Amino Acids (2006) 30: 397–402
DOI 10.1007/s00726-006-0332-z

Amino Acids
Printed in Austria

Classifying G protein-coupled receptors and nuclear receptors on the basis of protein power spectrum from fast Fourier transform

Y.-Z. Guo[1,2], **M. Li**[1,2], **M. Lu**[1], **Z. Wen**[1,2], **K. Wang**[1], **G. Li**[1], and **J. Wu**[1]

[1] College of Chemistry, Sichuan University, Chengdu, China
[2] State Key Laboratory of Chemo/Biosensing and Chemometrics, Hunan University, Changsha, China

Received November 23, 2005
Accepted January 4, 2006
Published online May 26, 2006;

Summary. As the potential drug targets, G-protein coupled receptors (GPCRs) and nuclear receptors (NRs) are the focuses in pharmaceutical research. It is of great practical significance to develop an automated and reliable method to facilitate the identification of novel receptors. In this study, a method of fast Fourier transform-based support vector machine was proposed to classify GPCRs and NRs from the hydrophobicity of proteins. The models for all the GPCR families and NR subfamilies were trained and validated using jackknife test and the results thus obtained are quite promising. Meanwhile, the performance of the method was evaluated on GPCR and NR independent datasets with good performance. The good results indicate the applicability of the method. Two web servers implementing the prediction are available at http://chem.scu.edu.cn/blast/Pred-GPCR and http://chem.scu.edu.cn/blast/Pred-NR.

Keywords: G-protein coupled receptors – Nuclear receptors – Hydrophobicity – Fast Fourier transform – Power spectrum – Support vector machine

Introduction

G-protein coupled receptors (GPCRs) belong to the largest superfamily of cell-surface receptors and they are characterized by seven transmembrane segments. They play a key role in the basic cellular processes such as vision, smell, taste, neurotransmission, and metabolism and so on. They are major therapeutic targets of numerous prescribed drugs and more than 50% of all medicines available today act through GPCR (Gudermann et al., 1995). The sequences of thousands of GPCRs have been known (Horn et al., 2003), however many receptors remain orphaned (i.e. with unknown ligand specificity), and to date the crystal structure of only one GPCR (bovine rhodopsin) is solved (Palczewski et al., 2000). So it is highly desirable to develop the computational methods to facilitate the identification and characterization of novel receptors only using sequence information.

Methods have been developed to predict GPCRs. The covariant discriminant algorithm was proposed to predict GPCRs (Chou, 2005a; Chou and Elrod, 2002; Elrod and Chou, 2002), and support vector machines (SVMs) were used to classify GPCRs at family and subfamily level (Bhasin and Raghava, 2004a; Karchin et al., 2002). The methods based on profile-hidden Markov model (HMM) have been developed (Bateman et al., 2004; Papasaikas et al., 2004; Qian et al., 2003), and there has been a method of bagging classification tree for the classification of GPCRs (Huang et al., 2004).

Nuclear receptors (NRs) are important transcription factors involved in many physiological functions like cell growth, differentiation and homeostasis (Gronemeyer and Laudet, 1995; Mangelsdorf et al., 1995). Many of them are important drug targets in designing drugs for diseases such as breast cancer and diabetes (Robinson-Rechavi and Laude, 2003). The simple similarity-based search tools like BLAST and FASTA (Altschul et al., 1990; Pearson and Lipman, 1988) can easily distinguish NRs from the genome sequences, but they are not always successful in classifying the subfamilies of NRs. To overcome this limitation, a SVM based method (Bhasin and Raghava, 2004b) has been developed for NR subfamily classification, but only four subfamilies.

Based on the concept of pseudo amino acid composition (Chou, 2001), the Fourier transform spectra has been used to predict membrane protein type (Liu et al., 2005a; Wang et al., 2004), particularly their low-frequency parts

(Chou, 1988), have been used to predict membrane protein types. This paper describes a new combination of fast Fourier transform with support vector machine for the classification of all GPCR classes and NR subfamilies based on the hydrophobicity of proteins. The similar method has been successfully used for the prediction of GPCR subfamilies (Guo et al., 2005). The primary amino acid sequences are translated into numerical sequences using the hydrophobicity and then the numerical series are transformed into uniform matrix according to fast Fourier transform. Last, taking the protein power spectrum as input, SVM is used to construct classifiers.

Materials and methods

Data set

On the basis of pharmacological knowledge, the GPCRDB and NucleaRDB information systems (Horn et al., 2001) classify GPCRs into six different families and NRs into eight subfamilies respectively. The sequence data of GPCRs were collected from GPCRDB (release 9.0, March 2005) and the data of NRs were obtained from NucleaRDB (release 5.0, April 2005). All sequences denoted as 'putative', 'hypothetical' or 'orphan' and fragmental sequences were removed. Meanwhile, it was assured that none of the sequences was identical to others. Next, all sequences are partitioned into two parts, the training dataset and the test dataset. The newly publicized sequences that are marked as 'new' in the two databases were used as the independent dataset. All the remaining sequences were used as the training dataset. The final training dataset contained 946 sequences belonging to the six GPCR families and 465 sequences belonging to the eight NR subfamilies. For Class A of GPCRs, we chose 540 sequences randomly through equal interval selection (one in four), but for other classes of GPCRs, all the eligible sequences were selected because of the fewer members. For all the subfamilies of NRs, all the eligible sequences were chosen. Considering the limited amount of data available for some classes, such as GPCR Class E and NR Nerve Growth factor IB like subfamily, the proteins with high sequence identity were not removed in order to provide enough sequences to develop a wide-range predictive system that can be applied to all GPCR families and NR subfamilies. The number of sequences for each GPCR family and NR subfamily is listed in Tables 1 and 2, respectively.

Substitution models

Three kinds of substitution models: hydrophobicity model, electron-ion interaction potential (EIIP) model (Cosic, 1994) and c-p-v model (Grantham, 1974), representing three principal properties of hydrophobicity, electronic property and bulk respectively, are used to transform the protein sequences into numerical sequences. Hydrophobicity of proteins is one of the most important factors in determining a protein's structure and function. However, with different experimental conditions, different organic solvents and computing approaches, hydrophobicity value per amino acid will be different. So, three hydrophobicity scales, including KDHΦ (Kyte and Doolittle, 1982), MHΦ (Mandell et al., 1997) and FHΦ (Fauchére and Pliška, 1983) were selected for optimization. EIIP value describes the average energy states of all valence electron of amino acids and c-p-v model includes the composition (c), polarity (p) and molecular volume (v) of each amino acid.

Protein power spectrum

The Fourier transform (FT) has been commonly used in bioinformatics (Hiramoto et al., 2002; Katoh et al., 2002; Shepherd et al., 2003; Trad et al., 2002) because the frequency content of signals is often of great importance. It is a good method in capturing the essence of data. In this paper, fast Fourier transform (FFT) was used to transform proteins of variable length into fixed length vectors. The power spectrum or power spectral density, a measurement of the power at various frequencies was taken as the input of SVMs by using 512-point FFT.

Support vector machine

The support vector machine (SVM) is a kind of learning machine based on statistical learning theory. A brief and clear description for how to use SVM to do classification has been given by Chou and Cai (see, e.g., Chou and Cai, 2002; Cai et al., 2003). For a two-class classification problem, only one SVM classifier needs to be constructed, but the classification of GPCRs and NRs is a multi-class problem, so we used the 'one versus rest' method (Hua and Sun, 2001) to transfer it into a two-class problem.

The radial basis function (RBF) was selected as the kernel function. All the parameters were kept constant except for C (regulatory parameter) and σ (kernel width parameter). In the training process, C and σ were optimized. The fixed length feature vector was obtained using the protein power spectrum with the fixed number of frequency points.

Performance evaluation

The performance of all classifiers was examined by jackknife test because it is the most rigorous and objective way to do cross-validation as elaborated in a comprehensive review (Chou and Zhang, 1995), and nowadays it has been adopted by more and more leading investigators in the area of statistical prediction (see, e.g., Cai and Chou, 2005; Chou, 1995, 2005b; Chou and Cai, 2004; Gao et al., 2005; Liu et al., 2005b; Shen and Chou, 2005a, b; Xiao et al., 2006; Zhou, 1998; Zhou and Assa-Munt, 2001; Zhou and Doctor, 2003). Each receptor is selected as the test receptor and the remaining receptors are used to train the SVMs. The prediction quality was evaluated using accuracy, total accuracy and Matthew's correlation coefficient (MCC) (Matthews, 1975).

$$\text{accuracy}(i) = \frac{p(i)}{\exp(i)} \tag{1}$$

$$\text{total accuracy} = \frac{\sum_i^K p(i)}{\exp(i)} \tag{2}$$

$$\text{MCC} = \frac{p(i)n(i) - u(i)o(i)}{\sqrt{(p(i)+u(i))(p(i)+o(i))(n(i)+u(i))(n(i)+o(i))}} \tag{3}$$

Here, K is the class number, N is the total number of sequences, $\exp(i)$ is the number of sequences observed in class i, $p(i)$ is the number of correctly predicted sequences of class i, $n(i)$ is the number of correctly predicted sequences not of class i, $u(i)$ is the number of under-predicted sequences, and $o(i)$ is the number of over-predicted sequences.

The measurement of prediction reliability is absolutely necessary when using the machine learning approaches for prediction. Here the index indicating the reliability of prediction (R) (Novic and Zupan, 1995) was used, as given by:

$$R(i) = \frac{2(\text{accuracy}(i) - \text{error}(i))}{1 + |\text{accuracy}(i) - \text{error}(i)|} \tag{4}$$

where, $\text{error}(i) = \frac{o(i)}{n(i)+o(i)}$

The reliability value $R(i)$ ranges from 1 to -1. In the best case, when all the receptors are correctly predicted, $R(i)$ is maximal (equal to 1), that is

when accuracy(i) = 1 and error(i) = 0. And $R(i)$ is -1 in the worst case when accuracy(i) = 0 and error(i) = 1.

Results and discussion

All the results of the following experiments are obtained from the datasets in which proteins with high sequence identity were not removed.

Selecting the optimal substitution model

To select the optimal substitution model for GPCRs and NRs, the performances of this method based on the three kinds of models were evaluated by two-fold cross validation respectively. For GPCRs, the total accuracies of this method based on the FHΦ, KDHΦ, MHΦ, c-p-v and EIIP models are 93.4%, 92.6%, 90.0%, 90.7%, 82.7% respectively, and for NRs, 95.1%, 92.9%, 91.2%, 91.6%, 92.6% respectively. We can see from the results that the method perform well on GPCRs and NRs using any one of the three hydrophobicity scales with all accuracies of $\geq$90.0%, but the method based on FHΦ achieves the highest accuracy. So in this work, the scheme FHΦ was chosen as coding scheme for GPCRs and NRs.

Model training and testing

Fourteen SVMs were constructed for six GPCR families and eight NR subfamilies using FHΦ. Each SVM was trained and validated using jackknife test. The results are summarized in Tables 1 and 2 respectively.

From Tables 1 and 2, the total results from the jackknife test are quite promising and prove the good performance of this method. Moreover, we can see that different classes have different prediction accuracies. For GPCRs and NRs, there is no apparent direct relationship between the prediction accuracy and class size. From Table 1, cAMP, the smallest family of GPCR that only contains 5 sequences, achieves the highest accuracy among all the families of GPCR. However, Fungal pheromone family that contains 21 sequences achieves the lowest accuracy (81.0%). From Table 2, 0A Knirps-like and 0B DAX-like both contain 7 receptors, but accuracies are 42.9% and 71.4% respectively. The smallest subfamily of NR is Germ cell nuclear factor-like that only contain 2 sequences, but it gives the highest accuracy. Nerve Growth factor IB-like subfamily contains 5 sequences, but only one sequence is incorrectly predicted.

Table 1. The performance of the method in classifying the six families of GPCRs using jackknife test based on hydrophobicity scale, FHΦ

GPCR family	No. of sequences	Accuracy (%)	MCC	R
Rhodopsin-like (Class A)	540	97.0	0.93	0.97
Secretin-like (Class B)	187	96.3	0.94	0.95
Metabotropic glutamate (Class C)	103	94.2	0.95	0.95
Fungal pheromone (Class D)	21	81.0	0.92	0.90
cAMP receptors (Class E)	5	100	1.0	1.0
Frizzled/smoothened (Class F)	90	95.6	0.94	0.94
Total	946	96.1	–	–

Table 2. The performance of the method in classifying the eight subfamilies of NRs using jackknife test based on hydrophobicity scale, FHΦ

NR subfamily	No. of sequences	Accuracy (%)	MCC	R
Thyroid hormone-like	165	95.8	0.95	0.97
HNF4-like	114	97.4	0.96	0.96
Estrogen-like	130	97.7	0.96	0.98
Fushitarazu-F1 like	35	94.3	0.97	0.97
Nerve growth factor IB-like	5	80.0	0.89	0.89
Germ cell nuclear receptor	2	100	1.0	1.0
0A Knirps-like	7	42.9	0.65	0.60
0B DAX-like	7	71.4	0.84	0.83
Total	465	95.3	–	–

Recognition of GPCRs from non-GPCR transmembrane proteins and NRs

The dataset of 1090 non-GPCR transmembrane protein sequences was collected from the Swiss-Prot (Release 46.5, 2005) and TrEMBL (Release 29.5, 2005). This dataset was excluded the sequences marked as 'putative', 'hypothetical' and 'fragment' but also contained proteins with high sequence identity. We constructed two SVM models based on hydrophobicity scale FHΦ for identifying GPCRs from non-GPCR transmembrane proteins and from NRs separately. The performance of each model was validated with 5-fold cross-validation test.

This method can differentiate GPCRs from non-GPCR transmembrane proteins with the accuracy, MCC and R of 95.0%, 0.88, 0.94, respectively and from NRs, 99.5%, 0.98, 0.99 respectively. Theoretically for GPCRs, selecting hydrophobicity of amino acids as the coding scheme seems only to stress the features of hydrophobic segments

and not to cover the features of extracellular domains. But the high accuracy of this method in classifying GPCRs and non-GPCR transmembrane proteins indicates that choosing hydrophobicity is reasonable in spite of the fact that GPCRs respond to a variety of ligands through their extracellular and transmembrane domains.

Performance on the independent dataset and comparison with other methods

It is necessary to check the practical application of the method using an independent dataset. From GPCRDB (release 9.0, March 2005) and NucleaRDB (release 5.0, April 2005), the newly publicized sequences (denoted as 'new') were collected as the independent datasets for unbiased evaluation of this method and Bhasin and Raghava's method (Bhasin and Raghava, 2004a, b) and Papasaikas et al.'s method (Papasaikas et al., 2004). There are 458 GPCRs and 128 NRs for the independent datasets. These sequences are not contained in the training datasets. We chose dipeptide composition based approach developed by Bhasin and Raghava (2004b) to predict the new NR sequences. The results are listed in Tables 3 and 4.

Table 3 indicates that this method can correctly predict 440 out of 458 sequences of GPCRs. Not only the largest family (Class A) but also other five smaller families are predicted with high accuracy. For the anterior five families (Class A–E), 421 out of 439 GPCRs are correctly predicted by this method. Bhasin and Raghava's method and Papasaikas et al.'s method also achieve good performance. Table 4 shows that both this method and Bhasin and Raghava's method can predict the anterior four subfamilies of NR well. Moreover this method predicts successfully 125 out of 128 NRs belonging to the anterior six subfamilies. For all GPCR families and NR subfamilies, we can draw a conclusion from the good prediction results obtained by our method that this method is not overfitted and has powerful prediction ability.

Table 3. Performance of our method, as compared to Bhasin and Raghava (2004a) and Papasaikas et al. (2004) on the GPCR independent dataset at class level

GPCR families	Total sequences	Correctly predicted sequences		
		This study	Bhasin and Raghava (2004a)	Papasaikas et al. (2004)
Rhodopsin-like	345	332 (96.2%)	345 (100%)	275 (79.7%)
Secretin-like	35	32 (91.4%)	26 (74.3%)	18 (51.4%)
Metabotropic glutamate	23	22 (95.7%)	13 (56.5%)	16 (69.6%)
Fungal pheromone	34	34 (100%)	14 (41.2%)	15 (44.1%)
cAMP receptors	2	1 (50%)	1 (50%)	1 (50%)
Total	439	421 (95.9%)	399 (91.7%)	325 (71.0%)
Frizzled/Smootened	19	19 (100%)	–	–
Total	458	440 (96.1%)	–	–

Table 4. Performance of our method and Bhasin and Raghava's method on the NR independent dataset at subfamily level

NR subfamilies	Total sequences	Correctly predicted sequences	
		This study	Bhasin and Raghava (2004)
Thyroid hormone-like	40	38	39
HNF4-like	35	35	35
Estrogen-like	41	41	38
Fushitarazu-F1-like	3	3	3
Total	118	116 (98.3%)	114 (96.6%)
Nerve Growth factor IB-like	9	8	–
Germ cell nuclear factor-like	1	1	–
0A Knirps-like	0	0	–
0B DAX-like	0	0	–
Total	128	125 (97.7%)	–

Conclusion

This paper describes a new method of FFT-based SVM for the classification of GPCRs and NRs. The information about the features of a protein is extracted from the power spectrum of FFT based on the hydrophobicity of proteins. The prediction results illustrate the powerful ability of the method to classify GPCRs and NRs, which testifies the attempt to try to mine the frequency-power features of proteins is feasible and successful.

The hydrophobic value sequences of variable length are transformed into the fixed-length vectors using FFT, which meets the requirement of SVM. In this study, we also selected the c-p-v model and EIIP model to transform the amino acid sequences into numerical sequences, but the prediction results were not comparable with those of the hydrophobicity scale. It is obvious that the substitution model will affect the prediction performance. So it is anticipant to develop the better substitution models for GPCRs and NRs. However, the establishment of such an accurate prediction method will facilitate the recognition of the novel GPCRs and NRs.

Acknowledgement

The authors would like to thank the anonymous reviewers for their patient review and constructive suggestions.

References

Altschul SF, Gish W, Miller W (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, Studholme DJ, Yeats C, Eddy SR (2004) The pfam protein families database. Nucleic Acids Res 32: D138–D141

Bhasin M, Raghava GPS (2004a) GPCRpred: an SVM-based method for prediction of families and subfamilies of G-protein coupled receptors. Nucleic Acids Res 32: W383–W389

Bhasin M, Raghava GPS (2004b) Classification of nuclear receptors based on amino acid composition and dipeptide composition. J Biol Chem 279: 23262–23266

Cai YD, Zhou GP, Chou KC (2003) Support vector machines for predicting membrane protein types by using functional domain composition. Biophys J 84: 3257–3263

Cai YD, Chou KC (2005) Using functional domain composition to predict enzyme family classes. J Proteome Res 4: 109–111

Chou KC (1988) Review: Low-frequency collective motion in biomacromolecules and its biological functions. Biophys Chem 30: 3–48

Chou KC (1995) A novel approach to predicting protein structural classes in a (20-1)-D amino acid composition space. Proteins 21: 319–344

Chou KC (2001) Prediction of protein cellular attributes using pseudo amino acid composition. Proteins 43: 246–255 (Erratum: ibid., 2001, 44: 60)

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

Chou KC (2005b) Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. Bioinformatics 21: 10–19

Chou KC, Cai YD (2002) Using functional domain composition and support vector machines for prediction of protein subcellular location. J Biol Chem 277: 45765–45769

Chou KC, Cai YD (2004) Predicting protein structural class by functional domain composition. Biochem Biophys Res Commun 321: 1007–1009 (Corrigendum: ibid., 2005, 329: 1362)

Chou KC, Elrod DW (2002) Bioinformatical analysis of G-protein-coupled receptors. J Proteome Res 1: 429–433

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

Cosic I (1994) Macromolecular bioactivity: is it resonant interaction between macromolecules? – Theory and applications. IEEE Trans Biomed Eng 41: 1101–1114

Elrod DW, Chou KC (2002) A study on the correlation of G-protein coupled receptor types with amino acid composition. Protein Eng 15: 713–715

Fauchére J, Pliška V (1983) Hydrophobic parameters Φ of amino-acid side chains from the partitioning of n-acetyl-amino-acid amides. Eur J Med Chem Chim Ther 18: 369–375

Gao Y, Shao SH, Xiao X, Ding YS, Huang YS, Huang ZD, Chou KC (2005) Using pseudo amino acid composition to predict protein subcellular location: approached with Lyapunov index, Bessel function, and Chebyshev filter. Amino Acids 28: 373–376

Grantham R (1974) Amino acid difference formular to help explain protein evolution. Science 185: 862–864

Gronemeyer H, Laudet V (1995) Transcription factors 3: nuclear receptors. Protein Profile 2: 1173–1308

Gudermann T, Nürnberg B, Schultz G (1995) Receptors and G proteins as primary components of transmembrane signal transduction. Part 1. G-protein-coupled receptors: Structure and function. J Mol Med 73: 51–63

Guo YZ, Li ML, Wang KL, Wen ZN, Lu ML, Liu LX, Jiang L (2005) Fast Fourier transform-based support vector machine for prediction of G-protein coupled receptor subfamilies. Acta Biochim Biophys Sin 37: 759–766

Hiramoto T, Nemoto W, Kikuchi T, Fujita N (2002) Construction of hypothetical three-dimensional structure of $P2Y_1$ receptor based on Fourier transform analysis. J Protein Chem 21: 537–545

Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, Vriend G (2003) GPCRDB information system for G protein coupled receptors. Nucleic Acids Res 31: 294–297

Horn F, Vriend G, Cohen FE (2001) Collecting and harvesting biological data: the GPCRDB and NucleaRDB information systems. Nucleic Acids Res 29: 346–349

Hua SJ, Sun ZR (2001) Support vector machine approach for protein subcellular localization prediction. Bioinformatics 17: 721–728

Huang Y, Cai J, Ji L, Li YD (2004) Classifying G-protein coupled receptors with bagging classification tree. Comput Biol Chem 28: 39–49

Karchin R, Karplus K, Haussler D (2002) Classifying G-protein coupled receptors with support vector machines. Bioinformatics 18: 147–159

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30: 3059–3066

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105–132

Liu H, Wang M, Chou KC (2005a) Low-frequency Fourier spectrum for predicting membrane protein types. Biochem Biophys Res Commun 336: 737–739

Liu H, Yang J, Ling JG, Chou KC (2005b) Prediction of protein signal sequences and their cleavage sites by statistical rulers. Biochem Biophys Res Commun 338: 1005–1011

Mandell AJ, Selz KA, Shlesinger MF (1997) Wavelet transformation of protein hydrophobicity sequences suggests their memberships in structural families. Physica (A) 244: 254–262

Mangelsdorf DJ, Thummel C, Beato M, Herrlich P, Schütz G, Umesono K, Blumberg B, Kastner P, Mark M, Chambon P, Evans RM (1995) The nuclear receptor superfamilily: the second decade. Cell 83: 835–839

Matthews BW (1975) Comparison of predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta 405: 442–451

Novic M, Zupan J (1995) Investigation of infrared spectra-structure correlation using kohonen and counterpropagation neural network. J Chem Inf Comput Sci 35: 454–466

Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M (2000) Crystal structure of rhodopsin: a G-protein coupled receptor. Science 289: 739–745

Papasaikas PK, Bagos PG, Litou ZI, Promponas VJ, Hamodrakas SJ (2004) PRED-GPCR: GPCR recognition and family classification server. Nucleic Acids Res 32: W380–W382

Pearson WR, Lipman DJ (1988) Improved tool for biological sequence comparison. Proc Natl Acad Sci USA 85: 2444–2448

Qian B, Soyer OS, Neubig RR, Goldstein RA (2003) Depicting a protein's two faces: GPCR classification by phylogenetic tree-based HMMs. FEBS Lett 554: 95–99

Robinson-Rechavi M, Laude V (2003) Bioinformatics of nuclear receptors. Methods Enzymol 364: 95–118

Shen HB, Chou KC (2005a) Using optimized evidence-theoretic K-nearest neighbor classifier and pseudo amino acid composition to predict membrane protein types. Biochem Biophys Res Commun 334: 288–292

Shen HB, Chou KC (2005b) Predicting protein subnuclear location with optimized evidence-theoretic K-nearest classifier and pseudo amino acid composition. Biochem Biophys Res Commun 337: 752–756

Shepherd AJ, Gorse D, Thornton JM (2003) A novel approach to the recognition of protein architecture from sequence using Fourier analysis and neural networks. Proteins 50: 290–302

Trad CH, Fang Q, Cosic I (2002) Protein sequence comparison based on the wavelet transform approach. Protein Eng 15: 193–203

Wang M, Yang J, Liu GP, Xu ZJ, Chou KC (2004) Weighted-support vector machines for predicting membrane protein types based on pseudo amino acid composition. Protein Eng Des Sel 17: 509–516

Xiao X, Shao SH, Ding YS, Huang ZD, Chou KC (2006) Using cellular automata images and pseudo amino acid composition to predict protein sub-cellular location. Amino Acids 30: 49–54

Zhou GP (1998) An intriguing controversy over protein structural class prediction. J Protein Chem 17: 729–738

Zhou GP, Assa-Munt N (2001) Some insights into protein structural class prediction. Proteins 44: 57–59

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

Authors' address: Prof. Menglong Li, College of Chemistry, Sichuan University, Chengdu, China,
Fax: +86-28-85412356; E-mail: liml@scu.edu.cn