Protein Sequence and Structure Relationship ARMA Spectral Analysis: Application to Membrane Proteins

Shaojian Sun* and R. Parthasarathy‡

*Department of Pharmaceutical Chemistry, University of California-San Francisco, San Francisco, California 94118, and ‡Biophysics Department, Roswell Park Cancer Institute, Buffalo, New York 14263 USA

ABSTRACT If it is assumed that the primary sequence determines the three-dimensional folded structure of a protein, then the regular folding patterns, such as α-helix, β-sheet, and other ordered patterns in the three-dimensional structure must correspond to the periodic distribution of the physical properties of the amino acids along the primary sequence. An AutoRegressive Moving Average (ARMA) model method of spectral analysis is applied to analyze protein sequences represented by the hydrophobicity of their amino acids. The results for several membrane proteins of known structures indicate that the periodic distribution of hydrophobicity of the primary sequence is closely related to the regular folding patterns in a protein's three-dimensional structure. We also applied the method to the transmembrane regions of acetylcholine receptor α subunit and Shaker potassium channel for which no atomic resolution structure is available. This work is an extension of our analysis of globular proteins by a similar method.

INTRODUCTION

Membrane proteins play an important role in modulating the activities of many cellular functions. Despite considerable efforts, very few membrane proteins have yielded crystals that diffract x rays to high resolution. A large number of predictive algorithms for the secondary structural topology of membrane proteins have been proposed (see Fasman, 1989; Fasman and Gilbert, 1990 for reviews), including statistical methods (Chou-Fasman, 1978; Garnier et al., 1978; Biou et al., 1988), Kyte-Doolittle hydropathy plots (Kyte and Doolittle, 1982), hydrophobic moment analysis (Eisenberg et al., 1984), hydrophobicity analysis (von Heijne 1991; Jähnig 1989), and direct Fourier transformation-based hydrophobicity analysis (Finer-Moore et al., 1989). New hydrophobicity scales developed for membrane proteins have also been proposed (Argos et al., 1982; Engelman et al., 1986).

It is physically understandable that the periodic distribution of physical properties of amino acids (such as hydrophobicity, charge, dipole moment, etc.) along the primary sequence may be one of the important factors determining the regular folding patterns, such as α-helix, β-sheet, and other ordered patterns, in the three-dimensional structure of a protein.

A primary sequence of a protein of length N amino acids can be represented by the physicochemical properties of individual amino acid along the chain,

$$\begin{pmatrix} p_1^1 \\ p_2^1 \\ \vdots \\ p_m^1 \end{pmatrix} \begin{pmatrix} p_1^2 \\ p_2^2 \\ \vdots \\ p_m^2 \end{pmatrix}, \cdots,$$

$$\begin{pmatrix} p_1^i \\ p_2^i \\ \vdots \\ p_m^i \end{pmatrix} \begin{pmatrix} p_1^{i+1} \\ p_2^{i+1} \\ \vdots \\ p_m^{i+1} \end{pmatrix}, \cdots, \begin{pmatrix} p_1^N \\ p_2^N \\ \vdots \\ p_m^N \end{pmatrix},$$

$$i \in [1, N]$$

where i is the index for the sequence number, and p_m^i is the mth kind physical property of the amino acid at ith position along the primary sequence. The physicochemical property p_m^i can be the hydrophobicity index, charge density, volume, etc., of an amino acid. We assume that the ordered folding patterns observed in the protein crystal structures, such as secondary and tertiary structures in most of proteins, should be characterized, to a large extent, by the periodic distribution of these quantities along their primary sequences. The question then is how to compute the hidden periodicity of these physicochemical properties from a given protein sequence. Instead of working with a vectorial physicochemical quantity in the computation of periodicity, the simplest approach is to represent a primary sequence by a scalar sequence, for example, the hydrophobicity sequence of a protein. The hydrophobic interaction has been identified as one of the most important interactions in determining the

Received for publication 17 September 1993 and in final form 29 December 1993.

Address reprint requests to Shaojian Sun, Department of Pharmaceutical Chemistry, Laurel Heights Campus, University of California-San Francisco, 3333 California Street, Room 102, San Francisco, CA 94118. Tel.: 415-476-8910; Fax: 415-476-1508; E-mail: shaojian@maxwell.ucsf.edu.

0006-3495/94/06/2092/15 $2.00

three-dimensional structures of proteins (Kauzmann, 1959; Dill, 1990). Although similar method can be developed in parallel for vectorial computation, we will only focus on the computational method for the scalar representation of protein primary sequence.

Spectral analysis methods compute the frequency domain representation of sequential signals. The peaks in the frequency domain spectrum correspond to the periodic oscillations of signal components in the original sequence. Direct Fourier analysis of the hydrophobicity sequence of a protein to reveal the possible regular folding patterns has been employed by several authors (McLachlan and Stewart, 1976; Finer-Moore et al., 1989; Sun and Parthasarathy, 1990). This method usually produces low resolution spectra because the hydrophobicity sequence of a protein is commonly a noisy sequence (Sun, 1993). Although filtering methods can be used in the direct Fourier analysis to partially remedy this sequence noise problem, it is well know that the computed spectra may be distorted by the addition of filters. Furthermore, the direct Fourier transform analysis uses a windowed sampling method, assuming zero data outside the window and, therefore, the unestimated autocorrelation sequence values outside the window are implicitly zero, normally an unrealistic assumption. In doing so, both the resolution and reliability of the estimated spectrum are sacrificed (Sun, 1993).

In a recent study, we proposed (Sun and Parthasarathy, 1993) an AutoRegressive Moving Average (ARMA) model method (Marple, 1987; Jenkins and Watts, 1968; Kay, 1987) to analyze the periodic distribution of hydrophobicities along the primary sequences of proteins. The ARMA model method of spectral estimation does not make a direct Fourier transform of the original hydrophobicity sequence data. Rather, it uses the hydrophobicity sequence to construct a mathematical model. The model is required to approximate the distribution that generated the observed hydrophobicity sequence data. The frequency-domain spectrum of a given sequence is computed from the parameters in the fitted ARMA model. It is well known that (Marple, 1987; Jenkins and Watts, 1968; Kay, 1987) ARMA-computed spectra have better resolution than classical periodograms or correlograms based on direct Fourier transforms. In this method of spectral analysis, the random component in the hydrophobicity sequence is filtered out automatically. Unlike spectral estimation by either periodogram or correlogram, the ARMA method overcomes this problem, because the power spectral density function of the sequence is totally described in terms of the model parameters and the variance of the white noise process (see Theory). Application of this spectral analysis method to a set of representative globular and fibrous proteins with different folding patterns (Sun and Parthasarathy, 1993; Sun, 1993) demonstrated that this method can produce spectra of much higher resolution than the direct Fourier transform method. We have found that there is a pronounced correlation between the ordered folding patterns in the three-dimensional structure of a protein and the computed ARMA spectrum of its hydrophobicity sequence. In the present study, we demonstrate that similar results hold for membrane proteins as well.

THEORY

In this section, we will define an ARMA process and provide the formulae for the power spectral density function expressed in terms of the ARMA model parameters. We will explain further how the model parameters and model orders can be chosen properly for a given set of data (in our case the hydrophobicity sequence). We refer readers to several standard textbooks (Marple, 1987; Jenkins and Watts, 1968; Kay, 1987) for further technical details.

Let $x[n]$, $n = 1, \ldots, N$ denote the hydrophobicity index of a protein primary sequence of length N amino acids according to a chosen hydrophobicity scale. This sequence can be modeled mathematically by a filter linear difference equation (Marple, 1987):

$$\begin{aligned} x[n] &= -\sum_{k=1}^{p} a[k]x[n-k] + \sum_{k=0}^{q} b[k]u[n-k] \\ &= \sum_{k=0}^{\infty} h[k]u[n-k], \end{aligned} \tag{1}$$

where p, q are called the autoregressive order and moving average order, respectively, $a[k]$, $k \in [1, p]$ and $b[k]$, $k \in [1, q]$ are the autoregressive coefficients and moving average coefficients, respectively (the assumption $b[0] = 1$ can be made without loss of generality because the input $u[n]$ can always be scaled to account for any filter gain). $u[i]$ is a white noise sequence of zero mean and variance ρ_w. (Obviously $h[k] = 0$ for $k < 0$). The model parameters p, q, $a[k]$, $k \in [1, p]$, $b[k]$, $k \in [1, q]$, and ρ_w are determined by the given hydrophobicity sequence derived from protein primary sequence. This ARMA model equation, with the determined model parameters, describes an infinite data process that has all of the characteristics of the finite observed data sequence (the hydrophobic sequence in our case). The ARMA power spectrum is computed from this model equation rather than directly computed from the original hydrophobicity sequence (as it would be in the direct Fourier transform method). This is why the ARMA spectrum can achieve a much higher spectral resolution than the direct Fourier transform method.

To compute the power spectral density function for the ARMA model equation, the system function $H(z)$ between the input and the output,

$$H(z) = \frac{B(z)}{A(z)}, \tag{2}$$

has to be used, where

$$A(z) = 1 + \sum_{k=1}^{p} a[k]z^{-k}, \tag{3}$$

$$B(z) = 1 + \sum_{k=1}^{q} b[k]z^{-k}, \tag{4}$$

$$H(z) = 1 + \sum_{k=1}^{\infty} h[k]z^{-k} \tag{5}$$

The z-transform (Oppenheim and Willsky, 1983) of the output sequence $x[n]$ is related to the z-transform of the input

random process $u[n]$ by

$$P_{xx}(z) = P_{uu}(z)H(z)H^*(1/z^*) = P_{uu}(z)\frac{B(z)B^*(1/z^*)}{A(z)A^*(1/z^*)}. \quad (6)$$

We will assume further that the sequence $u[i]$ is a white noise process of zero mean and variance ρ_w, so that $P_{uu}(z) = \rho_w$. The ARMA power spectral density function, $P_{ARMA}(f)$ is obtained by substituting $z = \exp(i2\pi fT)$ into Eq. 6 and scaling by the sample interval T ($T = 1$ in our case):

$$P_{ARMA}(f) = T\rho_\omega \left|\frac{B(f)}{A(f)}\right|^2 = T\rho\frac{e_q^H(f)bb^He_q(f)}{e_p^H(f)aa^He_p(f)}, \quad (7)$$

in which the polynomials are defined as

$$A(f) = 1 + \sum_{k=1}^{p} a[k]\exp(-i2\pi fkT)$$
$$B(f) = 1 + \sum_{k=1}^{q} b[k]\exp(-i2\pi fkT) \quad (8)$$

and

$$e_p(f) = \begin{pmatrix} 1 \\ \exp(i2\pi fT) \\ \vdots \\ \exp(i2\pi fpT) \end{pmatrix}, \quad a = \begin{pmatrix} 1 \\ a[1] \\ \vdots \\ a[p] \end{pmatrix}$$
$$e_q(f) = \begin{pmatrix} 1 \\ \exp(i2\pi fT) \\ \vdots \\ \exp(i2\pi fqT) \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ b[1] \\ \vdots \\ b[q] \end{pmatrix} \quad (9)$$

The ARMA power spectral density function is evaluated over the range $-1/2T \le f \le 1/2T$, and is symmetric for the real hydrophobicity sequence.

The model parameters in the ARMA(p, q) model Eq. 1 are related to the autocorrelation sequence of the known hydrophobic sequence of the given protein. Multiplying Eq. 1 by $x^*[n - m]$ and taking the expectation of the resulting equation, we have

$$\epsilon\{x[n]x^*[n-m]\} = -\sum_{k=1}^{p} a[k]\epsilon\{x[n-k]x^*[n-m]\} + \sum_{k=0}^{q} b[k]\epsilon\{u[n-k]x^*[n-m]\}, \quad (10)$$

or

$$r_{xx}[m] = -\sum_{k=1}^{p} a[k]r_{xx}[m-k] + \sum_{k=0}^{q} b[k]r_{ux}[m-k]. \quad (11)$$

The cross correlation $r_{ux}[i]$ between the input and the output can be expressed in terms of the $h[k]$ parameters of Eq. 5 by using Eq. 1 to yield

$$r_{ux}[i] = \epsilon\{u[n+i]x^*[n]\} = \epsilon\left\{u[n+i]\left(u^*[n] + \sum_{k=1}^{\infty} h^*[k]u^*[n-k]\right)\right\} = r_{uu}[i] + \sum_{k=1}^{\infty} h^*[k]r_{uu}[i+k]. \quad (12)$$

Assuming $u[k]$ is a white noise process, then

$$r_{ux}[i] = \begin{cases} 0 & \text{for } i > 0 \\ \rho_w & \text{for } i = 0. \\ \rho_w h^*[-i] & \text{for } i < 0 \end{cases} \quad (13)$$

The final relationship between the ARMA parameters and the autocorrelation sequence of the hydrophobicity sequence $x[n]$ is

$$r_{xx}[m] = \begin{cases} r_{xx}^*[-m] & \text{for } m < 0 \\ \left\{-\sum_{k=1}^{p} a[k]r_{xx}[m-k] + \rho_w\sum_{k=m}^{q} b[k]h^*[k-m]\right\} & \text{for } 0 \le m \le q \\ -\sum_{k=1}^{p} a[k]r_{xx}[m-k] & \text{for } m > q \end{cases} \quad (14)$$

where $h[0] = 1$ has been used.

It is easily seen that the autoregressive parameters of the ARMA model are related by a set of linear equations to the autocorrelation sequence of the known hydrophobicity sequence. Eq. 14 may be evaluated for the p lag indices $q + 1 \le m \le q + p$, for example, and grouped into the matrix expression

$$\begin{pmatrix} r_{xx}[q] & r_{xx}[q-1] & \cdots & r_{xx}[q-p+1] \\ r_{zz}[q+1] & r_{xx}[q] & \cdots & r_{xx}[q-p+2] \\ \vdots & \vdots & \vdots & \vdots \\ r_{xx}[q+p-1] & r_{xx}[q+p-2] & \cdots & r_{xx}[q] \end{pmatrix} \cdot \begin{pmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{pmatrix} = -\begin{pmatrix} r_{xx}[q+1] \\ r_{xx}[q+2] \\ \vdots \\ r_{xx}[q+p] \end{pmatrix} \quad (15)$$

These equations are called ARMA Yule-Walker normal equations. All of the diagonal elements are equal (Toeplitz form). This indicates that the autoregressive parameters may be found separately for the moving average parameters as the solution of the simultaneous Eq. 15. Unfortunately, the moving average parameters of an

ARMA model cannot be found simultaneously with the impulse response coefficients $h[k]$, as indicated by Eq. 14, resulting in a nonlinear relationship with the autocorrelation sequence.

Simultaneous estimate of the AR and MA parameters of an ARMA model is difficult and computationally intensive. The AR and the MA parameters are typically estimated separately. Usually, the AR parameters are estimated first, and then used to construct an inverse filter to apply to the original data. The residuals out of this filter should be representative of a moving average process, to which an MA parameter estimator can be applied. There are three main steps involved in this ARMA parameter estimation: 1) Estimate the autoregressive (AR) parameters by the least-squares modified Yule-Walker method; 2) filter the original data sequence with the estimated AR parameters; 3) estimate the moving average (MV) parameters from the filtered residuals. The selection of the autoregressive orders and moving average orders, p and q, is determined by the Akaike information criterion (Akaike 1974).

AR parameter $a[k]$ estimation

From Eq. 14, the autocorrelation sequence for an ARMA(p, q) model satisfies the relationship

$$r_{xx}[n] = -\sum_{k=1}^{p} a_p[k] r_{xx}[n-k] \tag{16}$$

for $n > q$. If the autocorrelation sequence are known exactly, then the p relations for $q + 1 \le n \le q + p$ can be set up as a set of simultaneous equations and solved for the AR parameters, as indicated by the Yule-Walker normal equations in (15). However, in practice, only finite data samples are available, so the autocorrelation estimates must be substituted for the unknown autocorrelation sequence. A least-squares modified Yule-Walker method has been developed (Mehra, 1971; Porat and Fridlander, 1985). The method uses more than p equations for lags greater than q and minimizes a squared error to fit the p parameters. Assuming autocorrelation estimates $\tilde{r}_{xx}[n]$ from lag 0 to lag M are calculated, where M is the largest lag index that can be accurately estimated, the $M - q$ equations (such that $M - q > p$)

$$\tilde{r}_{xx}[n] = -\sum_{k=1}^{p} a_p[k] \tilde{r}_{xx}[n-k] + \epsilon[n] \tag{17}$$

for $q + 1 \le n \le M$ are formed. $\epsilon[n]$ is the estimation error. The sum of squared error,

$$\rho = \sum_{n=q+1}^{M} |\epsilon[n]|^2 = \sum_{n=q+1}^{M} |r_{xx}[n] - \tilde{r}_{xx}[n]|^2, \tag{18}$$

is minimized with respect to the p AR parameters $a[k]$, and the normal equations to be solved are

$$T_p^H T_p \begin{pmatrix} 1 \\ a[1] \\ \vdots \\ a[k] \end{pmatrix} = \begin{pmatrix} \rho \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{19}$$

where T_p is a rectangular Toeplitz matrix of autocorrelation estimates

$$T_p = \begin{pmatrix} \tilde{r}_{xx}[q+1] & \cdots & \tilde{r}_{xx}[q-p+1] \\ \vdots & \ddots & \vdots \\ \tilde{r}_{xx}[M-p] & \cdots & \tilde{r}_{xx}[q+1] \\ \vdots & \ddots & \vdots \\ \tilde{r}_{xx}[M] & \cdots & \tilde{r}_{xx}[M-p] \end{pmatrix} \tag{20}$$

Filter the original data sequence

Once the AR parameters are estimated by solving Eq. 19, the next step is to use them to construct an inverse filter. Then the reverse filter is applied to obtain the parameters for the MA process. The filter system function is

$$\tilde{A}(z) = 1 + \sum_{k=1}^{p} \tilde{a}[k] z^{-k} \tag{21}$$

in which the $\tilde{a}[k]$ are the estimated AR parameters. The ARMA system function is $B(z)/A(z)$, so that

$$\frac{B(z)}{A(z)} \tilde{A}(z) \approx B(z). \tag{22}$$

Therefore, passing the measured data sample through the filter with the system function $\tilde{A}(z)$ will yield an approximate moving average process at the filter output. The filtered sequence of length $N - p$ is given by the convolution

$$z[n] = x[n] + \sum_{m=1}^{p} \tilde{a}[m] x[n-m], \tag{23}$$

for $p + 1 \le n \le N$.

MA parameter $b[k]$ estimation

The most obvious approach to estimate the MA parameters would be to solve the nonlinear Eq. 14 for $P = 0$ using the autocorrelation estimates made from the original sequence. Unfortunately, the solutions are very involved (Box and Jenkins, 1970). However, there is an alternative method based on a high-order AR approximation to the MA process that uses only linear operations (Kay, 1987; Marple, 1987). An MA(q) process has a system function

$$B(z) = 1 + \sum_{k=1}^{q} b[k] z^{-k}. \tag{24}$$

An AR(∞) process that is equivalent to a $MA(q)$ process has

a system function of the form $1/A_\infty(z)$, where

$$A_\infty(z) = 1 + \sum_{k=1}^{\infty} a[k]z^{-k}. \quad (25)$$

Therefore, we have

$$B(z)A_\infty(z) = 1. \quad (26)$$

An inverse z-transform of the Eq. 26 yields a convolution relation of the MA parameters with the AR parameters

$$a[m] + \sum_{n=1}^{q} b[n]a[m-n] = \delta[m] = \begin{cases} 1, & \text{if } m = 0; \\ 0, & \text{if } m \neq 0 \end{cases}. \quad (27)$$

where, by definition, $a[0] = 1$ and $a[k] = 0$, for $k < 0$. Thus, the moving average parameters can be determined from the parameters of an equivalent infinite-order AR model by solving any subset of q equations formed by (27). In practice, one can calculate a high-order $AR(M)$ ($M \gg q$) estimate from the filtered residual sequence given by (23). From the high-order $AR(M)$ parameters estimates $1, \tilde{a}_M[1], \cdots, \tilde{a}_M[M]$, one can construct

$$\epsilon_{MA} = \tilde{a}_M[m] + \sum_{n=1}^{q} b[n]\tilde{a}[m-n]. \quad (28)$$

So the estimates of MA parameters are obtained by the minimization of the squared error variance

$$\rho_q = \sum_{0 \le m \le M+q} |\epsilon_{MA}|^2/M. \quad (29)$$

Eq. 29 is solved by a least-squares technique described in the determination of the AR parameters.

Determination of the model orders (p, q)

Selecting the order for an ARMA model is not a simple procedure. One of the most frequently used techniques is Akaike information criterion (AIC) (Akaike, 1974). The AIC determines the model order by minimizing an information theoretic function

$$\mathrm{AIC}(p, q) = N \ln(\rho_{pq}) + 2(p + q), \quad (30)$$

in which ρ_{pq} is an estimate of the input white noise variance of the assumed $ARMA(p, q)$ model, and N is the number of sampled data. One could also check the whiteness of the residual output of the inverse filter with system function $B(z)/A(z)$ using the estimated AR and MA parameters. The estimated autocorrelation of the residual sequence should be approximately zero except at lag zero.

SPECTRAL ANALYSIS FOR MEMBRANE PROTEINS

The ARMA spectral analysis method presented in Theory requires the hydrophobicity sequence derived from a protein primary sequence as the only input. The periodicities of the hydrophobicity distribution along the primary sequence are then calculated from the ARMA model equation with the model parameters computed from the original hydrophobicity sequence. The ARMA spectral peaks correspond to the dominant periodic hydrophobicity distribution along the primary sequence. These dominant periodic hydrophobicity distributions correspond to possible regular folding patterns, such as α-helix, β strand, and other ordered folds.

Application of the ARMA method of spectral analysis to globular proteins with different folding patterns (Sun and Parthasarathy, 1993) reveals the following typical features that seem to appear in the membrane proteins studied here.

1) Global folding patterns: low frequency, 0.10–0.05 cycle/residue (10–20 residue/cycle), peaks in the spectral density function correspond to the large scale orders in the folded structures.
2) The α helix peak: the spectral peaks around 0.30–0.25 cycle/residue (3.3 to 4.0 residues/cycle) and their higher harmonics correspond to a possible α-helix conformation.
3) The β strand peak: the spectral peaks are typically greater than 0.35 cycle/residue (around 2.0 to 2.7 residues/cycle).
4) A randomly shuffled sequence of known protein sequences has no significant spectral peaks.

In this section, we will examine the computed ARMA power spectral density function (PSDF) for hydrophobicity sequences of several membrane proteins, and compare their spectral characteristics with their known structures. We will also discuss the structural implications of the computed ARMA PSDF for the transmembrane regions of the acetylcholine receptor α subunit and the *Shaker* potassium channel protein. In this study, the hydrophobicity scale of Rose et al. (1985) has been used. The Rose scale is derived from the solvent accessibility of amino acids in 23 x-ray crystallography-determined protein structures.

Melittin

The crystal structure of melittin (Terwilliger and Eisenberg, 1982), a 26-residue membrane bound protein, shows a bent α-helical structure. The bending appears at the middle of the helix due to a proline residue. Each of the helical section contains about 12 residues. Fig. 1 *A* is a stereo backbone plot of the x-ray crystallography-determined structure of melittin. Fig. 1 *B* plots the computed ARMA power spectral density function (PSDF) for the melittin primary sequence. The melittin PSDF has an α-helix peak at 0.277 cycle/residue (3.61 residue/cycle), indicating that melittin adopts an α-helical conformation. There are also two small peaks in the melittin PSDF plot. One peak at around 0.075 cycle/residue (13 residue/cycle), with a peak amplitude about 6.7% of the major one, corresponds to the segment length of the two α-helices. This is indeed true in the crystal structure as one

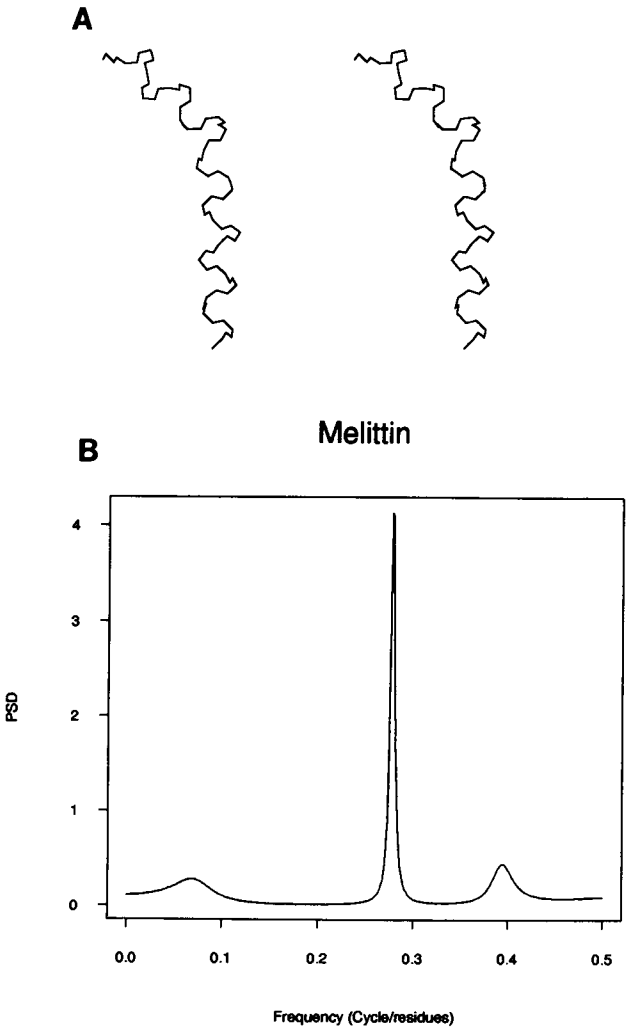

FIGURE 1 (*A*) Melittin crystal structure (backbone); (*B*) ARMA PSDF.

can see clearly in Fig. 1 *A*. We have no explanation for the second small peak at around 0.4 cycle/residue.

Porin

Porins are found in the outer membranes of gramnegative bacteria, mitochondria, and chloroplasts. They form channels for small hydrophilic molecules that are in most cases weakly ion-selective. Porins are trimeric and consist predominantly of β-pleated sheet structures determined by x-ray crystallographic diffraction (Weiss et al., 1991) (Fig. 2 *A*).

Fig. 2 *B* shows the computed ARMA power spectral density function (PSDF) for the entire porin primary sequence. There is a very strong and sharp peak at the frequency 0.375 cycle/residue (2.67 residue/cycle), which indicates the predominant content of β-sheet. This is in agreement with the crystal structure of porin. The ARMA spectral analysis method can also be applied to a part of a protein primary sequence. Fig. 2 *C* shows the PSDF for the porin sequence fragment 1–41. It has a predominant peak at 0.5 cycle/residue (2.0 residue/cycle) indicating a strong β-sheet structure tendency in this C-terminal segment. Fig. 2 *D* shows the PSDF for the sequence segment 227–301. There are four peaks. The two major peaks at 0.465 and 0.353 cycle/residue (2.15 and 2.83 residue/cycle) indicate that this segment is a β-sheet structure. The peak at 0.059 cycle/residue (17 residue/cycle) suggests an ordered folding pattern at about 17 residues in length. This pattern can been seen clearly in the three-dimensional crystallographic structure of porin. Fig. 2 *E* plots the ARMA PSDF for the sequence segment corresponding the loop region (74–116) in the three-dimensional structure of porin. There is a clear peak at 0.394 cycle/residue (2.54 residue/cycle), indicating a β-sheet-like structure. The crystal structure of this segment shows it is indeed a β-sheet like loop, running against the inner wall of the β-sheet barrel of porin.

Colicin

Colicins are antibiotic proteins produced by *Eschericia coli*, active against sensitive *E. coli* and closely related bacteria. The largest group of colicins comprises those that can form voltage-dependent channels in membranes, thereby destroying the the cell's energy potential. The pore-forming activity seems to be located at the carboxy terminus, which comprises about 200 residues (389–592 in the primary sequence of colicin A). This fragment is soluble in aqueous media, but nevertheless spontaneously inserts itself into lipid membranes. The crystal structure of this fragment shows (Parker et al., 1989) that colicin forms 10 α-helices (Fig. 3 *A*).

The ARMA PSDF (Fig. 3 *B*) for this sequence fragment (389–592) shows that it is exclusively α-helical. There is only one strong peak in the PSDF at 0.263 cycle/residues (3.80 residue/cycle), which is a typical α-helix peak. We also computed the ARMA PSDF for the sequence fragment 389–433, which corresponds to helices one and two in the crystal structure. The PSDF (Fig. 3 *C*) for this smaller sequence fragment shows a very strong α-helix peak at 0.273 cycle/residue (3.66 residue/cycle), indicating that this smaller fragment is in an α-helical conformation. A very minor peak at 0.1 cycle/residues (10 residue/cycle), which can hardly be seen in the linear scale (about 1% of the major peak), indicates a very week hydrophobicity oscillation along this smaller fragment.

Bacteriorhodopsin

The structure of bacteriorhodopsin was first elucidated by high resolution electron microscopy of two-dimensional crystals (Herderson and Unwin, 1975; Henderson et al., 1990). The 3-Å resolution-refined structure shows that bacteriorhodopsin contains seven transmembrane α-helices (Fig. 4 *A*). The ARMA PSDF for the bacteriorhodopsin sequence (Fig. 4 *B*) shows two spectral peaks at 0.10 and 0.256 cycle/residue (10.0 and 3.9 residue/cycle), respectively. The major peak at 3.9 residue/cycle indicates that bacteriorhodopsin forms mostly α-helical structure. The minor peak at around 10 residue/cycle corresponds to a typical hydrophobicity periodicity in these α-helical segments. If only the

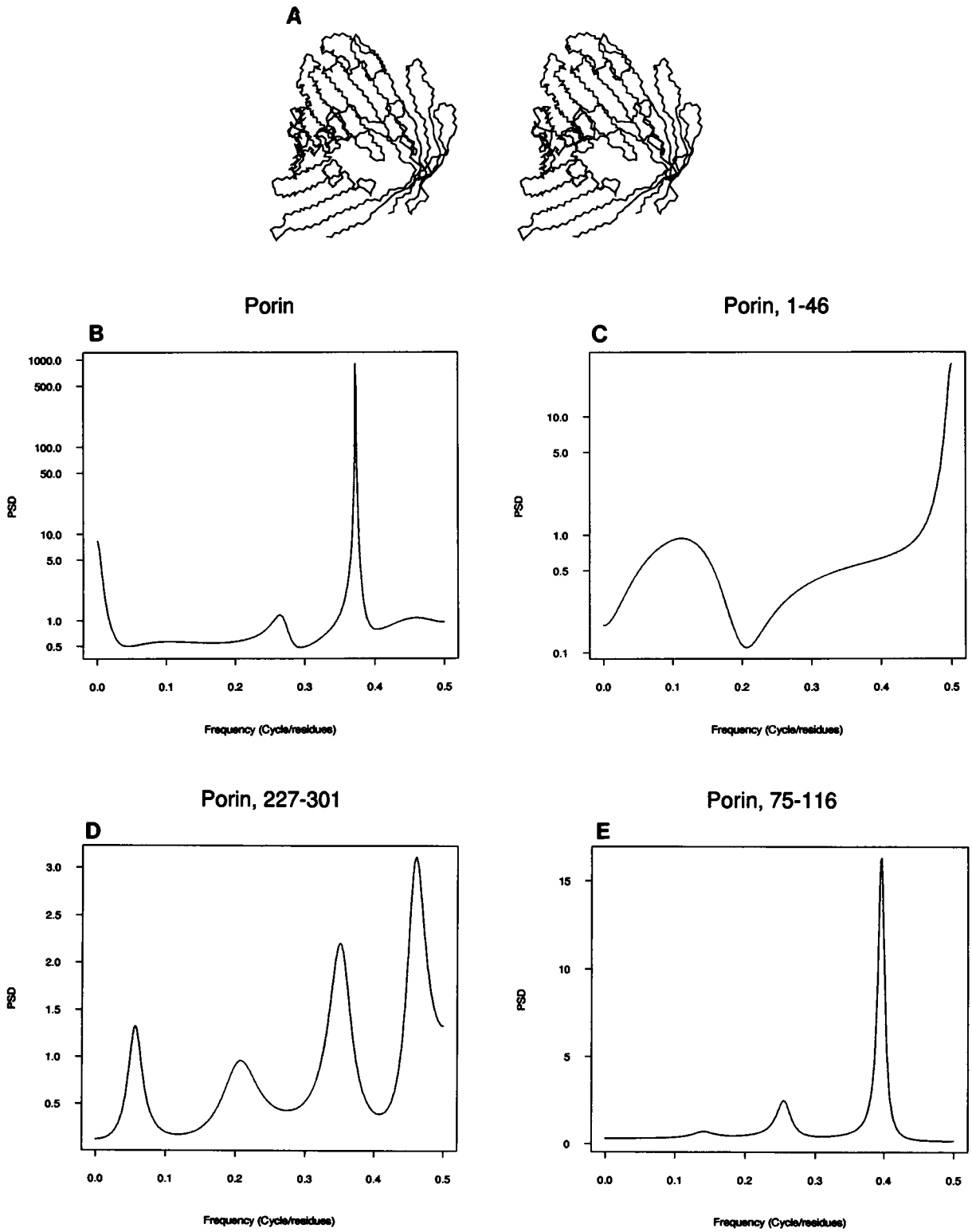

FIGURE 2 (*A*) Porin crystal structure (backbone); (*B*) ARMA PSDF for the entire Porin sequence; (*C*) ARMA PSDF for Porin sequence 1–46; (*D*) ARMA PSDF for Porin sequence 227–301; (*E*) ARMA PSDF for Porin sequence 75–116.

sequence segments of the transmembrane portion are used in the PSDF computation, we obtain similar spectral characteristics (Fig. 4 *C*), in which the two peaks are at 0.11 and 0.277 cycle/residue (9.0 and 3.6 residue/cycle). The amplitude of the minor peak is only about 2.5% of the major peak amplitude. The α-helical peak of the transmembrane sequences appears much stronger (plotted in logarithmic scale) than the α-helical peak for the entire sequence. The spectral density function clearly suggests that the transmembrane portion of bacteriorhodopsin is an α-helical structure.

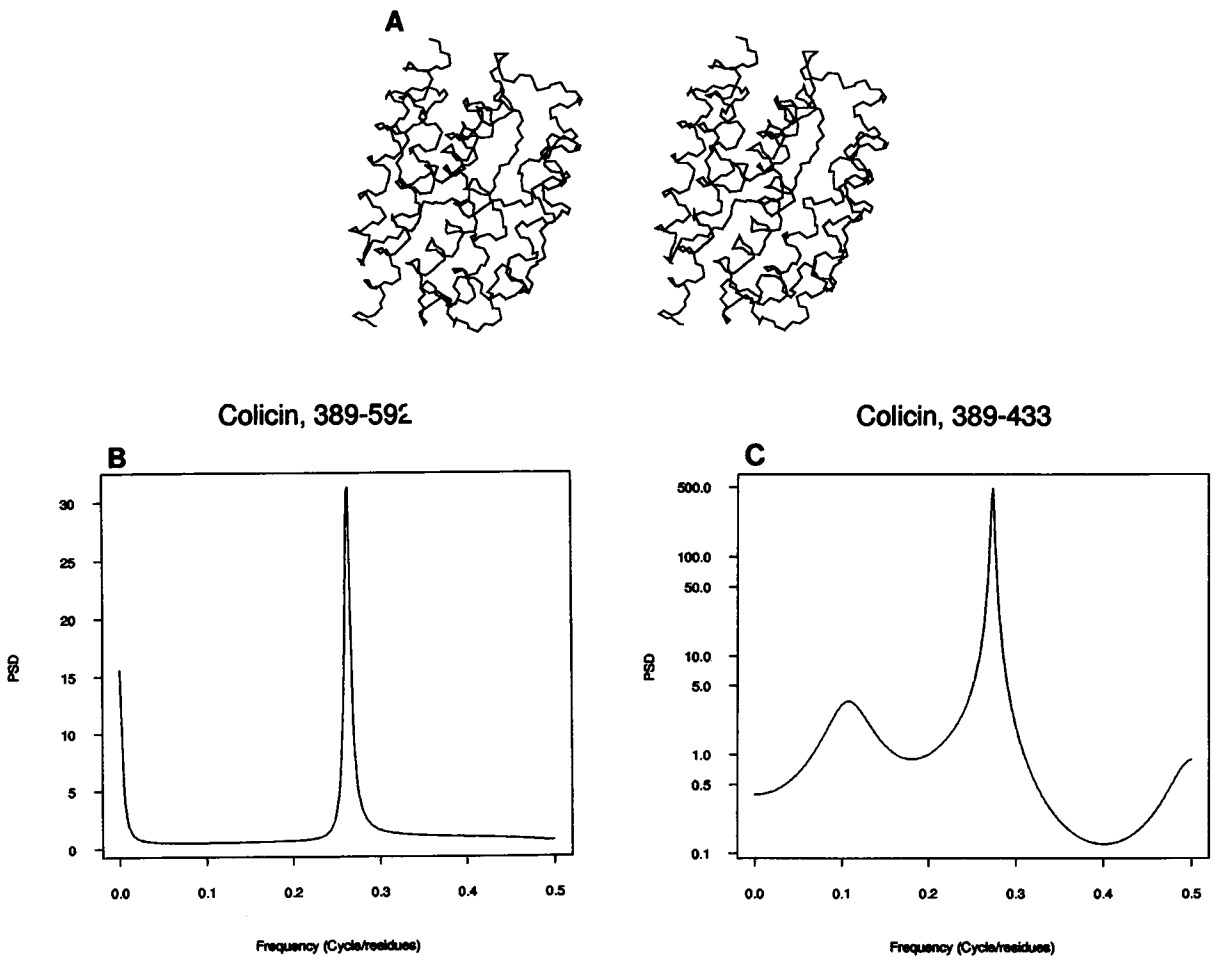

FIGURE 3 (*A*) Colicin crystal structure (backbone, transmembrane portion); (*B*) ARMA PSDF for the transmembrane portion of Colicin; (*C*) ARMA PSDF for Colicin sequence 389–433.

Photosynthetic reaction center L, M, and H domains

The photosynthetic reaction center is a complex of integral membrane proteins and cofactors (Deisenhofer et al., 1985) that use light energy to transport electrons across the membrane. The reaction center from the purple bacterium *Rhodopseudomonas viridis* consists of four protein subunits: L, M, H, and c-type cytochromes with four covalently linked haem groups (Prince et al., 1976). The structure of photosynthetic reaction center was determined by Deisenhofer et al. (1984, 1985). Consisting of mainly α-helical structures, subunits L and M form the major transmembrane portion of this protein complex. The amino-terminal helical segment of H subunit forms the only transmembrane portion of H subunit. Using the primary sequences of L, M, and H subunits, we calculated their corresponding ARMA spectral density functions. The results from the spectral analysis are in good agreement with the crystal structure.

Fig. 5 *A* is the stereo backbone plot for the L subunit. It has five transmembrane helices as well as several other nontransmembrane helical segments. Fig. 5 *B* shows the ARMA PSDF function for the entire L subunit sequence. The spectrum has a sharp peak at 0.256 cycle/residue (3.9 residue/cycle), indicating that the L subunit contains predominantly α-helical elements. We also plotted the spectrum (Fig. 5 *C*) for the transmembrane sequence segment L80–111. The ARMA spectrum for this segment clearly shows that it is an α-helical structure. The spectrum peaks at 0.278 cycle/residue (3.6 residue/cycle). This transmembrane segment corresponds to the second helix from the left in Fig. 5 *A*.

Fig. 6 *A* is the stereo backbone plot for the M subunit. Like the L subunit, it also has five transmembrane helices and several other nontransmembrane helical segments. Fig. 6 *B* shows the spectral density function for the entire M subunit sequence. The spectrum has a strong peak at 0.263 cycle/residue (3.8 residue/cycle), suggesting predominant α-helix content in the M subunit. Fig. 6 *C* shows the ARMA spectrum for the sequence segment M113–166. It has a sharp peak at 0.244 cycle/residue (4.1 residue/cycle), suggesting an α-helical structure for this transmembrane segment. In Fig. 6 *A*, the segment M113–166 corresponds to the second and the third transmembrane helices from the right.

The subunit H contains mainly β sheet and extended structures. It has only one transmembrane segment, which is α-helical in the crystal structure (Fig. 7 *A*). Fig. 7 *B* shows the the ARMA PSDF function for the entire H subunit sequence. There are three peaks in the spectrum at 0.0763 cycle/residue (13.1 residue/cycle), 0.20 cycle/residue (5.0

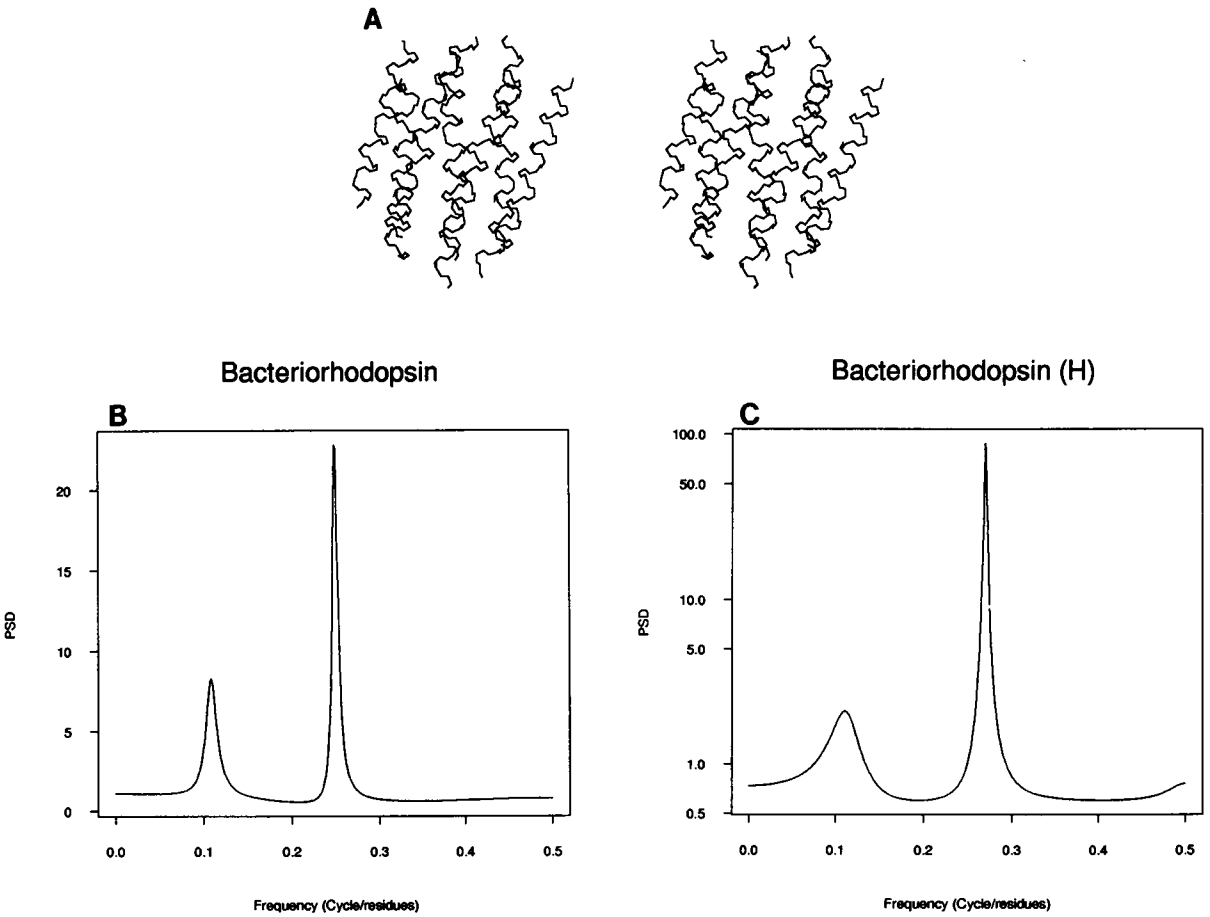

FIGURE 4 (*A*) Bacteriorhodopsin crystal structure (backbone, transmembrane portion); (*B*) ARMA PSDF for the entire Bacteriorhodopsin sequence; (*C*) ARMA PSDF for Bacteriorhodopsin transmembrane portion sequence.

residue/cycle) and 0.385 cycle/residue (2.6 residue/cycle), respectively. The major peak at 0.385 cycle/residue (2.6 residue/cycle) and the peak of its higher harmonics at 0.20 cycle/residue (5.0 residue/cycle) seem to suggest that the H subunit contains predominantly β sheet and extended structures. The peak at 13.1 residue/cycle suggests an average folding length scale. The ARMA spectrum for the transmembrane segment H11–38 (Fig. 7 *C*) shows a strong α-helix peak at 0.294 residue/cycle (3.4 residue/cycle). This is in full agreement with the crystal structure of the H subunit (the long helix at the bottom of Fig. 7 *A*). Fig. 7 *D* shows the ARMA spectrum for the sequence segment H124–211, a peripheral part of the H subunit on the cytoplasmic side of the membrane. There is a strong and major peak at 0.50 residue/cycle (2.0 residue/cycle) in the spectrum, suggesting a β sheet structure. Two small peaks (less than 5% of the major one in amplitude) are also shown in the spectrum. The peak at 0.385 cycle/residue (2.6 residue/cycle) suggests β-sheet conformation. This sequence analysis result agrees with the crystal structure very well. The segment H124–211 corresponds to the β-sheet region in the top of Fig. 7 *A*.

Our ARMA spectral analyses of the primary sequences of the photosynthetic reaction center complex are consistent with the crystal structures of these proteins.

Nicotinic acetylcholine receptor α subunit

The nicotinic acetylcholine receptor is a cation-selective, ligand-gated ion channel involved in signal transduction at the chemical synapse. It is composed of five homologous membrane-spanning subunits (α, α, β, γ, and δ). The sequence homology among different species is very high for each subunit. There is no high-resolution structure available for this receptor at the present time. It was proposed, based on the hydrophobicity plot and on the direct Fourier transform analysis (Stroud and Finer-Moore, 1985; Stroud et al., 1990; Galzi et al., 1991), that the transmembrane portion of each subunit has four transmembrane helices. Using the ARMA model, we have computed the spectra for these transmembrane sequence segments in the α-subunit (human). Here we assume that previous studies have correctly located the transmembrane region in the primary sequence of the α-subunit.

Fig. 8 *A* shows the ARMA PSDF for the entire human acetylcholine receptor α-subunit. It indicates the presence of both α-helix and β-sheet structures. Both of the characteristic spectral peaks are found in the PSDF. Fig. 8 *B–E* show the ARMA PSDF for M1(231–257), M2(263–281), M3-(297–318), and M4(429–447) transmembrane segments, respectively. There are two peaks in the PSDF for the M1

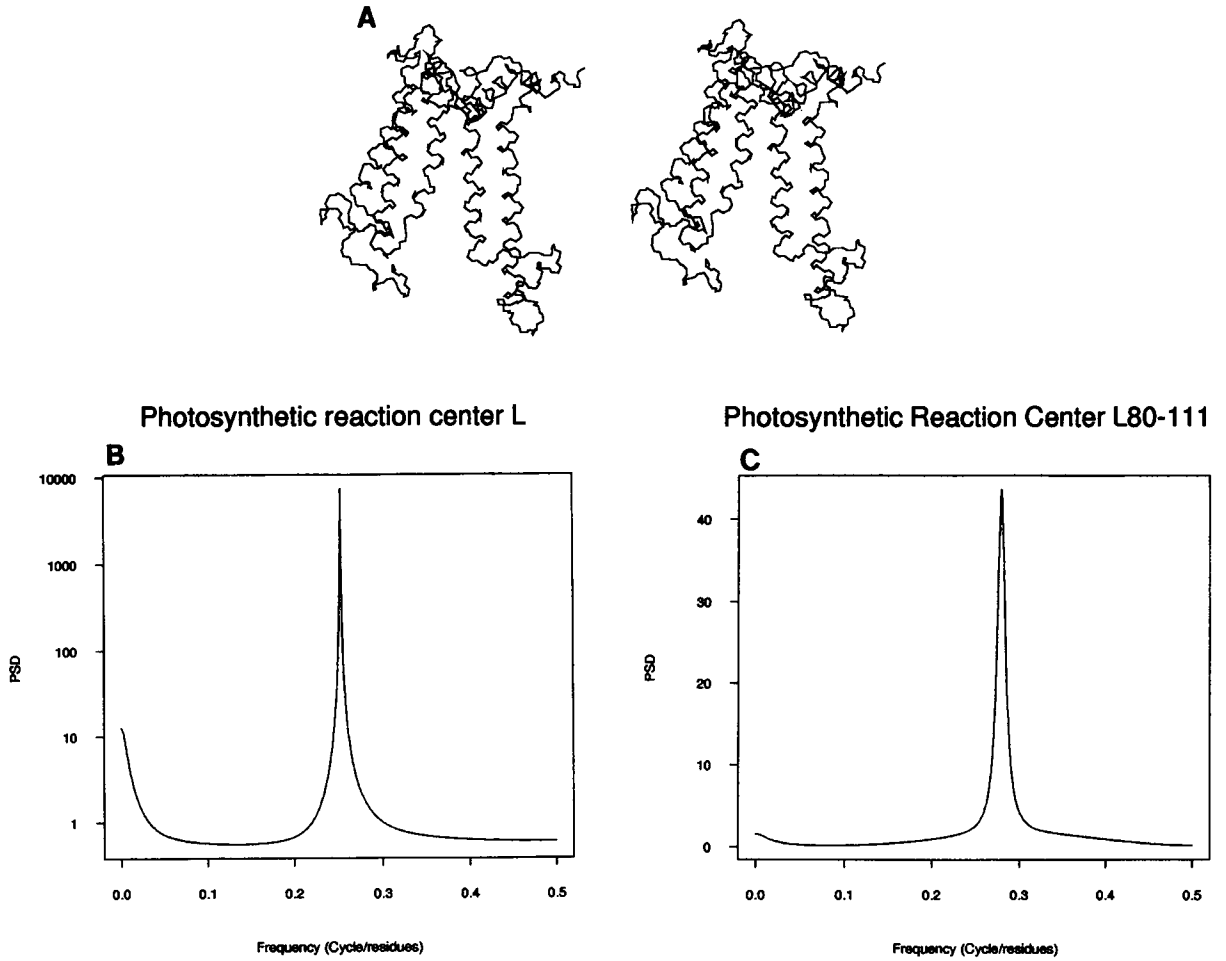

FIGURE 5 (*A*) Photosynthetic Reaction Center L subunit, crystal structure (backbone); (*B*) ARMA PSDF for the entire L sequence; (*C*) ARMA PSDF for transmembrane segment L80–111.

(Fig. 8 *B*). The major peak at 0.50 cycle/residue (2.0 residue/cycle) indicates that this segment may be a β-sheet structure. The minor peak at 0.217 cycle/residue (4.6 residue/cycle) suggests a higher harmonics of a typical β-sheet peak. The PSDF (Fig. 8 *C*) for the M2 region has a major peak at 0.50 cycle/residue (2.0 residue/cycle), suggesting a β-sheet structure. The minor peak (log scale in vertical) at 0.0476 cycle/residue (21.0 residue/cycle) suggests a structural motif length around 21 residues in the M2 region. The PSDF for the M3 region (Fig. 8 *D*) also indicates that this segment may be a β-sheet structure. There is a strong major peak at 0.50 cycle/residue (2.0 residue/cycle) in the spectrum. The minor peak at 0.185 cycle/residue (5.4 residue/cycle) hydrophobicity periodicity may corresponding to an internal folding length within the possible β-sheet structure. It is interesting to note that M1 and M2 have similar spectral shapes. The PSDF for the M4 region (Fig. 8 *E*) is very different from that of M1, M2, and M3. It has a strong major peak at 0.313 cycle/residue (3.20 residue/cycle), a significant α-helical peak, suggesting an α-helical structure structure for this segment. The above results, based on the ARMA spectral analysis of the primary sequence, suggest that the transmembrane portion of the α-subunit has three β-sheet regions and one α-helix region. These conclusions are different from the predictions based on the hydrophobicity plot and the windowed direct Fourier transform analysis (Stroud and Finer-Moore, 1985; Stroud et al., 1990), in which all the transmembrane segments have been assigned α-helical structures. However, our analysis seems to be in agreement with two recent experimental studies. Results using cysteine substitution (Akabas et al., 1992) in the M2 region indicate that the M2 region probably forms a β strand. Unwin has successfully carried out an electron microscopy analysis (Unwin, 1993) of acetylcholine receptor. He concluded, based on a 9 Å resolution EM structure, that there is only one transmembrane helix in the α-subunit, and that other three-transmembrane segments are in the β-sheet conformation. Unwin (1993) has assigned M2 to be the transmembrane helix. Our study suggests that M4 is the transmembrane helical segment.

Potassium channel

Voltage-gated potassium channels are integral membrane proteins that are fundamentally involved in the generation of bioelectric signals such as nerve impulses (Miller, 1991). There is presently no experimentally determined structural

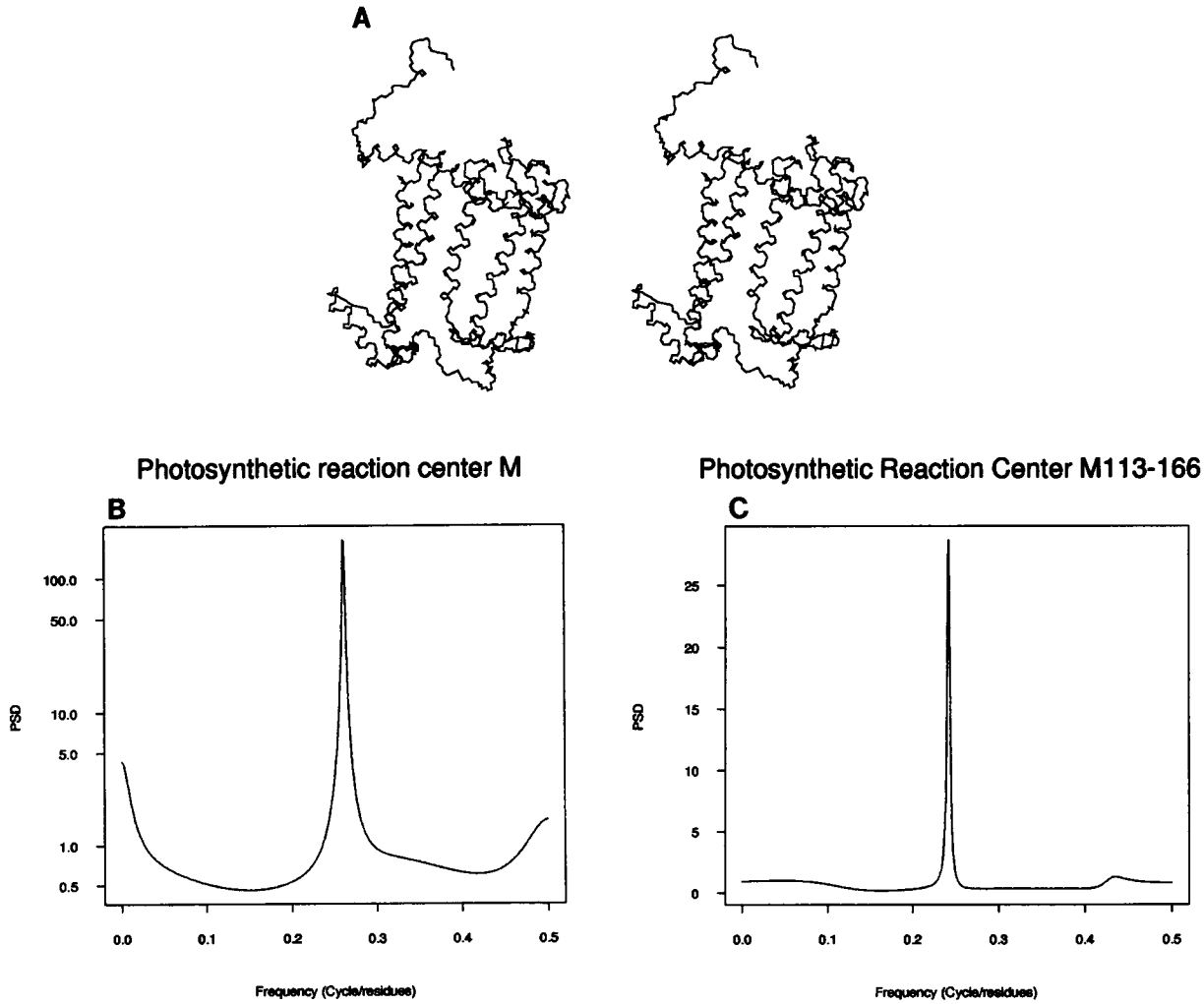

FIGURE 6 (*A*) Photosynthetic Reaction Center M subunit, crystal structure (backbone); (*B*) ARMA PSDF for the entire M sequence; (*C*) ARMA PSDF for transmembrane segment M113–166.

information about these proteins. The hydrophobicity plot for the *Shaker* potassium channel (Tempel et al., 1987) initially suggested that there were six transmembrane helices (S1–S6) in this protein (Jan and Jan, 1989). Recent experiments suggest that there is one more transmembrane region between the transmembrane helices S5 and S6 (Miller, 1991) (Fig. 9 *A*).

In this section, we will compute the spectral density function for all of the transmembrane sequence segments (the locations of the transmembrane sequence segments in Fig. 9 *A* are assumed to be approximately correct) and compare our results with the previously proposed model (Miller, 1991).

Fig. 9 *B* is the ARMA PSDF function plot for the entire sequence of the *Shaker* potassium channel. It has a sharp and strong peak at 3.23 residue/cycle, a strong indication that this protein contains predominantly α-helical structures. This is in agreement with the atomic model recently proposed (Durell and Guy, 1992). Fig. 9 *C* plots the PSDF function for the S1 sequence segment. The spectrum shows a very sharp peak at 0.294 cycle/residue (3.4 residue/cycle), suggesting an α-helical structure. The spectrum for the S2 (Fig. 9 *D*) shows both the β-sheet peak at 0.50 cycle/residue (2.0 residue/cycle), the major spectral peak, and the α-helix peak at 0.30 cycle/residue (3.33 residue/cycle) (about 10% of the major peak in amplitude). This suggests that the current S2 location in the sequence (279–300) is at a position of spatial structural transition. Indeed, if we plot the ARMA spectral for the sequence segment 290–305 (Fig. 9 *E*), we find that this sequence segment has the characteristic spectrum of a β sheet or an extended conformation because it has a sharp peak at 0.476 cycle/residue (2.1 residue/cycle). On the other hand, the spectrum for the sequence segment 275–290 (Fig. 9 *F*) indicates that this segment may be an α-helical structure, because there is a sharp peak at 0.27 cycle/residue (3.7 residue/cycle). This analysis suggests that either the current S2 location in the primary sequence has to be shifted toward N-terminal side for about 10 residues for it to be a helical region or it has to be shifted toward C-terminal in the sequence to have a β-strand conformation to avoid structure transition in the membrane. The spectral plots for S3 (peak at 0.294 cycle/residue (3.4 residue/cycle)), S4 (peak at 0.33 cycle/residue (3.0 residue/cycle)), S5 (peak at 0.244 cycle/

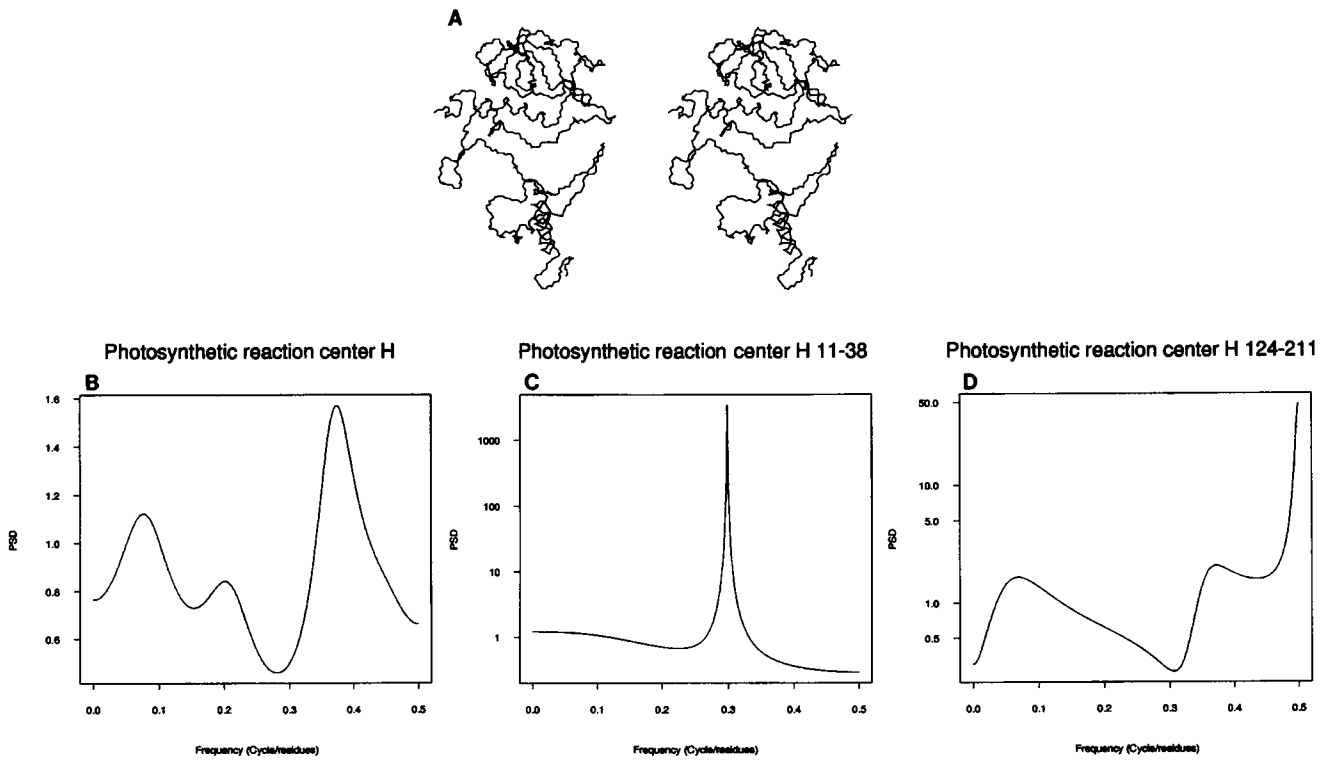

FIGURE 7 (*A*) Photosynthetic Reaction Center H subunit, crystal structure (backbone); (*B*) ARMA PSDF for the entire H sequence; (*C*) ARMA PSDF for transmembrane segment H11–38; (*D*) ARMA PSDF for H124–211.

residue (4.1 residue/cycle)), and S6 (peak at 0.313 cycle/residue (3.2 residue/cycle)) show that these four transmembrane segments have the characteristic spectra of α-helical structures. The PSDF function for the transmembrane sequence segment 430–450 has a sharp peak at 0.385 cycle/residue (2.6 residue/cycle), suggesting strongly a β-sheet conformation.

The ARMA sequence spectral analysis suggests that the *Shaker* potassium channel contains predominantly α-helical structures. The transmembrane segments S1, S3, S4, S5, S6 may be transmembrane helices. The current S2 location seems to be a structural transition region (from helix to sheet). Spectral analysis suggests that a 10-residue shift of the S2 location will make it a helical or a sheet structure in the membrane, depending on the direction of the shift in the primary sequence. The suggested pore-forming region, sequence segment 430–450, has the spectral characteristics of a β-sheet structure. A β-sheet pore conformation has also been proposed recently by a model at the atomic level (Bogusz et al., 1992).

CONCLUSION

The ARMA spectral analysis of the primary sequences of several membrane proteins shows a clear correlation between the periodicity of the hydrophobicity distribution along the amino acid primary sequences of membrane proteins and the ordered folding patterns in their three-dimensional structures. The assessment of the secondary structure elements in the studied membrane proteins (Melittin, Porin, Colicin, Bacteriorhodopsin, Photosynthetic reaction center) agrees with their crystal structures. We also analyzed the sequence spectra for the transmembrane regions of acetylcholine receptor α-subunit and the *Shaker* potassium channel. Our results on the acetylcholine receptor α-subunit are different from the predictions derived from the direct Fourier analysis (Stroud and Finer-Moore, 1985; Stroud et al., 1990). However, they are consistent with recent mutation experiments (Akabas et al., 1992) and low resolution EM studies (Unwin, 1993). The ARMA spectral analysis for the transmembrane regions of the *Shaker* potassium channel protein confirms the model proposed recently by other methods (Miller, 1991).

The power spectral density functions computed by ARMA method, in general, have much better resolution than those computed by direct Fourier transforms of the original sequences (Marple, 1987; Jenkins and Watts, 1968; Kay, 1987). Better resolution has also been demonstrated for hydrophobicity-represented protein sequences analyses (Sun and Parthasarathy, 1993; Sun, 1993). This is because the ARMA model equation, along with the model parameters, describes an infinite length of data sequence that has the characteristics of the original finite data sequence. In doing this ARMA model fitting for the original sequence data, the noise level in the estimated spectrum is much reduced and the resolution of the computed spectrum is enhanced. This method can be used not only to analyze the hydrophobicity sequence but also to reveal periodic distributions of other physical properties (or combinations of physical properties)

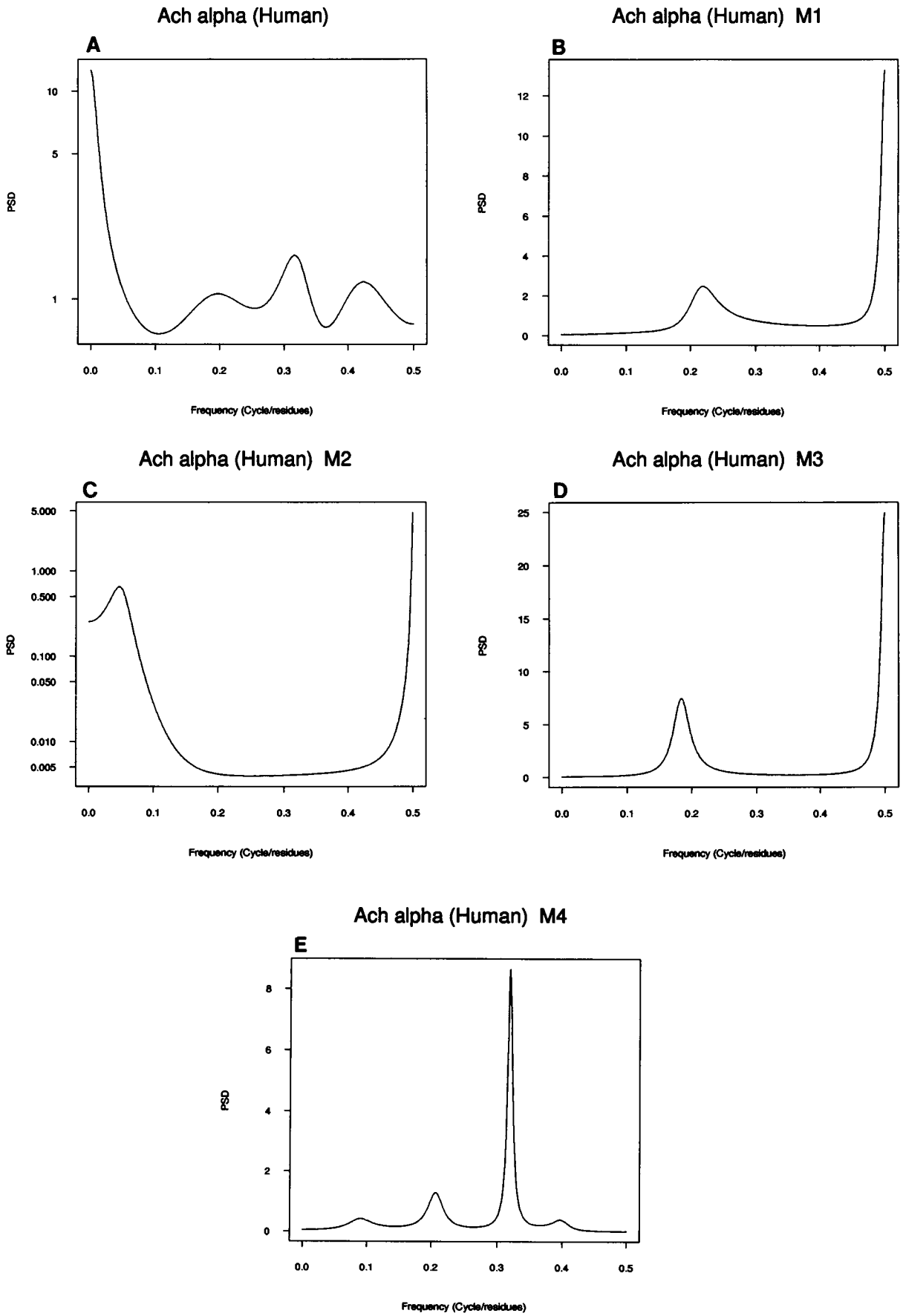

FIGURE 8 ARMA PSDF for entire AcH α subunit, M1, M2, M3, and M4 regions.

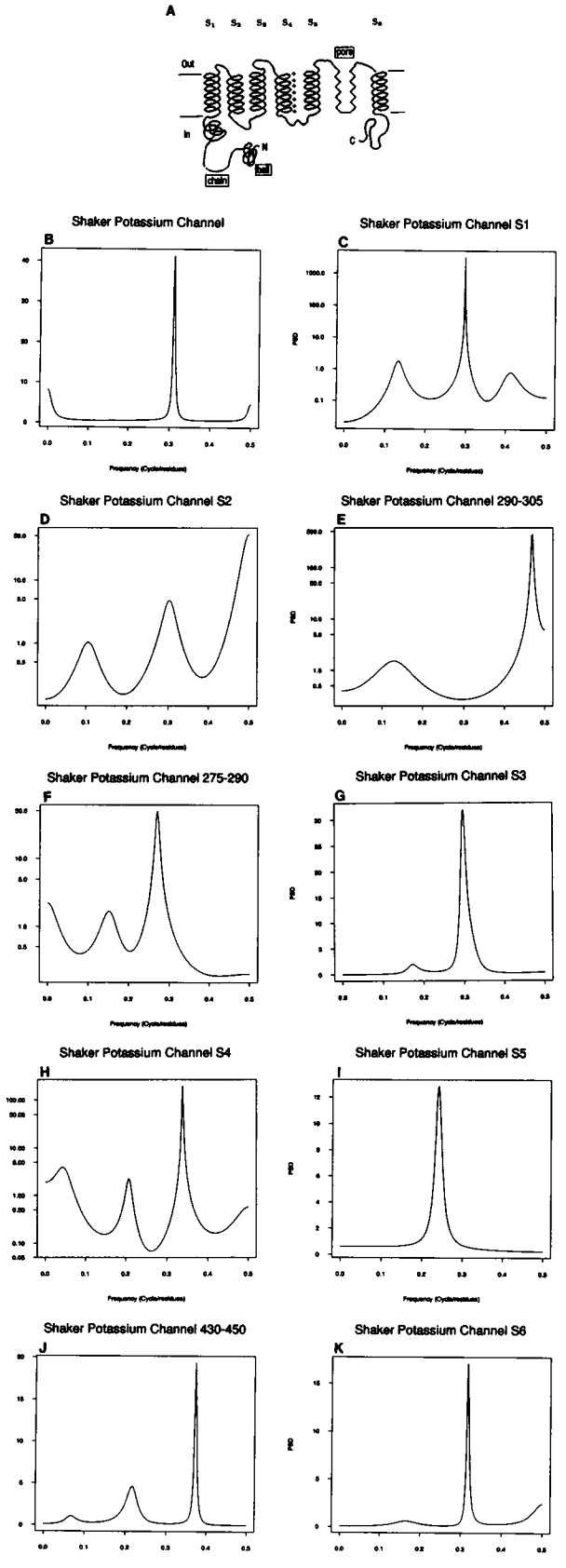

FIGURE 9 (*A*) Current model of potassium channel topology (after Miller 1991); (*B–K*) ARMA PSDF.

along the protein primary sequence, as long as a scale of physical properties is given. A scale that correctly combines several major physicochemical properties of amino acids will likely yield spectra that have even better correlation with the three-dimensional structures of proteins.

A single constant value hydrophobicity representation of an amino acid may not be good enough to characterize its interaction with the inhomogeneous protein environment around it. The variety of available hydrophobicity scales (Cornette et al., 1987) may be a reflection of this assessment. Different scales reflect the different behaviors of amino acid residues in inhomogeneous environments due to variations in their structural and physicochemical properties. Therefore, different hydrophobicity scales may result in quantitatively different (although qualitatively similar) computed spectra for a given primary sequence. Scales that better characterize the folded protein environments will generate ARMA PSDFs that correspond better to the actual three-dimensional structures (Sun, 1993). In general, if two scales have a high correlation coefficient, they will generate quantitatively similar spectra for the same protein sequence.

The current ARMA power spectral analysis method for computing the periodic distribution of physicochemical properties of amino acids along the primary sequence can be used to predict the secondary structure propensity of a given sequence. Different from the common methods of secondary structure prediction (Chou-Fasman, 1978; Garnier et al., 1978) that use local statistical sequence-structure information based on the secondary structures elements in known proteins, the ARMA power spectral analysis method does not require any statistical information that may be biased by the limited sampling of known protein structures. One of the obvious disadvantages of the spectral analysis method is that it is not sensitive enough to detect the starting and ending points of the secondary structure elements in a protein sequence, because it operates in the frequency domain. It seems to us that this method may be used to supplement the common methods of secondary structure prediction.

We thank Dr. Sarina Bromberg for her critical reading and editorial assistance with the manuscript. S. Sun is most grateful to Professor Ken Dill for discussions and suggestions.

S. Sun is partially supported by National Institutes of Health and Office of Naval Research grants to Professor Ken Dill.

REFERENCES

Akabas, M. H., D. A. Stauffer, M. Xu, and A. Karlin. 1992. Acetylcholine receptor channel structure probed in cysteine-substitution mutants. *Science.* 258:307–310.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* AC-19, 716–723.

Argos, P., J. K. Mohana-Rao, and P. A. Hargrave. 1982. Structural prediction of membrane bound proteins. *Eur. J. Biochem.* 128:565–75.

Biou, V., J. F. Gibrat, J. M. Levin, B. Robson, and J. Garnier. 1988. Secondary structure prediction: combination of three different methods. *Protein Eng.* 2:185–91.

Bogusz, S., A. Boxer, and D. D. Busath. 1992. An SS1-SS2 beta-barrel structure for the voltage-activated potassium channel. *Protein Eng.* 5:285–93.

Box, G. E. P., and G. M. Jenkins. 1970. Time Series Analysis, Forecasting and Control. Holden-Day Inc., San Francisco.

Chou, P. Y., and G. D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* 47:45–148.

Cornette, J. L., K. B. Cease, H. Margalit, J. L. Spouge, J. A. Berzofsky, C. DeLisi. 1987. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *J. Mol. Biol.* 195:659–85.

Deisenhofer, J., O. Epp, K. Miki, R. Huber, and H. Michel. 1984. X-ray structure analysis of a membrane protein complex. Electron density map at 3 A resolution and a model of the chromophores of the photosynthetic reaction center from Rhodopseudomonas viridis. *J. Mol. Biol.* 180: 385–98.

Deisenhofer, J., O. Epp, K. Miki, R. Huber, and H. Michel. 1985. Structure of the protein subunits in the photosynthetic reaction center of *Rhodopseudomonas viridis* at 3-Å resolution. *Nature.* 318:618–624.

Dill, K. A. 1990. Dominant forces in protein folding. *Biochemistry.* 29: 7133–55.

Durell, S. R., and H. R. Guy. 1992. Atomic scale structure and functional models of voltage-gated potassium channels. *Biophys. J.* 62:238–47.

Eisenberg, D., R. M. Weiss, and T. C. Terwilliger. 1984. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc. Natl. Acad. Sci. USA.* 81:140–144.

Engelman, D. M., T. A. Steitz, and A. Goldman. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* 15:321–53.

Fasman, D. G. 1989. Identification of membrane proteins and soluble protein secondary structural elements, domain structure, and packing arrangements by Fourrier-transform amphipathic analysis. *In* Principles of Protein Structure and Protein Folding. D. G. Fasman, editor. Plenum Press, New York.

Finer-Moore, J., F. Bazan, J. Rubin, and R. M. Stroud. 1989. Principles of Protein Stucture and Protein Folding. D. G. Fasman, editor. Plenum Press, New York.

Galzi, J. L., F. Revah, A. Bessis, and J. P. Changeux. 1991. Functional architecture of the nicotinic acetylcholine receptor: from electric organ to brain. *Annu. Rev. Pharmacol. Toxicicol.* 31:37–72.

Garnier, J., D. J. Osguthorpe, and B. J. Robson. 1978. Analysis of the accuracy and implications of simple method for predicting the secondary structure of globular protein. *Mol. Biol.* 120:97–120.

Henderson, R., J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann, and K. H. Downing. 1990. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* 213:899–929.

Henderson, R., and P. N. T. Unwin. 1975. Three-dimensional model of purple membrane obtained by electron microscopy. *Nature.* 257:28–32.

Jähnig, F. 1989. Structure prediction for membrane proteins. *In* Principles of Protein Structure and Protein Folding. D. G. Fasman, editor. Plenum Press, New York.

Jan, Y. N., and L. Y. Jan. 1989. Voltage-sensitive ion channels. *Cell.* 56: 13–25.

Jenkins, G. M., and D. G. Watts. 1968. Spectral analysis and its application. Holden-Day, Inc., San Francisco, CA.

Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. in Prot. Chem.* 14:1–63.

Kay, S. M. 1987. Modern Spectral Estimation. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105–32.

Marple, S. L., Jr. 1987. Digital Spectral Analysis with Applications. Prentice-Hall, Inc., Englewood Cliffs, NJ.

McLachlan, A. D., and M. Stewart. 1976. The 14-fold periodicity in alpha-tropomyosin and the interaction with actin. *J. Mol. Biol.* 103:271–98.

Mehra, R. K. 1972. On-line identification of linear dynamic systems with applications to Kalman filtering. *IEEE Trans. Autom. Control.* Vol. AC-16. 12–21.

Miller, C. 1991. Annus mirabilis of potassium channels. *Science.* 252:1092–6.

Oppenheim, A. V., and A. S. Willsky. 1983. Signal and Systems. Prentice-Hall Inc., Englewood Cliffs, NJ.

Parker, M. W., F. Pattus, A. D. Tucker, and D. Tsernoglou. 1989. Structure of the membrane-pore-forming fragment of colicin A. *Nature.* 337:93–6.

Porat, B., and B. Fridlander. 1985. Asymptotic analysis of the bias of the modified Yule-Walker estimator. *IEEE Trans. Autom. Control.* Vol AC-30. 765–767.

Prince, R. C., J. C. Leight, and L. Dutton. 1976. Thermodynamic properties of the reaction center of Rhodopseudomonas viridis. In vivo measurement of the reaction center bacteriochlorophyl-primary acceptor intermediary electron carrier. *Biochim. Biophys. Acta.* 440:622–36.

Rose, G. D., A. R. Geselowitz, G. J. Lesser, R. H. Lee, and M. H. Zehfus. 1985. Hydrophobicity of amino acid residues in globular proteins. *Science.* 229:834–8.

Stroud, R. M., and J. Finer-Moore. 1985. Acetylcholine receptor structure, function, and evolution. *J. Annu. Rev. Cell Biol.* 1:317–51.

Stroud, R. M., M. P. McCarthy, and M. Shuster. 1990. Nicotinic acetylcholine receptor superfamily of ligand-gated ion channels. *Biochemistry.* 29:11009–11023.

Sun, S. 1993. Protein structure prediction: power spectral analysis and reduced representation models. Ph.D. thesis. Department of Biophysical Science, State University of New York at Buffalo, Buffalo, NY.

Sun, S., and R. Parthasarathy. 1990. Preceedings of Converging Approaches in Computational Biology. Albany Conference. September.

Sun, S., and R. Parthasarathy. 1993. Periodic Hydropathy profile of primary sequence implies periodic structural patterns in protein: power spectral analysis. In press.

Tempel, B. L., D. M. Papazian, T. L. Schwarz, Y. N. Jan, and L. Y. Jan. 1987. Sequence of a probable potassium channel component encoded at *Shaker* locus of *Drosophila. Science.* 237:770–5.

Terwilliger, T. C., and D. Eisenberg. 1982. The structure of melittin. I. Structure determination and partial refinement. *J. Biol. Chem.* 257: 6010–22.

Unwin, N. 1993. Nicotinic acetylcholine receptor at 9 A resolution. *J. Mol. Biol.* 229:1101–24.

von Heijne, G. 1991. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* 225:487–94.

Weiss, M. S., A. Kreusch, E. Schiltz, U. Nestel, W. Welte, J. Weckesser, and G. E. Schulz. 1991. The structure of porin from *Rhodobacter capsulatus* at 1.8 A resolution. *FEBS Lett.* 280:379–82.