
FKNN[i]	95.3	96.7	100	91.7	95.9
SVM + PseAAC (This paper)	95.3	93.3	92.3	83.3	92.9

[a] Comes from Zhou and Doctor (2003). [b] Comes from Huang and Shi (2005). [c] Comes from Zhang et al. (2006). [d] Comes from Bulashevska and Eils (2006). [e] Comes from Zhou et al. (2008). [f] Comes from Chen and Li (2007a). [g] Comes from Chen and Li (2007b). [h] Comes from shi et al. (2008). [i] Comes from Ding and Zhang (2008)

for cytoplasmic proteins, membrane proteins, mitochondrial proteins and endoplasmic proteins, whereas for secreted proteins and nuclear proteins, the sensitivities of proposed method are lower than ID and FKNN. The overall predictive successful rate of proposed method is highest among other methods.

Table 3 exhibits the compared results with other methods for 98 apoptosis proteins. Here, by use of lots of examination, we select $\omega = 0.3$, $\lambda = 3$, $C = 1,000$ and $r = 0.08$ for this prediction. The results show that the predictive successful rate of proposed method is 92.9%.

The successful accuracies clearly indicate that the SVM combined PseAAC is a promising approach. We hope that the better results using novel descriptors or appropriate parameters will improve the performance of subcellular localization prediction of apoptosis proteins. The high accuracy is helpful for further drug development.

References

Adams JM, Cory S (1998) The Bcl-survival. Science 281:1322–1326. doi:10.1126/science.281.5381.1322

Bulashevska A, Eils R (2006) Predicting protein subcellular locations using hierarchical ensemble of Bayesian classifiers based on Markov chains. BMC Bioinformatics 7:298. doi:10.1186/1471-2105-7-298

Chang CC, Lin CJ (2001) LIBSVM: a library for support vector machines. Software available at (http://www.csie.ntu.edu.tw/_cjlin/libsvm)

Chen YL, Li QZ (2007a) Prediction of the subcellular location of apoptosis proteins. J Theor Biol 245:775–783. doi:10.1016/j.jtbi.2006.11.010

Chen YL, Li QZ (2007b) Prediction of apoptosis proteins subcellular location using improved hybrid approach and pseudo-amino acid composition. J Theor Biol 248:377–381. doi:10.1016/j.jtbi.2007.05.019

Chou KC (1988) Review: low-frequency collective motion in biomacromolecules and its biological functions. Biophys Chem 30:3–48. doi:10.1016/0301-4622(88)85002-6

Chou KC (1992) Energy-optimized structure of antifreeze protein and its binding mechanism. J Mol Biol 223:509–517. doi:10.1016/0022-2836(92)90666-8

Chou KC (1999) A key driving force in determination of protein structural classes. Biochem Biophys Res Commun 264:216–224. doi:10.1006/bbrc.1999.1325

Chou KC (2001) Prediction of protein cellular attributes using pseudo-amino acid composition. Proteins 43:246–255. doi:10.1002/prot.1035

Chou KC (2004) Review: structural bioinformatics and its impact to biomedical science. Curr Med Chem 11:2105–2134

Chou KC (2005) Prediction of G-protein-coupled receptor classes. J Proteome Res 4:1413–1418. doi:10.1021/pr050087t

Chou KC, Shen HB (2007a) Recent progress in protein subcellular location prediction. Anal Biochem 370:1–16. doi:10.1016/j.ab.2007.07.006

Chou KC, Shen HB (2007b) MemType-2L: a web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. Biochem Biophys Res Commun 360:339–345. doi:10.1016/j.bbrc.2007.06.027

Chou KC, Shen HB (2007c) Signal-CF: a subsite-coupled and window-fusing approach for predicting signal peptides. Biochem Biophys Res Commun 357:633–640. doi:10.1016/j.bbrc.2007.03.162

Chou KC, Shen HB (2008a) Cell-Ploc: a package of web servers for predicting subcellular localization of proteins in various organisms. Nat Protocols 3:153–162. doi:10.1038/nprot.2007.494

Chou KC, Shen HB (2008b) ProtIdent: a web server for identifying proteases and their types by fusing functional domain and sequential evolution information. Biochem Biophys Res Commun 376(2):321–325. doi:10.1016/j.bbrc.2008.1008.1125

Chou KC, Zhang CT (1995) Review: prediction of protein structural classes. Crit Rev Biochem Mol Biol 30:275–349. doi:10.3109/10409239509083488

Chou KC, Zhou GP (1982) Role of the protein outside active site on the diffusion-controlled reaction of enzyme. J Am Chem Soc 104:1409–1413. doi:10.1021/ja00369a043

Chou KC, Nemethy G, Scheraga HA (1984) Energetic approach to packing of a-helices: 2. General treatment of nonequivalent and nonregular helices. J Am Chem Soc 106:3161–3170. doi:10.1021/ja00323a017

Chou KC, Maggiora GM, Nemethy G, Scheraga HA (1988) Energetics of the structure of the four-alpha-helix bundle in proteins. Proc Natl Acad Sci USA 85:4295–4299. doi:10.1073/pnas.85.12.4295

Chou KC, Jones D, Heinrikson RL (1997) Prediction of the tertiary structure and substrate binding site of caspase-8. FEBS Lett 419:49–54. doi:10.1016/S0014-5793(97)01246-5

Chou JJ, Matsuo H, Duan H, Wagner G (1998) Solution structure of the RAIDD CARD and model for CARD/CARD interaction in caspase-2 and caspase-9 recruitment. Cell 94:171–180. doi:10.1016/S0092-8674(00)81417-8

Chou JJ, Li H, Salvessen GS, Yuan J, Wagner G (1999) Solution structure of BID, an intracellular amplifier of apoptotic signalling. Cell 96:615–624. doi:10.1016/S0092-8674(00)80572-3

Chou KC, Tomasselli AG, Heinrikson RL (2000) Prediction of the tertiary structure of a caspase-9/inhibitor complex. FEBS Lett 470:249–256. doi:10.1016/S0014-5793(00)01333-8

Chou KC, Wei DQ, Zhong WZ (2003) Binding mechanism of coronavirus main proteinase with ligands and its implication to drug design against SARS. (Erratum: ibid., 2003, Vol.310, 675). Biochem Biophys Res Commun 308:148–151

Chou KC, Wei DQ, Du QS, Sirois S, Zhong WZ (2006) Review: progress in computational approach to drug development against SARS. Curr Med Chem 13:3263–3270. doi:10.2174/092986706778773077

Ding YS, Zhang TL (2008) Using Chou's pseudo amino acid composition to predict subcellular localization of apoptosis proteins: an approach with immune genetic algorithm-based ensemble classifier. Pattern Recognit Lett 29:1887–1892. doi:10.1016/j.patrec.2008.06.007

Du QS, Huang RB, Chou KC (2008) Review: recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design. Curr Protein Pept Sci 9:248–259. doi:10.2174/138920308784534005

Evan G, Littlewood T (1998) A matter of life and cell death. Science 281:1317–1322. doi:
    10.1126/science.281.5381.1317

Gao QB, Wu CH, Ma XQ, Lu J, He J (2008) Classification of amine type G-protein coupled receptors
    with feature selection. Protein Pept Lett 15:834–842. doi:10.2174/092986608785203755

Huang J, Shi F (2005) Support vector machines for predicting apoptosis proteins types. Acta Biotheor
    53:39–47. doi:10.1007/s10441-005-7002-5

Jia P, Qian Z, Feng K, Lu W, Li Y, Cai Y (2008) Prediction of membrane protein types in a hybrid space.
    J Proteome Res 7:1131–1137. doi:10.1021/pr700715c

Jiang X, Wei R, Zhang T, Gu Q (2008) Using the concept of Chou's pseudo amino acid composition to
    predict apoptosis proteins subcellular location: an approach by approximate entropy. Protein Pept
    Lett 15:392–396. doi:10.2174/092986608784246443

Jin YH, Niu B, Feng KY, Lu WC, Cai YD, Li GZ (2008) Predicting subcellular localization with
    AdaBoost Learner. Protein Pept Lett 15:286–289. doi:10.2174/092986608783744234

Li FM, Li QZ (2008a) Using pseudo amino acid composition to predict protein subnuclear location with
    improved hybrid approach. Amino Acids 34:119–125. doi:10.1007/s00726-007-0545-9

Li FM, Li QZ (2008b) Predicting protein subcellular location using Chou's pseudo amino acid
    composition and improved hybrid approach. Protein Pept Lett 15:612–616. doi:10.2174/0929866
    08784966930

Li Y, Wei DQ, Gao WN, Gao H, Liu BN, Huang CJ, Xu WR, Liu DK, Chen HF, Chou KC (2007)
    Computational approach to drug design for oxazolidinones as antibacterial agents. Med Chem
    3:576–582. doi:10.2174/157340607782360362

Lin H (2008) The modified Mahalanobis discriminant for predicting outer membrane proteins by using
    Chou's pseudo amino acid composition. J Theor Biol 252:350–356. doi:10.1016/j.jtbi.2008.02.004

Lin H, Li QZ (2007a) Using pseudo amino acid composition to predict protein structural class:
    approached by incorporating 400 dipeptide components. J Comput Chem 28:1463–1466. doi:
    10.1002/jcc.20554

Lin H, Li QZ (2007b) Predicting conotoxin superfamily and family by using pseudo amino acid
    composition and modified Mahalanobis discriminant. Biochem Biophys Res Commun 354:548–551.
    doi:10.1016/j.bbrc.2007.01.011

Lin H, Ding H, Guo FB, Zhang AY, Huang J (2008) Predicting subcellular localization of mycobacterial
    proteins by using Chou's pseudo amino acid composition. Protein Pept Lett 15:739–744. doi:
    10.2174/092986608785133681

Munteanu CB, Gonzalez-Diaz H, Magalhaes AL (2008) Enzymes/non-enzymes classification model
    complexity based on composition, sequence, 3D and topological indices. J Theor Biol 254:476–482.
    doi:10.1016/j.jtbi.2008.06.003

Niu B, Jin YH, Feng KY, Liu L, Lu WC, Cai YD, Li GZ (2008) Predicting membrane protein types with
    bagging learner. Protein Pept Lett 15:590–594. doi:10.2174/092986608784966921

Raff M (1998) Cell suicide for beginners. Nature 396:119–122. doi:10.1038/24055

Reed JC, Paternostro G (1999) Postmitochondrial regulation of apoptosis during heart failure. Proc Natl
    Acad Sci USA 96:7614–7616. doi:10.1073/pnas.96.14.7614

Schulz JB, Weller M, Moskowitz MA (1999) Caspases as treatment targets in stroke and
    neurodegenerative diseases. Ann Neurol 45:421–429. doi:10.1002/1531-8249(199904)45:4<421::
    AID-ANA2>3.0.CO;2-Q

Shen HB, Chou KC (2007) EzyPred: a top-down approach for predicting enzyme functional classes and
    subclasses. Biochem Biophys Res Commun 364:53–59. doi:10.1016/j.bbrc.2007.09.098

Shen HB, Chou KC (2008a) PseAAC: a flexible web server for generating various kinds of protein pseudo
    amino acid composition. Anal Biochem 373:386–388. doi:10.1016/j.ab.2007.10.012

Shen HB, Chou KC (2008b) HIVcleave: a web-server for predicting HIV protease cleavage sites in
    proteins. Anal Biochem 375:388–390. doi:10.1016/j.ab.2008.01.012

Shi F, Chen QJ, Li NN (2008) Hilbert Huang transform for predicting proteins subcellular location.
    J. Biomed Sci Eng 1:59–63

Sirois S, Wei DQ, Du QS, Chou KC (2004) Virtual screening for SARS-CoV protease based on KZ7088
    pharmacophore points. J Chem Inf Comput Sci 44:1111–1122. doi:10.1021/ci034270n

Steller H (1995) Mechanisms and genes of cellular suicide. Science 267:1445–1449. doi:10.1126/
    science.7878463

Suzuki M, Youle RJ, Tjandra N (2000) Structure of Bax: coregulation of dimmer formation and
    intracellular location. Cell 103:645–654. doi:10.1016/S0092-8674(00)00167-7

Vapnik V (1998) Statistical learning theory. Wiley-Interscience, New York

Wang JF, Wei DQ, Chen C, Li Y, Chou KC (2008) Molecular modeling of two CYP2C19 SNPs and its implications for personalized drug design. Protein Pept Lett 15:27–32. doi:10.2174/092986608783330305

Zhang GY, Fang BS (2008) Predicting the cofactors of oxidoreductases based on amino acid composition distribution and Chou's amphiphilic pseudo amino acid composition. J Theor Biol 253:310–315. doi:10.1016/j.jtbi.2008.03.015

Zhang ZH, Wang ZH, Zhang ZR, Wang YX (2006) A novel method for apoptosis protein subcellular localization prediction combining encoding based on grouped weight and support vector machine. FEBS Lett 580:6169–6174. doi:10.1016/j.febslet.2006.10.017

Zheng H, Wei DQ, Zhang R, Wang C, Wei H, Chou KC (2007) Screening for new agonists against Alzheimer's disease. Med Chem 3:488–493. doi:10.2174/157340607781745492

Zhou GP, Doctor K (2003) Subcellular location prediction of apoptosis proteins. Proteins 50:44–48. doi:10.1002/prot.10251

Zhou XB, Chen C, Li ZC, Zou XY (2008) Improved prediction of subcellular location for apoptosis proteins by the dual-layer support vector machine. Amino Acids 35:383–388. doi:10.1007/s00726-007-0608-y