A Novel Approach to the Prediction of Transmembrane Proteins

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

[1] School of Chemistry and Chemical Engineering, Zhongshan University, Guangzhou 510275
[2] Department of Chemical Engineering, Pingxiang College, Pingxiang 337055

Abstract: A novel method based on continuous wavelet transform (CWT) for predicting the number and location of helices in membrane proteins is presented. The PDB code of 1yst is chosen as an example to describe the prediction of transmembrane helices (HTM) by using CWT. The results indicate that CWT is a promising approach for the prediction of HTM.

Keywords: Continuous wavelet transform, transmembrane helices, hydrophobicity, prediction.

Integral membrane proteins play important and functionally diverse roles in living cells. Due to the difficulties in crystallization of membrane proteins, we can not determine the experimental structure of these proteins by means of X-ray crystallography. Since prediction methods are the most convenient and least expensive ways for determining the protein structures, there is a great demand for developing efficient prediction methods.

In the early 1980s, wavelets have become a popular signal analysis tool due to their ability to elucidate simultaneously both spectral and temporal information within the signal. A mathematical introduction to the wavelet theory was reported by F.S. Yang[1]. In this work, a suite of continuous wavelet transform (CWT) analysis techniques were used to detect the location of helices in membrane proteins. This method consists of three main steps. First the amino acids of membrane protein are transformed into sequences of hydrophobic free energies per residue, the hydrophobic data utilized in this study are the Kyte-Doolittle hydrophobicity (HΦ) which, for each residue of the protein, provided the information derived from the protein structure and sequence[2]. Secondly, the hydrophobic profile is decomposed into wavelet coefficients, using CWT. At the third step the number and the relevant positions of HTM can be extracted easily and accurately according to the maxima of wavelet coefficient in corresponding continuous wavelet transform plot of hydrophobic value sequences with appropriate scale. This method does not need to set a data-dependent threshold for selecting wavelet coefficient and to recover the denoised signal by transforming these threshold coefficients[3].

In this section, 1yst was chosen from the Brookhaven Protein Databank (PDB) as an example to describe the prediction of the location of helices in membrane proteins by using CWT. **Figure 1(A)** depicts the input hydrophobicity data (HΦ) of 1yst; no obvious

---

* E-mail: ceszxy@zsu.edu.cn

evidence of HTM is apparent from visual inspection of the data.   While processing these HΦ data by CWT (using Morlet wavelet), the HΦ wavelet scalogram (**Figure 1(B)**) unfolds the distance-scale organization of the data for scales of 25 to 100.   At the range of 75 to 100 dilations, five dark regions on the scalogram indicate that the membrane protein has five HTM.   In order to further investigate the correlation between the wavelet coefficients and HTM of protein, we perform the wavelet coefficients with dilation a=32. From the plot, we can see clearly that there are 5 maxima, which correspond to the result of **Figure 2**, the wavelet coefficients are indicated on y axis; the x axis indicates the position along the sequence.   From the maxima of the plot we can easily identify the numbers and the positions of HTM, which could be clearly seen in **Table 1**.   **Table 1** shows the observed topology and locations of HTM segments of the sequence of 1yst, the prediction was made by using CWT.   In order to investigate the feasibility of our method, a test set of 83 membrane proteins has been used.   The percentage of HTM segments correctly predicted is 97.8%. The results demonstrate that the desired number and positions of transmembrane helices can be predicted conveniently and accurately by this method which promises a tremendous development foreground.

Table 1   Observed and predicted HTM segments

	TM1	TM2	TM3	TM4	TM5
Observed	32-55	83-111	116-138	171-198	225-250
Predicted	31-55	78-100	119-147	174-198	221-250

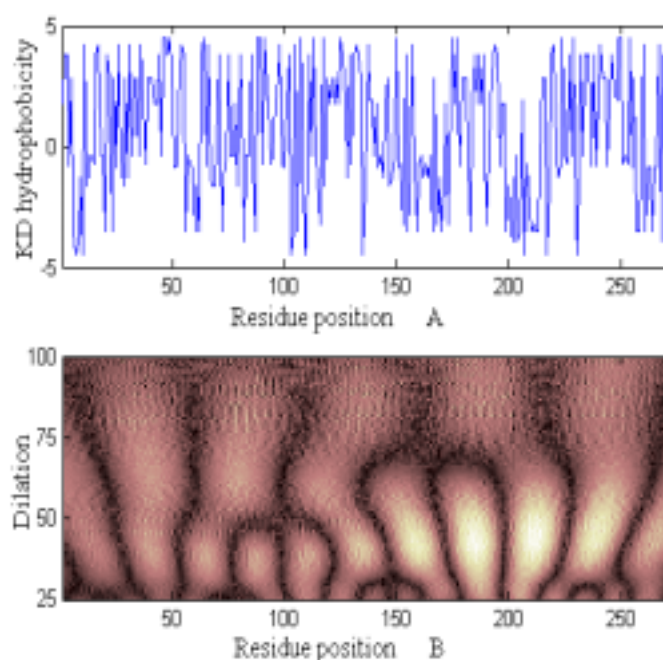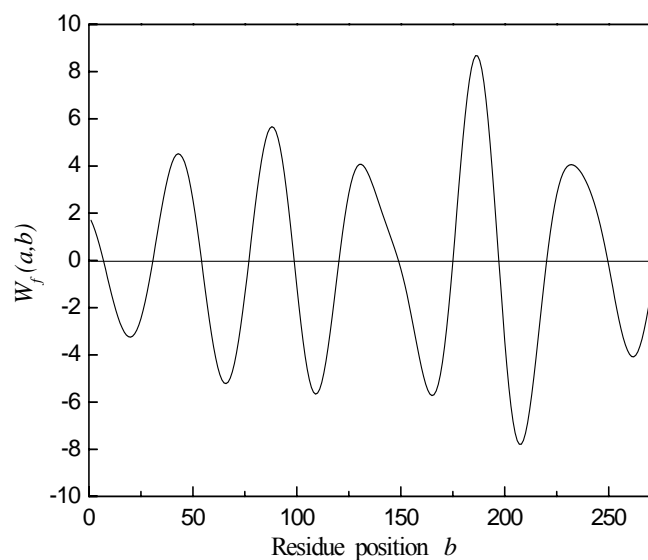**Figure 1**   The CWT of membrane protein 1yst (A) HΦ data; (B) HΦ scalogram by using CWT

Figure 2   The CWT coefficient plot (dilation a=32)

References

1.   F.S. Yang, "*The Engineering Analysis and Application of Wavelet Transform*", Science Press, Beijing, **1999**, pp. 1~26.
2.   J. Kyte, R.F. Doolittle. *J. Mol. Biol.*, **1982**, *157*, 105.
3.   P. Lio, M. Vannucci. *Bioinformatics.*, **2000**, *16*, 376.

Received 6 May, 2003