# Prediction of Transmembrane Proteins Based on the Continuous Wavelet Transform

Jianding Qiu,[†,‡] Ruping Liang,[†,‡] Xiaoyong Zou,*[,†] and Jinyuan Mo[†]

School of Chemistry and Chemical Engineering, Zhongshan University, Guangzhou 510275, People's Republic of China, and Department of Chemical Engineering, Pingxiang College, Pingxiang 337055, People's Republic of China

Received August 1, 2003

A novel method based on continuous wavelet transform (CWT) for predicting the number and location of helices in membrane proteins is presented. Two bacteria proteins are chosen as examples to describe the prediction of transmembrane helices (HTM) by using this method. Selections of an appropriate dilation and hydrophobicity data types are discussed in the text. The results indicate that CWT is a promising approach for the prediction of HTM.

## INTRODUCTION

Genome sequencing projects are generating thousands of sequences of new proteins from diverse organisms, of which a surprisingly high number contain membrane-embedded domains. Integral membrane proteins play important and functionally diverse roles in living cells. While the biological function of protein is mainly determined by its spatial structure, so the identification of the spatial structure of protein is the basis of the study of its biological function. Due to the difficulties in crystallization of membrane proteins (MP), the experimental structure of these proteins cannot efficaciously determinate by means of X-ray crystallography. Compared to the >9000 nonmembrane protein structures that have been solved, only a handful of membrane proteins have been characterized by high-resolution structural techniques. Since prediction methods are the most convenient and least expensive ways of determining protein structures based on the known structure of some proteins, there is a great demand for developing efficient prediction methods.

It is commonly accepted that topology prediction of membrane proteins is easier and results in higher accuracy than the prediction of the secondary structure of globular proteins. Hydropathy analysis has been used extensively as a standard approach to predict transmembrane (TM) segments from primary sequence in the absence of experimental data. Kyte and Doolittle[1] introduced a method which was based on the recognition of the existence of long hydrophobic segments in membrane protein, for the analysis of the hydropathy profile that used a sliding window of 19 residues. Segments with hydrophobicity above an average value were predicted to be membrane-spanning. A similar approach was adopted by Engelman et al.,[2] but using a different hydrophathy scale. von Heijne[3] proposed a trapezoid sliding window in his hydrophobic analysis of TM sequences in conjunction with a consideration of positive-changed residues interior to the membrane (the "positive-inside rule"). A hidden Markov model with special architecture was developed to search TM topology corresponding to the maximum likelihood among all the possible topologies of a given protein.[4] The predicted results suggested that the topologies of the TM proteins were determined by the maximum divergence of the amino acid composition of sequence segments.[5] As the number of experiments dealing with topology increased in the past few years, resulting in more reliable data, Jones et al.[6] used a dynamic programming algorithm to recognize the secondary structure and topology of integral membrane proteins based on the statistical data derived from well-characterized membrane proteins. Using the advantages of neural-network-based algorithms and combining prediction methods with multiple alignments, the accuracy of the prediction of the TM helices reached the 90−95% level.[7−10] These methods perform generally well because they have a very rich theoretical structure but require the specification of good models of TM segments and loop region.[11] Some of these prediction methods require the user to specify a minimal and maximal HTM segments length in order to fix the sequence scanning window; others make assumptions regarding the number of residues spanning the membrane.

Introduced 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.[12] This overcomes the basic shortcoming of Fourier analysis, which is based on functions that are localized in frequency, not in time, thus leading to location-specific features in the signal being lost. Using wavelets in proteins is relatively less and comparatively recent. Murray et al.[13] characterize and detect repeating motifs from hydropathy and accessibility data based on the continuous wavelet transform (CWT). Using the discrete wavelet transform (DWT) to predict the hydrophobic cores of globular proteins,[14] the prediction accuracy of the cores of nonhomologous proteins is 70%. Mandell et al.[15] apply the CWT to protein hydrophobicity sequences to unveil which structural family the sequence belongs to. Qiu et al.[16] adopt CWT to predict α-helices and connecting peptides based on the theory of hydrophobic minima. The wavelet transform is also introduced to measure the similarity between two or more protein

* Corresponding author fax: +86-20-84112245; e-mail: ceszxy@zsu.edu.cn. Address correspondence to the Zhongshan University address.
† Zhongshan University.
‡ Pingxiang College.

10.1021/ci0303868 CCC: $27.50 
Published on Web 02/12/2004

sequences at different resolution scales.[17] Lio and Vannucci[11] developed a discrete wavelet threshold technique to predict the location and topology of helices in TM proteins. However, setting a data-dependent threshold which is used to select wavelet coefficients that detect abrupt changes in the profile and reconstructing the denoised profile from the selected wavelet coefficients may be losing some useful information about the structures of membrane proteins. The information extracted by CWT about which dilation can be changed continuously is more abundant and integrated than that of DWT. In this work the continuous wavelet transform analysis techniques will be used to detect the number and locations of helices in membrane proteins. This method is not needed to set a data-dependent threshold for selecting wavelet coefficient and to recover the denoised signal by transforming these threshold coefficients. It is needed just to perform the hydrophobicity of amino acid sequences by CWT; the number and the locations of membrane proteins can be predicted conveniently under appropriate scale.

## THEORY

In this work the continuous wavelet is the preferred wavelet representation. The wavelet transform of a series of hydrophobic values, $H(n)$, is defined as

$$T(a,b) = \frac{1}{\sqrt{a}}\int_0^n H(n)\ \psi\left(\frac{n-b}{a}\right)\mathrm{d}n \qquad (1)$$

where $a$ is a dilation parameter and $b$ a translation parameter; they belong to the set of real numbers ($R$), and $a > 0$. $n$ is the amino acid sequence length of the protein, while $\psi((n - b)/a)$ is the analyzing wavelet function. The transform coefficients $T(a,b)$ are found for both specific locations on the signal, $n = b$, and for specific wavelet periods (which are a function of $a$). It is used to plot $T(a,b)$against $a$ and $b$ in a surface plot known as a scalogram, which is particularly suited to the detection of singularities. For $\psi$ we choose the continuous, symmetric, infinitely regular, modulated Gaussian Morlet wavelet:[18]

$$\psi(n) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-n^2}{2}\right)\exp(2\pi i f n) \qquad (2)$$

Note that the Morlet wavelet when used in the context of real data analysis is called the pseudo-Morlet wavelet because the admissibility condition is violated:

$$\int_{-\infty}^{+\infty}\psi(n)\,\mathrm{d}n = \exp(-2\pi^2 f^2) \neq 0 \qquad (3)$$

In practice this is not a problem because one can bring eq 3 arbitrarily close to 0 through the choice of $f$, which belongs to the set of real numbers.

The Morlet wavelet has real (modulus) and imaginary (phase) parts. Here, we are interested in the hydrophobic features of the sequence locations, frequencies, and hierarchically scaling transitions. So, we use the phase plot to detect and locate its singularities.

Here, the CWT will be used to predict the number and location of transmembrane helices (HTM) in membrane proteins. This method consists of three main steps. First the amine acids of membrane protein are transformed into sequences of hydrophobic free energies per residue. Second, the hydrophobic profile is decomposed into wavelet coefficients by using CWT. This is performed by mapping a one-dimensional amino acid sequence in the space domain into a two-dimensional space-frequency representation of the amino acid sequence. At the third step the number and the relevant positions of HTM can be extracted directly and easily according to the maxima of wavelet coefficient in a corresponding continuous wavelet transform plot of hydrophobic value sequences with appropriate scale.

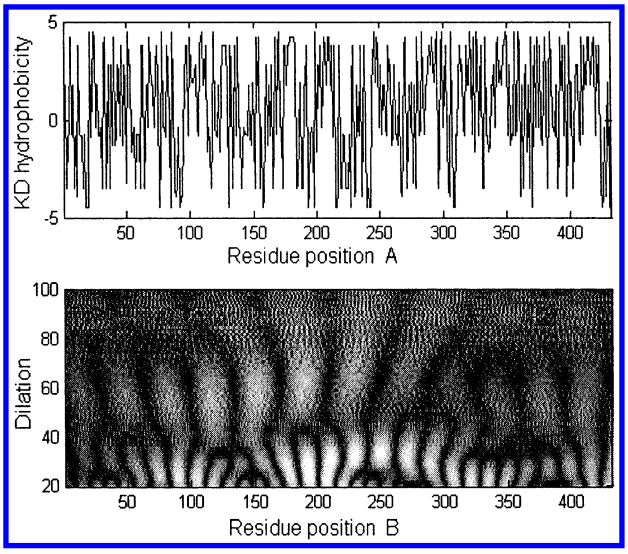


**Figure 1.** Kyte−Doolittle hydrophobic free energy series of membrane protein α-*ketoglutarate permease* (*kgtp_ecoli*) (A) and its continuous wavelet transformation (B). The Morlet wavelet was used.

Programs have been written in Matlab; software of Matlab 6.1 has been used.

## RESULTS

**Selection of an Appropriate Dilation.** We exemplify the CWT method on the task of detecting HTMs in bacteria proteins. Figure 1A depicts the input hydrophobicity data of *kgtp* (SwissProt accession: *kgtp_ecoli*) which has twelve HTMs. The hydrophobic data utilized in this study are the Kyte−Doolittle hydrophobicity (KD HΦ) which, for each residue of a protein, provides information derived from the protein structure and sequence.[1] The intensity of the signal is indicated on the $y$ axis; the $x$ axis indicates the position along the sequence. There are many sharp spikes in the hydrophobicity plot; no obvious evidence of HTMs is apparent from visual inspection of these data. While processing these HΦ data by CWT, the HΦ wavelet scalogram (Figure 1B) unfolds the distance-scale organization of the data for scales of 20−100. The abscissa of the Morlet wavelet transformation phase plot is the hydrophobic sequence of the amino acids, and the ordinate is the dilation. Each of the 80 dilation ranges was normalized to 100%, and the modular amplitudes gray scale were coded from black (relative minimum) to white (relative maximum).

In continuous wavelet transform, the larger dilation corresponds to the lower frequency region of the hydrophobicity sequence where the CWT has higher frequency resolution, so it can be used to study the relation between the topology structure of protein and its sequence and CWT can be used to investigate the regular secondary structure at

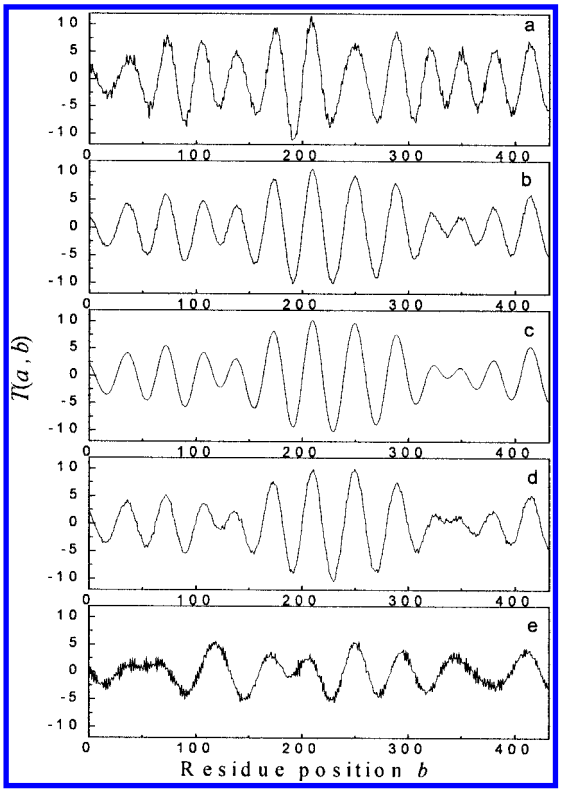


**Figure 2.** CWT coefficient plot of *kgtp_ecoli*: (a) dilation $a = 28$; (b) dilation $a = 31$; (c) dilation $a = 32$; (d) dilation $a = 33$; (e) dilation $a = 42$. The abscissa is the amino acid position along the protein backbone, and the ordinate is the magnitude of the CWT coefficient. The Morlet wavelet was used.

the intermediate scales; the smaller scales correspond to the high-frequency region of the hydrophobicity sequence, where the CWT has high resolution for the position of hydrophobicity sequence, and various constructions of proteins can be found and the position where the change happens can also be identified by CWT.[16] In this paper, we utilize the larger scales of CWT to predict the number and locations of HTM segments in membrane proteins.

At the larger scales in Figure 1B, there are 12 dark regions on the scalogram which correspond with the experimentally determined HTM segments number of the membrane protein. To further investigate the correlation between wavelet coefficients and HTMs of protein, an appropriate dilation $a$ is need to be selected. Selecting an appropriate dilation $a$ to perform CWT can not only accurately exhibit the HTMs of protein but also avoid submerging some useful information. The curves of wavelet coefficients under five different dilations are shown in Figure 2, which illustrates the process of dilation selection. In Figure 2, the middle curve (curve c) is relatively smooth. With deviation of this dilation, the curves are relatively coarse and a portion of the information is faint, which make the peak positions and the number of HTMs difficult to extract. The further the deviation is, the larger the predicted error will be. To further validate the reliability of the selected dilation for prediction of HTMs, we have used the data set of 83 membrane proteins[7] as the training set and another data set of 48 membrane proteins[11] as the testing set. The dilation parameter for prediction is determined to give the maximum average prediction accuracy at the training set. Through extensive sequential analysis, the optimal dilation $a = 32$ has been determined. By using this dilation, the HTMs are predicted to have 98.8% accuracy on average in the training set and 99.0% accuracy on average in the testing set. The accuracy is higher than that of other dilations. Therefore, the dilation $a = 32$ is selected as the appropriate dilation for the detection of HTM in this study.

From Figure 3, we can see clearly that there are 12 maxima which correspond with the result of Figure 1B, the wavelet coefficients are indicated on $y$ axis; the $x$ axis indicates the amino acid position along the sequence. The top row shows the locations of the experimentally determined HTM segments. In the plot, all regions with values above zero can be interpreted as HTMs. So we can easily identify the number and positions of HTMs, which can be clearly seen in Table 1. Table 1 shows the observed HTMs locations of *kgtp* (top row) and that of the predicted results by using CWT.

Figure 4 refers to the CWT of the *transport system permease protein hisQ* (SwissProt accession: *hisq_salty*). At the larger scales, there are five dark regions on the scalogram which correspond with the experimentally determined HTM segments number of the membrane protein.

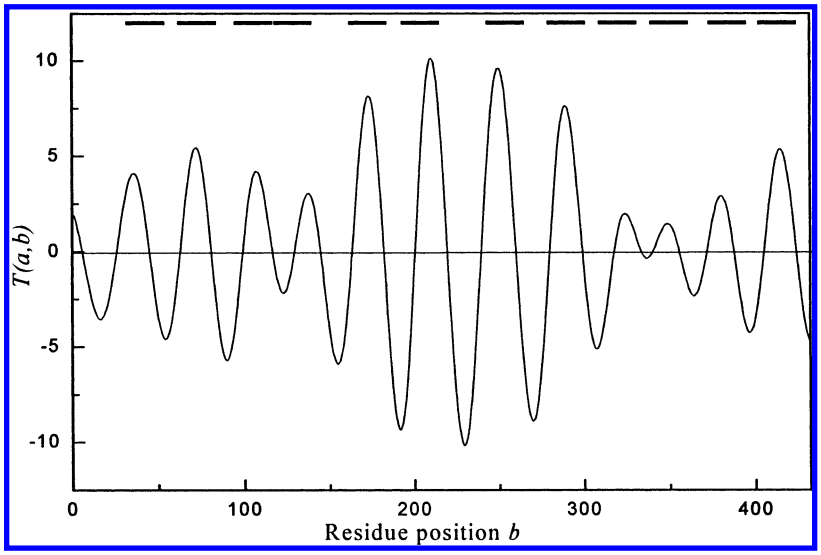


**Figure 3.** CWT coefficient plot of *kgtp_ecoli* (dilation $a = 32$). The top row indicates the positions of actual transmembrane helices (bars). The Morlet wavelet was used.

**Table 1.** Observed and Predicted HTM Segments of *kgtp_ecoli* Membrane Protein

| | TM1 | TM2 | TM3 | TM4 | TM5 | TM6 | TM7 | TM8 | TM9 | TM10 | TM11 | TM12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| observed | 33−53 | 63−83 | 96−116 | 119−139 | 163−183 | 194−214 | 244−264 | 280−300 | 310−330 | 340−360 | 374−394 | 403−423 |
| predicted | 27−47 | 63−82 | 98−116 | 129−144 | 163−183 | 199−218 | 240−260 | 280−300 | 317−332 | 341−356 | 373−390 | 404−423 |

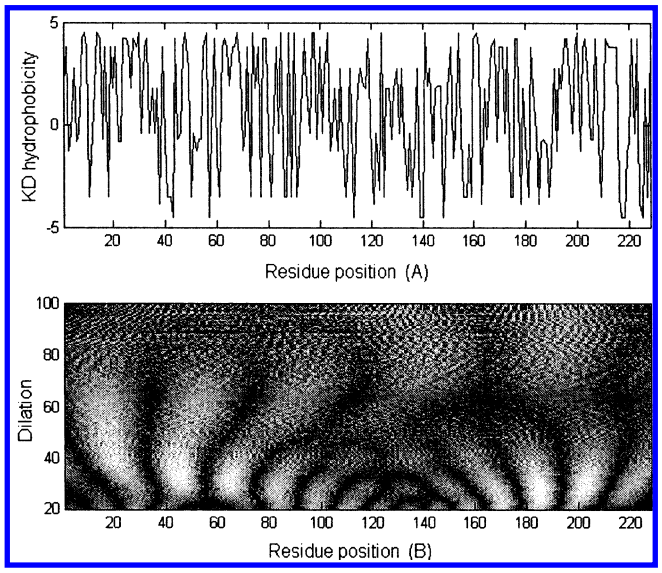


**Figure 4.** Kyte−Doolittle hydrophobic free energy series of membrane protein *transport system permease protein hisQ* (*hisq_salty*) (A) and its continuous Morlet wavelet transformation (B).

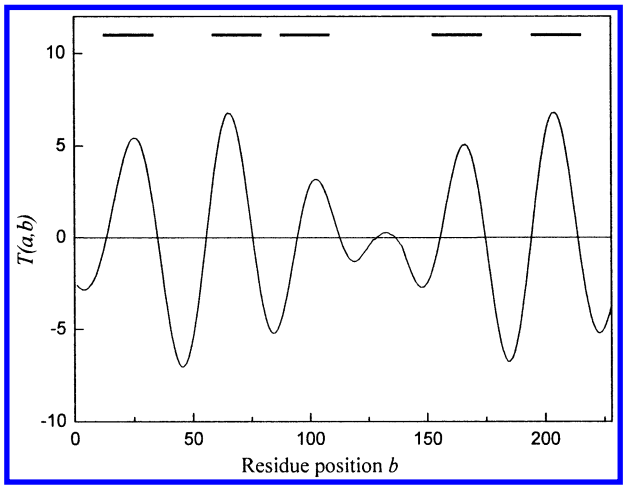


**Figure 5.** CWT coefficient plot of *hisq_salty* (dilation $a = 32$).

**Table 2.** Observed and Predicted HTM Segments of *hisq_salty* Membrane Protein

| | TM1 | TM2 | TM3 | TM4 | TM5 |
|---|---|---|---|---|---|
| observed | 13−33 | 59−79 | 88−108 | 153−173 | 195−215 |
| predicted | 13−34 | 56−76 | 93−110 | 154−174 | 194−214 |

Figure 5 shows the HTM organization and the CWT coefficients of *hisq_salty* with dilation $a = 32$. The prediction results and the observed topology are exhibited in Table 2.

**Hydrophobicity Data Types.** Proteins are one-dimensional chains composed of the 20 essential (nutrition-dependent) amino acids which are found in mammalian systems. Different side chains make different amino acids, and different amino acids have different properties; among various properties the most important one is hydrophobicity, which determines the stability of the protein structure.[19−24] Quantitative estimates of the hydrophobicity can be derived from their relative concentrations in organic versus water bulk phase of a binary solution. Different experimental conditions, solvents, and computational schemes have led to different sets of estimates of hydrophobicity. In fact, there are several amino acid hydrophobicity scales available to transform peptide sequences of transmembrance protein into real numbers. Three sets of hydrophobicity scales are listed in Table 3: Kyte−Doolittle[1] hydrophobicity scales (KD HΦ), Mandell[14] hydrophobicity scales (M HΦ), and Fauchereand[25] hydrophobicity scales (F HΦ). Figure 6 and Figure 7 are the CWT coefficient plots of *kgtp_ecoli* and *hisq_salty* membrane proteins, respectively, using the KD HΦ, the F HΦ, and the M HΦ at dilation $a = 32$. From Figure 6 and Figure 7, we can see that the varieties of wavelet coefficients are obvious and the predicted results tend to be accurate when using the KD HΦ. Compared with the M HΦ and the F HΦ, the KD HΦ performs the highest prediction power.

**Table 3.** Three Kinds of Hydrophobicity Scales

| amino acid | KD HΦ | M HΦ | F HΦ |
|---|---|---|---|
| isoleucine | 4.5 | 3.15 | 1.38 |
| valine | 4.2 | 1.87 | 1.08 |
| leucine | 3.8 | 2.17 | 1.06 |
| phenylalanine | 2.8 | 2.87 | 1.13 |
| cysteine | 2.5 | 1.52 | 0.29 |
| methionine | 1.9 | 1.67 | 0.64 |
| alanine | 1.8 | 0.87 | 0.62 |
| glycine | −0.4 | 0.00 | 0.48 |
| threonine | −0.7 | 0.07 | −0.05 |
| serine | −0.8 | 0.07 | −0.18 |
| tryptophan | −0.9 | 3.77 | 0.81 |
| tyrosine | −1.3 | 2.76 | 0.26 |
| proline | −1.6 | 2.77 | 0.12 |
| histidine | −3.2 | 0.87 | −0.4 |
| glutamane | −3.5 | 0.67 | −0.87 |
| glutamine | −3.5 | 0.00 | −0.78 |
| aspartate | −3.5 | 0.66 | −1.05 |
| asparagines | −3.5 | 0.09 | −0.85 |
| lysine | −3.9 | 1.64 | −1.35 |
| arginine | −4.5 | 0.85 | −1.37 |

**Comparison with the K−D Method.** To disclose the prediction power of the newly developed CWT, it is interesting to compare the results of more classical hydropathy profile smoothing methods. One of the most widely used is that of Kyte and Doolittle,[1] which is 20 years old, where the mean residue hydrophobicity values are calculated for consecutive 19-residue sequence spans; segments with hydrophobicity above an average value are predicted to be membrane-spanning. The Kyte−Dooloittle hydropathy plots of *kgtp* and *hisq* are shown in Figure 8 and Figure 9, respectively. Naturally, the hydropathy profiles using the Kyte−Doolittle method are noisier than that of the CWT method and some peaks which correspond with the experimentally determined HTM segments number of the two proteins cannot be distinguished clearly, which make the number and position of HTMs difficult to be extracted accurately. However, these problems can be overcome by CWT. CWT is a useful tool for analyzing the protein sequences both in time and in frequency domains localization, which is similar to mathematic microscopy and has the

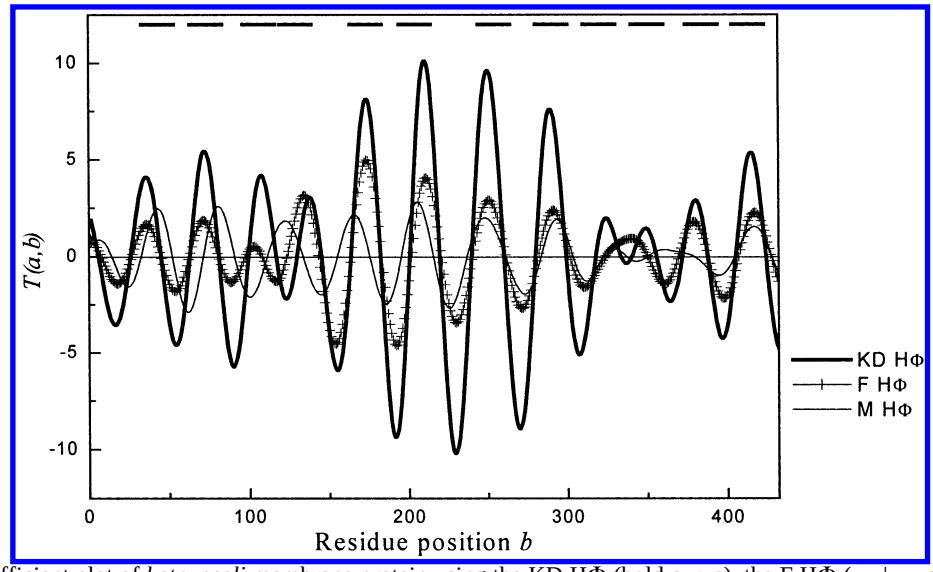


**Figure 6.** CWT coefficient plot of *kgtp_ecoli* membrane protein using the KD HΦ (bold curve), the F HΦ (− + − curve), and the M HΦ (thin curve) at dilation $a = 32$.

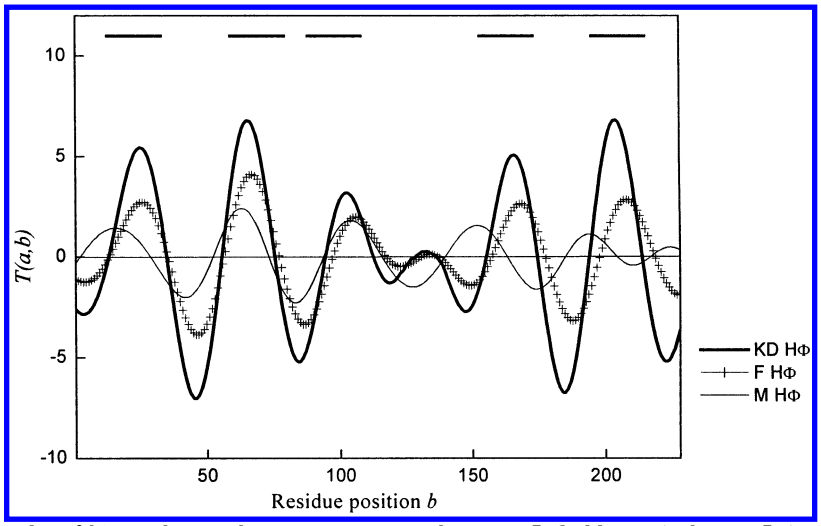


**Figure 7.** CWT coefficient plot of *hisq_salty* membrane protein using the KD HΦ (bold curve), the F HΦ (− + − curve), and the M HΦ (thin curve) at dilation $a = 32$.

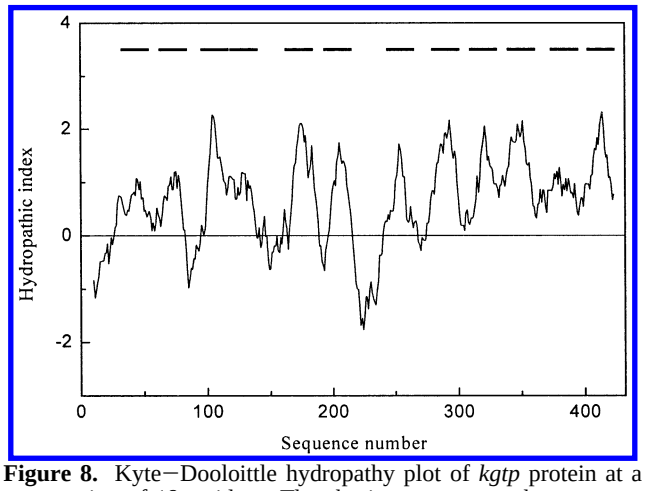


**Figure 8.** Kyte−Dooloittle hydropathy plot of *kgtp* protein at a span setting of 19 residues. The abscissa represents the sequence number along the protein backbone, and the ordinate is the magnitude of the hydropathic index. The top row indicates the positions of actual transmembrane helices (bars).

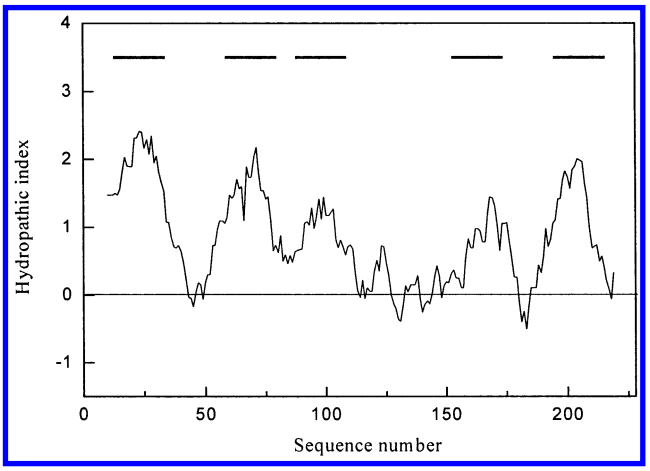


**Figure 9.** Kyte−Dooloittle hydropathy plot of *hisq* protein at a span setting of 19 residues.

ability of amplification and translation. In other words, CWT analysis, on one hand, can decompose the hydrophobic value sequences into coefficients at different dilations and remove the noise component from the hydrophobicity profiles at optimal dilation and on the other hand can give us local structures of sequences. So, results using the CWT method tend to be more accurate.

**Table 4.** Prediction Accuracies of Various Algorithms on Various Data Sets[a]

| data set | method | $N_{obs}$ | $N_{prd}$ | $N_{cor}$ | $Q_p$ (%) | $Q_{se}$ (%) | $Q_{sp}$ (%) | $Q_t$ (%) |
|---|---|---|---|---|---|---|---|---|
| 83TMP | CWT | 346 | 361 | 342 | 96.8 | 98.8 | 94.7 | 80.7 |
| | WCP1 | | 354 | 340 | 97.1 | 98.0 | 95.9 | 82.8 |
| | MEMSAT | | 351 | 336 | 96.4 | 97.1 | 95.7 | 78.3 |
| | TOPPRED | | 381 | 336 | 92.5 | 97.1 | 88.1 | 65.0 |
| | PHDhtm_ref | | 351 | 342 | 98.1 | 98.8 | 97.4 | 88.0 |
| | HMMTOP | | 353 | 344 | 98.4 | 99.4 | 97.4 | 86.7 |
| 48TMP | CWT | 194 | 202 | 192 | 97.0 | 99.0 | 95.1 | 77.1 |
| | WCP1 | | 198 | 189 | 96.4 | 97.4 | 95.0 | 73.2 |
| | MEMSAT | | 174 | 165 | 89.8 | 85.1 | 94.8 | 48.9 |
| | TOPPRED | | 200 | 193 | 98.0 | 99.4 | 96.5 | 53.0 |
| | PHDhtm_ref | | 192 | 192 | 99.2 | 99.0 | 99.0 | 89.0 |
| | HMMTOP | | 197 | 194 | 99.5 | 100.0 | 98.4 | 91.5 |

[a] $N_{obs}$, $N_{prd}$, and $N_{cor}$ are the number of observed, predicted, and correctly predicted transmembrane helices, respectively; $Q_p = 100 \times \sqrt{(N_{cor}/N_{obs})(N_{cor}/N_{prd})}$. $Q_{se}$, single HTM sensitivity: correctly predicted segments/true segments. $Q_{sp}$, single HTM specificity: correctly predicted segments/total predicted segments. $Q_t$ is the percentage of proteins for which all HTM were predicted correctly.

**Data Sets, Measure the Prediction Accuracy.** To investigate the feasibility of our method, two data sets of membrane proteins as a benchmark have been used. Following Rost et al.[7] and Lio and Vannucci,[11] we have used a test set of 83 membrane proteins and a blind set of 48 membrane proteins. For the two data sets the segments prediction accuracy is over 98% in each case by setting the CWT dilation $a = 32$ and using the KD hydrophobicity. For the 346 HTMs in 83 TMP 361 were predicted, of which 342 were predicted correctly, and for the 194 HTMs in 48 TMP 202 were predicted and 192 of them were correctly. Five other prediction methods were taken into consideration, TOPPRED,[3] HMMTOP,[4] MEMSAT,[6] PHDhtm_ref,[8] and WCP1,[11] and the accuracies of them on the two data sets are listed in Table 4. Note that our method has the same high prediction accuracies as WCP1, even a little better in 48 TMP. In addition, compared with more recent prediction methods that are in general more complex and require model assumptions, our nonparametric method manipulates simply and visually and performs reasonably well. However, there are also some discrepancies which are shown as follows: (1) With overprediction, there are no the experimentally determined HTM segments corresponding to the predicted HTM segments. (2) With underprediction, some of the experimentally determined HTM segments have not been predicted by this method. This is because of the following: (1) Although hydrophobicity of amino acid is the most important factor to determine the stability of membrane protein structure, it is not the only factor; besides hydrophobicity, there are hydrogen bonds, salt linkage, van der Waals contacts, and disulfide bonds at the interior of peptide chains. (2) Volume and change of molecular, contactable area and so on are all factors that affect the structure and the stability of protein. (3) The longer hydrophobic sequences in protein core regions may be predicted as HTMs. (4) Signal peptides may also be predicted as HTMs according to the hypothesis of signal peptide. Therefore, the regions shorter than 14 residues need to be further investigated as potential protein core regions or, if in the initial region, as potential signal peptide sequences. For the reasons above, finite discrepancies are permitted in predicting HTMs by using CWT under the appropriate scales after mapping amino acid sequences to hydrophobic value sequences per residue. Had all deterministic factors been included, we are sure of a higher capability for prediction.

## CONCLUSION

TM proteins have a fundamental importance in practical applications because a large portion of these proteins play key functional roles as drug receptors, transporters, immunological recognition targets, etc. So prediction of HTM locations is an important problem in molecular biology. As one of the time or space−frequency representations, CWT can be utilized to dissect, analyz, and interpret signals of amino acid sequences and further to detect and characterize the HTMs of membrane proteins. The results described in this paper indicate that KD HΦ data can provide useful information on the number and locations of transmembrane helices, and CWT analysis is an effective tool for extracting structural information on membrane proteins from sequences.

## ACKNOWLEDGMENT

This work was supported by Grants 29975033 from the National Natural Science Foundation of China and 001237 from the National Natural Science Foundation of Guangdong Province.

## REFERENCES AND NOTES

(1) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol*. **1982**, *157*, 105−132.
(2) Engelman, D. M.; Steiz, T. A.; Goldman, A. Identifying Nonpolar Transbilayer Helices in Amino Acid Sequences of Membrane Proteins. *Annu. Rev. Biophys. Chem*. **1986**, *15*, 321−353.
(3) von Heijne, G. Membrane Protein Structure Prediction−Hydrophobicity Analysis and the Positive-Inside Rule. *J. Mol. Biol*. **1992**, *225*, 487−494.
(4) Tusnády, G. E.; Simon, I. Principles Governing Amino Acid Composition of Integral Membrane Proteins: Application to Topology Prediction. *J. Mol. Biol*. **1998**, *283*, 489−506.
(5) Tusnády, G. E.; Simon, I. Topology of Membrane Proteins. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 364−368.
(6) Jones, D. T.; Taylor, W. R.; Thornton, J. M. A Model Recognition Approach to the Prediction of Allhelical Membrane Protein Structure and Topology. *Biochemistry* **1994**, *33*, 3038−3049.
(7) Rost, B.; Casadio, R.; Fariselli, P.; Sander, C. Transmembrane Helices Predicted at 95% Accuracy. *Protein Sci.* **1995**, *4*, 521−533.
(8) Rost, B.; Fariselli, P.; Casadio, R. Topology Prediction for Helical Transmembrane Proteins at 86% Accuracy. *Protein Sci.* **1996**, *5*, 1704−1718.
(9) Persson, B.; Argos, P. Topology Prediction of Membrane Proteins. *Protein Sci.* **1996**, *5*, 363−371.
(10) Lohmann, R.; Schneider, G.; Behrens, D.; Wrede, P. A Neural Network Model for the Prediction of Membrane-Spanning Amino Acid Sequences. *Protein Sci.* **1994**, *3*, 1597−1601.
(11) Lio, P.; Vannucci, M. Wavelet Change-Point Prediction of Transmembrane Proteins. *Bioinformatics* **2000**, *16*, 376−382.

(12) Burrus, C. S. *A Primer Introduction to Wavelet and Wavelet Transforms*; Prentice Hall: Upper Saddle River, NJ, 1998; pp 3−9.
(13) Murray, K. B.; Gorse, D.; Thornton, J. M. Wavelet Transforms for the Characterization and Detection of Repeating Motifs. *J. Mol. Biol*. **2002**, *316*, 341−363.
(14) Hirakawa, H.; Muta, S.; Kuhara, S. The Hydrophobic Cores of Proteins Predicted by Wavelet Analysis. *Bioinformatics* **1999**, *15*, 141−148.
(15) Mandell, A. J.; Selz, K. A.; Shlesinger, M. F. Wavelet Transformation of Protein Hydrophobocity Sequences Suggests Their Membership in Structural Families. *Physica A* **1997**, *244*, 254−262.
(16) Qiu, J. D.; Liang, R. P.; Zou, X. Y.; Mo, J. Y. Prediction of Protein Secondary Structures by Continuous Wavelet Transform. *Acta Chim. Sin*. **2003**, *61*, 748−754.
(17) Trad, C. H.; Fang, Q.; Cosic, I. Protein Sequence Comparison Based on the Wavelet Transform Approach. *Protein Eng.* **2002**, *15*, 193−203.
(18) Yang, F. S. *The Engineering Analysis and Application of Wavelet Transform*; Science Press: Beijing, China, 1999; pp 1−26.
(19) Rose, G. D. Prediction of Chain Turns in Globular Proteins on a Hydrophobic Basis. *Nature* **1978**, *272*, 586−590.
(20) Rose, G. D. Hydrogen Bonding, Hydrophobicity, Packing, and Protein Folding. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 381−415.
(21) Eisenberg, D.; Weiss, R. M.; Terwilliger, T. C. The Hydrophobic Moment Detects Periodicity in Protein Hydrophobicity. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 140−144.
(22) Eisenberg, D.; McLachlan, A. D. Solvation Energy in Protein Folding and Binding. *Nature* **1986**, *319*, 199−203.
(23) Mitaku, S. The Role of Hydrophobic Interaction in Phase Transition and Structure Formation of Lipid Membrane and Proteins. *Phase Transitions* **1993**, *45*, 137−155.
(24) He, Y. F.; Wang, E. K. Study on the Structure of Hemoglobin in Organic Solvents by in Situ STM. *Acta Chim. Sin.* **1997**, *55*, 801−805.
(25) Fauchereand, J. L.; Pliska, V. Transformational Homologies in Amino Acid Sequence. *Eur. J. Med. Chem.* **1983**, *18*, 369−375.

CI0303868