Journal of Theoretical Biology 262 (2010) 383–390

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Geometry, thermodynamics, and protein

Yi Fang [*], Junmei Jing

Centre for Bioinformation Science, Mathematical Sciences Institute, Australian National University, Canberra, ACT 0200, Australia

ARTICLE INFO

Article history:
Received 30 August 2008
Received in revised form
9 September 2009
Accepted 9 September 2009
Available online 23 October 2009

Keywords:
Hydrophobic effect
Hydrophobic core
Hydrophobic area
Volume
Area
Hydrogen bond
Constrained minimization
Free energy
Globular protein

ABSTRACT

We derive a new continuous free energy formula for protein folding. We obtain the formula first by adding hydrophobic effect to a classical free energy formula for cavities in water. We then obtain the same formula by geometrically pursuing the structure that fits best the well-known global geometric features of native structures of globular proteins: 1. high density; 2. small surface area; 3. hydrophobic core; 4. forming domains for long polypeptide chains. Conformations of a protein are presented as an all atom CPK model $P=\cup_{i=1}^{N}B(\mathbf{x}_i,r_i)$ where each atom is a ball $B(\mathbf{x}_i,r_i)$. All conformations satisfy generally defined steric conditions. For each conformation P of a globular protein, there is a closed thermodynamic system $\Omega_P \supset P$ bounded by the molecular surface M_P. Both methods derive the same free energy $aV(P)+bA(P)+cW(P)$, where $a,b,c>0$, $V(P)$, $A(P)$, and $W(P)$ are volume of Ω_P, area of M_P, and area of the hydrophobic surface $W_P \subset M_P$, which quantifies hydrophobic effect.

Minimizing $W(P)$ is sufficient to produce statistically significant native like secondary structures and hydrogen bonds in the proteins we simulated.

1. Introduction

Proteins are macromolecules consisting of amino acid sequences joined by peptide bonds. According to their side chains, the 20 amino acids are classified as *hydrophobic* or *hydrophilic*. The former avoid contact with water, the latter can form *hydrogen bonds* with water. Complicated interactions due to the locations of hydrophobic and hydrophilic moieties and surrounding water molecules contribute to free energy; the total contribution made is called the *hydrophobic effect* (Kauzmann, 1959; Tanford, 1978; Tanford and Reynolds, 2001; Finkelstein and Ptitsyn, 2002; Dill et al., 2008).

Proteins are nature's robots participating in every life phenomenon. Proteins' ability to perform such versatile functions depends on their specific 3-dimensional geometric shapes, their *native structures*. Proteins can only perform their functions in native structures (Branden and Tooze, 1998; Tanford and Reynolds, 2001). In physiological environments, among the infinite different shapes that an amino acid sequence may take, a protein always rapidly folds to its native structure automatically, though, some (larger ones) may need help.

The amino acid sequence of a protein is called its *primary structure*. Regular patterns of local (along the sequences) structures such as helix, strand, sheet and turn are called the *secondary structure* which contain many intramolecular hydrogen bonds. The global assembly of these secondary structures, connected by turns and irregular loops, is called the *tertiary structure*. For proteins having multiple amino acid sequences or structurally associated with other molecules there are also *quaternary structures* (Branden and Tooze, 1998; Finkelstein and Ptitsyn, 2002).

Generally speaking, why and how a protein's amino acid sequence can automatically fold to its native structure is called the *protein folding problem*. In Dill et al. (2008), it is summarized as: "the protein folding problem has come to be regarded as three different problems: (a) the folding code: the thermodynamic question of what balance of inter-atomic forces dictates the structure of the protein, for a given amino acid sequence; (b) protein structure prediction: the computational problem of how to predict a protein's native structure from its amino acid sequence; and (c) the folding process: the kinetics question of what routes or pathways some proteins use to fold so quickly."

In this paper we concentrate on part (a), the folding code, of the protein folding problem.

To an amino acid sequence $U=A_1A_2\cdots A_{n-1}A_n$, we define the *chain distance* of two atoms $\mathbf{a}$ and $\mathbf{b}$ in U by $d(\mathbf{a},\mathbf{b})=|i-j|$ if $\mathbf{a}\in A_i$ ($\mathbf{a}$ belongs to A_i) and $\mathbf{b}\in A_j$. Following Dill (1990), an interaction among atoms $\mathbf{a}_{i1},\mathbf{a}_{i2},\ldots,\mathbf{a}_{ik}$, $k=2,3,\ldots,$ is *local* if $\max_{1\le j,l\le k} d(\mathbf{a}_{ij},\mathbf{a}_{il})\le 4$; otherwise the interaction is non-local.

[*] Corresponding author. Tel.: +61 2 61250725, +61 2 62623623; fax: +61 2 61255549.
E-mail addresses: yi.fang3@gmail.com, yi.fang@anu.edu.au (Y. Fang).

0022-5193/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2009.09.013

We will say that an interaction is *global* if it involves all atoms in the molecule and the non-local ones collectively make a major contribution.

The CPK, or space-filling model of a molecule of N atoms, is a bundle of balls $P=\cup_{i=1}^{N}B(\mathbf{x}_i,r_i)$ where atom $\mathbf{a}_i$ is a ball $B(\mathbf{x}_i,r_i)=\{\mathbf{y}\in\mathbb{R}^3:|\mathbf{y}-\mathbf{x}_i|\le r_i\}$, $\mathbf{x}_i=(x_i^1,x_i^2,x_i^3)\in\mathbb{R}^3$ is the atomic center, $|\mathbf{y}-\mathbf{x}_i|=\sqrt{\sum_{k=1}^{3}(y_k-x_i^k)^2}$, and $r_i>0$ is the van der Waals radius. All CPK models in this article contain all atoms of the molecule, including hydrogen atoms.

A *conformation* is a CPK model P satisfying the following *steric conditions*: there are positive numbers ε_{ij}, δ_{ij}, and Δ_{ij}, $1\le i<j\le N$, such that for any two atoms $B(\mathbf{x}_i,r_i)$ and $B(\mathbf{x}_j,r_j)$ in $P=\cup_{k=1}^{N}B(\mathbf{x}_k,r_k)$,

$$\varepsilon_{ij}\le|\mathbf{x}_i-\mathbf{x}_j|,\quad \text{if } B(\mathbf{x}_i,r_i) \text{ and } B(\mathbf{x}_j,r_j) \text{ have no bond;}$$

$$\delta_{ij}\le|\mathbf{x}_i-\mathbf{x}_j|\le\Delta_{ij}<r_i+r_j,\quad \text{if } B(\mathbf{x}_i,r_i) \text{ and } B(\mathbf{x}_j,r_j) \text{ have a bond.} \tag{1}$$

We will denote the set of all conformations of U as $\mathcal{P}(U)$.

The steric conditions represent the totality of inter-atomic interactions within the molecule. We will discuss them more in Section 4.

The "thermodynamic hypothesis" in Anfinsen (1973) is: "the three-dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature, and other) is the one in which the Gibbs free energy of the whole system is lowest; that is, that the native conformation is determined by the totality of inter-atomic interactions and hence by the amino acid sequence, in a given environment." This is the consisensus in the protein folding community, called the *Thermodynamic Principle*, see for example, Tanford (1978), Dill (1990), Branden and Tooze (1998) and Tanford and Reynolds (2001).

Thus, any mathematical model of protein folding has to handle free energy. What is the free energy formula for a given protein? Various free energy formulae for the system consisting of protein and solvent (water) have been proposed, such as in Lazaridis and Karplus (1999, 2003). Here we form a mathematical model that gives a continuous free energy formula for every possible structure of an amino acid sequence. The solvent structure is not explicitly accounted for in the formula. Instead, like in Eisenberg and McLachlan (1986), and Lazaridis and Karplus (2003), we define the boundary of a thermodynamic system and use boundary data to express the solvent contribution to the free energy.

We will derive the continuous free energy formula independently through classical thermodynamics and through pure geometric calculus of variation. In the former we add the hydrophobic effect to a classic free energy formula for cavities in water; in the latter we imitate nature and pursue the structure that fits best the well-known global geometric features of native structures of globular proteins.

2. A free energy formula: thermodynamics

2.1. A classical free energy formula for cavities in water

Let $P=\cup_{i=1}^{N}B(\mathbf{x}_i,r_i)$ be a conformation and $\mathcal{T}\subset\mathbb{R}^3$ a closed thermodynamic system containing P with boundary $\Sigma=\partial\mathcal{T}$. Let $V(\mathcal{T})$ and $A(\Sigma)$ be the volume of $\mathcal{T}$ and the area of Σ. If we think of $\mathcal{T}$ as a cavity in water, then by classical thermodynamics the free energy for $\mathcal{T}$ is

$$G(\mathcal{T})=\sigma A(\Sigma)+pV(\mathcal{T}), \tag{2}$$

where σ is the surface tension and p the free energy per unit volume of the bulk liquid (Pippard, 1957; Southall et al., 2002), or pressure (Chandler, 2005).

Because a protein is not just a cavity in water, obviously, this would not work for proteins. We have put "the totality of inter-atomic interactions" (Anfinsen, 1973) of the protein into the steric conditions (1) such that it does not show in formula (2). But the totality of interactions of the protein molecule with the solvent, that is, the hydrophobic effect that is the main driving force of protein folding, is still missing. Thus, we should quantify hydrophobic effect and put it into the free energy formula (2).

2.2. Hydrophobic core

The hydrophobic effect causes the *hydrophobic core*: almost all globular proteins of known structure have a hydrophobic core—the interior of proteins are composed of clustered hydrophobic moieties, not polar side chains, nor ionized side chains in salt bridges. This is important in determining whether or not the driving force of protein folding is the hydrophobic effect (Tanford, 1978; Novotny et al., 1984, 1988; Richards and Lim, 1994; Branden and Tooze, 1998; Tanford and Reynolds, 2001; Lesk, 2001; Finkelstein and Ptitsyn, 2002).

2.3. Hydrophobic surface

Since the formation of the hydrophobic core is caused by the hydrophobic effect, we will quantify the hydrophobic core to indirectly quantify the hydrophobic effect.

Let U be a molecule, $H\subset U$ be the set of all hydrophobic moieties. If we label all atoms in the molecule by $\{\mathbf{a}_1,\ldots,\mathbf{a}_N\}$ then there is a subset $I\subset\{1,\ldots,N\}$ such that atom $\mathbf{a}_i$ belongs to H if and only if $i\in I$. Therefore, H is intrinsic, and, independent of conformations. Note that H contains all hydrophobic moieties, not just the hydrophobic amino acid side chains in a protein. In each conformation P, H's conformation is denoted as $H_P=\cup_{i\in I}B(\mathbf{x}_i,r_i)\subset P=\cup_{i=1}^{N}B(\mathbf{x}_i,r_i)\subset\mathbb{R}^3$. Define the *hydrophobic surface* $W_\Sigma\subset\Sigma$ as follows:

$$W_\Sigma=\{\mathbf{x}\in\Sigma:\operatorname{dist}(\mathbf{x},H_P)<\operatorname{dist}(\mathbf{x},P\backslash H_P)\}, \tag{3}$$

where $\operatorname{dist}(\mathbf{x},H_P)=\min_{\mathbf{y}\in H_P}|\mathbf{x}-\mathbf{y}|$ and $P\backslash H_P=\{\mathbf{y}\in P;\mathbf{y}$ does not belong to $H_P\}$. See Fig. 1

Let $W=W(\Sigma)=A(W_\Sigma)$ be the area of $W_\Sigma\subset\Sigma$, then $0\le W(\Sigma)\le A(\Sigma)$. If all hydrophobic moieties are in the core, i.e., surrounded by non-hydrophobic moieties, then $P\backslash H_P$ separates H_P and Σ, i.e., for any $\mathbf{x}\in\Sigma$ and any $\mathbf{y}\in H_P$, the straight line segment L between $\mathbf{x}$ and $\mathbf{y}$ will intersect $P\backslash H_P$. Thus, there will be at least one $\mathbf{z}\in P\backslash H_P$ lying on L, hence, $|\mathbf{z}-\mathbf{x}|<|\mathbf{y}-\mathbf{x}|$. We have

$$\operatorname{dist}(\mathbf{x},P\backslash H_P)=\min_{\mathbf{z}\in P\backslash H_P}|\mathbf{z}-\mathbf{x}|\le\min_{\mathbf{y}\in H_P}|\mathbf{y}-\mathbf{x}|=\operatorname{dist}(\mathbf{x},H_P),$$

by definition, $\mathbf{x}$ does not belong to W_Σ. Since this is true for any $\mathbf{x}\in\Sigma$, we have $W_\Sigma=\emptyset$ and $W(\Sigma)=0$. Similarly on the other hand, if all non-hydrophobic moieties are surrounded by hydrophobic moieties, then for any $\mathbf{x}\in\Sigma$, $\operatorname{dist}(\mathbf{x},H_P)<\operatorname{dist}(\mathbf{x},P\backslash H_P)$; therefore, $W_\Sigma=\Sigma$ and $W(\Sigma)=A(\Sigma)$.

Between these two extremes, the less hydrophobic moieties are exposed to water, the smaller the $W(\Sigma)$ is and the better the formation of the hydrophobic core. Hence, $W(\Sigma)$ can be used to quantify the hydrophobic core.

Putting $W(\Sigma)$ into (2) we have the free energy formula for a protein contained in the closed thermodynamic system $\mathcal{T}$

$$G(\mathcal{T})=pV(\mathcal{T})+\sigma A(\Sigma)+hW(\Sigma);\quad p,\sigma,h>0. \tag{4}$$

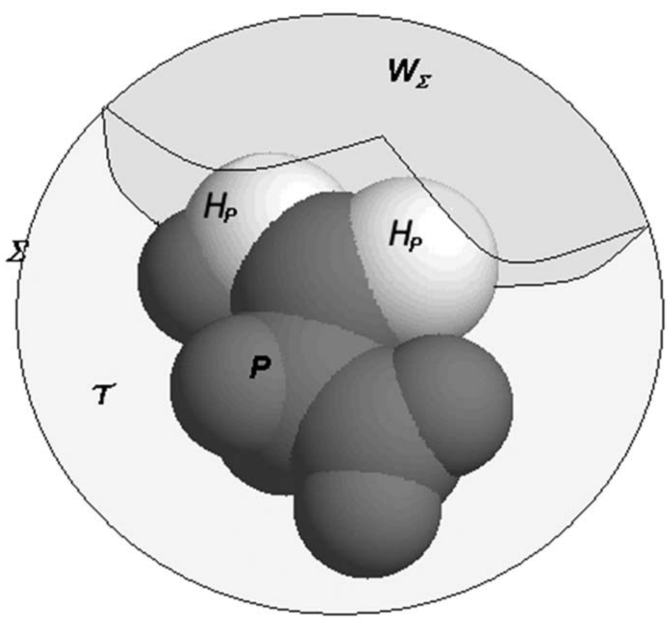


Fig. 1. Sketch of hydrophobic surface on a sphere. Any boundary Σ, usually not a sphere, totally surrounds the conformation P.

This is still not satisfactory, since for a conformation P, there are infinite systems $\mathcal{T}$ containing it. The question remains as to which one should be used.

2.4. *Boundary Σ of the thermodynamic system $\mathcal{T}$*

For the conformation P, there are some natural surfaces that are determined uniquely by P, some of them have been treated as the natural boundary of P.

The conformation P's own boundary ∂P is called the van der Waals surface. But it is not suitable to be $\partial\mathcal{T}$ since it does not count on the interactions with the surrounding water. If we add a certain length, say the rough radius of a water molecule about $r_p = 1.6\,\text{Å}$ to the various van der Waals radii r_i, i.e., make $R_i = r_i + r_p$, then the van der Waals surface of $P' = \cup_{i=1}^{N} B(\mathbf{x}_i, R_i)$ is called the *solvent accessible surface* of P (Lee and Richards, 1971) See Fig. 2. Many believe that the area of the solvent accessible surface of a protein structure is correlated to the free energy, for example, in Bolen and Rose (2008), it is treated as if it were truthfully the free energy.

Molecular surface is generated by rolling a probe sphere of radius r_p on the van der Waals surface (Richards, 1977). It has two parts (not necessarily connected), the convex part called contact surface (those sphere caps in the conformation $P = \cup_{i=1}^{N} B(\mathbf{x}_i, r_i)$ where the probe sphere touches only one sphere in P), and reentrant surface (the probe sphere simultaneously touches two or three or more spheres in P) (Connolly, 1983). Since there may be inner cavities in P large enough to allow the probe sphere to roll inside, the molecular surface may have separate components (Connolly, 1983). We will take M_P as the outermost and also the largest component and still call it the molecular surface of P.

Rolling the probe sphere can be viewed as water surrounding the protein, thereby viewing the force or energy exchanges on the molecular surface as the solvent effect. In this view, we can see why selecting the correct boundary surface will dramatically affect the correct treatment of the hydrophobic effect as shown in Tuñón et al. (1992).

In fact, molecular surface area reflects the free energy better than accessible surface area does, see Tuñón et al. (1992) and Jackson and Sternberg (1993, 1994, 1995). While the non-smooth part of the accessible surface caused the gap between experimental data of free energy and the surface area (Jackson and Sternberg, 1993, 1994, 1995), a molecular surface does not have non-smooth part.

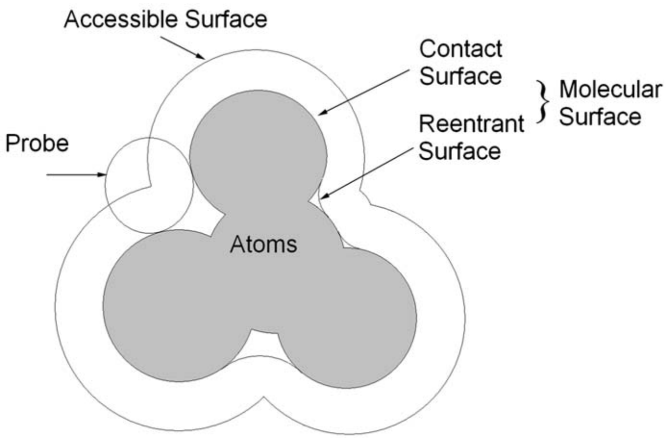


Fig. 2. Two dimensional sketch of various surfaces. Shaded area is the conformation P, its boundary is the van der Waals surface. Molecular surface is in between the van der Waals surface and accessible surface and is smoother than both.

For the above considerations and practical purposes, we use the molecular surface M_P as the boundary of the thermodynamic system $\Omega_P = \mathcal{T}$, i.e., we take $\Sigma = \partial\mathcal{T} = \partial\Omega_P = M_P$.

Since such M_P is uniquely determined by P (once we fixed the probe radius) and Ω_P is uniquely determined by M_P, we can write $V(\Omega_P) = V(P)$, $A(M_P) = A(P)$. We will denote $W_P = W_{M_P}$ and $W(P) = A(W_P)$.

The free energy of the conformation P thus can be written as

$$G(P) = pV(P) + \sigma A(P) + hW(P); \quad p, \sigma, h > 0. \tag{5}$$

3. The same free energy formula: geometry

We now derive the free energy formula (5) from a pure geometric consideration. The fact that geometry enters the picture is not surprising at all. Anfinsen already stated in Anfinsen (1973) that “biological function appears to be more a correlate of macromolecular geometry than of chemical detail.”

We may need new perspectives on the problem. We take a phenomenological point of view, i.e., mathematically express the results of observed phenomena without paying detailed attention to their fundamental significance (Thewlis, 1973, p. 248).

What are the “observed phenomena”? We look at the well-known global geometric features of native structures of globular proteins.

3.1. *Global geometric features of native structures of globular proteins*

Bond length, bond angle, and torsion angle are local geometric quantities. Native structures of globular proteins have well-known global features depending on the structure as a whole:

3.1.1. *High density*

It was observed that protein interiors are closely packed and the interior of lysozyme and ribonuclease have a packing density of 0.75 (Richards, 1974). The packing density is quite uniform or homogenous, such that water molecules are excluded from the interior (Lee and Richards, 1971; Richards, 1974, 1977, 1979). Unfilled cavities large enough to accommodate a water molecule only appear in very few cases (Richards, 1974, 1979; Richards and Lim, 1994; Branden and Tooze, 1998).

3.1.2. Small surface area

The solvent accessible surface (Lee and Richards, 1971) and the molecular surface (Richards, 1977) are designed as boundaries of protein structures. It is observed that the native structure of a globular protein has smaller surface area than that of other conformations (Lee and Richards, 1971; Janin, 1976; Richards, 1977, 1979; Novotny et al., 1984, 1988).

3.1.3. Hydrophobic core

As described in Section 2.2.

3.1.4. Long amino acid sequences fold into domains

If the amino acid sequence $U=A_1A_2\cdots A_n$ of a globular protein is long, say $n\geq 400$, then the native structure of U will divide into several domains connected by loops. Each domain looks like the native structure of a smaller globular protein (Richards, 1977; Branden and Tooze, 1998; Tanford and Reynolds, 2001; Lesk, 2001; Finkelstein and Ptitsyn, 2002).

3.2. Pursing a structure that fits these features best

Let U be an amino acid sequence. Our goal is to quantify all these global geometric features of the native structures of globular proteins and then try to find a structure (conformation) $Q\in\mathcal{P}(U)$ such that Q best fits these features.

Again we consider a conformation $P=\cup_{i=1}^{N}B(\mathbf{x}_i,r_i)\in\mathcal{P}(U)$ being contained in a closed thermodynamic system $\mathcal{T}$, $P\subset\mathcal{T}\subset\mathbb{R}^3$. We will require that the boundary $\Sigma=\partial\mathcal{T}$ of $\mathcal{T}$ is a simply connected closed surface in $\mathbb{R}^3$ (A sphere is simply connected while a tyre surface is not. A mathematical fact is that all simply connected closed surfaces in $\mathbb{R}^3$ are homeomorphic to a sphere). Thus we have the quantities $V(\mathcal{T})$, $A(\Sigma)$, and $W(\Sigma)=A(W_\Sigma)$, where the hydrophobic surface W_Σ is defined in (3).

3.2.1. Smaller $V(\mathcal{T})$ means higher density

Let V_P be the volume of P, then the density of P in $\mathcal{T}$ is given by

$$0\leq \text{Density}=\frac{V_P}{V(\mathcal{T})}\leq 1,$$

because $P\subset\mathcal{T}$. Since P satisfies the steric conditions (1), V_P is almost a constant in different conformations. Thus, the smaller the volume $V(\mathcal{T})$ is, the higher the density will be.

3.2.2. Smaller $A(\Sigma)$ gives better global-like shape

The area $A(\Sigma)$ can be seen as a shape optimizer of conformations in the sense that among all surfaces enclosing the same volume, the sphere has the minimal surface area (Almgren, 1986). In our case, because all conformations must satisfy the steric conditions (1), $\Sigma=\partial\mathcal{T}$ can never be a sphere. But the general principle is still true, the smaller the $A(\Sigma)$ is, the more global-like the Σ will be.

3.2.3. Smaller $W(\Sigma)$ gives better hydrophobic core

We have already seen in Section 2.2 that the hydrophobic surface area $W(\Sigma)=A(W_\Sigma)$ measures the hydrophobic core. The smaller the $W(\Sigma)$, the better the formation of the hydrophobic core.

Let $P\in\mathcal{P}(U)$, define $\mathcal{S}(P)$ as the set of all simply connected closed surfaces $\Sigma\subset\mathbb{R}^3$ such that $\Sigma=\partial\mathcal{T}$, where $\mathcal{T}$ is a closed region such that $P\subset\mathcal{T}\subset\mathbb{R}^3$ and $V(\mathcal{T})<\infty$. We want to find a conformation $Q\in\mathcal{P}(U)$ and a $\mathcal{L}\supset Q$, $\partial\mathcal{L}=S\in\mathcal{S}(Q)$ such that

$$u(V(\mathcal{L}),A(S),W(S))=\min_{P\in\mathcal{P}(U)}\min_{\Sigma\in\mathcal{S}(P)}u(V(\mathcal{T}),A(\Sigma),W(\Sigma)), \tag{6}$$

where $u(x,y,z)$ is a smooth function such that

$$\frac{\partial u}{\partial x}>0,\quad \frac{\partial u}{\partial y}>0,\quad \frac{\partial u}{\partial z}>0. \tag{7}$$

Thus, u is increasing for any of its three variables, any shrinking of $V(\mathcal{T})$, or $A(\Sigma)$, or $W(\Sigma)$ will reduce the value of $u(V(\mathcal{T}),A(\Sigma),W(\Sigma))$. This unknown function u reflects the requirement that we have to simultaneously reduce the volume, surface area, and hydrophobic area in a coherent way.

Note that under the variational constraint, i.e., the steric conditions (1), the surface $S\in\mathcal{S}(Q)$ in (6) can never be a sphere. But it is certain in mathematics that such a conformation Q in (6) will fit best the global geometric features of the native structures of globular proteins among all conformations. An analogue is that as long as a surface has the minimal area among all surfaces enclosing the same volume, it must be a sphere (Almgren, 1986).

Since these global geometric features essentially characterize the native structures of globular proteins, we can reasonably hypothesize that Q in (6) actually should be the native structure, of course, if we have the correct function u.

The uncertainty of u can be resolved by linear approximation. Let $Q\in\mathcal{P}(U)$ and $S=\partial\mathcal{L}\in\mathcal{S}(Q)$ be the minimizer in (6) and $\mathbf{x}_0=(x_0,y_0,z_0)=(V(\mathcal{L}),A(S),W(S))$, $u_0=u(\mathbf{x}_0)$, and $(\mu^1,\mu^2,\mu^3)=(\partial u/\partial x,\partial u/\partial y,\partial u/\partial z)(\mathbf{x}_0)$. Then near $\mathbf{x}_0$ the first approximation of $u(x,y,z)$ can be written as

$$u(x,y,z)\cong u_0+\mu^1(x-x_0)+\mu^2(y-y_0)+\mu^3(z-z_0),$$
$$\mu^i>0,\quad i=1,2,3.$$

Since a constant is irrelevant in minimization, (6) can be approximately written as

$$\mu^1V(\mathcal{L})+\mu^2A(S)+\mu^3W(S)=\min_{P\in\mathcal{P}(U)}\min_{\Sigma\in\mathcal{S}(P)}\mu^1V(\mathcal{T})+\mu^2A(\Sigma)+\mu^3W(\Sigma). \tag{8}$$

3.3. System boundary revisited

The two-tier minimizations in (6) and (8) actually gives us a way to theoretically determine the ideal thermodynamic system $\mathcal{T}\supset P$. For any $P\in\mathcal{P}(U)$, let $\partial\mathcal{T}_P=\Sigma_P\in\mathcal{S}(P)$ such that

$$\mu^1V(\mathcal{T}_P)+\mu^2A(\Sigma_P)+\mu^3W(\Sigma_P)=\min_{\Sigma\in\mathcal{S}(P)}\mu^1V(\mathcal{T})+\mu^2A(\Sigma)+\mu^3W(\Sigma). \tag{9}$$

This special surface Σ_P then is the boundary of a tailor-made thermodynamic system $\mathcal{T}_P\supset P$. It has been proven that such a Σ_P is a piecewise constant mean curvature smooth surface (Fang, 2005).

Recall that what is called mean curvature in mathematics is called surface tension in physics. It has been discovered that many interfaces between two different materials are constant mean curvature surfaces. The first example, the A/B block copolymer consisting of two macromolecules bonded together, was discussed in the pioneering work Thomas et al. (1988). Hence the fact that Σ_P has piecewise constant mean curvature is not a surprise at all. On the other hand, because Σ_P has to be a closed simply connected surface and not a sphere, it can never be a constant mean curvature surface.

Using Σ_P, we can write $V(P)=V(\mathcal{T}_P)$, $A(P)=A(\Sigma_P)$, and $W(P)=A(W_{\Sigma_P})$ and (8) can be written as

$$\mu^1 V(Q)+\mu^2 A(Q)+\mu^3 W(Q) \\ =\min_{P\in\mathcal{P}(U)} \mu^1 V(P)+\mu^2 A(P)+\mu^3 W(P), \quad \mu^i>0,\ i=1,2,3. \tag{10}$$

Compared with (5), we see that what we are pursing in geometry is actually minimizing the free energy; a wonderful coincidence! The advantage of thermodynamic consideration is that we can infer that $\mu^1=p$, the pressure. But the geometric consideration also suggests how to pursue the best boundary of a protein structure. It is a piecewise constant mean curvature (surface tension) smooth surface.

The theoretical boundary Σ_P is hard to handle. The good news is that the molecular surface M_P (see Section 2.4) is a good approximation to Σ_P. In fact M_P is smooth and has a large portion with piecewise constant mean curvature (Fang and Jing, 2008). Moreover, there exists ready software for calculating molecular surfaces, for example, Connolly (2002). Thus from now on we will replace Σ_P with M_P and $\mathcal{T}_P$ with Ω_P, $\partial\Omega_P=M_P$, $V(P)=V(\Omega_P)<\infty$.

4. Geometry and thermodynamics lead to the same mathematical model

From both classical thermodynamics and geometry, we have achieved the same free energy formula for a conformation $P\in\mathcal{P}(U)$ of a protein U, i.e.,

$$G(P)=\mu^1 V(P)+\mu^2 A(P)+\mu^3 W(P); \quad \mu^i>0,\ i=1,2,3. \tag{11}$$

To count the feature in Section 3.1.4, long amino acid sequences fold into domains, we assume that the weights μ^i in (11) may depend on chain length n. Hence, we modify (11) to get a free energy formula $G_n(P)$ for a single chain globular protein of length n, for each conformation $P=\cup_{i=1}^{N} B(\mathbf{x}_i,r_i)$, after a normalization,

$$G_n(P)=\mu_n^1 V(P)+\mu_n^2 A(P)+\mu_n^3 W(P); \quad \mu_n^i>0, \\ i=1,2,3,\ \mu_n^1+\mu_n^2+\mu_n^3=1. \tag{12}$$

4.1. Mathematical model: constrained minimization

By the thermodynamic principle of protein folding, the native structure should have minimum free energy; by the global features of native structures of globular proteins, the native structure should simultaneously have smaller volume, smaller surface area, and smaller hydrophobic area.

Therefore, we should carry out a constrained minimization under both considerations, i.e., the native structure $Q\in\mathcal{P}(U)$ should satisfy

$$G_n(Q)=\min_{P\in\mathcal{P}(U)} G_n(P). \tag{13}$$

Note that since all $P\in\mathcal{P}(U)$ satisfies the steric conditions (1), the constraints in (13) are the steric conditions (1).

4.2. The role and setting of steric conditions

Setting steric conditions is another phenomenological treatment. Simply due to the large quantity of atoms in a protein, the pairwise interactions among these atoms are too complicated to be accurately calculated by empirical formulae, so we take the effect and respect the observed facts, setting them as constraint conditions in minimization.

The steric conditions are defined via the allowed minimal atomic distances, such that for non-bonding atoms, the allowed minimal distances are: shorter between differently charged or polarized atoms; a little longer between non-polar ones; and much longer (generally greater than the sum of their radii) between the same charged ones, etc. For example, we allow minimal distance between sulfur atoms in Cysteines to form disulfide bonds.

Various bond lengths and angles have standard ranges; we can just follow them and place them under the steric conditions. As for non-bonding atoms, we will preset the allowed constant minimal distances among all moieties of all amino acids for different chain distances and then list them in tables labeled by chain distances. For example, for the central carbon group $C^{\alpha}H$ and the Alanine side chain group CH_3, in the table for chain distance 5 we preset, say, $D(C^{\alpha}H, CH_3)\geq 2$ Å. Then let $U=A_1\cdots A_n$ be an amino acid sequence such that $\mathbf{a}_i$ is the carbon or one of the three hydrogen atoms of CH_3 in the side chain of an Alanine residue, $\mathbf{a}_j$ is either C^{α} or H of $C^{\alpha}H$ such that $d(\mathbf{a}_i,\mathbf{a}_j)=5$, then we set $\varepsilon_{ij}=2$ in (1). Consider another example, the atoms C in residue A_k and N in residue A_l will have the bond relation $\delta_{ij}\leq|\mathbf{x}_i-\mathbf{x}_j|\leq\Delta_{ij}<r_i+r_j$ only when $l=k+1$; otherwise, we should use the non-bonding relation $\varepsilon_{ij}\leq|\mathbf{x}_i-\mathbf{x}_j|$.

Another important role of steric conditions is to make sure that different amino acid sequences correspond to different constrained minimization problems (13). For example, two proteins U_1 and U_2 may have the same chain length and even the same amino acid distributions, i.e., each amino acid appears in the sequences of U_1 and U_2 exactly the same number of times. Or precisely, there is a permutation $\sigma:\{1,2,\ldots,n\}\to\{1,2,\ldots,n\}$ such that $U_1=A_1A_2\cdots A_n$ and $U_2=A_{\sigma(1)}A_{\sigma(2)}\cdots A_{\sigma(n)}$. Because of the order difference, steric conditions in (1) are still able to distinguish between them, so the conformation spaces $\mathcal{P}(U_1)$ and $\mathcal{P}(U_2)$ are different and the minimizations in (13) for U_1 and U_2 are different problems.

5. Simulation of hydrophobic area reduction

For both classical thermodynamics and geometric considerations, we have emphasized the role of hydrophobic effect via the hydrophobic surface area $W(P)$. Thus, it is important to verify that $W(P)$ indeed represents the hydrophobic effect.

We will carry out simulations that reduce the hydrophobic area $W(P)$ alone to see whether or not it can produce secondary structures. If it does so, then we can conclude that the $W(P)$ is really correlated to hydrophobic effect. Since $W(P)$ is not free energy, just like in Chan and Dill (1989, 1990) (HP lattice model), and (Maritan et al., 2000) (tube model), our simulations are solely for the purpose of searching for the folding code, not the prediction of protein structures.

We will use the fastest descending method to reduce the hydrophobic area $W(P)$. Let U be a protein molecule and $b_i, i=1,\ldots,L$ the rotatable torsion axis of U, ϕ_i the current torsion angle of b_i, $\mathbf{\Phi}=(\phi_1,\ldots,\phi_L)$. Then for fixed standard bond lengths and bond angles, we can identify the conformation $P\in\mathcal{P}(U)$ by specifying all of its torsion angles $\mathbf{\Phi}=(\phi_1,\ldots,\phi_L)$. Thus $W(P)=W(\mathbf{\Phi})=W(\phi_1,\ldots,\phi_L)$, and the gradient is $DW(P)=(\partial W/\partial\phi_1,\ldots,\partial W/\partial\phi_L)$.

From a random initial conformation, we change $\mathbf{\Phi}$ simultaneously proportional to $-DW$, i.e., new angles will be $\mathbf{\Phi}(t)=\mathbf{\Phi}-tDW$, where $t>0$ is adjustable to make the new conformation satisfy the steric conditions. Repeating this until we find a conformation that either has $DW=\overrightarrow{0}$ or can no longer be moved along $-DW$ without violating steric conditions. The former is an *interior stationary conformation*, the latter is a *boundary stationary conformation*. We then record this conformation and start anew until we have accumulated enough stationary conformations. The gradient DW is calculated as

$$\frac{\partial W}{\partial \phi_i} = -2\int_{W_P} H(X_i \bullet N)\, \mathrm{d}\mathcal{H}^2 + \int_{\Gamma} \left(\eta \bullet X_i - \frac{\dfrac{\mathrm{d}f_P}{\mathrm{d}t}}{|\nabla_{M_P} f_P|} \right) \mathrm{d}\mathcal{H}^1, \quad i = 1, \ldots, L, \tag{14}$$

where X_i is the vector field in $\mathbb{R}^3$ induced by the rotation around b_i, N the inward unit normal vector of M_P, $\bullet$ the inner product in $\mathbb{R}^3$, H the mean curvature of M_P, Γ the boundary of W_P, η the unit outward conormal vector of Γ (tangent to W_P and perpendicular to Γ), $f_P(\mathbf{x}) = \mathrm{dist}(\mathbf{x}, H_P) - \mathrm{dist}(\mathbf{x}, P \backslash H_P)$, $\mathbf{x} \in \mathbb{R}^3$, $\nabla_{M_P} f_P$ the projection of the gradient ∇f_P to tangent planes of M_P, and $\mathcal{H}^1$ and $\mathcal{H}^2$ the Haussdorff measures. The mathematical proof of (14) will be given elsewhere. The integrated analytic formulae such as (14) on the molecular surface M_P are given in Fang and Jing (2008).

The amino acid sequences used in simulations are HYPOTHETICAL PROTEIN RPA1889, of 123 residues, PDB code 2i9c; HYPOTHETICAL PROTEIN SP_1588, of 127 residues, PDB code 2hng; and CONSERVED HYPOTHETICAL ALANINE RICH PROTEIN RV2844, of 162 residues, PDB code 2ib0. They were CASP7 targets t0382, t0383, t0385, the simulation results were obtained before the experiment data were published.

The same simulations also applied recently to 1poa, PHOSPHOLIPASE A2, of 118 residues; 1aac, AMICYANIN, of 105 residues; 1rro, RAT ONCOMODULIN, of 108 residues; 2end, ENDONUCLEASE V, of 137 residues; and 4fgf, BASIC FIBROBLAST GROWTH FACTOR, of 124 residues.

We used the common structure analysis software RasMol (2009) to qualify the secondary structure of conformations generated by our minimization procedure. Figs. 3–5 show the RasMol outputs to three of such conformations.

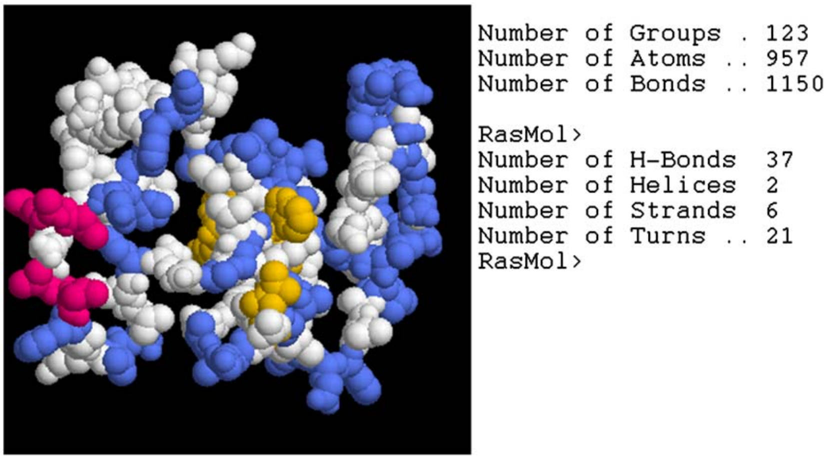
It is significant that these secondary structures and hydrogen bonds are generated by reducing a major part of a global continuous free energy function alone. Moreover, the presentation of conformation is the most realistic all atom CPK model. We neither calculated the Φ_i and Ψ_i angles nor paid any special attention to the C^{α}s' positions and residue propensities to secondary structures, nor pursued hydrogen bond formation. Under such a simulation, secondary structures and hydrogen bonds can still be produced and be recognized by objective software such as RasMol. Such a result has never been reported before. Indeed, pairwise interaction models without explicitly pursuing hydrogen bonds cannot produce secondary structures (Hubner and Shakhnovic, 2005).

The appearance of hydrogen bonds while reducing the hydrophobic area may contribute to resolving a decades long dispute concerning the dominant force in protein folding: intramolecular hydrogen bonds or hydrophobic effect? Since hydrogen bonds automatically appear while reducing hydrophobic area alone, we may infer that assuming that intramolecular hydrogen bonds are the driving force of protein folding is unnecessary. For the history of this debate, see for example: Pauling et al. (1951), Kauzmann (1959), Tanford (1978), Chan and Dill (1989, 1990), Dill (1990), Chandler (2005), Bolen and Rose (2008) and Dill et al. (2008).

To rule out the possibility that these secondary structures and hydrogen bonds are just random products, we made statistical significance tests.

6. Statistical significance tests of our simulation results

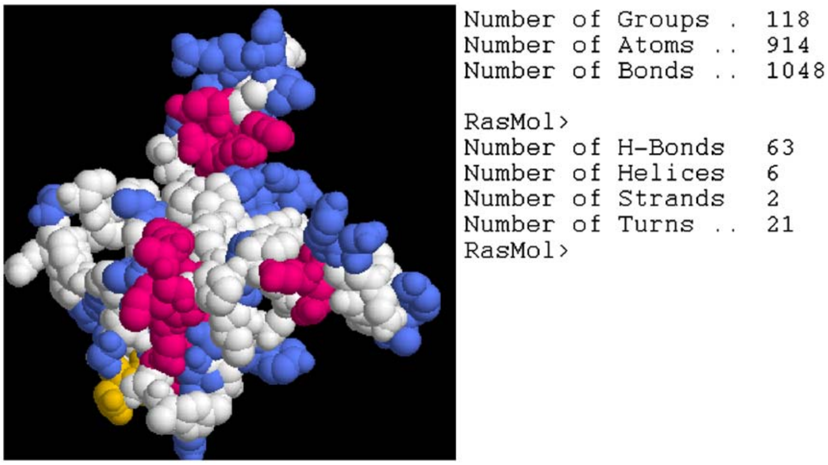
We perform a random simulation to get the random data for comparison. Both random and hydrophobic area reduction simulations use the change of dihedral angles $\mathbf{\Phi}(t) = \mathbf{\Phi} - tDW$ to get new conformations, where $t > 0$ is adjustable to make the new conformation satisfy the steric conditions. The difference is that for the hydrophobic area reduction simulation, the gradient $DW(P) = (\partial W/\partial \phi_1, \ldots, \partial W/\partial \phi_L)$ is calculated by formula (14) and for random simulation $DW = D\mathbf{\Phi} = (d\phi_1, \ldots, d\phi_L)$ are just random values. Both simulations generated 500 structures for 2i9c. We selected the best 50 structures for each of the two kinds of simulations. Selection is done by manually checking the results and picking out those with near globular shape.



```
Number of Groups . 123
Number of Atoms .. 957
Number of Bonds .. 1150

RasMol>
Number of H-Bonds  37
Number of Helices  2
Number of Strands  6
Number of Turns .. 21
RasMol>
```

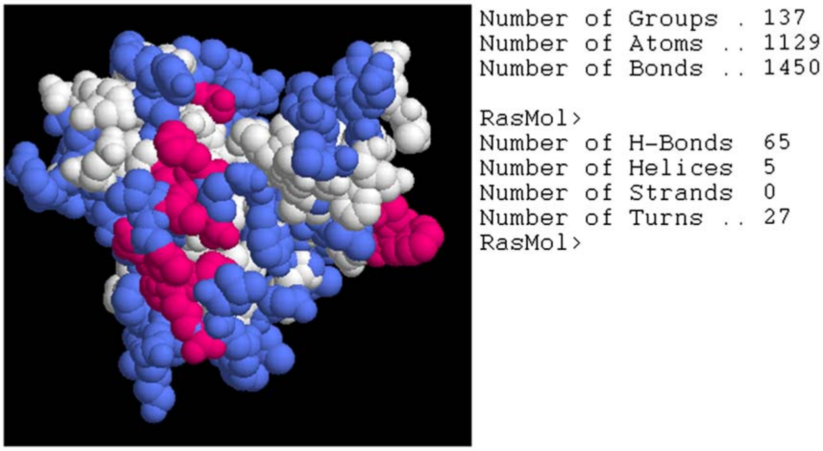
Fig. 3. RasMol output of a resulting structure of simulation of hydrophobic area reduction. It shows that in this 123 residue protein chain, there are 2 helices, 6 strands, 21 turns, and 37 hydrogen bonds in the structure. Helices were anticipated because rotation of a torsion angle to shrink the local hydrophobic area will push hydrophobic moieties in consecutive residues to turn and hide behind the hydrophilic moieties; exactly what is required to form a helix. Since the simulation paid no attention to any local geometry and no effort was tried to imitate any chemical relationship besides the steric conditions, the appearance of hydrogen bonds is extra awarding. Like the HP model and tube model inferred from their simulations, our simulation results prove unequivocally that hydrophobic effect pushes hydrophobic moieties to hide from water. The hydrophilic moieties covalently bonded to them then form intramolecular hydrogen bonds. Therefore, secondary structures are produced to stabilize the whole structure. This result may contribute to resolving the long standing dispute about whether hydrophobic effect or intramolecular hydrogen bonding is the main driving force in protein folding. This output is for Target T0382 of CASP7, HYPOTHETICAL PROTEIN RPA1889, PDB code 2i9c. The simulation was done before the target structure was published. The simulation was run on a 686 CPU with clock speed of 730-Mhz; the time used from a random conformation to a stationary conformation is less than 1 h.



```
Number of Groups . 118
Number of Atoms .. 914
Number of Bonds .. 1048

RasMol>
Number of H-Bonds  63
Number of Helices  6
Number of Strands  2
Number of Turns .. 21
RasMol>
```

Fig. 4. RasMol output of a resulting structure of simulation of 1poa reducing hydrophobic area alone. It shows that there are 63 hydrogen bonds, 6 helices, 2 strands, and 21 turns.



```
Number of Groups . 137
Number of Atoms .. 1129
Number of Bonds .. 1450

RasMol>
Number of H-Bonds  65
Number of Helices  5
Number of Strands  0
Number of Turns .. 27
RasMol>
```

Fig. 5. RasMol output of a resulting structure of simulation of 2end reducing hydrophobic area alone. It shows that there are 65 hydrogen bonds, 5 helices, 0 strands, and 27 turns.

Table 1
Statistical summary of 50 hydrophobic area reducing simulation structures of 2i9c.

	Min.	First Qu.	Median	Mean	Third Qu.	Max.
HBond	13.00	28.50	36.00	37.50	46.75	63.00
Helices	0.00	0.00	1.00	1.62	3.00	4.00
Strands	0.00	0.00	2.00	3.04	6.00	10.00
Turns	9.00	17.25	20.00	18.72	21.00	24.00
Res in helices	0.00	0.00	4.50	5.78	11.75	18.00
Res in strands	0.00	0.00	4.00	3.96	6.00	17.00
Longest helix	0.00	0.00	3.00	2.78	4.00	9.00
Longest strand	0.00	0.00	1.00	1.08	2.00	3.00

Table 2
Statistical summary of 50 random simulation structures of 2i9c.

	Min.	First Qu.	Median	Mean	Third Qu.	Max.
HBond	18.00	26.00	30.00	29.64	33.00	43.00
Helices	0.00	0.00	1.00	0.72	1.00	2.00
Strands	0.00	0.00	3.00	2.24	3.00	7.00
Turns	18.00	21.25	23.50	23.56	25.00	31.00
Res in helices	0.00	0.00	0.00	1.08	2.00	4.00
Res in strands	0.00	0.00	0.00	1.12	2.00	4.00
Longest helix	0.00	0.00	3.00	2.00	3.00	4.00
Longest strand	0.00	0.00	0.00	0.50	1.00	2.00

Table 3
Comparison results. $D-$value is defined as $D=\max_{-\infty<x<\infty}|S_{N_1}(x)-S_{N_2}(x)|$, where S_{N_i} are the cumulative distribution functions of the simulated and random samples.

	$D-$value	$p-$value	Different
HBond	0.44	0.000125	Yes
Helices	0.36	0.003068	Yes
Strands	0.4	0.000671	Yes
Turns	0.54	9.311×10^{-7}	Yes
Res in helices	0.42	0.000295	Yes
Res in strands	0.48	1.99×10^{-5}	Yes
Longest helix	0.2	0.270	No
Longest strand	0.32	0.01195	Yes

The $p-$value is calculated by $Q_{KS}\left(\sqrt{\frac{N_1N_2}{N_1+N_2}}D\right)$, where $Q_{KS}(\lambda)=2\sum_{j=1}^{\infty}(-1)^{j-1}e^{-2j^2\lambda^2}$. In our case $N_1=N_2=50$. Yes or no is according to whether or not $p<\alpha=0.05$.

We compare the two kinds of simulations in following aspects: 1, number of hydrogen bonds; 2, number of helices; 3, number of strands; 4, number of turns; 5, number of residues in helices; 6, number of residues in strands; 7, length of the longest helices; 8, length of the longest strands. Statistical summaries of these 8 items are listed in Tables 1 and 2 above.

Since we do not know the statistical distributions of these samples, we use the two-sample Kolmogorov–Smirnov test that neither requires the distributions to be known nor that the distributions be normal, nor the variances be equal (Press et al., 1992, pp. 623–625).

For each of the 8 characters listed above, the null hypothesis is that there is no significant difference between the two kinds of simulations, while the alternative hypothesis is that they are significantly different. The significance level is set to be $\alpha = 0.05$.

A summary of testing results is given in Table 3. We see that except for the length of the longest helix having no difference, all tests show that there are significant differences between the two kinds of simulation structures.

7. Conclusion

A continuous free energy formula for globular proteins and a mathematical model based on it are discussed. Hydrophobic effect plays an essential role. Its main effect, hydrophobic core, is quantified by hydrophobic surface area $W(P)$. The pairwise interactions among the atoms in a protein molecule are put into steric conditions (1) which play the important constraint role in the minimization process of the mathematical model. Another role of the steric conditions is to distinguish similar amino acid sequences.

All conformations are given as an all atom (including hydrogen atoms) CPK model $P = \cup_{i=1}^{N} B(\mathbf{x}_i, r_i)$, each atom is a ball $B(\mathbf{x}_i, r_i)$. Hence the model is not a coarse grained model.

Minimizing $W(P)$ shows that it is sufficient to produce statistically significant native like secondary structures and hydrogen bonds in the proteins simulated.

The free energy formula is independently derived both from the classical thermodynamic formula for cavities in water plus the hydrophobic effect; and from geometric consideration of pursuing the conformation that fits best the global geometric features of native structures of globular proteins. The free energy formula achieved by both considerations, thermodynamics and geometry, are coincidentally identical. This is not purely coincidence.

Acknowledgments

Yi Fang sincerely thanks Jiri Novotny who introduced him to the protein folding problem. In simulations we used programs created by Andrew Butterfield. Xiaohan Fang helped us polish our English. We appreciate the suggestions and encouragement from the reviewers.

References

Almgren, F., 1986. Optimal isoperimetric inequalities. Indiana Univ. Math. J. 35 (3), 451–547.

Anfinsen, C.B., 1973. Principles that govern the folding of protein chains. Science 181, 223–230.

Bolen, D.W., Rose, G.D., 2008. Structure and energetics of the hydrogen-bonded backbone in protein folding. Annu. Rev. Biochem. 77, 339–362.

Branden, C., Tooze, J., 1998. Introduction to Protein Structure, second ed. Garland, New York.

Chan, H.S., Dill, K.A., 1989. Compact polymers. Macromolecules 22, 4559–4573.

Chan, H.S., Dill, K.A., 1990. Origins of structures in globular proteins. Proc. Natl. Acad. Sci. USA 87, 6388–6392.

Chandler, D., 2005. Interfaces and the driving force of hydrophobic assembly. Nature 437, 640–647.

Connolly, M.L., 1983. Analytical molecular surface calculation. J. Appl. Cryst. 16, 548–558.

Connolly, M.L., 2002. Molecular surfaces. 〈http://www.biohedron.com/〉.

Dill, K.A., 1990. Dominant forces in protein folding. Biochemistry 29, 7133–7155.

Dill, K.A., Ozkan, S.B., Shell, M.S., Weikl, T.R., 2008. The protein folding problem. Annu. Rev. Biophys. 37, 289–316.

Eisenberg, D., McLachlan, A.D., 1986. Solvation energy in protein folding and binding. Nature 319 (16), 199–203.

Fang, Y., 2005. Mathematical protein folding problem. In: Hoffman, D. (Ed.), Global Theory of Minimal Surfaces. Proceedings of the Clay Mathematical Proceedings, vol. 2, pp. 611–622.

Fang, Y., Jing, J., 2008. Implementation of a mathematical protein folding model. Int. J. Pure Appl. Math. 42 (4), 481–488.

Finkelstein, A.V., Ptitsyn, O.B., 2002. Protein Physics: A Course of Lectures. Academic Press, Amsterdam An imprint of Elsevier Science.

Hubner, I.A., Shakhnovic, E.I., 2005. Geometric and physical considerations for realistic protein models. Phys. Rev. E 72, 022901 10.1103/PhysRevE.72.022901.

Jackson, R.M., Sternberg, M.J.E., 1993. Protein surface area defined. Nature 366, 638.

Jackson, R.M., Sternberg, M.J.E., 1994. Application of scaled particle theory to model the hydrophobic effect: implications for molecular association and protein stability. Protein Eng. 7, 371–383.

Jackson, R.M., Sternberg, M.J.E., 1995. A continuum model for protein–protein interactions: application to the docking problem. J. Mol. Biol. 250, 258–275.

Janin, J., 1976. Surface area of globular proteins. J. Mol. Biol. 105, 13–14.

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

Lazaridis, T., Karplus, M., 1999. Effective energy function for proteins in solution. Proteins 35, 133–152.

Lazaridis, T., Karplus, M., 2003. Thermodynamics of protein folding: a microscopic view. Biophys. Chem. 100, 367–395.

Lee, B., Richards, F.M., 1971. The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. 55, 379–400.

Lesk, A.M., 2001. Introduction to Protein Architecture. Oxford University Press, Oxford.

Maritan, A., Micheletti, C., Trovato, A., Banavar, J.R., 2000. Optimal shapes of compact strings. Nature 406, 287–290.

Novotny, J., Bruccoleri, R., Karplus, M., 1984. An analysis of incorrectly folded protein models. Implications for structure predictions. J. Mol. Biol. 177, 787–818.

Novotny, J., Rashin, A.A., Bruccoleri, R., 1988. Criteria that discriminate between native proteins and incorrectly folded models. Proteins 4, 19–30.

Pauling, L., Corey, R.B., Branson, H.R., 1951. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. Proc. Natl. Acad. Sci. USA 37, 205–211.

Pippard, A.A., 1957. Classical Thermodynamics. Cambridge University Press, Cambridge.

Press, W.H., Teukolsky, S.A., Vetteriling, W.T., Fannery, B.P., 1992. Numerical Recipes in C: The Art of Scientific Computing, second ed. Oxford University Press, Oxford.

RasMol, 2009. Molecular visualization freeware. 〈http://www.umass.edu/microbio/rasmol/〉.

Richards, F.M., 1974. The interpretation of protein structures: total volume, group volume distributions and packing density. J. Mol. Biol. 82, 1–14.

Richards, F.M., 1977. Areas, volumes, packing, and protein structure. Ann. Rev. Biophys. Bioeng. 6, 151–176.

Richards, F.M., 1979. Packing defects, cavities, volume fluctuations, and access to the interior of proteins. Including some general comments on surface area and protein structure. Carlsberg Res. Commun. 44, 47–63.

Richards, F.M., Lim, W.A., 1994. An analysis of packing in the protein folding problems. Q. Rev. Biophys. 26, 423–498.

Southall, N.T., Dill, K.A., Haymet, A.D.J., 2002. A view of the hydrophobic effect. J. Phys. Chem. B 106, 521–533.

Tanford, C., 1978. Hydrophobic effect and the organization of living matter. Science 200, 1012–1018.

Tanford, C., Reynolds, J., 2001. Nature's Robots: A history of Proteins. Oxford University Press, Oxford.

Thewlis, J., (Ed.), 1973. Concise Dictionary of Physics. Pergamon Press, Oxford.

Thomas, E.L., Anderson, D.M., Henkee, C.S., Hoffman, D., 1988. Periodic area-minimizing surfaces in block copolymers. Nature 334, 598–601.

Tuñón, I., Silla, E., Pascual-Ahuir, J.L., 1992. Molecular surface area and hydrophobic effect. Protien Eng. 5 (8), 715–716.