PROTEINS: Structure, Function, and Genetics 28:99–108 (1997)

Disposition of Amphiphilic Helices in Heteropolar Environments

Kou-Chen Chou,[1]* Chun-Ting Zhang,[2] and Gerald M. Maggiora[1]
[1]*Computer-Aided Drug Discovery, Pharmacia and Upjohn, Kalamazoo, Michigan 49007-4940*
[2]*Department of Physics, Tianjin University, Tianjin, China*

ABSTRACT **It is known that α helices in globular proteins usually consist of two types of residues, hydrophobic and hydrophilic, with the number of each type being roughly equal. Except for many transmembrane helices, α-helices are generally amphiphilic to some degree. This is not entirely surprising because α-helices typically reside in heteropolar environments that arise from the polar aqueous solution that surrounds a protein and the apolar "hydrophobic core" located at its center. The packing of α-helices in such heteropolar environments is driven by the minimization of free energy brought about by placing hydrophobic sidechains into apolar environments and hydrophilic sidechains into polar environments. The interface between the two environments can be characterized by an interfacial plane, called the demarcation plane, that optimally separates the two classes of residues. The inclination angle Ω between the axis of the helix and the demarcation plane provides a measure of the degree of amphiphilicity of an α-helix. For highly amphiphilic helices, Ω ≈ 0. The inclination angle provides a new measure of amphiphilicity that complements the hydrophobic moments of Eisenberg et al. Based on the simple physical model described above, an algorithm is developed for predicting the helix inclination angle. The calculated results show that the inclination angle for most α-helices extracted from globular proteins is less than 25° in magnitude. This suggests that helices found in globular proteins tend to be reasonably amphiphilic with half their face dominated by hydrophobic residues and the other half by hydrophilic residues. A new two-dimensional representation that characterizes the disposition of hydrophobic and hydrophilic residues in α-helices, called a "wenxiang diagram," is presented. The wenxiang diagram can also be used as an important element to represent a protein molecule. Proteins 28:99–108, 1997**

Key words: hydrophobic centroid; hydrophilic centroid; wenxiang diagram; polar-apolar interface

INTRODUCTION

It is widely recognized that the interaction of the amino acid side chains with water is a major factor in determining the native structure of proteins.[1–6] The side chains of hydrophilic residues seek contact with water, whereas the side chains of hydrophobic residues avoid contact with water. Except for transmembrane α-helices, whose residues are basically hydrophobic, most α-helices in proteins consist of both hydrophilic and hydrophobic residues.[7–9] Thus, the following two questions naturally arise. How are the two types of residues distributed in the α-helices? For an α-helix with a given distribution of constituent residues, how does the helix align itself with respect to the polar and apolar domains of the protein-water environment? This study describes a quantitative procedure that addresses these questions.

Eisenberg et al.[1] introduced the concept of hydrophobic moment and hydrophobic moment plots. As noted by these authors, hydrophobic moments are exact analogues of electric dipole moments except that they measure the asymmetry of the hydrophobicity or amphiphilicity. It was shown that the hydrophobic properties of an α-helix can be classified by means of a hydrophobic moment plot. Thus, hydrophobic moments and hydrophobic moment plots provide at least a partial answer to the above two questions. However, the approach by Eisenberg et al.[1] has the following problems. First, according to the definition described by them, a helical hydrophobic moment is always perpendicular to the axis of a helix. It is true only when the interfacial plane of a helix, as described later, is parallel to its axis. However, the actual cases for most amphiphilic helices are not like that. Second, in hydrophobic moment plots the property of an α-helix depends on the position of its associated point in that plot. However, the boundary between the different clustering regions in such plots is rather ambiguous because of the intrinsically qualitative nature of the method itself. The method described below may serve as a complement to the use of hydrophobic moments and hydrophobic moment plots.

The computer code for the prediction algorithm is available on request.

*Correspondence to: Dr. Kou-Chen Chou, Computer-Aided Drug Discovery, Pharmacia and Upjohn, Kalamazoo, MI 49007-4940.

Received 9 August 1996; Accepted 24 October 1996

FORMULATION

In the coordinate system (x, y, z) used in this work, the z-axis coincides with the axis of the α-helix, as defined in Chou et al.[10] The x-axis and the y-axis as well as the origin of the coordinate system are chosen so that the C^{α} atom of the first residue is situated at $x = r$, $y = 0$, and $z = 0$, where r is the radius of the α-helix shown in Figure 1. Suppose the α-helix has n residues of which n^+ residues are hydrophobic and n^- residues are hydrophilic. Obviously, $n^+ + n^- = n$. The hydrophobicity of an amino acid is expressed by H, where hydrophobic amino acids have $H > 0$ and hydrophilic amino acids have $H < 0$. According to Tanford[11] and Jones,[12] the hydrophobicity for each of the 20 amino acids in proteins may be given by

$$\begin{Bmatrix} \text{A} & \text{C} & \text{D} & \text{E} & \text{F} \\ 0.62 & 0.29 & -0.90 & -0.74 & 1.19 \\ \text{G} & \text{H} & \text{I} & \text{K} & \text{L} \\ 0.48 & -0.40 & 1.38 & -1.50 & 1.06 \\ \text{M} & \text{N} & \text{P} & \text{Q} & \text{R} \\ 0.64 & -0.78 & 0.12 & -0.85 & -2.53 \\ \text{S} & \text{T} & \text{V} & \text{W} & \text{Y} \\ -0.18 & -0.05 & 1.08 & 0.81 & 0.26 \end{Bmatrix}. \quad (1)$$

The coordinates and hydrophobicity for each hydrophobic residue of the α-helix are given by $x^+(i)$, $y^+(i)$, $z^+(i)$, and $H^+(i)$, respectively, where $i = 1, 2, 3, \ldots, n^+$. Similarly, the coordinates and hydrophobicity for each hydrophilic residue in the α-helix are given by $x^-(i)$, $y^-(i)$, $z^-(i)$, and $H^-(i)$, respectively, where $i = 1, 2, 3, \ldots, n^-$.

The coordinates of the C^{α} atoms in an α-helix are not distributed randomly but are distributed according to the following equation:[13]

$$\begin{cases} x(j) = r\cos(j\theta), \\ y(j) = r\sin(j\theta), \quad (j = 0, 1, 2, \ldots, N-1) \\ z(j) = jh \end{cases} \quad (2)$$

where

$$\begin{cases} \theta = 100^\circ \\ r = 2.3\ \text{Å} \\ h = 1.5\ \text{Å}, \end{cases} \quad (3)$$

and $x(j)$, $y(j)$, and $z(j)$ are the coordinates for the j-th C^{α} atom in the helix. It is useful to introduce the following new concepts.

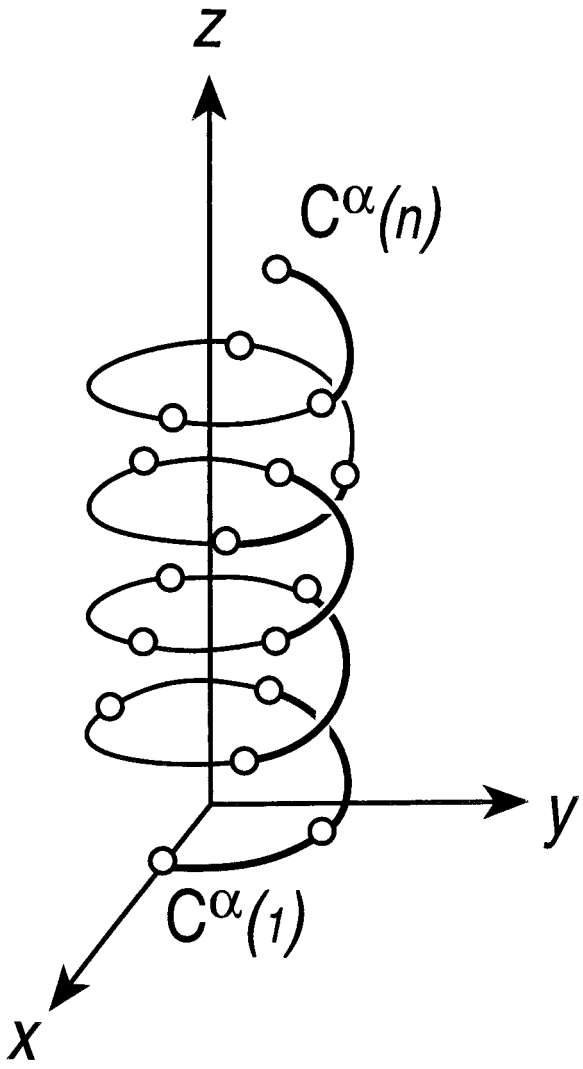

Fig. 1. Schematic representation of an α-helix and the (x, y, z) coordinate system used in this study. The z-axis is the helix axis, defined as the least-squares line through all C^{α} atoms of the helix.[10] The x and y axes are defined as C^{α} (1), the C^{α} atom of the first residue, is situated at $x = r$ and $y = 0$, where r is the radius of the α-helix. The z-axis points in the usual direction from the NH_2- to the COOH-terminal of the helix.

Definition 1: The Hydrophobic Centroid of an α-Helix

Its coordinates are defined by

$$\begin{cases} x_0^+ = \dfrac{1}{H_{tot}^+}\displaystyle\sum_{i=1}^{n^+} H^+(i)x^+(i) \\ y_0^+ = \dfrac{1}{H_{tot}^+}\displaystyle\sum_{i=1}^{n^+} H^+(i)y^+(i) \\ z_0^+ = \dfrac{1}{H_{tot}^+}\displaystyle\sum_{i=1}^{n^+} H^+(i)z^+(i), \end{cases} \quad (4)$$

where

$$H_{tot}^{+} = \sum_{i=1}^{n^{+}} H^{+}(i). \tag{5}$$

Definition 2: The Hydrophilic Centroid of an α-Helix

Its coordinates are defined by

$$\begin{cases} x_0^- = \dfrac{1}{H_{tot}^-} \displaystyle\sum_{i=1}^{n^-} |H^-(i)| x^-(i) \\ y_0^- = \dfrac{1}{H_{tot}^-} \displaystyle\sum_{i=1}^{n^-} |H^-(i)| y^-(i) \\ z_0^- = \dfrac{1}{H_{tot}^-} \displaystyle\sum_{i=1}^{n^-} |H^-(i)| z^-(i), \end{cases} \tag{6}$$

where

$$H_{tot}^{-} = \sum_{i=1}^{n^-} |H^-(i)|. \tag{7}$$

Definition 3: The Overall Hydropathic Centroid of an α-Helix

Its coordinates are defined by

$$\begin{cases} x_0 = \dfrac{H_{tot}^+ x_0^+ + H_{tot}^- x_0^-}{H_{tot}^+ + H_{tot}^-} \\ y_0 = \dfrac{H_{tot}^+ y_0^+ + H_{tot}^- y_0^-}{H_{tot}^+ + H_{tot}^-} \\ z_0 = \dfrac{H_{tot}^+ z_0^+ + H_{tot}^- z_0^-}{H_{tot}^+ + H_{tot}^-}. \end{cases} \tag{8}$$

Suppose that a given α-helix is located in a heteropolar environment consisting of a polar phase (such as water) and an apolar phase (such as the interior of a globular protein or biological membrane). A key problem is how the helix aligns itself in the interfacial region between the two phases so as to attain its lowest free energy. To reach such an optimal free energy state, the helix should be divided longitudinally into two halves: one essentially hydrophobic and one essentially hydrophilic, so that as much of its hydrophobic part as possible resides within the apolar lipid-like environment and as much of its hydrophilic part as possible resides within the polar water environment. Below, we propose an algorithm for determining such a plane, called the interfacial plane S, that optimally separates a given helix into its hydrophobic and hydrophilic parts. The algorithm is based upon the linear discriminant analysis developed by Fisher[14] many years ago.

In the coordinate system (x, y, z) described above a plane S is given by the following equation

$$C_0 + C_1 x + C_2 y + C_3 z = 0 \tag{9}$$

where C_i $(i = 0, 1, 2, 3)$ are the parameters that uniquely define the plane. The desired interfacial plane can be derived according to the following three criteria:

- Criterion 1. Project the hydrophobic and hydrophilic centroids of an α-helix as given by Definitions 1 and 2, respectively, onto the normal **n** to the plane S. The distance between these two projected points should be kept as large as possible.
- Criterion 2. Along the normal **n** the sum of the squares of the distances between the hydrophobic centroid and the C^{α} atoms of the hydrophobic residues should be as small as possible. The same is true for the distances between the hydrophilic centroid and the C^{α} atoms of the hydrophilic residues.
- Criterion 3. The desired plane should pass through the overall centroid (x_0, y_0, z_0) as given by Definition 3.

If plane S satisfies all of the above three criteria for a given α-helix, it will become the desired interfacial plane that uniquely defines the inclination angle and intercept of the helix to a polar-apolar interface. To calculate the corresponding parameters C_i $(i = 0, 1, 2, 3)$ that determine the interfacial plane, all of the above three criteria must be utilized. Note that the conditions specified in Criterion 1 and Criterion 2 are analogous to those used by Fisher[14] in the development of his discriminant function approach to classification. Also note that Criterion 1 and Criterion 2 emphasize two opposite aspects of an entity, i.e., that on one hand the hydrophobic and hydrophilic "clusters" of residues should be separated as far as possible, but on the other hand each of the two clusters should be as compact as possible. Hence, the optimal arrangement of a helix with respect to the polar-apolar interface is actually the result of a balance driven by two "opposing forces."

According to analytic geometry, the square of the projected distance, D^2, between the hydrophobic and the hydrophilic centroids along the normal **n** to the plane S is

$$D^2 = \frac{[C_1(x_0^+ - x_0^-) + C_2(y_0^+ - y_0^-) + C_3(z_0^+ - z_0^-)]^2}{C_1^2 + C_2^2 + C_3^2} = \frac{\tilde{D}^2}{C_1^2 + C_2^2 + C_3^2}, \tag{10a}$$

where

$$\tilde{D}^2 = [C_1(x_0^+ - x_0^+) + C_2(y_0^+ - y_0^-) + C_3(z_0^+ - z_0^-)]^2. \tag{10b}$$

On the other hand, the sum of the hydrophobicity-weighted squares of the projected distances between the hydrophobic centroid and the C^{α} atoms of the hydrophobic residues along the normal **n** may be expressed as

$$F_+^2 = \frac{\tilde{F}_+^2}{C_1^2 + C_2^2 + C_3^2}, \tag{11a}$$

where

$$\tilde{F}_+^2 = \sum_{i=1}^{n^+} w^+(i)\{C_1[x^+(i) - x_0^+] + C_2[y^+(i) - y_0^+] + C_3[z^+(i) - z_0^+]\}^2, \quad (11b)$$

and

$$w^+(i) = \frac{H^+(i)}{H_{tot}^+} \quad (11c)$$

reflects the hydrophobicity-weighted contribution from each residue. Similarly, the sum of the hydrophilicity-weighted squares of the projected distances between the hydrophilic centroid and the C^α atoms of the hydrophilic residues along the normal **n** may be expressed as

$$F_-^2 = \frac{\tilde{F}_-^2}{C_1^2 + C_2^2 + C_3^2}, \quad (12a)$$

where

$$\tilde{F}_-^2 = \sum_{i=1}^{n^-} w^-(i)\{[C_1(x^-(i) - x_0^-] + C_2[y^-(i) - y_0^-] + C_3[z^-(i) - z_0^-]\}^2, \quad (12b)$$

and

$$w^-(i) = \frac{|H^-(i)|}{|H_{tot}^-|} \quad (12c)$$

reflects the hydrophilicity-weighted contribution from each residue.

Criterion 1 requires that the value of D^2 be as large as possible, whereas Criterion 2 requires that the value of $F_+^2 + F_-^2$ be as small as possible. To find the desired plane S, define

$$\Gamma(C_1, C_2, C_3) = \frac{D^2}{F_+^2 + F_-^2} = \frac{\tilde{D}^2}{\tilde{F}^2}, \quad (13a)$$

where

$$\tilde{F}^2 = \tilde{F}_+^2 + \tilde{F}_-^2. \quad (13b)$$

Thus, the combined requirement of both Criterion 1 and Criterion 2 is equivalent to the condition that the value Γ be a maximum. Accordingly, the values of C_1, C_2, and C_3 can be derived as follows:

$$\frac{\partial \Gamma}{\partial C_i} = 0, \quad (i = 1, 2, 3) \quad (14a)$$

or

$$\frac{\partial \Gamma}{\partial C_i} = \frac{\tilde{F}^2 \partial \tilde{D}^2 / \partial C_i - \tilde{D}^2 \partial \tilde{F}^2 / \partial C_i}{\tilde{F}^4} = 0 \quad (14b)$$

Substituting Eq. 13a into Eq. 14b, we have

$$\frac{1}{\Gamma}\frac{\partial \tilde{D}^2}{\partial C_i} = \frac{\partial \tilde{F}^2}{\partial C_i}, \quad (i = 1, 2, 3) \quad (15)$$

The above equation leads to the following set of equations for C_1, C_2, and C_3:

$$\begin{cases} S_{xx}C_1 + S_{xy}C_2 + S_{xz}C_3 = \left[\dfrac{C_1 d_1 C_2 d_2 + C_3 d3}{\Gamma(C_1, C_2, C_3)}\right] d_1 \\ S_{yx}C_1 + S_{yy}C_2 + S_{yz}C_3 = \left[\dfrac{C_1 d_1 + C_2 d_2 + C_3 d_3}{\Gamma(C_1, C_2, C_3)}\right] d_2 \\ S_{zx}C_1^* + S_{zy}C_2^* + S_{zz}C_3^* = \left[\dfrac{C_1 d_1 + C_2 d_2 + C_3 d_3}{\Gamma(C_1, C_2, C_3)}\right] d_3 \end{cases} \quad (16a)$$

where

$$\begin{cases} d_1 = x_0^+ - x_0^- \\ d_2 = y_0^+ - y_0^- \\ d_3 = z_0^+ - z_0^- \end{cases} \quad (16b)$$

and

$$\begin{cases} S_{xx} = \sum_{i=1}^{n^+} w^+(i)[x^+(i) - x_0^+]^2 - \sum_{i=1}^{n^-} w^-(i)[x^-(i) - x_0^-]^2 \\ S_{yy} = \sum_{i=1}^{n^+} w^+(i)[y^+(i) - y_0^+]^2 - \sum_{i=1}^{n^-} w^-(i)[y^-(i) - y_0^-]^2 \\ S_{zz} = \sum_{i=1}^{n^+} w^+(i)[z^+(i) - z_0^+]^2 - \sum_{i=1}^{n^-} H^-(i))[z^-(i) - z_0^-]^2 \\ S_{xy} = S_{yx} = \sum_{i=1}^{n^+} w^+(i)[x^+(i) - x_0^+][y^+(i) - y_0^+] - \sum_{i=1}^{n^-} w^-(i)[x^-(i) - x_0^-][y^-(i) - y_0^-] \\ S_{xz} = S_{zx} = \sum_{i=1}^{n^+} w^+(i)[x^+(i) - x_0^+][z^+(i) - z_0^+] - \sum_{i=1}^{n^-} w^-(i)[x^-(i) - x_0^-][z^-(i) - z_0^-] \\ S_{yz} = S_{zy} = \sum_{i=1}^{n^+} w^+(i)[y^+(i) - y_0^+][z^+(i) - z_0^+] - \sum_{i=1}^{n^-} H^-(i)[y^-(i) - y_0^-][z^-(i) - z_0^-]. \end{cases} \quad (16c)$$

Solving Eq. 16a directly for the desired plane S is difficult because it contains a set of nonlinear equations. To overcome this difficulty, convert Eq. 16a to the following form by dividing both sides of the equation by $(C_1d_1 + C_2d_2 + C_3d_3)/[\Gamma(C_1, C_2, C_3)]$:

$$\begin{cases} S_{xx}C_1^* + S_{xy}C_2^* + S_{xz}C_3^* = d_1 \\ S_{yx}C_1^* + S_{yy}C_2^* + S_{yz}C_3^* = d_2 \\ S_{zx}C_1^* + S_{zy}C_2^* + S_{zz}C_3^* = d_3, \end{cases} \quad (17a)$$

where

$$C_i^* = \left[\frac{\Gamma(C_1, C_2, C_3)}{C_1d_1 + C_2d_2 + C_3d_3}\right]C_i, \quad (i = 1, 2, 3) \quad (17b)$$

The solution for Eq. 17a is

$$C_1^* = \frac{\begin{vmatrix} d_1 & S_{xy} & S_{xz} \\ d_2 & S_{yy} & S_{yz} \\ d_3 & S_{zy} & S_{zz} \end{vmatrix}}{\begin{vmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{vmatrix}}, \quad C_2^* = \frac{\begin{vmatrix} S_{xx} & d_1 & S_{xz} \\ S_{yx} & d_2 & S_{yz} \\ S_{zx} & d_3 & S_{zz} \end{vmatrix}}{\begin{vmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{vmatrix}},$$

$$C_3^* = \frac{\begin{vmatrix} S_{xx} & S_{xy} & d_1 \\ S_{yx} & S_{yy} & d_2 \\ S_{zx} & S_{zy} & d_3 \end{vmatrix}}{\begin{vmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zx} \end{vmatrix}}. \quad (18)$$

Thus, according to analytic geometry the direction cosines of the desired plane S can be derived through Eqs. 17b and 18 as shown below. Here the positive direction of the normal is taken to be the direction that points to the apolar side from the polar side of the interfacial plane. The angles between the positive direction of the normal vector and the x, y, and z axes, respectively, denoted by α, β, and γ, are given by

$$\begin{cases} \alpha = \arccos\left(\dfrac{C_1}{\sqrt{C_1^2 + C_2^2 + C_3^2}}\right) \\ \quad = \arccos\left(\dfrac{C_1^*}{\sqrt{C_1^{*2} + C_2^{*2} + C_3^{*2}}}\right) \\ \beta = \arccos\left(\dfrac{C_2}{\sqrt{C_1^2 + C_2^2 + C_3^2}}\right) \\ \quad = \arccos\left(\dfrac{C_2^*}{\sqrt{C_1^{*2} + C_2^{*2} + C_3^{*2})}}\right) \\ \gamma = \arccos\left(\dfrac{C_3}{\sqrt{C_1^2 + C_2^2 + C_3^2}}\right) \\ \quad = \arccos\left(\dfrac{C_3^*}{\sqrt{C_1^{*2} + C_2^{*2} + C_3^{*2}}}\right). \end{cases} \quad (19)$$

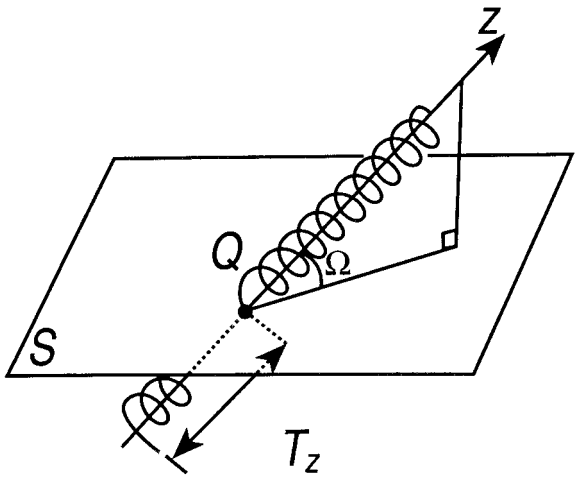

Fig. 2. Schematic drawing showing the inclination angle Ω and the intercept T_z of the interfacial plane S. Q is the intersection point of the helix axis z with the interfacial plane S. As shown in the figure, Ω is the angle between the helix axis Z and its projection on the plane S, whereas T_z is the distance from the projection of the first C^α on the helix axis to Q.

Similar to the vertical projected angle used to study the packing between an α-helix and a β-sheet (see Fig. 3 of Chou et al.[15]), here we define the inclination angle, Ω, as follows. Specifically, Ω is defined as the angle formed by the axis of the helix and its projected line on the plane S (Fig. 2). In contrast, the angle γ given in Eq. 19 is the angle between the helix and **n**, the normal to the plane S. The inclination angle thus defined for an α-helix should be in the range of $-90° \le \Omega \le 90°$, with $\Omega > 0$ when its COOH-terminal is located in the apolar side of the interface and $\Omega < 0$ when its NH_2-terminal is located in the apolar side of the interface (Fig. 3). Thus, we have $\Omega = 90° - \gamma$, or

$$\Omega = \arcsin\left(\frac{C_3^*}{\sqrt{C_1^{*2} + C_2^{*2} + C_3^{*2}}}\right) \quad (20)$$

According to analytical geometry, the arrangement between a plane and a line will be uniquely defined if the inclination angle of the line to the plane and the intercept of the plane on the line are both given. The former can be derived from Eq. 20, whereas the latter can be obtained as follows. According to Criterion 3 and Eq. 9, the parameter C_0 can be expressed by

$$C_0 = -C_1x_0 - C_2y_0 - C_3z_0 \quad (21)$$

where x_0, y_0, and z_0 are given in Eq. 8. Thus, the intercept, T_z, of the interfacial plane S on the z-axis, i.e., the axis of the α-helix, is given by (Fig. 2)

$$T_z = -\frac{C_0}{C_3} = -\frac{C_1^*}{C_3^*}X_0 - \frac{C_2^*}{C_3^*}Y_0 - Z_0. \quad (22)$$

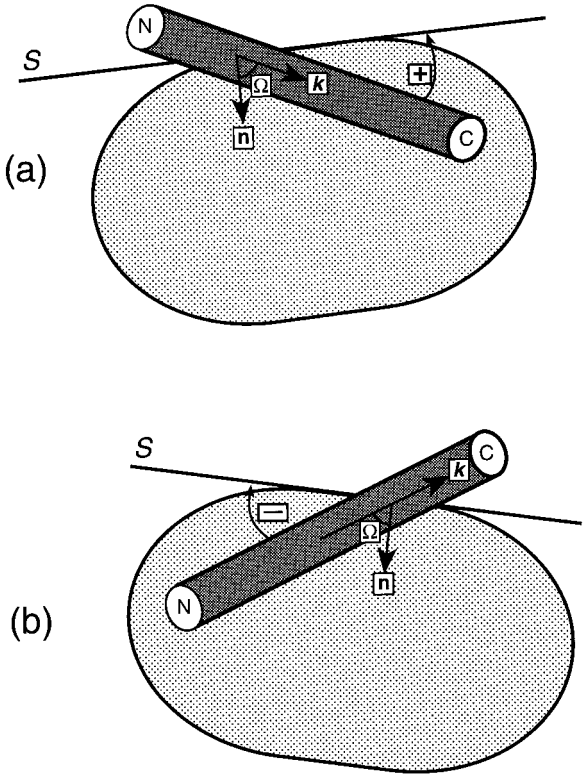

Fig. 3. An illustration showing the sign of the inclination angle Ω. The Ω value of an α-helix is defined to be positive when its COOH-terminal is located in the apolar or lipid side (i.e., protein side) of the interface and negative when its COOH-terminal is located away from the apolar phase. The tangent interfacial plane, *S*, lies along the protein surface and intersects the helix axis; the normal **n** to *S* points from the polar water side toward the apolar side. The helix axis points from the NH_2- to COOH-terminal of the helix, as indicated by the unit vector **k**.

Obviously, when the axis of a helix is parallel to its interfacial plane *S*, one has $T_z = \pm\infty$, the special case based on which the helical hydrophobic moment was defined by Eisenberg et al. was based.[1]

RESULTS AND DISCUSSION

As a concrete example, let us consider a protein of known structure. The three-dimensional structure of an apolipoprotein isolated from the African migratory locust *Locusta migratoria* has been determined by X-ray analysis to a resolution of 2.5 Å.[16] The protein consists of five long α-helices connected by short loops, as shown by a ribbon drawing in Figure 4. It is known that these helices are distinctly amphiphilic, with the hydrophobic residues facing the inside of the protein and the hydrophilic side chains facing outward. Therefore, it can be anticipated that these helices are almost parallel to the polar-apolar interface; i.e., their inclination angles as defined in this study should be small (close to 0°). The amino acid sequences for each of the five α-helices are as follows:

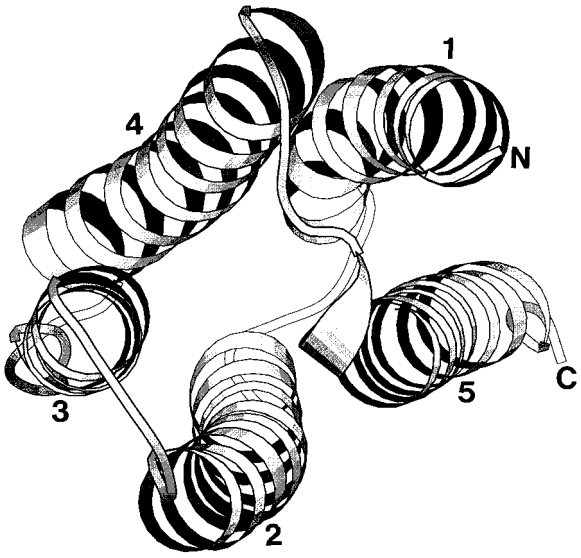

Fig. 4. The ribbon drawing of an apolipoprotein based on the X-ray data by Breiter et al.[16]

$$\begin{cases} \alpha_1\text{: NIAEAVQQLNHTIVNAAHELHETLGL;} \\ \alpha_2\text{: PDEALNLLTEQANAFKTKIAEVTTSLK} \\ \quad \text{QEAEK;} \\ \alpha_3\text{: SVAEQLNRFARNLNNSI;} \\ \alpha_4\text{: PADQLNSLQSALTNVGHQW} \\ \quad \text{QTSQPRPS;} \\ \alpha_5\text{: PVGSALQEAAEKTKEAAA} \\ \quad \text{NLQNSIQSAV.} \end{cases} \tag{23}$$

Listed in Table I are the inclination angles Ω and intercepts T_z calculated by the current method for the above helices. Indeed, as shown in Table I the inclination angles for all five helices are small ($|\Omega| < 3°$), indicating that each helix possesses significant amphiphilic character. Also, we can determine from the data for T_z in the table that none of these helices has its axis really parallel to its interfacial plane *S*, as originally assumed by Eisenberg et al.[1] in defining the helical hydrophobic moment.

It is instructive to introduce a two-dimensional diagram generated by conical projection of an α-helix. An illustration of the conical projection is given in Figure 5, where the projected image of a hollow cylinder Φ onto a plane perpendicular to it by conical projection will become a ring Ψ, with the outer and inner circles corresponding to the top and bottom circles of the hollow cylinder, respectively. Accordingly, the projected image obtained for a helix will become a planar spiral with a continuously varying radius. The larger the radius a point has in the spiral, the further the corresponding point in the

TABLE I. The Inclination Angles Ω of the Five Amphiphilic α-Helices From Apolipoprotein

Helix mark*	Residue numbers*	Inclination angle Ω (deg)	Intercept T_z (Å)	Helix length (Å)
α_1	7–32	2.2	23.6	37.5
α_2	35–66	−1.7	18.6	46.5
α_3	70–86	−1.7	10.9	24.0
α_4	95–121	−2.6	13.9	39.0
α_5	129–156	0.7	36.3	40.5

*These data are from Breiter et al.[16]

helix is from the projection plane and vice versa. If a helix is set with its axis pointing from the apex to the projection plane, in the corresponding two-dimensional diagram the COOH-terminal of the helix lies near the center while its NH_2-terminal lies at the outer rim of the diagram. For example, the five α-helices of the apolipoprotein, α_1, α_2, α_3, α_4, and α_5 (Eq. 23), can thus be expressed by five two-dimensional diagrams as shown in Figure 6a–e, respectively. This type of diagram will be called a "wenxiang diagram" because it looks like the Chinese wenxiang, a coil-like incense widely used in China to repel mosquitos. In these figures each residue is represented by a circle with a letter to indicate its code: a hydrophobic residue is denoted by a filled circle with a white code symbol, whereas a hydrophilic residue is denoted by an open circle with a black code symbol. On observing the wenxiang diagrams of the five helices of apolipoprotein a common feature emerges, i.e., most filled circles (hydrophobic residues) are distributed in one-half of each diagram and most open circles (hydrophilic residues) are distributed in the other half. Thus, the wenxiang diagrams provide an easily visualizable representation of the five α-helices in apolipoprotein which clearly shows that each helix shares the features of a typical amphiphilic helix, i.e., half of the helix flank is dominated by hydrophobic residues and the other half by hydrophilic residues. Therefore, their inclination angles are all very close to 0°, as noted in Table I and as observed in Figure 4, fully consistent with the requirement of minimum free energy in a polar-apolar system.

The wenxiang diagram can also be used as an important component to represent a protein molecule. For example, the 5-helix bundle of apolipoprotein can be expressed by a combination of five two-dimensional wenxiang diagrams, as shown in Figure 7. As shown, most filled circles (hydrophobic residues) are buried inside the bundle and most open circles (hydrophilic residues) are exposed to the aqueous environment that surrounds the protein. Although the wenxiang diagram is a two-dimensional representation of an α-helix, it distinguishes the helix ends from one another. In other words, it provides the information of the directionality of a

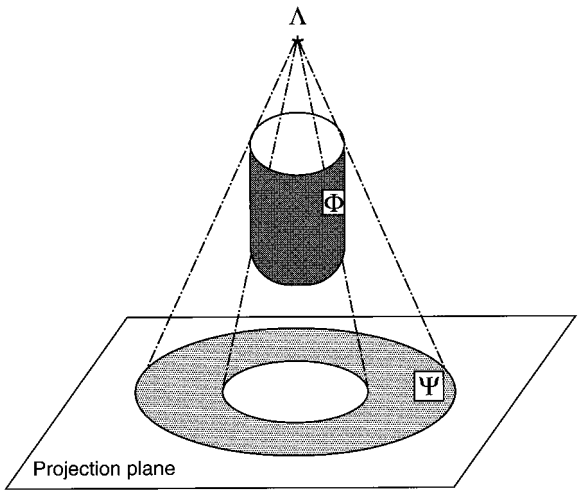

Fig. 5. Conical projection of a hollow cylinder Φ onto a plane perpendicular to the cylinder. The dot-dash lines (—·—·—) represent the radiating lines from the apex Λ. The distance Λ from the projection plane is arbitrary and is chosen to ensure that the resulting projection graph clearly portrays the relevant information. The image of the hollow cylinder on the projection plane is the ring Ψ.

helix as well. Therefore, when a protein is represented by combining several wenxiang diagrams, the direction of each of these diagrams must conform to the direction of its corresponding helix in the protein (Fig. 7).

Similar calculations have been conducted for the helices of several other proteins, such as myoglobin, cytochrome *c′*, hydrolase, and triose phosphate isomerase. Myoglobin and cytochrome *c′* belong to the all α, hydrolase to the α + β, and triose phosphate isomerase to the α/β structural classes.[17] It was found that of the 26 α-helices examined, 25 have an inclination angle with $|\Omega| < 25°$ (Table II). The only exception is the F_2 helix in triose phosphate isomerase. As shown in Table II, its inclination angle is $\Omega = 46.5°$, significantly deviating from 0°. This can be easily explained by looking at its wenxiang diagram in Figure 8, where it is shown that there is no clear-cut demarcation line that separates most of the filled circles from most of the open circles, implying that the hydrophobic and hydrophilic residues are mingled with each other over the entire face of the helix. In such cases, the minimum free energy requirement will not lead to arrangements in which half of the face of a helix is buried inside a protein while the other half is exposed to water, as reflected by a small inclination angle Ω. Moreover, the F_2 helix is short, with only eight residues, and it is contiguous to the F_1 helix (Table II). If the two contiguous helices, F_1 and F_2, are combined into a single helix F, it was found that the inclination angle of F is $\Omega = 3.9°$, as observed for most of amphiphilic helices in proteins.

Finally, it should be realized that use of the algorithm to predict the inclination angle of an α-helix in a water-lipid environment is limited by the requirement that the helix concerned should be an amphiphilic one. In this sense, a pure hydrophobic or pure hydrophilic helix, i.e., one consisting of only hydrophobic or hydrophilic residues, will lie outside of the capability of the current algorithm. As mentioned at the beginning of this study, except for most transmembrane helices, α-helices in proteins are to some degree amphiphilic and can thus be treated by the algorithm.

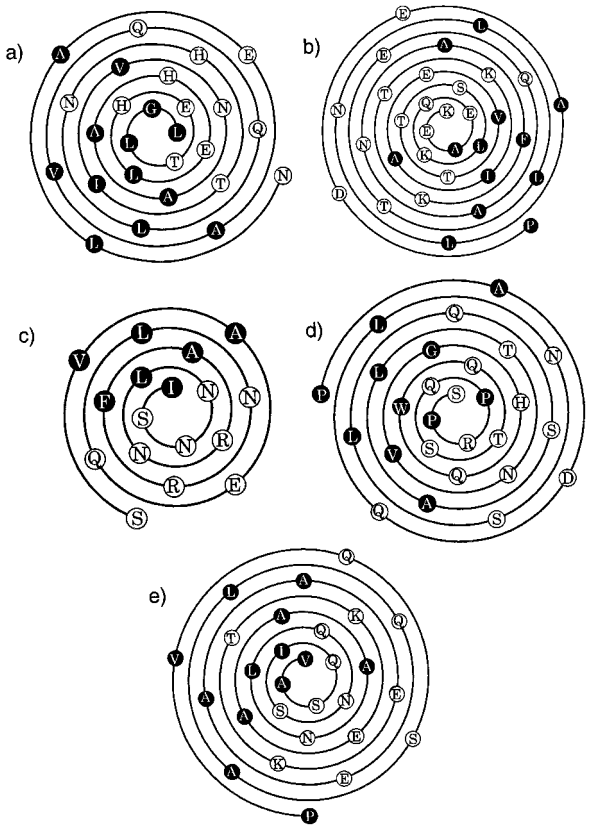

Fig. 6. The two-dimensional wenxiang diagram for the 1st (**a**), 2nd (**b**), 3rd (**c**), 4th (**d**), and 5th (**e**) α-helix of apolipoprotein. The diagram is depicted by a conical projection of the helix onto a plane perpendicular to its axis. The projected image thus obtained looks like the Chinese wenxiang, a coil-like incense used widely in China to repel mosquitos, which has the geometry of a planar spiral with a continuously varying radius. The larger the radius an amino acid residue has in the spiral, the farther above the projection plane is the corresponding amino acid in the helix and vice versa. A hydrophobic residue is denoted by a filled circle with its code symbol in white, whereas a hydrophilic residue is denoted by an open circle with its code symbol in black. As shown in the figure, the majority of hydrophobic residues are located on one side of the helix, whereas the majority of hydrophilic residues are located on the opposite side, a typical feature of an amphiphilic helix. Note also that if the helix is set with its axis pointing from the apex to the plane for conical projection as done here, then in the two-dimensional wenxiang diagram the COOH-terminal of the helix lies near the center while its NH_2-terminal lies at the outer rim of the diagram.

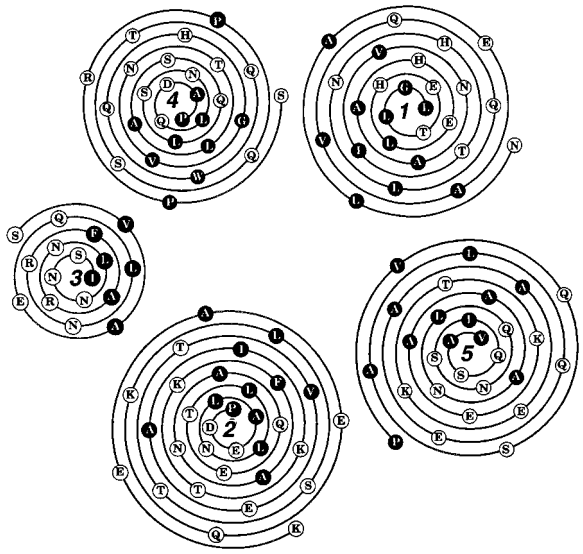

Fig. 7. A combination of the two-dimensional wenxiang diagram to express the 5-helix bundle of apolipoprotein when looking down the composite axis of the whole molecule as shown in Figure 4. As shown, most filled circles (hydrophobic residues) are buried inside the bundle and most open circles (hydrophilic residues) are exposed to the aqueous environment that surrounds the protein. Note the difference between helix 2 of this figure and the one in Figure 6b. Although both represent the second helix of apolipoprotein, the former is drawn with its axis pointing to an opposite direction to the latter because of the spatial constraint by the loop. There is a similar difference between helix 4 of this figure and the one in Figure 6d for the same reason.

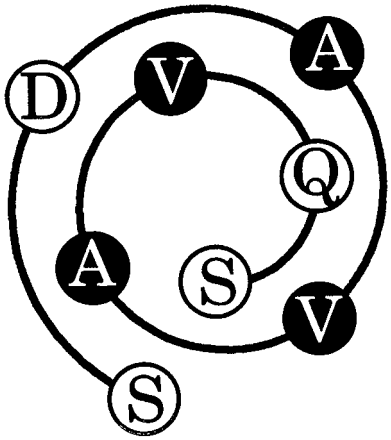

Fig. 8. The two-dimensional wenxiang diagram for the F_2 helix of the triose phosphate isomerase. Compared with Figure 6a–e, this one exhibits a remarkable difference, i.e., the filled and open circles are interweaved with each other and a clear-cut demarcation line separating the filled circles from the open ones is not found. Such a helix will, as observed, have a relatively large inclination angle.

CONCLUSION

Except for transmembrane α-helices, where the amino acid composition is basically hydrophobic, the α-helices in globular proteins are usually amphiphilic. If an amphiphilic helix is put into a water-

TABLE II. The Inclination Angles Ω for 26 Amphiphilic α-Helices From a Selection of Other Proteins

Helix mark	Residue numbers	Inclination angle Ω (deg)	Intercept T_z (Å)	Helix length (Å)
A*	3–18*	0.6	28.4	22.5
B	20–35	−0.9	32.9	22.5
C	36–42	−9.9	3.8	9.0
D	51–57	12.1	3.4	9.0
E	58–77	6.0	15.5	28.5
F	86–95	−2.0	14.1	13.5
G	100–118	−10.4	12.1	27.0
H	124–149	−3.5	20.2	37.5
A†	5–30†	9.4	19.2	37.5
B	40–53	12.9	9.3	19.5
C	79–102	−0.6	24.9	34.5
D	104–125	−2.0	23.7	31.5
H_1‡	54–68‡	−4.5	16.5	21.0
H_2	98–106	−23.1	7.7	12.0
H_3	121–135	−1.2	−6.4	21.0
A§	17–31§	17.8	9.4	21.0
B	44–55	−17.2	10.5	16.5
C	79–87	−8.4	5.7	12.0
D	105–120	8.5	11.8	22.5
E_1	130–137	21.2	3.3	10.5
E_2	138–154	−4.4	14.7	24.0
F_1	177–196	3.6	14.9	28.5
F_2	197–204	46.5	5.3	10.5
G	213–223	−5.4	10.2	15.0
H_1	232–236	−9.6	0.8	6.0
H_2	237–246	18.4	7.8	13.5

*The following 8 entries are taken from sperm whale myoglobin[19].
†The following 4 entries are taken from the A chain of cytochrome *c*′[20].
‡The following 3 entries are taken from hydrolase[21].
§The following 11 entries are taken from the A chain of triose phosphate isomerase[22].

lipid environment, the inclination angle Ω of the helical axis with respect to the water-lipid interface will be governed by the distribution of the hydrophobic and hydrophilic residues in the α-helix. In terms of Criterion 1 and Criterion 2, the inclination angle of a given α-helix with respect to the water-lipid interface can be calculated. Moreover, in terms of Criterion 3, the relative position between the given α-helix and the two-phase interface can be uniquely calculated as well. Therefore, according to these three criteria the geometric arrangement for a given helix in a polar-apolar environment can be approximately determined.

The calculated results for a number of proteins show that the inclination angles for most helices therein are small ($|\Omega| < 25°$). This suggests that driven by the free energy an amphiphilic helix will tend to seek its own arrangement in a protein such that approximately half of its face is buried in the protein, whereas the other half is exposed to the aqueous environment that surrounds the protein.

The wenxiang diagram provides an easily visualizable picture that characterizes the disposition of hydrophobic and hydrophilic residues in α-helices. Because the wenxiang diagram is generated by a conical projection of a helix onto a plane, the location of each residue in a helix is not only defined by an angle around the diagram's center, but is also defined by its radius from the center. Therefore, in principle, the wenxiang diagram can be used to represent an α-helix of any length. In contrast with this, if the helical wheel diagram[18] is used to represent an α-helix longer than 20 residues, they must crowd into a very limited space or overlap one another and can hardly be discerned from each other. Also, for the same reason, the wenxiang diagram can provide much more information of the constituent amino acids in an α-helix than the helical wheel diagram. In other words, if needed, various information such as hydrophobility, polarity, and structural characters of each individual residue in an α-helix can all be illustrated in a single wenxiang diagram. This is obviously impossible for a helical wheel diagram. Consequently, compared with the helical wheel diagram[18] widely used in the literature, the wenxiang diagram has the following advantages: 1) it is easier to discern the relative locations of the amino acids of an α-helix, especially when the helix concerned is long; 2) it distinguishes the helix ends from one another, i.e., contains the information of the directionality of an α-helix as well; and 3) it has great potential in providing more information about each of the constituent amino acid residues in an α-helix. All of these features are very useful when the wenxiang diagram is used as a component to represent a protein.

ACKNOWLEDGMENTS

Illuminating discussions with Dr. Ferenc J. Kézdy are gratefully acknowledged. We also thank Dr. Wei-Zhu Zhong for help drawing the wenxiang diagrams and Diane M. Ulrich and Cynthia A. Ludlow for help in drawing the figures of this study.

REFERENCES

1. Eisenberg, D., Schwarz, E., Komaromy, M., Wall, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. 179:125–142, 1984a.
2. Eisenberg, D., Weiss, R.M., Terwilliger, T.C. The hydrophobic moment detects periodicity in protein hydrophobicity. Proc. Natl. Acad. Sci. U.S.A. 81:140–144, 1984b.
3. Loof, H.D., Rosseneu, M., Brasseur, R., Ruysschaert, J.M. Use of hydrophobicity profiles to predict receptor binding domains on apolipoprotein E and the low density lipoprotein apolioprotein B-E receptor. Proc. Natl. Acad. Sci. U.S.A. 83:2295–2299, 1986.
4. Brasseur, R. Calculation of the three-dimensional structure of *Saccharomyces cerevisiae* cytochrome *b* inserted in a lipid matrix. J. Biol. Chem. 263:12571–12575, 1988.
5. Brasseur, R. Differentiation of lipid-associating helices by use of three-dimensional molecular hydrophobicity potential calculations. J. Biol. Chem. 266:16120–16127, 1991.
6. Talmud, P., Lins, L., Brasseur, R. Prediction of signal peptide functional properties: A study of the orientation

and angle of insertion of yeast invertase mutants and human apolipoprotein B signal peptide varients. Protein Eng. 9:317–321, 1996.
7. Kaiser, E.T., Kézdy, F.J. Secondary structures of proteins and peptides in amphiphilic environments: A review. Proc. Natl. Acad. Sci. U.S.A. 80:1137–1143, 1983.
8. Cornette, J.L., Cease, K.B., Margalit, H., Spouge, J.L., Berzofsky, J.A., DeLisi, C. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. J. Mol. Biol. 195:659–685, 1987.
9. Jones, M.K., Anantharamaiah, G.M., Segrest, J.P. Computer programs to identify and classify amphipathic α-helical domains. J. Lipid Res. 33:287–290, 1992.
10. Chou, K.C., Némethy, G., Scheraga, H.A. Energetic approach to the packing of α-helices. J. Am. Chem. Soc. 106:3161–3170, 1984.
11. Tanford, C. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. J. Am. Chem. Soc. 84:4240–4274, 1962.
12. Jones, D.D. Amino acid properties and side-chain orientation in proteins: A cross correlation approach. J. Theor. Biol. 50:167–183, 1995.
13. Schulz, G.E., Schirmer, R.H. "Principles of Protein Structure." chapt. 5. New York: Springer-Verlag, 1979.
14. Fisher, R.A. Theory of statistical estimation. Proc. Cambridge Philos. Soc. 22:700–725, 1925.
15. Chou, K.C., Némethy, G., Rumsey, S., Tuttle, R.B., Scheraga, H.A. Interactions between an α-helix and a β-sheet: Energetics of α/β packing in proteins. J. Mol. Biol. 186:591–609, 1985.
16. Breiter, D.R., Kanost, M.R., Benning, M.M., Wesenberg, G., Law, J.H., Wells, M.A., Rayment, I., Holden, H.M. Molecular structure of an apolipoprotein determined at 2.5 A resolution. Biochemistry 30:603–608, 1991.
17. Chou, K.C., Zhang, C.T. Prediction of protein structural classes. Crit. Rev. Biochem. Mol. Biol. 30:275–349, 1995.
18. Schiffer, M., Edmundson, A.B. Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. Biophys. J. 7:121–135, 1967.
19. Kuriyan, J., Wilz, S., Karplus, M., Petsko, G.A. X-ray structure and refinement of carbon-monoxy (Fe II)-myoglobin at 1.5 Å resolution. J. Mol. Biol. 192:133–154, 1986.
20. Finzel, B.C., Weber, P.C., Hardman, K.D., Salemme, F.R. Structure of ferricytochrome C′ from rhodopspirillum molishianun at 1.67 Å resolution. J. Mol. Biol. 186:627–643, 1985.
21. Loll, P.J., Lattman, E.E. The crystal structure of the ternary complex of staphylococcal nuclease, Ca^{+}_{2}, and the inhibitor pdTp, refined at 1.65 Å. Proteins 5:183–201, 1989.
22. Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Wilson, I.A. Atomic coordinates for triose phosphate isomerase from chicken muscle. Biochem. Biophys. Res. Comm. 72:146–155, 1976.