Prediction of Hydrophobic Cores of Proteins Using Wavelet Analysis

Hideki Hirakawa
hirakawa@grt.kyushu-u.ac.jp

Satoru Kuhara
kuhara@grt.kyushu-u.ac.jp

Graduate School of Genetic Resources Technology, Kyushu University
6-10-1 Hakozaki, Higashi-ku, Fukuoka 812, Japan

Abstract

Information concerning the secondary structures, flexibility, epitope and hydrophobic regions of amino acid sequences can be extracted by assigning physicochemical indices to each amino acid residue, and information on structure can be derived using the sliding window averaging technique, which is in wide use for smoothing out raw functions. Wavelet analysis has shown great potential and applicability in many fields, such as astronomy, radar, earthquake prediction, and signal or image processing. This approach is efficient for removing noise from various functions. Here we employed wavelet analysis to smooth out a plot assigned to a hydrophobicity index for amino acid sequences. We then used the resulting function to predict hydrophobic cores in globular proteins. We calculated the prediction accuracy for the hydrophobic cores of 88 representative set of proteins. Use of wavelet analysis made feasible the prediction of hydrophobic cores at 6.13% greater accuracy than the sliding window averaging technique.

Keywords: wavelet analysis, hydrophobic cores, PDB, interior/exterior

1 Introduction

According to Anfinsen's report (1973), amino acid sequences of proteins carry all the information needed to form their three-dimensional structures. Thus, the protein structure theoretically can be predicted based solely on amino acid sequences. Structural or biological information such as secondary structure (Rose, 1978; Qian, 1996), flexibility (Vihinen et al., 1994), epitope (Hopp and Woods, 1981), hydrophobic region (Hubbard and Blundell, 1987; Swindells, 1995; Desjarlais and Handel, 1995) are derived by assigning a physicochemical index to an amino acid sequence. In these prediction methods, the sliding window averaging technique has been used to smooth out the raw function. The physicochemical value for each residue inside the window is summed up and given for residues in the middle of the window. Then the hydrophobicity values alternate locally maximal and minimal values, and extrema partitions the amino acid sequence into structural segments. However, the relationship between segments and structure does not always correspond. As extracting structural information from amino acid sequences alone is difficult, various prediction methods have been developed (Rost et al., 1994a). By obtaining more sequential and structural data, the attainment of structural information from amino acid sequences is facilitated.

Using a hydrophobicity index, prediction of hydrophobic regions from amino acid sequences was previously investigated (Kuntz, 1972) and it was shown that hydrophobicity tended to be low at the loop region. The minimal hydrophobicity profile corresponded to the loop region, and the turn region could be predicted with about 70% accuracy. Highly hydrophobic regions tended to form an α-helix. Prediction of hydrophobic regions is important because the hydrophobic effect is the major factor that drives a protein molecule toward folding. During this process, residues with apolar side chains associate to form a solvent-shielded hydrophobic core (Creamer et al., 1995).

Previously, methods for hydrophobic region prediction using only amino acid sequences have been reported (Qian, 1996; Creamer et al., 1995). These studies suggested that hydrophobic residues were

buried in the core of protein, and secondary structures could be predicted using a hydrophobicity index. Rose and Roy (1980) showed that the hydrophobicity profile and the packing density profile correlated, cores of a globular protein were relatively tightly packed, and surface residues were generally more movable due to fewer stabilizing interactions. They also compared these results with temperature factors from x-ray refinement to show that both the minimum packing densities and hydrophobicity profiles corresponded to conformationally flexible regions. They interpolated the hydrophobicity plot using spline functions. This simple method gives a smooth, differentiable curve that systematically removes dispersion without affecting the position of dominant local extrema. Using this procedure, the segments which corresponded to hydrophobic regions were easily identified.

Kyte and Doolittle (1982) investigated the span of window size which reflected the interior and exterior portions of proteins. Hydrophobicity profiles using the shortest spans were too noisy, and employing spans less than seven residues was generally unsatisfactory. On the other hand, long spans tended to lose structural segments. The best agreement between hydrophobic region and position of amino acid residues, whether interior or exterior, was obtained with a nine for globular proteins. The problem with this technique is noise in the smoothed profiles, which makes it particularly difficult to find segments in case of globular proteins.

Fourier analysis has been applied to predict secondary structures from a sequential dotted hydrophobicity index (Finer-Moore and Stroud, 1984; Cornette et al., 1987). The utilities of the Fourier transform lie in its ability to analyze a signal in the time domain for its frequency content. The transform works by first translating a function in the time domain into a function in the frequency domain. The signal can then be analyzed for its frequency content because the Fourier coefficients of the transformed function represent the contribution of each sine and cosine function at each frequency. Although the Fourier analysis is useful for acquiring structural information, this method tends to cause positional error.

Wavelets are mathematical functions that divide data into different frequency components. Wavelet analysis is now applied to astronomy, acoustics, nuclear engineering, signal or image processing, speech discrimination, earthquake prediction, radar, and so on. This approach has advantages over traditional Fourier methods in analyzing data where the signal contains discontinuities or high frequency noise. With wavelet analysis, one can use approximating functions that are contained neatly in finite domains. Wavelet analysis is well-suited for approximating data to remove noise from functions (Arneodo and Grasseau, 1988; Argoul et al., 1989; Yamada and Ohkitani, 1990, 1991).

As mentioned above, numerous attempts have been made by researchers to define the relation between interior and exterior regions directly from the amino acid sequence. A high hydrophobic region tends to form hydrophobic cores. However, it is difficult to divide interior and exterior position of amino acid residues by assigning a hydrophobicity threshold, because of unacceptable noise level.

We attempted to apply wavelet analysis to the structural analysis of proteins. To investigate how accurately one can predict hydrophobic cores using hydrophobicity index, we removed the high frequency at each level from the raw function.

2   System and Methods

There are important differences between Fourier analysis and wavelet analysis. Fourier analysis is based on functions that are localized in frequency, not in time. Small frequency changes in the Fourier transform will produce changes everywhere in the time domain. Wavelets are local in both frequency/scale (via dilations) and in time (via translations).

There are numerous mother wavelets such as Haar, Daubechies, Coiflet, Symlet and so on. Once the mother wavelet $\varphi(x)$ is fixed, one can form the basis for translations and dilations of the mother wavelet $\varphi((x-b)/a)$. Wavelet transform of $f(x)$ with mother wavelet $\varphi(x)$, $T(a,b)$ is defined as follows,
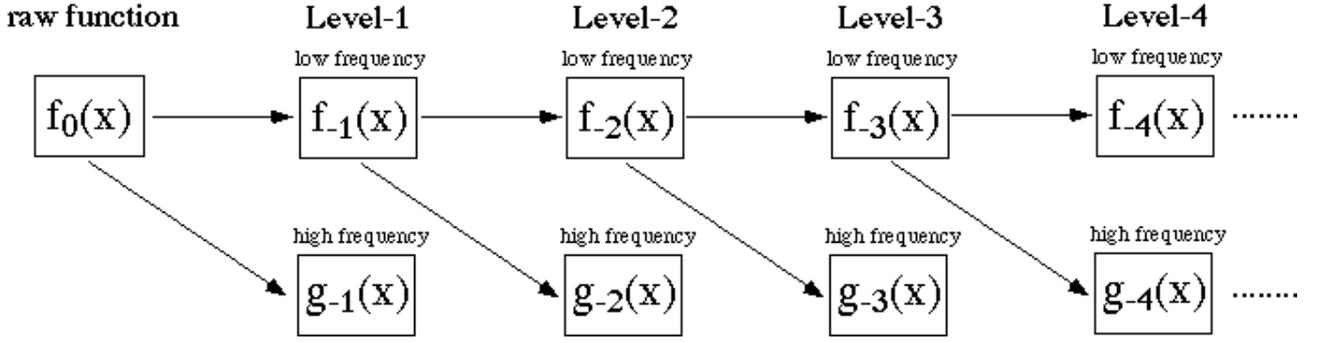
Figure 1: Decomposition of raw function to low and high frequencies at each level by wavelet analysis. $f_0(x)$: raw function. $f_{-1}(x)$: low frequency function, $g_{-1}(x)$: high frequency function (level-1). $f_{-2}(x)$: low frequency function, $g_{-2}(x)$: high frequency function (level-2). $f_{-3}(x)$: low frequency function, $g_{-3}(x)$: high frequency function (level-3). $f_{-4}(x)$: low frequency function, $g_{-4}(x)$: high frequency function (level-4).

$$T(a,b) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{|a|}} \overline{\varphi(\frac{x-b}{a})} f(x) dx. \tag{1}$$

It is convenient to use special values for a and b in defining the wavelet basis: $a = 2^{-j}$ and $b = k2^{-j}$, where k and j are integers. J is the dilation parameter and k indicates the temporal position of time $k2^{-j}$. This choice of a and b will give a sparse basis,

$$T(a,b) = d_k^{(j)} = \int_{-\infty}^{\infty} \overline{\varphi(2^j x - k)} f(x) dx. \tag{2}$$

This modification is called the *Discrete Wavelet Transform* (DWT). The inverse discrete wavelet transform is given by

$$f(x) \sim \sum_j \sum_k d_k^{(j)} \varphi(2^j x - k). \tag{3}$$

In this study, we refer to j as "level" which comprises part of the given function. If $\sum_k d_k^{(j)} \varphi(2^j x - k)$ is given by $g_j(x)$ in Eq. (3), it is written,

$$f_j(x) = g_{j-1}(x) + g_{j-2}(x) + ... \tag{4}$$

Here, $g_j(x)$ is considered to be high frequency at level j. In this study, DWT was performed from level-1 to level-4, and the raw function $f_0(x)$ is transformed into each $g_j(x)$ (j=1,2,3,4). Eq. (4) is thus converted,

$$f_j(x) = g_{j-1}(x) + f_{j-1}(x). \tag{5}$$

Eq. (5) shows that one function at level j is transformed into the high frequency at level $j-1$ and low frequency at level $j-1$. Fig. 1 shows the time-frequency decomposition by DWT. The raw function $f_0(x)$ is decomposed to the low frequency $f_{-1}(x)$ and the high frequency $g_{-1}(x)$. In this manner, the original function is decomposed to frequencies at each level. Therefore, one can extract the low frequency at level $j-1$ by subtracting the high frequency at level $j-1$ from low frequency at level j.

In this study, the Daubechies were selected from families of mother wavelets. A mother wavelet is defined by a scaling function which is satisfied by a two-scale difference equation. Each mother wavelet is smoothed by increasing a natural number N. We defined N as "order". The mother wavelet is smoothed by increasing the number of N. In this study, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20 were employed as values for N. Here, we used Daubechies16 as mother wavelet.

2.1 Application to protein structure

The protein sequential and structural data were from the Brookhaven's Protein Data Bank (PDB). The solvent accessibility (Acc) value which describes the number of water molecules in contact with each residue was calculated with the DSSP program (Kabsch and Sander, 1983).

A crude approximation for residue accessibility has been used: a projection onto two states, buried or exposed. Such a projection of surface area onto two states reduces by about one-half the information concerning solvent accessibility. Additionally, the problem arises as to how to define the threshold to distinguish between two states (Rost and Sander,1994b). Buried residues in protein are determined by the value of relative accessibility (RelAcc),

$$RelAcc = (Acc/MaxAcc) * 100. \tag{6}$$

The value of RelAcc was calculated using a maximum accessibility (MaxAcc), as shown in Table 1. The amino acid residue with RelAcc below 16% is considered a buried residue, and RelAcc above 16% as an exposed residue (Rost and Sander, 1994b).

Amino acid	Hydrophobicity value	MaxAcc($\mathring{A}$)
Isoleucine	4.5	169
Valine	4.2	142
Leucine	3.8	164
Phenylalanine	2.8	136
Cysteine	2.5	135
Methionine	1.9	188
Alanine	1.8	106
Glycine	-0.4	84
Threonine	-0.7	142
Serine	-0.8	130
Tryptophan	-0.9	227
Tyrosine	-1.3	222
Proline	-1.6	136
Histidine	-3.2	184
Asparatic acid	-3.5	163
Asparagine	-3.5	157
Glutamic acid	-3.5	194
Glutamine	-3.5	198
Lysine	-3.9	205
Arginine	-4.5	248

Table 1: Hydrophobicity values and MaxAcc values. The numerical hydrophobicity values which were calculated by Kyte and Doolittle and assigned to the 20 amino acids commonly found in proteins. Maximal accessibility (measured in $\mathring{A}$) for the 20 amino acids.

To avoid redundant data in PDB files, we used the representative set of globular proteins selected by Rost and Sander (1993). There are 126 globular proteins. It is necessary to be complete in amino acid sequences for applying wavelet analysis. Therefore we removed 38 globular proteins whose amino acid sequences were incomplete. Then we investigated remained 88 globular proteins. To quantify the prediction accuracy of hydrophobic cores, we defined the function as follows:

$$F_{accu} = (N_{rbu} + N_{rex}) * 100/N_{all}, \tag{7}$$

where N_{rbu} indicates the number of residues which are predicted to be correctly buried, and N_{rex} indicates the number of residues predicted to be correctly exposed, and N_{all} indicates the total number of residues.

To determine the optimum hydrophobic threshold to separate interior and exterior regions, we defined the cost function, as follows,

$$F_{cost} = (N_{rbu}/N_{bu}) * (N_{rex}/N_{ex}), \tag{8}$$

where N_{bu} indicates the number of residues predicted to be buried, and N_{ex} indicates the number of residues predicted to be exposed. The threshold which corresponded to the maximum point was considered optimal for the prediction of hydrophobic cores.

To apply the wavelet analysis for protein structure analysis, we attempted to predict hydrophobic cores in globular proteins, using low frequencies. The sliding window averaging technique and wavelet analysis were then compared based on corresponding results.

For the present study, we used the Mac Wavelet ver2.0 program for wavelet analysis.

3   Results

3.1   Prediction of hydrophobic cores in globular proteins

In this section, bovine intestine ($3ICB\_A$) was chosen as a representative globular protein.

Fig. 2 shows a hydrophobicity profile smoothed by the sliding window averaging technique and low frequencies extracted using wavelet analysis at levels -1 to -4. The top on Fig. 4a showed buried residues defined by binary model as a bar. Sharp spikes remained in the low frequency at level-1 (Fig. 2b). The frequency at level-2 (Fig. 2c) was lower than at level-1, thus it was difficult to find segments related to hydrophobic cores. At level-3 (Fig. 2d), the frequency was lower. The frequency at level-4 (Fig. 2e) was too low, thus was unsuitable to search for hydrophobic segments. The frequencies decreased with increasing levels, so the peak of frequency at level-1 tended to be satisfactory for each segment. On the other hand, the profile of the sliding window averaging technique (Fig. 2a) was hard to find segments corresponded to buried residues.

In order to determine a suitable level for detecting hydrophobic cores, we searched for hydrophobic thresholds which could distinguish interior and exterior residues by evaluating the cost function F_{cost}.

3.2   Calculation of prediction accuracy for 88 representative set of proteins

We calculated the maximum value of the F_{cost} when the buried region and exposed region matched. Table2 shows the average prediction accuracy of hydrophobic cores by the sliding window averaging technique and wavelet analysis using each order at levels-1 to -4 by suitable hydrophobic thresholds. By the sliding window averaging technique, the hydrophobic cores were predicted with 55.6% accuracy. By wavelet analysis at level-1, the average prediction accuracy was 61.9%, and the prediction accuracy of this condition was the highest of all tested cases. At highers levels, the frequency was lower. Because the predicted hydrophobic regions tended to be broader. Detection of dotted hydrophobic residues at higher levels proved difficult.
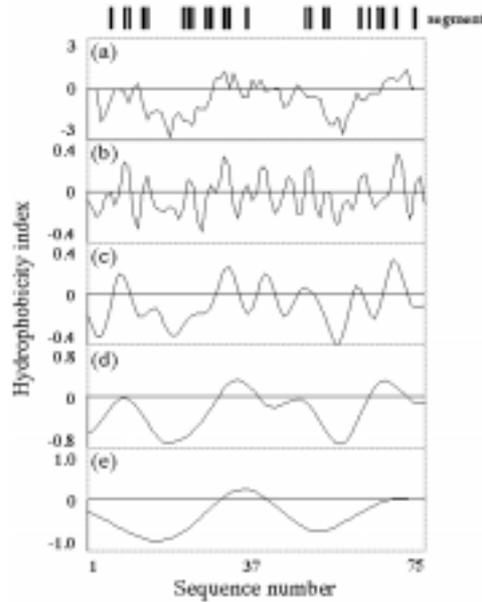
Figure 2: Smoothed curve using sliding window averaging technique and low frequencies at each level for bovine intestine. (a) sliding window averaging technique (b) level-1, (c) level-2, (d) level-3, (e) level-4.

sliding window averaging technique

value of cost function	0.301
hydrophobic threshold	−0.24
prediction accuracy	55.6

wavelet analysis using Daubechies16

	level-1	level-2	level-3	level-4
value of cost function	0.381	0.342	0.301	0.234
hydrophobic threshold	−0.01	−0.01	−0.01	−0.04
prediction accuracy	61.9	60.0	59.8	57.5

Table 2: Prediction accuracy of sliding window averaging technique and wavelet analysis.

The histogram of prediction accuracy at the best condition for each prediction method is shown in Fig. 3. Fig. 3a shows the result of the sliding window averaging technique and Fig. 3b to Fig. 3e the result of wavelet analysis (level-1 to level-4), respectively. We calculated sample variance of prediction accuracy for 88 representative set of proteins. The histogram of levels-1 -2 concentrated at a high prediction accuracy, with each sample variance being 5.68 and 5.81 respectively. The histogram of wavelet analysis at levels-3 -4 concentrated at a low accuracy, with sample distributions being 6.86 and 9.71 respectively. The sample distribution of the sliding window averaging technique was 7.77. This indicates that low frequencies at level-1 in wavelet analysis are suitable for predicting hydrophobic cores.
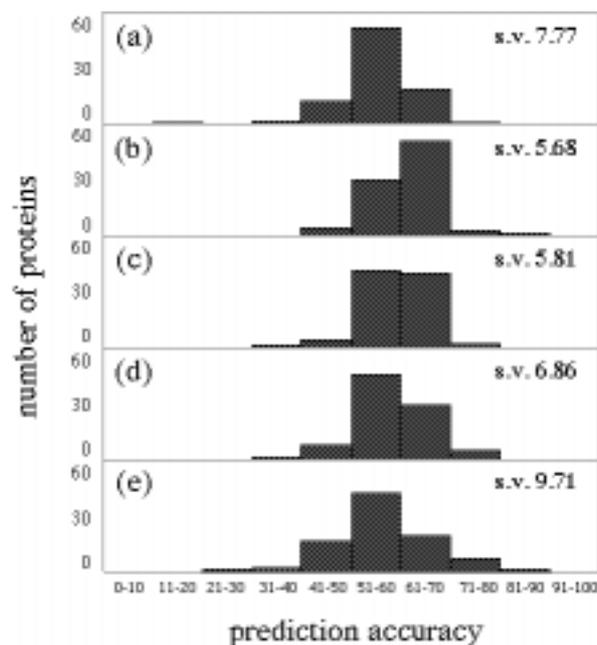
Figure 3: Histogram of prediction accuracy of hydrophobic cores at optimal conditions.
(a) sliding window averaging technique: hydrophobic threshold = -0.20. (b) level-1 : hydrophobic threshold$=-0.01$. (c) level-2 : hydrophobic threshold$=-0.01$. (d) level-3 : hydrophobic threshold$=-0.01$. (e) level-4 : hydrophobic threshold$=-0.04$. s.v: sample variance

We calculated average prediction accuracy in case of changing window size from 1 to 5. The result was shown in Table 3. Use of window size 2, the prediction accuracy was 60.6%. Detecting hydrophobic cores using the sliding window averaging technique, window size 2 is suitable.

	window size 1	window size 2	window size 3	window size 4	window size 5
value of cost function	0.352	0.318	0.301	0.303	0.274
hydrophobic threshold	-0.27	0.01	-0.23	-0.31	-0.31
prediction accuracy(%)	58.8	60.6	55.6	55.2	52.9

Table 3: Prediction accuracy of sliding window averaging technique using window size 1 to 5.

3.3 Structural comparison of hydrophobic cores predicted by each smoothing method

We compared hydrophobic cores predicted using each suitable hydrophobic threshold, as shown in Table 2, with a binary model. Fig. 4a shows the buried residues defined by the binary model. Fig. 4b to Fig. 4f shows the predicted hydrophobic cores.

By wavelet analysis, the hydrophobic core predicted by wavelet analysis at level-1 (Fig. 4c) corresponded well with the binary model, with a prediction accuracy of 81.3%. At lower levels, hydrophobic cores were predicted incorrectly and overall prediction accuracy was poor. With the sliding window averaging technique, the hydrophobic residues continued in protein cores were difficult to detect, because there was remained noise in the smoothed profile. For these reasons, predicted hydrophobic cores were dotted, and the prediction accuracy was poor. Judging from these results, the hydrophobic cores defined by a binary model were predicted with high accuracy at level-1, and prediction accuracy decreased with increasing levels. Above level-2, dotted hydrophobic residues were difficult to find.
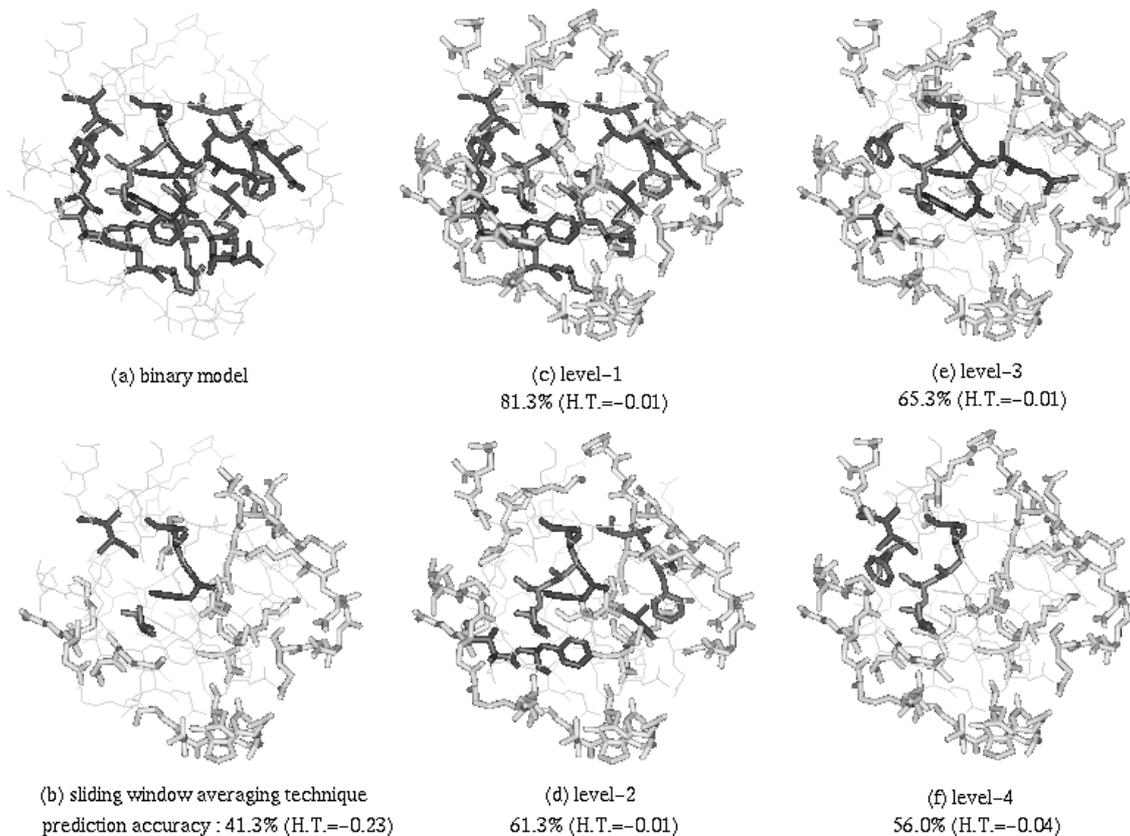
Figure 4: Prediction accuracy of hydrophobic cores of bovine intestine. (a) thick line: buried residues. (b) to (f) thick and dark line: predicted correctly as buried, thin and light line: predicted correctly as exposed. H.T.: hydrophobicity threshold.

Table 2 shows the average prediction accuracy analyzed using Daubechies16. Previously we investigated that each order of mother wavelet could predict hydrophobic cores. From these results, we determined that the optimal condition for prediction of hydrophobic cores was at level-1, hydrophobic threshold= -0.01.

4 Discussion

The profile assigning an amino acid index to a sequence is inherently noisy. To eliminate noise from raw functions, various methods have been proposed. Although the sliding window averaging technique is in wide use, the precise region of the hydrophobic core is difficult to determine because of the averaging calculation. Extracting structural information from amino acid sequences using spline function is suitable for smoothing hydrophobicity profile. However, it is necessary to smooth previously using the sliding window averaging technique. Thus, the spline function cannot be applied directly to a raw function which is assigned a hydrophobicity index.

Rost and Sander (1994b) introduced a neural network system that predicted relative solvent accessibility using evolutionary profiles of amino acid substitutions derived from multiple sequence alignment. When they applied sequence alignment for protein families, the buried residues were predicted with

high accuracy. Using structural alignment, they could predict with higher accuracy. They suggested that relative accessibility was less conserved for three-dimensional homologues than with the secondary structure. This indicates that it is useful for sequence alignment and also for modeling protein three-dimensional structures. This method is useful for hydrophobic core prediction, but selecting data for input is difficult using this approach.

Wako and Blundell (1994) predicted buried residues of protein families using sequence alignment. They used amino acid environment-dependent substitution tables and conformational propensities. The mean prediction accuracy was 76.5% by their definition. By wavelet analysis, we could predict hydrophobic cores of their protein families dataset with about 70% accuracy. Using other mother wavelet or hydrophobic index, we may improve prediction accuracy.

Fourier analysis was applied for smoothing the hydrophobicity plot (Finer-Moore and Stroud, 1984). If the high frequency was removed from the original profile by this method, the height of each peak was lower, and tended to affect another region. For these reasons, it is difficult to get positional information. The most interesting dissimilarity between the Fourier transform and the wavelet transform is that individual wavelet functions are localized in space. This localization makes many given functions approximate by transformation into the wavelet domain. Thus, wavelet analysis is useful for data compression, detecting features in images, and removing noise from time series. For these reasons, the result of wavelet analysis hardly affects other data and we attempted to use this approach for interior/exterior prediction of proteins to correctly detect hydrophobic segments.

We extracted the low frequencies from raw data using wavelet analysis, and investigated the relationship between low frequencies and structural information of proteins. For globular proteins, optimal hydrophobic thresholds which could divide interior and exterior residues were determined using defined cost function, and the information on hydrophobic cores was extracted. The frequency at level-1 is corresponded with hydrophobic segments. Thus, the optimal condition for prediction of hydrophobic cores was hydrophobic threshold $= -0.01$, and level-1.

The efficiency of wavelet analysis for interior/exterior prediction indicated the detection of segments related to a hydrophobic core at a single peak, thus facilitating the accurate prediction of buried residues directly from amino acid sequences. Moreover, wavelet analysis is beneficial in that the noise is eliminated without causing positional error. The hydrorophobic residues tend to be dotted periodically on the surface of proteins, and their side chains face the interior region of the protein. In wavelet analysis, the width of peak in high frequency at level-1 tended to be corresponded with the buried segments. Then, the average prediction accuracy at level-1 was the highest of all tested cases. At lower levels, buried residues were difficult to detect, because the sharp spikes associated with buried residues were removed. We suggest that wavelet analysis may also be effective for prediction of hydrophobic cores and protein structure analysis, such as membrane-spanning segments directly from the amino acid sequence. Wavelet analysis is more useful for detecting continuous segments, such as secondary structure, flexibility and epitope regions. This method may be also be effective for extracting many structural information on proteins.

Acknowledgments

References

[1] Anfinsen, C.B. (1973) Principles that govern the folding of protein chains. *Science*, **181**(96), 223–230.

[2] Argoul, F., Arneodo, A., Grasseau, G., Gagne, Y., Hopfinger, E.J. and Frisch, U. (1989) Wavelet analysis of turbulence reveals the multifractal nature of the Richardson cascade. *Nature*, **338**, 51–53.

[3] Arneodo, A. and Grasseau, G. (1988) Wavelet transform of multifractals. *Phys. Rev. Lett.*, **61**, 2281–2284.

[4] Cornette, J.L., Cease, K.B., Margalit, H., Spouge, J.L., Berzofsky, J.A. and DeLisi, C. (1987) Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *J. Mol. Biol.*, **195**, 659–685.

[5] Creamer, T.P., Srinivasan, R. and Rose, G.D. (1995) Modeling unfolded states of peptides and proteins. *Biochemistry*, **34**, 16245–16250.

[6] Desjarlais, J.R. and Handel, T.M. (1995) De novo design of the hydrophobic cores of proteins. *Protein Science*, **4**, 2006–2018.

[7] Finer-Moore, J. and Stroud, R.M. (1984) Amphipathic analysis and possible formation of the ion channel in an acetylcholine receptor. *Proc. Nat. Acad. Sci., U.S.A.*, **81**, 155–159.

[8] Hopp, T.P. and Woods, K.R. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proc. Nat. Acad. Sci., U.S.A.*, **78**(6), 3824–3828.

[9] Hubbard, T.J.P. and Blundell, T.L. (1987) Comparison of solvent-inaccessible cores of homologous proteins: definitions useful for protein modelling. *Protein Eng.*, **1**(3), 159–171.

[10] Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

[11] Kuntz, I.D. (1972) Protein folding. *J. Am. Chem. Soc.*, **94**, 4009–4012.

[12] Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–132.

[13] Qian, H. (1996) Prediction of α-helices in proteins based on thermodynamic parameters from solution chemistry. *J. Mol. Biol.*, **256**, 663–666.

[14] Rose, G.D. (1978) Prediction of chain turns in globular proteins on a hydrophobic basis. *Nature*, **272**, 586–590.

[15] Rose, G.D. and Roy, S. (1980) Hydrophobic basis of packing in globular proteins. *Proc. Nat. Acad. Sci., U.S.A.*, **77** (8), 4643–4647.

[16] Rost, B. and Sander, C. (1993) Prediction of protein secondary structure at better than 70% Accuracy. *J. Mol. Biol.*, **232**, 584–599.

[17] Rost, B., Sander, C. and Schneider, R. (1994a) Redefining the goals of protein structure prediction. *J. Mol. Biol.*, **235**, 13–26.

[18] Rost, B. and Sander, C. (1994b) Combining evolutionary information and neural networks to predict secondary structure. *PROTEINS: Structure, Function, and Genetics*, **19**, 55–72.

[19] Swindells, M.B. (1995) A procedure for the automatic determination of hydrophobic cores in protein structures. *Protein Science*, **4**, 93–102.

[20] Vihinen, M., Torkkila, E. and Riikonen, P. (1994) Accuracy of protein flexibility predictions. *PROTEINS: Structure, Function, and Genetics*, **19**, 141–149.

[21] Wako, H. and Blundell, T.L. (1994) Use of amino acid environment-dependent substitution tables and conformational propensities in structure prediction from aligned sequences of homologous proteins. I.Solvent Accessibility classes. *J. Mol. Biol.*, **238**, 682–692.

[22] Yamada, M. and Ohkitani, K. (1990) Orthonormal wavelet expansion and its application to turbulence. *Prog. Theor. Phys.*, **83**(5), 819–823.

[23] Yamada, M. and Ohkitani, K. (1991) An identification of energy cascade in turbulence by orthonormal wavelet analysis. *Prog. Theor. Phys.*, **86**(4), 799–815.