Ben-Naim's ``Pitfall'': Don Quixote's Windmill

Article *in* Open Journal of Biophysics · January 2013

DOI: 10.4236/ojbiphy.2013.31002

1 author:

Yi Fang
Australian National University
33 PUBLICATIONS **153** CITATIONS

SEE PROFILE

Open Journal of Biophysics, **2013, 3, 13-21**
http://dx.doi.org/10.4236/ojbiphy.2013.31002 Published Online January 2013 (http://www.scirp.org/journal/ojbiphy)

Ben-Naim's "Pitfall": Don Quixote's Windmill

Yi Fang
Department of Mathematics, Nanchang University, Nanchang, China
Email: yifang@ncu.edu.cn, yi.fang3@gmail.com

Received October 7, 2012; revised November 15, 2012; accepted November 23, 2012

ABSTRACT

Ben-Naim in three articles dismissed and "answered" the Levinthal's paradox. He announces there are pitfalls caused by the "misinterpretation" of thermodynamic hypothesis. He claims no existence of Gibbs free energy formula $G(\boldsymbol{X})$ where the variable is a protein's conformation $\boldsymbol{X}$. His Gibbs energy functional is $G(T,P,N;P(\boldsymbol{R}))$, where the variable is probability distributions $P(\boldsymbol{R})$ of the conformations. His "minimum distribution P_{eq}" is wrong. By carefully establishing thermodynamic systems, we demonstrate how to apply quantum statistics to derive Gibbs free energy formula $G(\boldsymbol{X})$. The formula of the folding force $-\nabla G(\boldsymbol{X})$ is given.

Keywords: Protein Folding; Gibbs Free Energy; Quantum Mechanics; Statistical Mechanics

1. Introduction

In [1], Levinthal pointed out that assuming a protein folds by randomly searching its native structure it will need time longer than the age of the universe to achieve its native structure. Based on this contradiction, he then concluded that the natural protein folding must be cause-based, that is, the native structure has the (local) minimum value of the Gibbs free energy. Because of too involved in the random thinking of target-based mentality, many people would not understand the proof by contradiction of Levinthal's mathematical style argument. Instead, they felt that there exists a Levinthal's paradox that Levinthal never raised. In [2] Ben-Naim dismissed the so called Levinthal's paradox.

But Ben-Naim invents a new "pitfall": "This misinterpretation (of thermodynamic hypothesis) has inspired many scientists to search for a global minimum in the Gibbs energy as a function of the conformation of the protein, sometimes referred to as the Gibbs energy landscape. Such a minimum in the Gibbs energy is different from the minimum required by the Second Law of Thermodynamics" [3].

Trying to answer the so called Levinthal's paradox in [4] Ben-Naim gives the following inference:

"The following two statements are true:

a) The native stable structure of the protein must be at a minimum of the GEL (Gibbs Energy Landscape).

b) Upon releasing a constraint within the system, specified by the variables: *T*, *P* ***N***, the Gibbs energy of the system will reach a single absolute minimum".

Ben-Naim's conclusion is: "From the two true statements a) and b), people have concluded that the stable state of the protein must be in a global minimum in the GEL. Unfortunately, this conclusion is invalid... The reason so many people fell into this pitfall is that in making statements a) and b), we have not specified the variables with respect to which the Gibbs energy has a minimum".

Here Ben-Naim implies that conformation of a protein should not be the variable of the Gibbs energy. To answer the question of what is the variable in the Gibbs energy Ben-Naim states in [4]: "For a system characterized by the variables *T*, *P* and ***N***", (respectively the temperature, pressure, and the number of particles) "we can write the Gibbs energy function of the system as $G(T,P,N;\boldsymbol{R})$. If we start with a system having one particle at a fixed position, say $\boldsymbol{R}=\boldsymbol{R}_0$, then releasing the constraint on ***R***, but keeping *T*, *P*, and ***N*** fixed, the system's Gibbs energy will always decrease by the amount:

$$\Delta G = k_B T \ln \rho\Lambda^3 < 0\text{".}$$

So Ben-Naim confirms here that the variable of the Gibbs energy is not conformation ***R***. In [4], Ben-Naim continues to state the variable should be probability distributions *P* of the conformations: "Note again that $G(T,P,N;\boldsymbol{R})$ is not a monotonic decreasing function of *R*, and that there exists no value of *R*, for which *G* is minimal. Instead, the *functional* $G(T,P,N;P(\boldsymbol{R}))$ has a single minimum with respect to all possible distributions $P(\boldsymbol{R})$. The distribution $P_{eq}(\boldsymbol{R})$, for which *G* is minimal, is given in Equation (1)". Ben-Naim's Equation

(1) in [4] is as follows:

$$P_{eq} \to P_{eq}(\boldsymbol{R}) = \frac{1}{V}, \text{ for any } \boldsymbol{R} \qquad (1)$$

Unfortunately, Ben-Naim's solution of the single minimum (maximum) distribution P_{eq} at equilibrium is wrong, either for the Gibbs energy functional $G(T,P,N;P(\boldsymbol{R}))$ or for the entropy function $S(U,V,N;P(\boldsymbol{R}))$ in [4]. Because in physiological environment, almost all proteins are in the native structure, *i.e.*, the native structure has much larger opportunity to appear than any other conformation.

But even someone can get a correct minimum distribution for Ben-Naim, Ben-Naim's shifting from statement (a) to statement (b) is still a misleading, or a real pitfall. Because it shifts the study of protein structure to the study of probability distribution of conformations. The two are different problems and answer to one would not automatically solve the other problem. For example, even knowing what is Ben-Naim's minimum distribution, we still do not known what is the three-dimensional shape of the native structure.

In this article, why Ben-Naim falls into a pitfall is analyzed. We will also demonstrate how to derive Gibbs free energy formula $G(\boldsymbol{X})$ from quantum statistics to show how to get out of Ben-Naim's "pitfall", where we have omitted the environment parameters *T* and *P*, since they do not vary in nature protein folding process. Where $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_i, \cdots, \boldsymbol{x}_M) \in \boldsymbol{R}^{3M}$ is a conformation of the protein $\mathfrak{U}$, equivalent to Ben-Naim's $\boldsymbol{R}$, and $\boldsymbol{x}_i \in \mathbb{R}^3$ is the atomic center of the atom $\boldsymbol{a}_i$, supposing that the molecule has total *M* atoms. Denying the existence of such $G(\boldsymbol{X})$ (it is equivalent to Ben-Naim's $G(T,P,1;R)$) is one of the reasons that Ben-Naim claims "pitfall". The negative gradient $-\nabla G$, is the force that forces the portein to fold. Formulas of ∇G are given. The details of the derivation of $G(\boldsymbol{X})$ is given in Section 7.

2. Where Comes the "Pitfall"

To analysize Ben-Naim's "pitfall" and look for the reason why there is a "pitfall", we should recall what is the thermodynamic principle (Anfinsen called it modestly the thermodynamic hypothesis in [5]). Anfinsen stated in [5] clearly that "This hypothesis states that 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"; What did Anfinsen mean by the "whole system"? It seems from beginning to present, nobody has really specified it. But all assume that in it there are many conformations of the same protein molecule among other things. Ben-Naim's molecule number *N* is no exception. But look at what Anfinsen continued in [5]: "That is, that the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment". Here without any ambiguity the "totality" is "interatomic interactions" of a single protein molecule. Unfortunately, nobody really paid attention to these.

All previous attempts of deriving the Gibbs free energy formula, including Ben-Naim's, missed the goal of identifying "the three-dimensional structure of a native protein" that Anfinsen had emphasized in above quotation. By their derivation, the whole system consists of $N \geq 1$ conformations of the same protein molecule, each is only a point in the $\mathbb{R}^{3M}$ Euclidean space, supposing that the protein has *M* atoms. Each micro state of the system, the *N* points in $\mathbb{R}^{3M}$, is structureless if we consider the three-dimensional conformation. In this kind of treatment, statistical mechanics cannot tell us anything about "the three-dimensional conformation of a native protein". Once realized this, one should stop using such kind of systems and start to look for systems that can answer the problem of what is the three-dimensional shape of the native structure.

But many just followed the standard setting of statistical mechanics that successfully treated objects such as ideal gas. Instead of telling "the three-dimensional conformation of a native protein", they shift the problem to that what is the share of the native structure in the probability distribution of conformations. This problem is also interesting and important, but it is a different problem, and as afore mentioned, its resolution tells us nothing about "the three-dimensional conformation of a native protein". One has to be careful when making inferences between these two different problems. Ben-Naim's "pitfall" comes exactly from the misplaced inference, *i.e.*, even knowing what is the correct "minimum distribution P_{eq}" (Ben-Naim's is wrong) would not help us to know what is "the three-dimensional conformation of a native protein", not even one iota.

Our understanding of the thermodynamic principle is that under the physiological environment, for each conformation $\boldsymbol{X}$ of the peptide chain of the protein molecule $\mathfrak{U}$ there is a Gibbs free energy $G(\boldsymbol{X})$. The native structure $\boldsymbol{X}_N$ has the minimum value of this Gibbs free energy function G. The only uncertainty is that $\boldsymbol{X}_N$ might just correspond a local minimum of G, as asserted by Levinthal in [1]. Then the initial conformation $\boldsymbol{X}_I$ becomes important, because it will determine which local minimum conformation is the native structure $\boldsymbol{X}_N$.

But, to answer the question of what is "the three dimensional conformation of a native protein"? as Anfinsen emphasized, we have to make the transition of conformations in $\mathbb{R}^{3M}$ to conformations in $\mathbb{R}^3$. Based on the

three-dimensional geometry of each conformation $\boldsymbol{X}$, a thermodynamic system $\mathcal{T}_X \subset \mathbb{R}^3$ should be established, in which among other particles, contain exactly only one protein molecule with the conformation $\boldsymbol{X}$. Then one can apply statistical mechanics, classical or quantum, to get the Gibbs free energy of the system $\mathcal{T}_X$, denoted as $G(\boldsymbol{X})$.

Kinetically, in the physiological environment, an individual protein molecule takes an initial conformation $\boldsymbol{X}_1$ with a Gibbs free energy $G(\boldsymbol{X}_1)$. With the totality of interatomic interactions of the protein molecule, (we have to add that plus the interaction with its immediate environment), the conformation changes to a series conformations $\boldsymbol{X}_i$, with Gibbs free energy $G(\boldsymbol{X}_i)$. At last the conformation changes to the native structure $\boldsymbol{X}_N$ with $G(\boldsymbol{X}_N) \le G(\boldsymbol{X}_i)$. The "whole" system is the series of systems $\mathcal{T}_{X_i}$ in (time) series. Searching the native structure $\boldsymbol{X}_N$ then becomes the mathematical problem of solving the minimization problem

$$G(\boldsymbol{X}_N) = \min_{\text{All } \boldsymbol{X}} G(\boldsymbol{X}) \tag{1}$$

The solution of (1) will not only tell us what is the value $G(\boldsymbol{X}_N) = \min_{\text{All } X} G(\boldsymbol{X})$ (which is not important) but also will tell us what is $\boldsymbol{X}_N$ (which is the most important). This is one way to answer the question that what is "the three dimensional conformation of a native protein", *i.e.*, making protein structure prediction.

So if we want to resolve the protein folding problem (PFP), for any individual conformation $\boldsymbol{X}$ we should create a tailored thermodynamic system $\mathcal{T}_X$ and derive from it the Gibbs free energy formula $G(\boldsymbol{X})$. Given a native protein's amino acid sequence, searching for global minimum of $G(\boldsymbol{X})$ is truly following the thermodynamic hypothesis as Anfinsen stated it. Unable to derive such $G(\boldsymbol{X})$ should not be labeled as "misinterpretation of the (thermodynamic) hypothesis" [3]. Lacking of Gibbs free energy function $G(\boldsymbol{X})$ explains the question in [4]: "why an answer to this problem (PFP) has been elusive for so long". The fact that many, including Ben-Naim, in trying to establishing $G(\boldsymbol{X})$ have shifted the variable of G from $\boldsymbol{X}$, the conformation, to $P(\boldsymbol{X})$, the probability distribution of $\boldsymbol{X}$s, partially explains that why for so long such formula $G(\boldsymbol{X})$ has not been discovered. In particular, one common point of all previous theoretical treatment of protein folding is setting the thermodynamic system contains $N \ge 1$ copies of the same protein molecule, for example [6], thus failed to obtain $G(\boldsymbol{X})$.

On the other hand, since 1990's many techniques for probing individual molecules were developed and experimentally observing and testing single molecule is currently a common practice, see [7,8] for example. Theory anyway should not lagged too far behind experiment in single molecule protein folding study.

3. Thermodynamic System $\mathcal{T}_X$ and the Gibbs Free Energy Formula $G(\boldsymbol{X})$

3.1. The Systems

The thermodynamic system $\mathcal{T}_X$ occupies a region in $\mathbb{R}^3$. Given $\boldsymbol{X} \in \mathbb{R}^{3M}$, how to put it into a space region $\mathcal{T}_X \subset \mathbb{R}^3$? And actually, what is $\mathcal{T}_X$? To resolve this we have to use $\boldsymbol{X}$'s three dimensional structure. Assume that each atom has the shape of a ball with van der Wals radius r_i, $B(\boldsymbol{x}_i, r_i) \subset \mathbb{R}^3$, the three dimensional structure of $\boldsymbol{X}$ is $P_X = \bigcup_{i=1}^{M} B(\boldsymbol{x}_i, r_i) \subset \mathbb{R}^3$. The $\mathbb{R}^3$ is the real space or *behavior space* while the $\mathbb{R}^{3M}$ is only the *control space* of the protein conformation, [9].

To establish $\mathcal{T}_X$ we need some geometric preparation, although it may sounds too mathematical, it is no surprise at all. In fact, Anfinsen stated as early as in 1973 that "biological function appears to be more a correlate of macromolecular geometry than of chemical detail" [5]. Unfortunately, so far, nobody has taken it seriously.

Although the shape of each atom in $\mathfrak{U}$ is well defined by the theory of atoms in molecules [9,10], what concerning us here is the overall shape of the structure P_X. The cutoff of electron density $\rho \ge 0.001$ au [9,10], gives the overall shape of a molecular structure that is just like P_X, a bunch of overlapping balls. Moreover, the boundary of the $\rho \ge 0.001$ au cut off is almost the same as the molecular surface M_X (**Figure 1**) which was defined by Richards in 1977 [11] and was shown to be a more suitable boundary surface of P_X than other surfaces in 1992 and 1993 [12,13].

In mathematics, for any closed surface (compact and connected) $\Sigma \subset \mathbb{R}^3$, there are a bounded domain Ω_Σ and a un-bounded domain Ω'_Σ such that

$$\mathbb{R}^3 = \Omega_\Sigma \bigcup \Sigma \bigcup \Omega'_\Sigma, \ \partial\Omega_\Sigma = \partial\Omega'_\Sigma = \Sigma \tag{2}$$

Let d_w be the diameter of a water molecule and M_X be the molecular surface of P_X with the probe

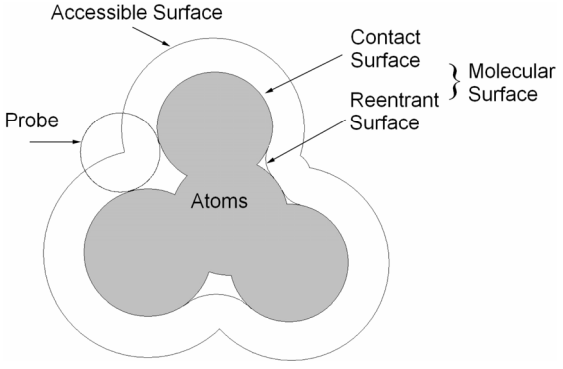

Figure 1. Two dimensional presenting of molecular surface [11] and solvent accessible surface [14]. This figure was originally in [15].

radius $d_w/2$. If M_X is connected, then we can use $\Sigma = M_X$ in (2). If M_X has multiple connected components S_i, $1 \le i \le m$, such that S_1 is the largest component, *i.e.*, all other components of M_X are contained in Ω_{S_1}. Then denote $\Omega_X = \Omega_{S_1} \bigcap \left(\bigcap_{i \ne 1} \overline{\Omega'_{S_i}} \right)$ and $\Omega'_X = \Omega'_{S_1} \bigcup \left(\bigcup_{i \ne 1} \Omega_{S_i} \right)$. Thus, we always have

$$\mathbb{R}^3 = \Omega_X \bigcup M_X \bigcup \Omega'_X, \ \partial\Omega_X = \partial\Omega'_X = M_X \tag{3}$$

Let $dist(\boldsymbol{x}, C) = \inf_{\boldsymbol{y} \in C} |\boldsymbol{x} - \boldsymbol{y}|$ be the distance from a point $\boldsymbol{x}$ to a subset $C \subset \mathbb{R}^3$. Define

$$\mathcal{T}_X = \{ \boldsymbol{x} : \text{dist}(\boldsymbol{x}, \Omega_X) \le d_w \} \tag{4}$$

as our thermodynamic system. While

$$\mathcal{R}_X = \mathcal{T}_X / \Omega_X \tag{5}$$

is the first hydration shell surrounding P_X.

To be simple, we only consider single peptide chain, self-folding globular proteins here. Hence in the system $\mathcal{T}_X$, except P_X, there are only water molecules and electrons. We have $P_X \subset \bar{\Omega}_X = \Omega_X \bigcup M_X$ and all nuclear centers of water molecules in $\mathcal{T}_X$ are contained in $\mathcal{R}_X$. Moreover, since Ω_X is bounded, it has a finite volume $V(\boldsymbol{X}) = V(\Omega_X)$.

The thermodynamic system $\mathcal{T}_X$ will be an open system, *i.e.*, electrons and water molecules can enter and leave $\mathcal{T}_X$. Therefore, the numbers N and N_e, of water molecules and electronics in $\mathcal{T}_X$ are variables. According Anfinsen [5], the protein folding process is after the peptide chain synthesis. Therefore, part of the totality of the "interatomic interactions", as emphasized by Anfinsen in [5], has already contributed to form correct chemical bonds. In the folding process, "chemical details" may be represented by the forming of intra-molecular hydrogen bonds and the interactions with the immediate environment, in our case, the solvent consisting of water molecules.

Ben-Naim claims that "in the author's opinion, the main hindrance to finding a solution to the protein folding problem has been the adherence to the hydrophobic (HOO) dogma, which states that various HOO effects (both solvation and interaction) are the dominant forces in protein folding" and "an exhaustive analysis of all the solvent induced effects on protein folding reveals that the hydrophilic (HOI) effects are much more important than the corresponding HOO effects" [2].

In [15] a simulation of enlarging the hydrophobic core alone, whose forming is considered the main effect of HOO, not only produced secondrary structures, but also produced the intra-molecular hydrogen bonds. This result shows that HOO should not be dismissed so simply.

But no matter the driving force of protein folding is HOO or HOI, a common essence for them is that in a protein there are many different moieties or atom groups with different levels of ability of forming hydrogen bonds (hydrophobic levels). Simply classifying amino acids as hydrophobic or hydrophilic is an over simplification [16]. In fact, since each atom belongs to a particular moiety or atom group, it can be assigned a hydrophobic level as the level of the moiety or atom group. Suppose we classify the atoms into H hydrophobic levels H_i, $i = 1, \cdots, H$, such that $\bigcup_{i=1}^{H} H_i = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_i, \cdots, \boldsymbol{a}_M)$. For example, in [16] there are $H = 5$ classes, C, O/N, O^-, N^+, S. If a hydrogen atom is bonded with an atom in H_i, we will put it in H_i.

Let $I_i \subset \{1, 2, \cdots, M\}$ be the subset such that $\boldsymbol{a}_j \in H_i$ if and only if $j \in I_i$. Define $P_{Xi} = \bigcup_{j \in I_i} B(\boldsymbol{x}_j, r_j) \subset P_X$ and as shown in **Figure 2**,

$$\mathcal{R}_{Xi} = \left\{ \boldsymbol{x} \in \mathcal{R}_X : \text{dist}(\boldsymbol{x}, P_{Xi}) \le \text{dist}(\boldsymbol{x}, P_X / P_{Xi}) \right\}, \quad 1 \le i \le H, \tag{6}$$

Let $V(\Omega)$ be the volume of $\Omega \subset \mathbb{R}^3$, then

$$\mathcal{R}_X = \bigcup_{i=1}^{H} \mathcal{R}_{Xi}, \quad V(\mathcal{R}_X) = \sum_{i=1}^{H} V(\mathcal{R}_{Xi}), \quad \text{and for } i \ne j, \ V(\mathcal{R}_{Xi} \bigcap \mathcal{R}_{Xj}) = 0. \tag{7}$$

Define the hydrophobicity subsurface M_{Xi}, $1 \le i \le H$, as

$$M_{Xi} = M_X \bigcap \mathcal{R}_{Xi}. \tag{8}$$

Let $A(\Sigma)$ be the area of a surface $\Sigma \subset \mathbb{R}^3$, then

$$M_X = \bigcup_{i=1}^{H} M_{Xi}, \quad A(M_X) = \sum_{i=1}^{H} A(M_{Xi}), \quad \text{and if } i \ne j, \text{ then } A(M_{Xi} \bigcap M_{Xj}) = 0. \tag{9}$$

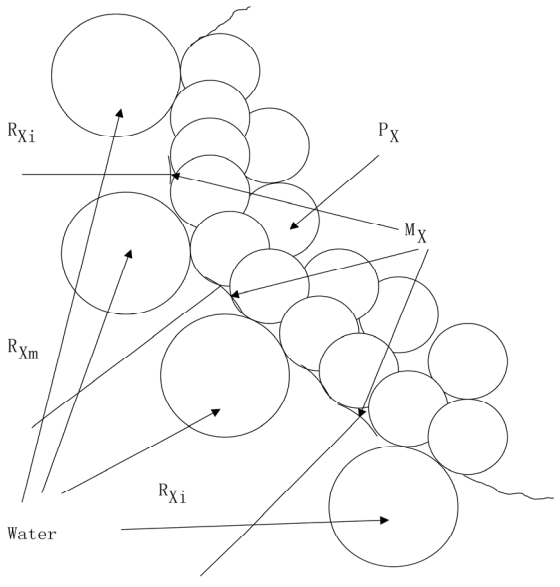

Figure 2. Note that R_{Xi} generally are not connected, *i.e.*, having more than one block.

3.2. The Formulas

In our open thermodynamic system, there will be N_i water molecules in $\mathcal{R}_{Xi}$, and N_e electrons in $\mathcal{T}_X$, thus we will denote the variable N as a vector

$$N=\left(N_e,N_1,\cdots,N_H\right).$$

After statistical treatment, the mean number of N_e and N_i will be denoted as $N_e(\boldsymbol{X})$, $N_i(\boldsymbol{X})$, $i=1,\cdots,H$. Each water molecule in $\mathcal{R}_{Xi}$ will contact M_{Xi}. The chemical potential reflecting the contacting energy will be denoted as μ_i. Similarly, the energy for an electron kept in $\mathcal{T}_X$ will be the chemical potential μ_e. With these preparation, arguing in quantum statistics via the grand canonic ensemble we derive the Gibbs free energy of protein folding $G(\boldsymbol{X})$ as follows (see Section 7 for the detailed derivation, also see [17,18] for further discussions):

$$G(\boldsymbol{X})=\mu_e N_e(\boldsymbol{X})+\sum_{i=1}^{H}\mu_i N_i(\boldsymbol{X}). \tag{10}$$

Note that in the folding process, each intermediate structure $\boldsymbol{X}$ is not in a stationary state, it is rather a system of quasi-equilibrium states of the folding. So that it is not the case that $\mu_i=\mu_j$, as in equilibrium state. Rather, the chemical potentials will be constants during the folding, as the environment is kept unchanged.

Formula (10) is not easy to calculate, we can convert it into a geometric form that is not only calculable but also coincident to a mathematically derived formula appeared in [15,19].

Since every water molecule in $\mathcal{R}_{Xi}$ has contact with the surface M_{Xi} and the curvature of M_X is uniformly bounded, $N_i(\boldsymbol{X})$ is proportional to the area $A(M_{Xi})$. That is, there are $\nu_i>0$, independent of $\boldsymbol{X}$, such that

$$\nu_i A(M_{Xi})=N_i(\boldsymbol{X}),1\le i\le H. \tag{11}$$

Similarly, there will be a $\nu_e>0$, independent of $\boldsymbol{X}$, such that $\nu_e V(\mathcal{T}_X)=N_e(\boldsymbol{X})$.

By the definition of $\mathcal{T}_X$ and Ω_X, we have roughly $V(\mathcal{T}_X/\Omega_X)=d_w A(M_X)$. Thus

$$\begin{aligned}N_e(\boldsymbol{X})&=\nu_e V(\mathcal{T}_X)=\nu_e\left[V(\Omega_X)+V(\mathcal{T}_X/\Omega_X)\right]\\&=\nu_e V(\Omega_X)+\nu_e d_w A(M_X).\end{aligned} \tag{12}$$

Substitute (11) and (12) into (10), we get

$$\begin{aligned}G(\boldsymbol{X})&=\nu_e\mu_e V(\Omega_X)+d_w\nu_e\mu_e A(M_X)\\&\quad+\sum_{i=1}^{H}\nu_i\mu_i A(M_{Xi}).\end{aligned} \tag{13}$$

This Gibbs free energy function $G(\boldsymbol{X})$ really should be written as $G(E_n;\boldsymbol{X})$, where E_n is environment, its parameters including the temperature T and pressure P which will affect the values of chemical potential μ_e and μ_i. Since protein folding is in a fixed physiological environment, we can omit E_n in this stage.

It should be emphasized here that since we assumed that the proteins are single peptide chain, self-folding globular proteinsins, the first hydration of P_X contains only water molecules and electrons, no presence of other components at all, this Gibbs free energy function $G(\boldsymbol{X})$ should be only suitable to these proteins. For other kinds of proteins, the presence of other components such as chaperonins must be considered in the thermodynamic system $\mathcal{T}_X$. Then, the geometry of $\mathcal{T}_X$ will become more complicated.

4. Applying and Testing the Thermodynamic Hypothesis

Anfisen had shown that the protein folding is a spantenously process [5], thus the thermodynamic hypothesis should be treated as thermodynamic principle. A direct application of it, also a real test of it, is the *ab initio* prediction of a protein's native structure as in (1). However, without control of overlapping of the balls $B(\boldsymbol{x}_i,r_i)$, we may get a single ball with all other balls collapsed in it as a minimum structure, a disaster for a prediction. The pairwise potentials used for force fields will prevent the collapsing happen. Why the pairwise potential energy among atoms of the protein $\mathfrak{U}$ does not show in formulae (10) and (13)? The reason is that according to Anfinsen [5], protein folding is after the synthesis of the whole peptide chain. So that during the folding process all covalent bonds in the main chain and each side chain are already formed and non-bonding atoms keep a certain distance from each other. That is, the potential energy has already played its role during the synthesis of the peptide chain. This reality forces us to restrict what $\boldsymbol{X}$ can be treated as a conformation, *i.e.*, a conformation should satisfy the steric conditions below.

There are $\varepsilon_{ij}>0$, $1\le i<j\le M$ such that for nuclear centers $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ in $\boldsymbol{X}$,

$$\begin{aligned}&\varepsilon_{ij}\le\left|\boldsymbol{x}_i-\boldsymbol{x}_j\right|,\ \text{no covalent bond between } \boldsymbol{a}_i \text{ and } \boldsymbol{a}_j;\\&d_{ij}-\varepsilon_{ij}\le\left|\boldsymbol{x}_i-\boldsymbol{x}_j\right|\le d_{ij}+\varepsilon_{ij},\\&d_{ij}\ \text{is the standard bond length between } \boldsymbol{a}_i \text{ and } \boldsymbol{a}_j.\end{aligned} \tag{14}$$

We will denote all conformations satisfying (14) as $\mathfrak{X}$. Then the minimization will become:

$$G(\boldsymbol{X}_N)=\inf_{\boldsymbol{X}\in\mathfrak{X}}G(\boldsymbol{X}), \tag{15}$$

or, at least, within $\mathfrak{X}$, $\boldsymbol{X}_N$ corresponds to a local minimum of G.

With the steric conditions we avoided the collapsing problem. But the steric conditions turn the minimization problem (1) into a constrained minimization problem (15). Mathematically the latter is much more difficult to

solve. To avoid the constraint in minimization for non-bonding atoms, we can use the van der Waals force to modify the formula as:

$$G_v(\boldsymbol{X}) = \nu_e\mu_e V(\Omega_{\boldsymbol{X}}) + d_w\nu_e\mu_e A(M_{\boldsymbol{X}}) + \sum_{i=1}^{H}\nu_i\mu_i A(M_{\boldsymbol{X}i}) + \sum_{\text{non-bonding } i,j} E_{ij}\left[\left(\frac{r_{ij0}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{ij0}}{r_{ij}}\right)^{6}\right], \tag{16}$$

where $E_{ij} > 0$ is the corresponding energy and $r_{ij} = |\boldsymbol{x}_i - \boldsymbol{x}_j|$ and r_{ij0} the ideal distance between the atoms $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$. Before using G_v to eliminate the constraint of $\boldsymbol{X} \in \mathfrak{X}$, we take a more convenient coordinate of the conformation $\boldsymbol{X}$. We require that all bond lengths and angles (denoted as one angle-length pattern) are kept as obtained from a conformation $\boldsymbol{X} \in \mathfrak{X}$ and from $\boldsymbol{X}$ calculate the values of all rotatable dihedral angles $\Phi = (\phi_1, \cdots, \phi_L) \in (-\pi, \pi)^L$ (including all the main chain Φ_i, Ψ_i s). In fact, new conformations obtained by changing Φ will keep the same angle-length pattern and all conformations with the same angle-length pattern as $\boldsymbol{X}$ are obtained by choose suitable Φ values. The function $G_v(\boldsymbol{X})$ then can be written as

$$G_v(\boldsymbol{X}) = G_v(\Phi) = G_v(\phi_1, \cdots, \phi_L); \quad \Phi \text{ is induced from } \boldsymbol{X} \in \mathfrak{X}. \tag{17}$$

Let $\boldsymbol{X}_N$ have the dihedral angles $\Phi^N = (\phi_1^N, \cdots, \phi_L^N)$, then the constraint in (15) will be relaxed and we will have a minimization problem without any constraint:

$$G_v(\boldsymbol{X}_N) = G_v(\Phi^N) = \inf_{\Phi \in \mathbb{R}^L \text{ induced from } \boldsymbol{X} \in \mathfrak{X}} G_v(\Phi). \tag{18}$$

5. The Force That Forces the Protein to Fold

Ben-Naim correctly emphasizes that the protein folding is a cause-based process, "One can imagine that at each stage of the folding process, there are strong solvent-induced forces exerted on the various groups along the protein. These forces will force the protein to fold along a narrow range of pathways..." [2], and the folding force actually is the negative of the gradient of the Gibbs free energy function, that is $-\nabla G$, "we need to know the forces acting on each of the M groups of the protein being at the conformation $\boldsymbol{R}^M$. This force is obtained by taking the gradient of the Gibbs energy with respect to each of the $\boldsymbol{R}_i$" [4].

However, with only a "minimum distribution P_{eq}" Ben-Naim cannot tell what is the garden ∇G. With formula (13), it is easy to write down mathematical formula of ∇G. For example, in the coordinates $\Phi \in \mathbb{R}^L$, the folding force is

$$-\nabla G_v(\Phi) = -\left(\frac{\partial G_v}{\partial \phi_1}, \cdots, \frac{\partial G_v}{\partial \phi_i}, \cdots, \frac{\partial G_v}{\partial \phi_L}\right). \tag{19}$$

5.1. Newton's Fastest Descending Method

Before giving the formula of ∇G_v, we will point out that if it is calculable, then we can apply the fastest descending method to pursue the minimum value of $G_v(\Phi)$. That is, starting from a Φ_0, the immediate next conformation Φ_t will be chosen such that

$$\Phi_t = \Phi_0 - t\nabla G_v(\Phi_0), \tag{20}$$

where $t > 0$ is a suitable step length. When t is small, it is guaranteed that $G(\Phi_t) < G(\Phi_0)$. Any (local) minimum $\boldsymbol{X}_m$ would have that $\nabla G(\boldsymbol{X}_m) = \boldsymbol{0}$.

5.2. The Formula of ∇G_v

We will give the analytic formula of $\nabla G_v(\boldsymbol{X})$ here without mathematical proof. It is:

$$\frac{\partial G_v}{\partial \phi_i}(\boldsymbol{X}) = \nu_e\mu_e \frac{\partial V}{\partial \phi_i}(\Omega_{\boldsymbol{X}}) + d_w\nu_e\mu_e \frac{\partial A}{\partial \phi_i}(M_{\boldsymbol{X}}) + \sum_{j=1}^{H}\nu_j\mu_j \frac{\partial A}{\partial \phi_i}(M_{\boldsymbol{X}j}) + \sum_{\text{non-bonding } j,k} E_{jk}\frac{\partial}{\partial \varphi_i}\left[\left(\frac{r_{jk0}}{r_{jk}}\right)^{12} - 2\left(\frac{r_{jk0}}{r_{jk}}\right)^{6}\right]. \tag{21}$$

It should be mentioned here that bond in $\mathfrak{U}$ is rotatable if it is a single bond and if we cut this bond, all nuclear centers in $\boldsymbol{X}$ can be divided into two (non-empty) groups, such that we can fix one group and rotate around the bond axis the other group. Let $\wedge$ be the outer product in $\mathbb{R}^3$. Let $\boldsymbol{x}_i\boldsymbol{y}_i$ be the bond, then $\boldsymbol{b}_i = (\boldsymbol{x}_i - \boldsymbol{y}_i)/|\boldsymbol{x}_i - \boldsymbol{y}_i|$ will be the rotation axis and $\boldsymbol{L}_i$ the rotation vector field, *i.e.*, $\boldsymbol{L}_i(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{y}_i) \wedge \boldsymbol{b}_i$ if $\boldsymbol{x}$ is a rotated nuclear center; and $\boldsymbol{L}_i(\boldsymbol{x}) = \boldsymbol{0}$ if $\boldsymbol{x}$ is a fixed nuclear center. Furthermore,

$$\frac{\partial V}{\partial \phi_i}(\Omega_{\boldsymbol{X}}) = -\int_{M_{\boldsymbol{X}}} \boldsymbol{L}_i \cdot \boldsymbol{N} \mathrm{d}\mathcal{H}^2, \quad \frac{\partial A}{\partial \phi_i}(M_{\boldsymbol{X}}) = -2\int_{M_{\boldsymbol{X}}} H(\boldsymbol{L}_i \cdot \boldsymbol{N}) \mathrm{d}\mathcal{H}^2, \tag{22}$$

and

$$\frac{\partial A}{\partial \phi_i}(M_{\boldsymbol{X}j}) = -2\int_{M_{\boldsymbol{X}j}} H(\boldsymbol{L}_i \cdot \boldsymbol{N}) \mathrm{d}\mathcal{H}^2 + \int_{\partial M_{\boldsymbol{X}j}} \left[\boldsymbol{L}_i \cdot \boldsymbol{\eta} - \frac{\mathrm{d}f_{0,j}}{|\nabla_{M_{\boldsymbol{X}}} f_{0,j}|}\right] \mathrm{d}\mathcal{H}^1, \tag{23}$$

where $\boldsymbol{N}$ and H are the outer unit normal and the mean curvature of $M_{\boldsymbol{X}}$, $\mathcal{H}^2$ and $\mathcal{H}^1$ the Hausdorff measures of dimensions 2 and 1. Let $\boldsymbol{X}^t$ be the family of conformations such that $\boldsymbol{X}^0 = \boldsymbol{X}$ and $\boldsymbol{x}_k^t = \boldsymbol{x}_k + t\boldsymbol{L}_i(\boldsymbol{x}_k)$, $k = 1, \cdots, M$. Define $f_{t,j} : \mathbb{R}^3 \to \mathbb{R}$ as $f_{t,j}(\boldsymbol{x}) = \mathrm{dist}(\boldsymbol{x}, M_{\boldsymbol{X}^t j}) - \mathrm{dist}(\boldsymbol{x}, M_{\boldsymbol{X}^t}/M_{\boldsymbol{X}^t j})$, and

denote

$$\nabla_{M_X} f_{0,j} = \nabla f_{0,j} - \left(\nabla f_{0,j} \cdot \boldsymbol{N}\right)\boldsymbol{N},\ f'_{0,j} = \left.\frac{\partial f_{t,j}}{\partial t}\right|_{t=0},$$
$$\frac{\mathrm{d}f_{0,j}}{\mathrm{d}t} = \boldsymbol{L}_i \cdot \nabla_{M_X} f_{0,j} + f'_{0,j}, \tag{24}$$

Finally, if $\boldsymbol{x}_k$ is rotated and $\boldsymbol{x}_j$ is fixed, then

$$\frac{\partial r_{kj}}{\partial \phi_i} = \frac{\left(\boldsymbol{x}_k - \boldsymbol{x}_j\right)\cdot \boldsymbol{L}_i\left(\boldsymbol{x}_k\right)}{r_{kj}}, \tag{25}$$

if $\boldsymbol{x}_k$ and $\boldsymbol{x}_j$ are both rotated or both fixed, then we have $\frac{\partial r_{kj}}{\partial \phi_i} = 0$.

The integration of above formulae on the molecular surface M_X are given in [20].

6. Conclutions

The Ben-Naim's pitfall of "misinterpretation of thermodynamic hypotheses" is dismissed as a Don Quixote's windmill by demonstrating the existence of Gibbs free energy formulas (10) and (13), pursuing of them were claimed by Ben-Naim as fallen into a pitfall. The formulae themselves need detailed geometric formulation of the thermodynamic system to present them, is a realization of Anfinsen's insight that "biological function appears to be more a correlate of macromolecular geometry than of chemical detail" [5]. Contrary to Ben-Naim's claims that "In the author's opinion, the main hinderence to finding a solution to the protein folding problem has been the adherence to the hydrophobic (HOO) dogma" [2], the derivation of (10) and (13) heavily depends on the concept of hydrophobicity.

In Section 7, the quantum statistical derivation of formula (10) is given, the convertion of (10) to (13) is demonstrated in Section 3.2.

Ben-Naim's minimization at $P_{eq}\left(\boldsymbol{R}\right)$ is analyzed and dismissed because it predicts that at equilibrium every possible conformation $\boldsymbol{R}$ will have the same probability to be the structure of a native protein. That is, Ben-Naim claims that $P_{eq}\left(\boldsymbol{R}\right) = 1/V$ for any conformation $\boldsymbol{R}$. In fact, in the contrary, in the physiological environment the native structure is dominate.

The reason of why calculable formulas such as (10) and (13) have not appeared so far is discussed, blindly imitating successful classical examples of applying statistical mechanics and ignoring Anfinsen's insight are two main reasons.

The force that forces the protein to fold is identified as $-\nabla G\left(\boldsymbol{X}\right)$ by general physical law, that Ben-Naim has correctly pointed out. The calculable formula of ∇G is given.

7. Derivation of Formula (10)

7.1. The Shrödinger Equation

For any conformation $\boldsymbol{X} \in \mathfrak{X}$, let $\boldsymbol{W} = \left(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_i, \cdots, \boldsymbol{w}_N\right) \in \mathbb{R}^{3N}$ be the nuclear centers of oxygen atoms in water molecules in $\boldsymbol{\mathcal{R}}_X$ and $\boldsymbol{E} = \left(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_i, \cdots, \boldsymbol{e}_L\right) \in \mathbb{R}^{3L}$ be electronic positions of all electrons in $\boldsymbol{\mathcal{T}}_X$. Then the Hamiltonian for the system $\boldsymbol{\mathcal{T}}_X$ is:

$$\hat{H} = \hat{T} + \hat{V} = -\sum_{i=1}^{M}\frac{\hbar^2}{2m_i}\nabla_i^2 - \frac{\hbar^2}{2m_w}\sum_{i=1}^{N}\nabla_i^2 - \frac{\hbar^2}{2m_e}\sum_{i=1}^{L}\nabla_i^2 + \hat{V}\left(\boldsymbol{X}, \boldsymbol{W}, \boldsymbol{E}\right), \tag{26}$$

where m_i is the nuclear mass of atom $\boldsymbol{a}_i$ in $\mathfrak{U}$, m_w and m_e the masses of water molecule and electron, ∇_i^2 the Laplacian in corresponding $\mathbb{R}^3$, and V the potential.

7.2. The First Step of the Born-Oppenheimer Approximation

Depending on the shape of P_X, for each i, $1 \le i \le H$, the maximum numbers N_{Xi} of water molecules contained in $\boldsymbol{\mathcal{R}}_{Xi}$ vary. Theoretically we consider all cases, *i.e.*, there are $0 \le N_i \le N_{Xi}$ water molecules in $\boldsymbol{\mathcal{R}}_{Xi}$, $1 \le i \le H$. Let $M_0 = 0$ and $M_i = \sum_{j \le i} N_j$ and $\boldsymbol{W}_i = \left(\boldsymbol{w}_{M_{i-1}+1}, \cdots, \boldsymbol{w}_{M_{i-1}+j}, \cdots, \boldsymbol{w}_{M_i}\right) \in \mathbb{R}^{3N_i}$, $1 \le i \le H$, and $\boldsymbol{W} = \left(\boldsymbol{W}_1, \boldsymbol{W}_2, \cdots, \boldsymbol{W}_{M_H}\right) \in \mathbb{R}^{3M_H}$ denote the nuclear positions of water molecules in $\boldsymbol{\mathcal{R}}_X$. As well, there will be all possible numbers $0 \le N_e < \infty$ of electrons in $\boldsymbol{\mathcal{T}}_X$. Let $\boldsymbol{E} = \left(\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_{N_e}\right) \in \mathbb{R}^{3N_e}$ denote their nuclear positions. For each fixed $\boldsymbol{X}$ and $N = \left(N_1, \cdots, N_H, N_e\right)$, the Born-Oppenheimer approximation has the Hamiltonian

$$\hat{H}_X = -\frac{\hbar^2}{2}\left\{\frac{1}{m_w}\sum_{j=1}^{M_H}\nabla_j^2 + \frac{1}{m_e}\sum_{\nu=1}^{N_e}\nabla_\nu^2\right\} + \hat{V}\left(\boldsymbol{X}, \boldsymbol{W}, \boldsymbol{E}\right).$$

The eigenfunctions $\psi_i^{X,N}\left(\boldsymbol{W}, \boldsymbol{E}\right) \in L_0^2\left(\prod_{i=1}^{H}\boldsymbol{\mathcal{R}}_{Xi}^{N_i} \times \boldsymbol{\mathcal{T}}_X^{N_e}\right) = \mathcal{H}_{X,N}$, $1 \le i < \infty$, comprise an orthonormal basis of $\mathcal{H}_{X,N}$. Denote their eigenvalues (energy levels) as $E_{X,N}^i$, then

$$\hat{H}_X \psi_i^{X,N} = E_{X,N}^i \psi_i^{X,N}.$$

7.3. Grand Partition Function and Grand Canonic Density Operator

In the following we will use the natotions and definitions in [21, Chapter 10]. Let k_B be the Bolzmman constant, set $\beta = 1/k_B T$. Since the numbers N_i and N_e vary, we should adopt the grand canonic ensemble. Let μ_i be the chemical potentials, that is, the Gibbs free energy per water molecule in $\boldsymbol{\mathcal{R}}_{Xi}$. Let μ_e be electron chemical

potential. The grand canonic density operator is ([21, 22])

$$\hat{\rho}_X = \exp\left\{-\beta\left[\hat{H}_X - \sum_{i=1}^{H}\mu_i\hat{N}_i - \mu_e\hat{N}_e - \Omega(\boldsymbol{X})\right]\right\}.$$

where the grand partition function is

$$\begin{aligned}&\exp\left[-\beta\Omega(\boldsymbol{X})\right]\\ &= \text{Trace}\left\{\exp\left[-\beta\left(\hat{H}_X - \sum_{i=1}^{H}\mu_i\hat{N}_i - \mu_e\hat{N}_e\right)\right]\right\}\\ &= \sum_{i,N} \mathrm{e}^{-\beta\left[E^i_{\boldsymbol{X},N} - \sum_{i=1}^{H}\mu_i N_i - \mu_e N_e\right]}.\end{aligned}$$

7.4. The Gibbs Free Energy $G(\boldsymbol{X})$

According to [21, p. 273], under the grand canonic ensemble the entropy $S(\boldsymbol{X}) = S(\mathcal{T}_X)$ of the system $\mathcal{T}_X$ is

$$\begin{aligned}S(\boldsymbol{X}) &= -k_B\text{Trace}(\hat{\rho}_X \ln\hat{\rho}_X) = -k_B\langle\ln\hat{\rho}_X\rangle\\ &= k_B\beta\left\langle\hat{H}_X - \Omega(\boldsymbol{X}) - \sum_{i=1}^{H}\mu_i\hat{N}_i - \mu_e\hat{N}_e\right\rangle\\ &= \frac{1}{T}\left[\left\langle\hat{H}_X\right\rangle - \langle\Omega(\boldsymbol{X})\rangle - \sum_{i=1}^{H}\mu_i\left\langle\hat{N}_i\right\rangle - \mu_e\left\langle\hat{N}_e\right\rangle\right]\\ &= \frac{1}{T}\left[U(\boldsymbol{X}) - \Omega(\boldsymbol{X}) - \sum_{i=1}^{H}\mu_i N_i(\boldsymbol{X}) - \mu_e N_e(\boldsymbol{X})\right].\end{aligned} \tag{27}$$

Here we denote $\left\langle\hat{N}_i\right\rangle = N_i(\boldsymbol{X})$ the mean numbers of water molecules in $\mathcal{R}_{Xi}$, $1 \le i \le H$, and $\left\langle\hat{N}_e\right\rangle = N_e(\boldsymbol{X})$ the mean number of electrons in $\mathcal{T}_X$. The inner energy $\langle\hat{H}_X\rangle$ of the system $\mathcal{T}_X$ is denoted as:

$$U(\mathbf{X}) = U(\mathcal{T}_X).$$

The term $\Omega(\boldsymbol{X})$ is a state function with variables $T, V, \mu_1, \cdots, \mu_H$, and μ_e, and is called the grand canonic potential ([21, p. 27]) or the thermodynamic potential ([22, p. 33]). By the general thermodynamic equations [22, pp. 5-6]:

$$\begin{aligned}d\Omega(\boldsymbol{X}) &= -SdT - PdV - \sum_{i=1}^{H}N_i d\mu_i - N_e d\mu_e,\\ \lambda\Omega(\boldsymbol{X}) &= \Omega(\boldsymbol{X})(T, \lambda V, \mu_1, \cdots, \mu_H, \mu_e),\end{aligned}$$

we see that

$$\Omega(\boldsymbol{X})(T, V, \mu_1, \cdots, \mu_H, \mu_e) = -PV(\boldsymbol{X}),$$

where $V(\boldsymbol{X}) = V(\mathcal{T}_X)$ is the volume of the thermodynamic system $\mathcal{T}_X$. Thus by (27) we obtain the Gibbs free energy $G(\boldsymbol{X}) = G(\mathcal{T}_X)$ in (10):

$$\begin{aligned}G(\boldsymbol{X}) &= G(\mathcal{T}_X) = PV(\boldsymbol{X}) + U(\boldsymbol{X}) - TS(\boldsymbol{X})\\ &= \sum_{i=1}^{H}\mu_i N_i(\boldsymbol{X}) + \mu_e N_e(\boldsymbol{X}).\end{aligned}$$

REFERENCES

[1] C. Levinthal, "How to Fold Graciously," *Mössbauer Spectroscopy in Biological Systems Proceedings*, Vol. 67, No. 41, 1969, pp. 22-26. http://www-miller.ch.cam.ac.uk/levinthal/levinthal.html

[2] A. Ben-Naim, "Levinthal's Paradox Revisited, and Dismissed," *Open Journal of Biophysics*, Vol. 2, No. 2, 2012, pp. 23-32. doi:10.4236/ojbiphy.2012.22004

[3] A. Ben-Naim, "Pitfalls in Anfinsen's Thermodynamic Hypothesis," *Chemical Physics Letters*, Vol. 511, No. 1-3, 2011, pp. 126-128. doi:10.1016/j.cplett.2011.05.049

[4] A. Ben-Naim, "Levinthal's Question Revisited, and Answered," *Journal of Biomolecular Structure and Dynamics*, Vol. 30, No. 1, 2012, pp. 113-124. doi:10.1080/07391102.2012.674286

[5] C. B. Anfinsen, "Principles that Govern the Folding of Protein Chains," *Science*, Vol. 181, No. 4096, 1973, pp. 223-230. doi:10.1126/science.181.4096.223

[6] I. Lazaridis and M. Karplus, "Thermodynamics of Protein Folding: A Microscopic View," *Biophysical Chemistry*, Vol. 100, No. 1-3, 2003, pp. 367-395. doi:10.1016/S0301-4622(02)00293-4

[7] A. Borgia, P. M. Williams and J. Clarke, "Single-Molecule Studies of Protein Folding," *Annual Review of Biochemistry*, Vol. 77, 2008, pp. 101-125. doi:10.1146/annurev.biochem.77.060706.093102

[8] J. Stigler, F. Ziegler, A. Gieseke, J. C. M. Gebhardt and M. Rief, "The Complex Folding Network of Single Calmodulin Molecules," *Science*, Vol. 334, No. 6055, 2011, pp. 512-516. doi:10.1126/science.1207598

[9] R. F. W. Bader, "Atoms in Molecules: A Quantum Theory," Clarendon Press, Oxford, 1990.

[10] P. Popelier, "Atoms in Molecules: An Introduction," Prentice Hall, Upper Saddle River, 2000.

[11] F. M. Richards, "Areas, Volumes, Packing, and Protein Structure," *Annual Review of Biophysics and Bioengineering*, Vol. 6, 1977, pp. 151-176. doi:10.1146/annurev.bb.06.060177.001055

[12] I. Tuñón, E. Silla and J. Pascual-Ahuir, "Molecular Surface Area and Hydrophobic Effect," *Protien Engineering*, Vol. 5, No. 8, 1992, pp. 715-716. doi:10.1093/protein/5.8.715

[13] R. M. Jackson and M. J. E. Sternberg, "Protein Surface Area Defined," *Nature*, Vol. 366, No. 6456, 1993, p. 638. doi:10.1038/366638b0

[14] B. Lee and F. M. Richards, "The Interpretation of Protein Structures: Estimation of Static Accessibility," *Journal of Molecular Biology*, Vol. 55, No. 3, 1971, pp. 379-400. doi:10.1016/0022-2836(71)90324-X

[15] Y. Fang and J. Jing, "Geometry, Thermodynamics, and Protein," *Journal of Theoretical Biology*, Vol. 262, No. 3, 2010, pp. 382-390. doi:10.1016/j.jtbi.2009.09.013

[16] D. Eisenberg and A. D. McLachlan, "Solvation Energy in Protein Folding and Binding," *Nature*, Vol. 319, No. 6050, 1986, pp. 199-203. doi:10.1038/319199a0

[17] Y. Fang, "A Gibbs Free Energy Formula for Protein Folding Derived from Quantum Statistics," *SCIENCE*

CHINA, *Physics*, *Mechanics & Astronomy*, 2013.

[18] Y. Fang, "Gibbs Free Energy Formula for Protein Folding," In: R. Morales-Rodriguez, Ed., *Thermodynamics*: *Fundamentals and Its Application in Science*, INTECHOPEN.COM, Novi Sad, 2012, pp. 47-82. www.intechopen.com

[19] Y. Fang, "Mathematical Protein Folding Problem," In: D. Hoffman, Ed., *Global Theory of Minimal Surfaces*, *Proceedings of the Clay Mathematics*, Vol. 2, 2005, pp. 611-622.

[20] Y. Fang and J. Jing, "Implementation of a Mathematical Protein Folding Model," *International Journal of Pure and Applied Mathematics*, Vol. 42, No. 4, 2008, pp. 481-488.

[21] W. Greiner, L. Neise and H. Stöker, "Thermodynamics and Statistical Mechanics," Spriger-Verlag, New York, 1994.

[22] X. Dai, "Advanced Statistical Physics," Fudan University Press, Shanghai, 2007.

