Chinese Chemical Letters Vol. 16, No. 12, pp 1645-1648, 2005 1645
http://www.imm.ac.cn/journal/ccl.html

Sequence Hiberarchy Evolution Measurement Method and its Application

Jian Ding QIU[1]*, Ru Ping LIANG[1], Jin Yuan MO[2], Xiao Yong ZOU[2]

[1] Department of Chemistry, Nanchang University, Nanchang 330047
[2] School of Chemistry and Chemical Engineering, Zhongshan University, Guangzhou 510275

Abstract: A novel method based on discrete wavelet transform (DWT) and cross-covariance for revealing the evolution of species at different spatial resolutions is presented. The trypsin proteins of different species are chosen as an example to describe the evolution relationship according to the evolution vectors by using this method. The results indicated that this method is a promising approach to reveal species evolution at different spatial resolutions.

Keywords: Sequence hiberarchy evolution measurement, protein, evolution, hydrophobicity.

Life is a complex system evolving from basic composition (such as protein and nucleic acid) over a long time. It is very important to study molecular evolution to reveal the origin and evolution mechanism of life. Proteins that have descended from the same ancestor retain the memory in its sequence, structure and function. Strong sequence similarity alone is considered to be a sufficient evidence for the common ancestry. The inability of current sequence alignment to analyze the proteins with similarity less than 20% makes it very difficult for the current phylogenetic methods to analyze proteins evolution.

The wavelet transform (WT) is an efficient signal-processing tool for multi-resolution analysis and local feature extraction of non-stationary signals[1-3]. The sequence hiberarchy evolution measurement (SHEM) introduced here is based on the discrete wavelet transform (DWT) and the cross-correlation analysis of numerical representation of protein sequences. The compared sequences are initially converted into numerical series using KD hydrophibicity[4-5]. Then they are decomposed to j levels with details from level 1 to level j and an approximation at level j by DWT (using *Bior*3.3 wavelet). Because a correlation function quantifies the degree of interdependence of one process upon another or establishes the evolution between one set of data and another, the cross-correlation coefficients are calculated at each level to establish and quantify the evolution between the two compared protein sequences. There are a total of j+1 correlation coefficients lie between –1 and +1; +1 means 100% correlation in the same sense and –1 means 100% correlation in the opposing sense. The cross-

* E-mail: jdqiu@ncu.edu.cn

correlation coefficients are defined as follows[6]:

$$\rho^{12}(k) = \frac{r^{12}(k)}{\frac{1}{N}[\Sigma_{n=0}^{N-1} s_1^2(n) \Sigma_{n=0}^{N-1} s_2^2 s(n)]^{\frac{1}{2}}} \quad k = 0, \pm 1, \pm 2, \pm 3, \qquad (1)$$

Where N is the signal length, k is the number of lag and $r_{xy}(n)$ is an estimate of the cross-covariance and defined as:

$$r^{12}(k) = \frac{1}{N} \Sigma_{n=0}^{N-1} s_2(n) s_1(n-k) \qquad (2)$$

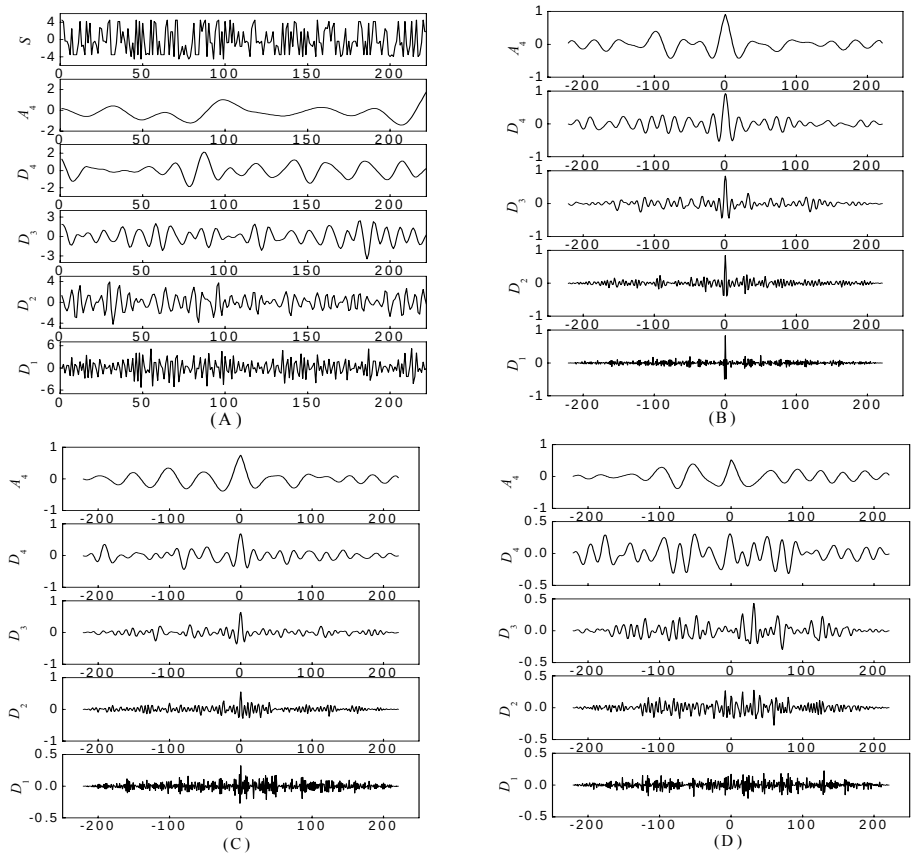
The maximum absolute value of the correlation coefficients at each decomposition level is regarded as the evolution score for the two proteins at that level. Therefore, a total of j+1 maximum values are taken out to form the sequence hiberarchy evolution vectors. The sequence hiberarchy evolution vectors depict the evolution of two proteins at different scales or different frequency bands. More specifically, the vectors describe the correlation with a multi-resolution point of view.

In this section, trypsin proteins of different species are chosen as an example to describe how to use the sequence hiberarchy evolution measurement (SHEM) to reveal the evolution of species. The discrete wavelet transform up to level 4 of a protein signal, trypsin human, is shown in **Figure 1 (A)**. The original protein signal is represented by S, A_n denotes the approximation at level n and D_n denotes the details at level n. Once the multi-level decompositions of two species proteins sequences have been obtained by such a way, their cross-correlation coefficients at different details and approximation can be calculated using Equation (1).

Figure 1 (B) shows the cross-correlation coefficients between the DWTs of trypsin human and bovine. The abscissa is the amino acid position along the protein backbone and the ordinate is the magnitude of the cross-coefficient. For biomedical signals, it is deemed strongly correlated if the correlation coefficient exceeds ± 0.7 and weakly correlated if the correlation coefficient is between ± 0.7 and ± 0.5 [7]. The computed sequence hiberarchy evolution vectors, which are 0.9083, 0.9242, 0.8391, 0.8426, 0.8313, following the order from A_4, D_4, D_3, D_2 to D_1, reveal a strong correlation at all resolution levels between the two species proteins. **Figure 1(C)** shows the sequence hiberarchy evolution analysis of trypsin human and salmon. The evolution vectors are 0.7450, 0.6890, 0.6401, 0.5466, 0.3260, revealing one strongly correlated frequency band A_4 and three weakly correlated frequency bands D_4, D_3, and D_2. **Figure 1(D)** shows the sequence hiberarchy evolution analysis of trypsin human and streptomyces griseus. There is no strong cross-correlation but a weak correlation at A_4. We also use this method to reveal the evolution relationship between trypsin human and pig, human and rat, human and fusarium oxysporum, and the results are shown in **Table 1**. From **Table 1** we can see clearly that there is a close correlation between the magnitude of cross-correlation coefficient and the evolution position of different species, that is, the farther the evolution position between each other, the smaller the cross-correlation coefficient is, along with the weaker correlation. Therefore, the evolution relationship can be determined conveniently and accurately according to the evolution vectors. Concretely, comparing with human, the evolution relationship is in order: bovine and pig

> rat > salmon > fusarium oxysporum > streptomyces griseus, which is consistent with that of sequence alignment[8]. Hemoglobin and cytochrome c of different species are also investigated by this method and the same results are obtained.

Figure 1 The DWT of trypsin human protein and the cross-correlation coefficients plots



(A) The DWT of trypsin human protein; (B) the cross-correlation coefficients between human and bovine; (C) human and salmon; (D) human and streptomyces griseus

Table 1 The computed evolution vectors of trypsin proteins between human and other species using sequence hiberarchy evolution measurement *

	h-b	h-p	h-r	h-s	h-f	h-st
A_4	0.9083	0.8436	0.8340	0.7450	0.5391	0.5261
D_4	0.9242	0.8455	0.8037	0.6890	0.5861	0.3100
D_3	0.8391	0.9034	0.7859	0.6401	0.1501	0.4303
D_2	0.8426	0.9138	0.6373	0.5466	0.0723	0.1851
D_1	0.8313	0.9031	0.5528	0.3260	0.3305	0.1710

* h-b: human and bovine; h-p: human and pig; h-r: human and rat; h-s: human and salmon; h-f: human and fusarium oxysporum; h-st: human and streptomyces griseus

The results confirm that our method based on DWT and the cross-correlation analysis can be established as a novel and convenient approach to reveal species evolution at different spatial resolutions, which promising a tremendous development foreground.

Acknowledgment

We thank the National Natural Science Foundation of China (Project No.29975033) and the Education Office Program of Jiangxi province ([2005] 242) for financial support.

References

1. J. D. Qiu, R. P. Liang, X. Y. Zou, *et al*., *Chin. Chem. Lett.,* **2004**, *15*(*6*), 711.
2. J. B. Zheng, R. Zhao, H. Q. Zhang, *et al*., *Chinese J. Anal. Chem*. , **1999**, *27*(7), 855.
3. X. G. Shao, C. Y. Pang, L. Sun. *Progress in Chemistry*, **2000**, *12*(3), 233.
4. J. Kyte, R. F. Doolittle. *J. Mol. Biol*., **1982**, *157*, 105.
5. J. D. Qiu, R. P. Liang, X. Y. Zou, *et al*., *J. Chem. Inf. Comput. Sci*., **2004,** *44*(2), 741.
6. Y. W. Zhang, "*Mathematics Method to Prediction*", National Defence Industry Press, Beijing, **1991**.
7. C. K. Oyster, W. O. Hanten, L. A. Liorence, "*Introduction to Research: A Guide for The Health Science Professional*", Lippincott, Oxford, **1987**.
8. Z. R. Sun, Y. Wang, S. M. Hu, *et al*., *Acta Biophysica Sinica*, **1999**, *15*, 530.

Received 9 May, 2005