

Thermodynamic Principle Revisited: Theory of Protein Folding

Yi Fang

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

Received 10 January 2015; accepted 25 January 2015; published 29 January 2015

Copyright © 2015 by author and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Anfinsen's thermodynamic hypothesis is reviewed and misunderstandings are clarified. It really should be called the thermodynamic principle of protein folding. Energy landscape is really just the mathematical graph of the Gibbs free energy function $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$, a very high dimensional hyper surface. Without knowing it any picture of the Gibbs free energy landscape has no theoretical base, including the funnel shape claims. New insight given by newly obtained analytic Gibbs free energy function $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$ of protein folding derived via quantum statistical mechanics are discussed. Disputes such as target-based or cause-based; what is the folding force, hydrophobic effect or hydrophilic force? Single molecule or ensemble of molecules to be used for the statistical physics study of protein folding, are discussed. Classical observations of 1970's and 1980's about global geometric characteristics of native structures of globular proteins turn out to have grabbed the essence of protein folding, but unfortunately have been largely forgotten.

Keywords

Protein Folding, Gibbs Free Energy, Quantum Statistic, Single Molecule

1. Introduction

1.1. The Second Law of Thermodynamics

The Second Law of Thermodynamics states that in an isolated system the entropy will increase. For a spontaneous process in a system of constant temperature T , pressure P , and composition, the equivalent statement of the second law of thermodynamics states that the Gibbs free energy will be lowered, and at the new equilibrium state it will be at a minimum. Indeed, this applies to any process or any chemical reaction with constant temperatures and pressure [1]. The Gibbs free energy has the form $G = U + PV - TS$, where U is the internal

energy, P the pressure, V the volume, T the temperature, and S the entropy, of the system. Even T and P are not uniform in the system, moreover even they are not well defined in the system, the available free energy $A = U + P_0V - T_0S$ will be lowered and goes to minimum as long as the heat bath of the system has well defined temperature and pressure and on the boundary of the system they are constants T_0 and P_0 respectively [2]. This fundamental principle was known long before.

1.2. Anfinsen's Thermodynamics Hypothesis of Protein Folding

A good example of applying this fundamental principle to lift experimental results to guiding theory of further research is Anfinsen's Thermodynamic Hypothesis of protein folding [3]. After many years of experimental work proved that the refolding process of ribonuclease is spontaneous, Anfinsen summarised: "The studies on the renaturation of the fully denatured ribonuclease required many supporting investigations to establish, finally, the generality which we have occasionally called the 'thermodynamic hypothesis'. 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; that is that the native conformation is determined by the totality of the inter atomic interactions, and hence by the amino acid sequence, in a given environment." [3].

Once we know that protein folding is a spontaneous process, the thermodynamic hypothesis should be upgraded to thermodynamic principle. The thermodynamic principle makes the protein folding problem a pure physics problem, all biological knowledge needed is how to specify the physiological environment for a particular protein and how to reasonably simplify the environment for the further study.

But this has not been recognised so far, instead, it is thought that in biological problems such as protein folding theoretical consideration is unpractical. Coincidentally, in the early 1970's, the same time when the thermodynamic principle of protein folding was established by Anfinsen, computer entered research and played more and more important role in protein folding research. Theory was neglected, simulations became essential as if they were experiments, but many cannot satisfy the essential requirement to experiments in experimental sciences, the reproducibility, see [4]. Furthermore, theoretical background justification of these simulations were rarely questioned. One wonders that if the increasing computer power were really guided by the thermodynamics principle, perhaps today the mystery of protein folding phenomenon would not be really a mystery anymore.

1.3. Reasons of the Thermodynamic Principle of Protein Folding Is Neglected

Why the thermodynamics principle were not actively and persistently pursued?

1.3.1. Do Not Believe It

First, some do not think that the thermodynamic principle is correct. For example, in [5] it is claimed that Anfinsen's theory was disapproved for long time because "other complexities of biological systems for example solvents of different compositions may affect the folding/unfolding of proteins, the role of high dielectric constant of water, chaperone assisted folding of proteins and existence of stable folding intermediates."

All the reasons listed above to "disapprove" the thermodynamic principle belong to neglecting that in the thermodynamic principle of protein folding environment plays the same important role as the peptide chain of a protein. In fact, Anfinsen never claimed "that the primary amino acid sequence of polypeptides contains all of the necessary information to direct their folding into functional native proteins" [6]. Instead, Anfinsen stressed that "in a given environment". Solvents of different compositions, particular properties of water, chaperones to assist folding, etc., are constitutes of environment. For example, globular proteins have the simplest environment which can be simplified as only water molecules surrounding a protein molecule. For proteins needing chaperones to assist folding, chaperone molecules must appear in the environment. For membrane proteins, the environment must be described as including three layers, the middle one is hydrophobic, and the other two are mainly water molecules. Some proteins would not fold in environments not including certain constitutes, does not disapprove the thermodynamic principle, rather the proteins are not in their "normal physiological milieu". This is just one example that the generality of the thermodynamic principle is often misunderstood and then thought as wrong.

1.3.2. Misunderstandings Caused by Energy Landscape "Theory"

Second, the main reason of the thermodynamic principle was not pursued enough is because of confusions

caused by the energy landscape “theory”, both the EL (potential energy landscape) and GEL (Gibbs energy landscape) “theories”. Indeed to minimise the Gibbs free energy one should have a Gibbs free energy function $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$, where the variable \mathbf{X} is a conformation of the protein \mathcal{U} , and the parameter \mathcal{E}_N is the physiological environment in which the protein is folding. Although many have tried to derive $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$, for example, [7], all failed.

Without knowing $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$, landscape “theories” take place, such as the GEL “theory”. In fact, the GEL “theory” really has no theory, and in principle cannot explain anything [8]. All its formulae for calculating Gibbs free energy are ad hoc, without any theoretical base. Terms such as random energy formula, minimally frustrated principle, are only borrowings from other field without discussion of justification, see [8]. Especially, although proteins fold in a fixed environment, GEL has several different temperatures to be used to calculate Gibbs free energies of different conformations, [9], something wrong in principle. That is, if one invents a theory to explain a natural phenomenon, one cannot add something that is not in the natural phenomenon.

In fact, the GEL is just the graph of $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$, a very high dimensional hyper-surface (the dimension n is at least more than 200 for a 100 residue peptide chain). Advocators of GEL fully understand this, for example, in [10], it is stated that “In the field of physical chemistry, the energy landscape of a protein-solvent system is defined as an energy function $F(x) = F(x_1, x_2, \dots, x_n)$, where x_1, x_2, \dots, x_n are variables specifying the protein microscopic states”. The GEL “theory” trying to produce pictures of the very high dimensions hypo-surface. Of course, nobody can penetrate the inhibiting high dimension of this hypo-surface, trying to show it as a two dimension surface give many misleading metaphors such as “funnel shaped”, cause more confusion than understanding.

Thus, if we know $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$, depicting its graph only make trouble, we really do not need the GEL. If we do not know what is $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$, GEL “theory” only makes *ad hoc* calculations of the Gibbs free energy of various conformations without theoretical base and without consistency. No wonder it caused more misunderstanding than understanding of the thermodynamic principle of protein folding.

For example, in [1] it is claimed that pursuing of the thermodynamic principle (equivalented to GEL) leads to pitfalls, and the thermodynamic principle will not help to solve the protein folding problem [5], [11].

1.3.3. Lack of Mathematical Training

One of the main reasons of GEL, in fact, the thermodynamic principle, will not help solving the protein folding problem is that the second law of the thermodynamics cannot guarantee that the Gibbs free energy $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$ will have a global minimum [6]. Even though we do not have explicit formula, a little mathematical knowledge will help clarify this situation. For example, a lower semi-continuous function can always achieve its minimum on a compact set is a theorem in mathematics. Since any conformation’s diameter is uniformly bounded, the definition domain of \mathbf{X} certainly is contained in a compact set in higher dimensional Euclidean space. It is hard to imagine that a energy function is worse than lower semi-continuous, hence $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$ must have a global minimum. Besides, recently an analytic formula $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$ for monomeric globular proteins was derived via quantum statistics, [12]-[15], and this function is certainly continuous, see (6). In [6], to refute the funnel shape claim of GEL by suspecting the existence of global minimum is not a good argument.

1.3.4. Should the Native Structure Be Only a Local Minimum?

Another main reason of GEL will not help solving the protein folding problem is that the native structure of a protein maybe is only at a local minimum, instead of the global minimum of Gibbs free energy. This is possible, but in circumstances not against the second law of thermodynamics, hence will not negate the thermodynamic principle. In this case, the initial conformation will determine which minimal points will be achieved by the native structure. We should know more of the conformation of newly synthesised poly peptide chain in a cell. Is it always the same conformation or does it vary with each individual molecule? If it is the former, then starting with other initial conformations may lead to local or global minima different to the native structure. If it is the later, then perhaps the native structure is really the unique global minimum, since starting from all initial conformations lead to the same native structure. Judged from the experiment results of ribonuclease denaturation/renaturation, denatured ribonuclease still hold about 1% biological function. Since there are 105 patterns of fulfilling the 4 disulphide bonds, we may infer that perhaps each of the 105 patterns has the similar percentage in the denatured state. Yet, all these initial conformations refold to the same native structure in which the protein has 100%

biological function. Thus we can infer that for ribonuclease the $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$ really has a unique global minimum. Since this was the entire knowledge Anfinsen had been known, so he hypothesised that “lowest” Gibbs free energy. Considering certain initial conformation leads to certain minimiser, local or global, modifying the thermodynamic principle to admit local minimum will not harm the principle.

1.3.5. Mistaking Environment

A really legitimate concern about the thermodynamic principle is argued in [16]. It says that

“According to this hypothesis, if we define G_N as the Gibbs free energy G of a folded protein in its native state N , G_N is the global minimum of the protein’s free energy functional \hat{G} . However, G_N can only be reached if N is the current equilibrium state for the native thermodynamic conditions $\Omega(N)$. We describe the condition set $\Omega(X)$ as $\Omega(T_X, P_X, Q_X)$, where T_X and P_X are the equilibrium pressure and temperature for a protein in a state X with the conformation C_X . We assume constant T and P and use the only microscopic solvent composition Q_X to define the present conditions for X . The Anfinsen’s thermodynamic hypothesis, therefore, seems to make sense. Indeed, from the Second Law (at constant T and P), a free energy change $\Delta G_{XN}^{\text{relax}} = G_N - \tilde{G}_X < 0$ should be obtained for any thermodynamic pathway to relax the non-equilibrium state X to the folded native equilibrium state N with the respective free energies $\tilde{G}_X = \hat{G}(C_X, Q_N)$ and $G_N = \hat{G}(C_N, Q_N)$. C_N and Q_N are the native conformation and solvent composition, respectively. The possible pitfall is that \tilde{G}_X is a non-equilibrium free energy because \tilde{G}_X is not at equilibrium for Q_N . The real free energy change that has to be considered in a pathway where an intermediate state X has enough time to reach equilibrium is $\Delta G_{XN} = G_N - G_X$, where $G_X = \hat{G}(C_X, Q_X)$ (We think this is an error for $\hat{G}(C_X; Q_X)$, otherwise $G_X = \tilde{G}_X$) is the equilibrium free energy for Q_X . The Anfinsen’s thermodynamic hypothesis can, therefore, only hold with a good likeliness if $\tilde{G}_X \cong G_X$. However, cases where G_X is a deep minimum with $G_X < \tilde{G}_X$ cannot be excluded, which may lead to $\Delta G_{XN} = G_N - G_X > 0$.”

The point is that the solvent composition Q_X is really varying with the conformation C_X , *i.e.*, there is not any common solvent composition Q_N , see the next section on the formula $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$, where Q_X is the first water layer of the conformation \mathbf{X} and is part of the thermodynamic system \mathcal{T}_X for each conformation \mathbf{X} , and $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$ is really the Gibbs free energy of the thermodynamic system \mathcal{T}_X . As for equilibrium, the protein folding process should be considered as a quasi-static process, and as mentioned before, we only need the heat bath has a constant temperature T_0 and pressure P_0 , and in this case the second law of thermodynamics is that the available energy $A = U + P_0V - T_0S$ is getting its minims as stated in the beginning according to [2].

After clarifying these suspicions on the thermodynamics principle, we will demonstrate what is the $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$ and look at some new insight it gives.

2. The Formula $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$

We will not give the derivation of $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$, which was done in [12]-[15], with the same idea but progressively better understanding of the quantum physics. We concentrate on the rationals coming from the understanding of the thermodynamic principle of protein folding.

Our understanding of the thermodynamic principle is that it emphasises holistic view, it requires a single molecule method and quantum statistics instead of classical statistics to derive the Gibbs free energy formula $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$.

2.1. The Function $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$ Cannot Be a Sum of Local Contributions

Unlike the potential energy function, the Gibbs free energy function, or, the GEL, is not pairwise additive as has been pointed in [6]. In fact, we cannot first consider local contributions and then sum them up to get the Gibbs free energy. This is emphasised by Anfinsen in the statement “that is that the native conformation is determined by the totality of the inter atomic interactions, and hence by the amino acid sequence, in a given environment.” [3].

So that when trying to derive $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$ by the first principle, we cannot divide \mathbf{X} into several parts, consider each part, and sum up Gibbs free energies of all these parts. In fact, we even cannot take a coarse grained model of conformation to try to derive $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$, because an atom’s contribution to the whole cannot be separated. Hence for us, a conformation is the atomic centres’ coordinates of all atoms $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N)$ of the given protein \mathfrak{U} , witting as $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \in \mathbb{R}^{3N}$. For proteins with more than one

poly peptide chains, all chains should be consider together, *i.e.*, let $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$, $\mathcal{U} = (\mathcal{U}_1, \dots, \mathcal{U}_n)$, $n > 1$, then $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N) \neq \sum_{i=1}^n G(\mathbf{X}_i; \mathcal{U}_i, \mathcal{E}_N)$.

2.2. Single Molecule Treatment Is Necessary

Like any computer simulation of protein folding, we describe only one protein molecule in various conformations \mathbf{X} , not an ensemble of (the same) protein molecules each taking a conformation. To derive $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$, it is nature that one needs adopt the statistical physics. When applying statistics, naturally one thinks that there should be many copies of the same object, such as a protein molecule, to form an ensemble. This was pursued by many, see for example, [7], where integrations on all molecular conformations of the ensemble except one \mathbf{X} was performed to get $E(\mathbf{X})$. With a not integrable integrand (in fact, it is in exponential form) and without clear delimitation of the integral domain, the obtained formula $E(\mathbf{X})$ is a complicated unknown function buried in multi-dimensional integral. Worse still, $E(\mathbf{X})$ is even not the Gibbs free energy. Nevertheless, the authors of [7] called it the effective energy and used it in many places as if is following the thermodynamic principle. This is a perfect example of starting from the thermodynamic principle and end up with metaphor expressions and endless computer simulations to cover up theoretical poverty. Following this trend, the protein folding problem will never be resolved.

Since one of the tasks of protein folding problem is to figure out the individual protein's native structure, but in an ensemble of molecules, all available methods are actually neglecting the structures of individual molecule, we cannot use the ensemble method. Therefore, we have to take a single molecule \mathcal{U} , consider an arbitrary conformation \mathbf{X} of it, and to figure out a thermodynamic system $\mathcal{T}_{\mathbf{X}}$ which is tailor made for the conformation \mathbf{X} and contains the immediate surrounding environment of \mathbf{X} . Finally, the Gibbs free energy $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$ is the Gibbs free energy of the thermodynamic system $\mathcal{T}_{\mathbf{X}}$.

2.3. Classical or Quantum Statistics?

We have to figure out how to do statistics on this thermodynamic system $\mathcal{T}_{\mathbf{X}}$. Both classical and quantum statistics were tried, see [12], with the classical result missing the volume contribution in formula (6). Consider that classical mechanics does not fit to describe physics of objects of molecule and even macro molecule size, we choose quantum statistics. Moreover, quantum statistics will allow we do further theoretical studies of protein folding, if we can handle the electronic density function defined in [17], which we cannot do presently. This function holds the origins of our hydrophobicity level classification in subsection 2.6.

2.4. The Importance of Environment

Our tailor made thermodynamic system $\mathcal{T}_{\mathbf{X}}$ in fact contains the immediate surrounding environment of the conformation \mathbf{X} . Biological knowledge comes here to help us describe and make necessary and rationale simplifications of this immediate environment. For example, it is known for globular proteins, we can simply assume that in the physiological condition only water molecules immediately surrounding a conformation. For membrane proteins, the immediate environment should have at least three parallel layers, water molecules in the outer two layers and the middle is hydrophobic. For proteins needing chaperones' help to fold, these chaperones must be contained in the immediate environment of the conformation \mathbf{X} . This is also the holistic view, without the chaperones, the protein is in a wrong environment and will either fold to another structure, or no structure at all, meaning many different conformations achieve the minimum of Gibbs free energy.

Since except for monomeric globular proteins, we have not figured out how to handle environment, our present function $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$ is only for monomeric globular proteins.

2.5. The Thermodynamic System $\mathcal{T}_{\mathbf{X}}$ for Monomeric Globular Proteins

Since only globular proteins allow us to simplify their physiological environment as consisting of only water molecules, we will only work on monomeric globular proteins here. A conformation of a polymeric globular protein is $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$, $n > 1$. Theoretically, our function $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$ can be generalised without any theoretical difficulty to polymeric globular proteins. But to apply it we will face the docking problem, *i.e.*, consider $\mathcal{T}_{\mathbf{X}} = \bigcup_{i=1}^n \mathcal{T}_{\mathbf{X}_i}$ and the relative positions between $\mathcal{T}_{\mathbf{X}_i}$'s. This is too hard for now.

First, a conformation \mathbf{X} by definition lives in \mathbb{R}^{3N} , but the textbook definition of a thermodynamic system is that it is a region in \mathbb{R}^3 , see, for example, [18]. To create $\mathcal{T}_{\mathbf{X}}$, we consider $P_{\mathbf{X}} = \bigcup_{i=1}^N B(\mathbf{x}_i, r_i) \subset \mathbb{R}^3$, where $B(\mathbf{x}, r)$ is solid ball of radius r and centred at $\mathbf{x} \in \mathbb{R}^3$. Essentially the $P_{\mathbf{X}}$ together with its first layer of water molecules will be our $\mathcal{T}_{\mathbf{X}}$. Here we assumed that each atom \mathbf{a}_i 's shape is a solid ball with van der Waals radius r_i . Although the shape of each atom in \mathcal{U} is well defined by the theory of atoms in molecules [17], what concerning us here is the overall shape of the structure $P_{\mathbf{X}}$. The cutoff of electron density $\rho \geq 0.001$ au in [17], gives the overall shape of a molecular structure that is just like $P_{\mathbf{X}}$, a bunch of overlapping balls. Moreover, the boundary of the $\rho \geq 0.001$ au cutoff is almost the same as the **molecular surface** $M_{\mathbf{X}}$ which was defined by Richards in 1977 [19] and in 1992 and 1993, [20] and [21] was shown to be a more suitable boundary surface of $P_{\mathbf{X}}$ than other surfaces

To explain the formula $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$, we have to describe $\mathcal{T}_{\mathbf{X}}$ in details.

In general, any closed surface (connected, bounded, and has no boundary, for example, a sphere) $S \subset \mathbb{R}^3$ will divide \mathbb{R}^3 into three parts,

$$\mathbb{R}^3 = \Omega_S \cup S \cup \Omega'_S, \quad \partial\Omega_S = \partial\Omega'_S = S, \quad (1)$$

where Ω_S is a bounded domain (connected open set) and Ω'_S a un-bounded domain, $\partial\Omega_S$ means the boundary of Ω_S .

Rolling a sphere of radius r on the boundary surface $\partial P_{\mathbf{X}}$ of $P_{\mathbf{X}}$ will produce a molecular surface $M_r(\mathbf{X})$ [19]. Let d_w be the diameter of a water molecule and denote the molecular surface $M_{\frac{d_w}{2}}(\mathbf{X})$ as $M_{\mathbf{X}}$. If $M_{\mathbf{X}}$ is connected, then we can use $S = M_{\mathbf{X}}$ in (1). If $M_{\mathbf{X}}$ has multiple connected components S_i , $1 \leq i \leq m$, such that S_1 is the largest component, *i.e.*, all other components of $M_{\mathbf{X}}$ are contained in Ω_{S_1} (this is the case that $P_{\mathbf{X}}$ has $m-1$ cavities Ω_{S_i} , $i = 2, \dots, m$, each is large enough to hold a water molecule), then denote $\Omega_{\mathbf{X}} = \Omega_{S_1} \cap \left(\bigcap_{i=1} \Omega'_{S_i} \right)$ and $\Omega'_{\mathbf{X}} = \Omega'_{S_1} \cup \left(\bigcup_{i=1} \Omega_{S_i} \right)$. Thus, we always have

$$\mathbb{R}^3 = \Omega_{\mathbf{X}} \cup M_{\mathbf{X}} \cup \Omega'_{\mathbf{X}}, \quad \partial\Omega_{\mathbf{X}} = \partial\Omega'_{\mathbf{X}} = M_{\mathbf{X}}, \quad P_{\mathbf{X}} \subset \overline{\Omega_{\mathbf{X}}} = \Omega_{\mathbf{X}} \cup M_{\mathbf{X}}. \quad (2)$$

Let

$$\mathcal{R}_{\mathbf{X}} = \left\{ \mathbf{x} \in \mathbb{R}^3 : \text{dist}(\mathbf{x}, M_{\mathbf{X}}) \leq d_w \right\} \setminus \Omega_{\mathbf{X}}, \quad (3)$$

be the first hydration shell surrounding $P_{\mathbf{X}}$. Then the tailor made thermodynamic system for the conformation \mathbf{X} is

$$\mathcal{T}_{\mathbf{X}} = \overline{\Omega_{\mathbf{X}}} \cup \mathcal{R}_{\mathbf{X}}. \quad (4)$$

2.6. Hydrophobicity Levels

Any Gibbs free energy formula should not only have fairly general form for all proteins, or at least a class of proteins such as monomeric globular proteins, but also must be able to distinguish different proteins. That means that if \mathcal{U}_1 and \mathcal{U}_2 are two different proteins with the same number of atoms, say N . Then even $\mathbf{X} \in \mathbb{R}^{3N}$ simultaneously appears as conformations of both \mathcal{U}_1 and \mathcal{U}_2 , $G(\mathbf{X}; \mathcal{U}_1, \mathcal{E}_N)$ and $G(\mathbf{X}; \mathcal{U}_2, \mathcal{E}_N)$ are not identical in any \mathbb{R}^{3N} neighbourhood of \mathbf{X} . In particular, even $G(\mathbf{X}; \mathcal{U}_1, \mathcal{E}_N)$ and $G(\mathbf{X}; \mathcal{U}_2, \mathcal{E}_N)$ have the same minimum values, \mathcal{U}_1 and \mathcal{U}_2 should have different natives structures, if they have native structures at all.

Hence, we should find a way to distinguish proteins by their peptide chains. The hardest task is that given a peptide chain $P = A_1 \cdots A_n$, let n_i be the number of amino acid i appears in P , shuffle the amino acid A_i 's around we will have $S(P) = n! / \prod_{i=1}^{20} n_i!$ different amino acid sequences, the formula $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$ has to be able to distinguish all of these $S(P)$ peptide chains. For example, there should be $S(P)$ different minimisation problems in (9), though the minimisers may vary just slightly for some of them.

To this purpose, we divide atoms in a protein according to their hydrophobicity levels. Atoms in a protein molecule are naturally existing in atom groups or moieties which have different physicochemical properties. One of these properties is the electronic charge distributions caused the tendency of forming hydrogen bonds either with other moieties (intra-molecular) or with other molecules in the environment (inter-molecular). Accordingly, we can divide these atom groups or moieties into different levels of hydrophobicity, from the most hydrophobic (cannot form hydrogen bond) to the most hydrophilic, say there are H levels H_1, \dots, H_H , $H \geq 2$. Then we can assign an atom \mathbf{a}_k into one hydrophobic level H_i if \mathbf{a}_k belongs to an H_i atom group. For example, we may assume that the classification is as in [22], there are $H = 5$ classes, C, O/N, O⁻, N⁺, S. Unlike in [22], we also classify every hydrogen atom into one of the H hydrophobicity level groups. Note that this classification is independent of conformations, it only depends on the peptide chain.

For any compact (closed and bounded) set $U \subset \mathbb{R}^3$, let $\text{dist}(\mathbf{x}, U) = \min_{z \in U} |\mathbf{x} - \mathbf{z}|$ be the distance between the point \mathbf{x} and the subset U . Define compact sets $P_{X_i} = \bigcup_{\mathbf{a}_j \in H_i} B(\mathbf{x}_j, r_j)$, $i = 1, \dots, H$, then $P_X = \bigcup_{i=1}^H P_{X_i}$. Define subsurfaces in M_X ,

$$M_{X_i} = \{\mathbf{x} \in M_X : \text{dist}(\mathbf{x}, P_{X_i}) < \text{dist}(\mathbf{x}, P_X \setminus P_{X_i})\}, \quad i = 1, \dots, H, \quad (5)$$

where $P_X \setminus P_{X_i}$ is the set of points \mathbf{x} that belong to P_X but do not belong to P_{X_i} .

To each hydrophobic level H_i , there is a chemical potential μ_i , such that a water molecule touching M_{X_i} will gain a Gibbs free energy μ_i . Similarly, there is a chemical potential μ_e for electrons inside \mathcal{T}_X . Let v_i be the average number of water molecules that can simultaneously touching M_{X_i} in a unit area, then $\omega_i = v_i \mu_i$ will be chemical potential per unit area of M_{X_i} . Moreover, since the curvature of M_X is uniformly bounded for all conformations \mathbf{X} , ω_i 's do not depend on conformations \mathbf{X} .

2.7. The Formula $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$

Let $V(\Omega)$ be the volume of $\Omega \subset \mathbb{R}^3$ and $A(S)$ the area of a surface $S \subset \mathbb{R}^3$. Now we can write the analytic Gibbs free energy formula $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$.

Theorem 1 Let \mathfrak{U} be a monomeric globular protein with N atoms $(\mathbf{a}_1, \dots, \mathbf{a}_N)$ and $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ be a conformation. Let q_A be the electronic charges in the nucleus of \mathbf{a}_A .

$$G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N) = \omega_e V(\Omega_X) + d_w \omega_e A(M_X) + \sum_{i=1}^H \omega_i A(M_{X_i}) + \sum_{1 \leq A < B \leq N} \frac{q_A q_B}{4\pi \epsilon_0 |\mathbf{x}_A - \mathbf{x}_B|}. \quad (6)$$

In [12]-[15], the quantum statistical derivation first get a intermediate formula, which is much familiar but with new meaning for $N_i(\mathbf{X})$:

$$G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N) = \mu_e N_e(\mathbf{X}) + \sum_{i=1}^H \mu_i N_i(\mathbf{X}) + \sum_{1 \leq A < B \leq N} \frac{q_A q_B}{4\pi \epsilon_0 |\mathbf{x}_A - \mathbf{x}_B|}, \quad (7)$$

where $N_e(\mathbf{X})$ and $N_i(\mathbf{X})$'s are mean numbers of electrons in \mathcal{T}_X and water molecules that touches M_{X_i} . Moreover, there is a $v_e > 0$ such that $v_e N_e(\mathbf{X}) = V(\mathcal{T}_X)$ so define $\omega_e = v_e \mu_e$. Using $N_i(\mathbf{X}) = v_i A(M_{X_i})$ and the volume decomposition

$$V(\mathcal{T}_X) \approx V(\Omega_X) + d_w A(M_X), \quad (8)$$

We get formula (6) from (7).

3. New Insight

3.1. Structure Prediction

With theoretically established $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$, *ab initio* structure prediction not only becomes possible, but also simple. It is a pure mathematical problem of seeking the minimisers of $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$. That is, let \mathbf{X}_N be the native structure, then

$$G(\mathbf{X}_N; \mathfrak{U}, \mathfrak{E}_N) = \min_{\text{All conformations } \mathbf{X}} G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N). \quad (9)$$

Or, in case that \mathbf{X}_N is only a local minimiser, it must satisfy:

$$\nabla G(\mathbf{X}_N; \mathfrak{U}, \mathfrak{E}_N) = \vec{\mathbf{0}}. \quad (10)$$

As discussed before, in this situation, initial conformation \mathbf{X}_0 is important. If the biological knowledge including the conformation of the nascent peptide chain, we should use it.

To solve (9) there is no need of searching landscapes as seen so important in GEL “theory” [9]. Just following the $-\nabla G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$ (the negative gradient of G at \mathbf{X}) from any initial conformation \mathbf{X}_0 to $\mathbf{X}_{i+1} = \mathbf{X}_i - s \nabla G(\mathbf{X}_i; \mathfrak{U}, \mathfrak{E}_N)$, $i = 0, 1, \dots$, and s is a small positive number. If the GEL is really funnel shaped, the native structure will be reached eventually by this classical Newton’s fastest descending method.

When the native structure may take only a local minimum instead a global one, we have to try different initial conforms \mathbf{X}_0 , to get as many as possible local minimisers \mathbf{X}_L (such that $\nabla G(\mathbf{X}_L; \mathfrak{U}, \mathfrak{E}_N) = \vec{\mathbf{0}}$) by above Newton’s fastest descending method. The native structure must be one of the \mathbf{X}_L ’s. Other information is needed to determine which one is \mathbf{X}_N .

Of course we can use another set of variables, *i.e.*, the dihedral angles $\Phi = (\phi_1, \dots, \phi_L)$, including every routable dihedral angles, in main chain or side chains. Dihedral angles corresponding to a covalent bond inside a Bunsen ring is an example of not rotatable dihedral angle.

In fact, the dihedral angles are the most efficient variables in solving (9) and (10). For the explicit derivative formulae of $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$, please see [23] and [13].

3.2. Understanding the Folding Process

Theoretically derived $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$ can explain various phenomena in the folding process. For example, since the folding force is $-\nabla G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$, *i.e.*, the natural folding process will follow Newton’s fastest descending path, the initial conformation \mathbf{X}_0 determines the pathway of the folding. To test the shape of GEL, one can select as many as possible of initial conformations and apply the fastest descending method to find a solution of (10), *i.e.*, a conformation \mathbf{X}_C such that $\nabla G(\mathbf{X}_C; \mathfrak{U}, \mathfrak{E}_N) = 0$. If all initial conformations lead to the same solution \mathbf{X}_C , then the landscape GEL is really funnel shaped. If we get many different \mathbf{X}_C ’s, then which one is the native structure need to be further discussed with more information. But in this case if the conformation of the nascent peptide chain is always the same, we still have a single folding pathway and even can also observe a classical two phase folding phenomenon.

Now consider an ensemble of M conformations of the same protein, $\mathbf{X}_1, \dots, \mathbf{X}_M$. This ensemble case is more familiar, and in fact is the only thermodynamic system appeared in protein folding literature so far. Single molecule phenomena, such as what is the shape of native structure, can never be deduced from such an ensemble system. Instead, the concern should be concentrated on knowing the collective phenomena such as folding rate and folding time, etc. Then indeed we should consider the distribution $P(\mathbf{X}_1, \dots, \mathbf{X}_M; \mathfrak{U}, \mathfrak{E}_N)$. For example, the second law of thermodynamics indicates that there is a distribution $P_{\text{eq}}(\mathbf{X}_1, \dots, \mathbf{X}_M; \mathfrak{U}, \mathfrak{E}_N)$ such that when the proteins at full function state, for any distribution P , it will be

$$G(P_{\text{eq}}(\mathbf{X}_1, \dots, \mathbf{X}_M); \mathfrak{U}, \mathfrak{E}_N) \leq G(P(\mathbf{X}_1, \dots, \mathbf{X}_M); \mathfrak{U}, \mathfrak{E}_N), \quad (11)$$

where $G(P(\mathbf{X}_1, \dots, \mathbf{X}_M); \mathfrak{U}, \mathfrak{E}_N)$ is the Gibbs free energy of the ensemble under distribution $P(\mathbf{X}_1, \dots, \mathbf{X}_M)$. The problem is, nobody knows P_{eq} , as Ben-Naim admitted in [24]. One suggestion is that we apply $G(\mathbf{X}) = G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$ and try the Boltzmann distribution

$$P_{\text{eq}}(\mathbf{X}_i; \mathfrak{U}, \mathfrak{E}_N) = \frac{\exp\left(-\frac{G(\mathbf{X}_i)}{kT}\right)}{\sum_{j=1}^M \exp\left(-\frac{G(\mathbf{X}_j)}{kT}\right)}. \quad (12)$$

The rational is that any conformation \mathbf{X} with $G(\mathbf{X}) \gg G(\mathbf{X}_N)$ will have much smaller $P_{\text{eq}}(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$, thus less chance to appear in the full function state of the ensemble. Of course, this is only a conjecture and it is not so important to know, at least not as a claim made in [24]: “If one knew this distribution, then one could tell which conformations are more probable than the others under the given environment.” In fact, we now all know that in physiological situation the native structure is “more probable than the others under the given environment”, but, we still do not know the shape of the native structure. So to solve the protein folding problem, at least for the prediction of native structure from the knowledge of amino acid sequence, we have to know what is $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$ and solve (9) or (10) with whatever mathematical method.

3.3. Force of Folding

In [6], the folding force is claimed as $-\nabla G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$. In [6] it is also claimed that search the minimiser in GEL, such as in (9), is target-based, and identify the folding force is cause-based. In fact, target-based and cause-based are only artificial distinctions. We see that in solving (9), we do not have a target to subjectively approaching and if we use the fastest deciding method to find the minimiser, we are really cause-based, because we are explicitly using the force $-\nabla G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$.

3.4. Understanding Denaturation and Refolding

If theoretically derived $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E})$ is known for any specific environment \mathfrak{E} , for example, environments differ only in temperature values, $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E})$ can explain the denaturation and protein refolding by changing environments. Changing the environment \mathfrak{E}_N to, say \mathfrak{E}_D , the native structure \mathbf{X}_N of \mathfrak{U} will no longer be a minimiser of $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_D)$ in problem (9) or (10), therefore, \mathbf{X}_N is unstable in \mathfrak{E}_D . Moreover, $\nabla G(\mathbf{X}_N; \mathfrak{U}, \mathfrak{E}_D) \neq \vec{0}$. By the same second law of thermodynamics, or thermodynamic principle, $-\nabla G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_D)$ will force a minimisation of $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_D)$, resulting different minimisers other than \mathbf{X}_N . This is a theoretical explanation of denaturation. For certain types of proteins, when the environment \mathfrak{E}_D changes back \mathfrak{E}_N , or similar ones, refolding happens as the same procedures as in problem (9).

Furthermore, various thermodynamic functions, such as the entropy S , can be obtained by the family $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E})$. For example,

$$S(\mathbf{X}; \mathfrak{U}, \mathfrak{E}) = -\frac{\partial G(\mathbf{X}; \mathfrak{U}, \mathfrak{E})}{\partial T}. \quad (13)$$

3.5. Hydrophobic, Hydrophilic, Which Is the Folding Force?

There is a hot debate in [6] on which one is the main folding force, hydrophobic effect or hydrophilic force? Once we know $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$, moving along the force direction $-\nabla G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$, G will become smaller. The qualitative character of ω_i is that if the class H_i is hydrophobic, then $\omega_i > 0$, and if H_j is hydrophilic, then $\omega_j < 0$. Thus reducing hydrophobic areas $A(M_{\mathbf{X}_i})$ (hydrophobic effect) or enlarging hydrophilic areas $A(M_{\mathbf{X}_j})$ (hydrophilic force?) will reduce the Gibbs free energy G .

In terms of the force $-\nabla G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$, the situation is much more complicated. Consider the case of $\frac{\partial A(M_{\mathbf{X}_i})}{\partial \phi_k} > 0$ and $\frac{\partial A(M_{\mathbf{X}_j})}{\partial \phi_k} > 0$, they contribute to $-\frac{\partial G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)}{\partial \phi_k}$ a force pushing \mathbf{X} to a new conformation \mathbf{X}' such that $A(M_{\mathbf{X}'_i}) < A(M_{\mathbf{X}_i})$ and $A(M_{\mathbf{X}'_j}) > A(M_{\mathbf{X}_j})$. This well explains that both hydrophobic effect and hydrophilic force playing their roles in reducing the Gibbs free energy. But these are only two terms in $-\frac{\partial G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)}{\partial \phi_k}$, other terms' values will eventually determine the sign of $-\frac{\partial G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)}{\partial \phi_k}$. And if

$\frac{\partial A(M_{x_i})}{\partial \phi_k} < 0$, or $\frac{\partial A(M_{x_j})}{\partial \phi_k} < 0$, what will happen? So we have to adopt a holistic view of the folding force, both hydrophobic effect and hydrophilic force play their role, which is dominated depends on concrete circumstances.

To simplify the discussion of hydrophobic and hydrophilic effects (or force), we consider the simplest classification of hydrophobicity levels, *i.e.*, $H = 2$, there are only hydrophobic and hydrophilic atom groups. Denote H_h as the hydrophobic class, H_p as the hydrophilic class. In this case, we have

$$A(M_{\mathbf{x}}) = A(M_{\mathbf{x},H_h}) + A(M_{\mathbf{x},H_p}). \quad (14)$$

Thus for fixed $A(M_{\mathbf{x}})$, enlarge $A(M_{\mathbf{x},H_p})$ is equivalent to reduce $A(M_{\mathbf{x},H_h})$, no distinction between hydrophobic effect and hydrophilic force. In this case,

$$\begin{aligned} G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N) &= \omega_e V(\Omega_{\mathbf{x}}) + d_w \omega_e A(M_{\mathbf{x}}) + \omega_h A(M_{\mathbf{x},H_h}) + \omega_p A(M_{\mathbf{x},H_p}) + \sum_{1 \leq A < B \leq M} \frac{q_A q_B}{4\pi \epsilon_0 |\mathbf{x}_A - \mathbf{x}_B|} \\ &= \omega_e V(\Omega_{\mathbf{x}}) + (d_w \omega_e + \omega_p) A(M_{\mathbf{x}}) + (\omega_h - \omega_p) A(M_{\mathbf{x},H_h}) + \sum_{1 \leq A < B \leq M} \frac{q_A q_B}{4\pi \epsilon_0 |\mathbf{x}_A - \mathbf{x}_B|}. \end{aligned} \quad (15)$$

Since $\omega_h - \omega_p > 0$, the combination of hydrophobic effect and hydrophilic force is the effect of reducing the hydrophobic area $A(M_{\mathbf{x},H_h})$. In [25] a simulation of reducing $A(M_{\mathbf{x},H_h})$ alone was performed, the result is very interesting. Secondary structures, such as α helices, β strands, β turns, and hydrogen bonds appeared in obtained conformations with statistical significance [25]. Note that, when reducing $A(M_{\mathbf{x},H_h})$, nothing about secondary structure and hydrogen bond were considered. No calculation of the dihedral angles. No testing of positions of the donor and acceptor groups, let alone any intent to push them closer to form a hydrogen bond. Yet nevertheless, secondary structures and hydrogen bond appeared in statistical significance. Before this simulation, it was recognised that hydrogen bond must be explicitly modelled for helix formation and pairwise simulation without specifying hydrogen bonding cannot produce secondary structures [26].

3.6. Continue a Classics Global View

If $d_w \omega_e + \omega_p > 0$, then $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$ in the version (15) is an realisation of classical global view of the characteristics of global geometry of native structures of globular proteins. These characteristics are: 1) the native structure has smaller volume; 2) the native structure has smaller area; 3) the native structure has a better hydrophobic core or smaller hydrophobic area $A(M_{\mathbf{x},H_h})$, [19], [21] and [27]-[31]. These important observations were forgot after 1990's, people were more interested in computer simulations based on force field, GEL, EL, etc. Based on these characteristics, a phenomenological model was formed with exactly the form of (15) with three positive coefficients, thus claim that simultaneously reducing volume, area, and hydrophobic area will eventual lead us to the native structure, see [32] and [25]. Moreover, the first two terms of formula (15) is the formula of the Gibbs free energy of a cavity in water, the two coefficients are pressure and surface tension [18]. First principle and phenomenology, both derived the same Gibbs free energy formula, can it be just a coincidence?

4. Conclusion

The reasons that over four decades Anfinsen's thermodynamic hypothesis has been dismissed as leading to pitfalls, as disapproved, as no importance at all, are analysed and clarified. They are due to misunderstanding and inability in deriving a Gibbs free energy formula of protein folding. The misunderstandings mainly come from neglecting environment's role that Anfinsen so emphasised. The inability of deriving a Gibbs free energy function of protein folding comes from using ensemble of conformations that neglected individual conformation's 3-dimensional shape. The no importance dismiss came coincidentally with various of computer simulation without theoretical discuss of their theoretical bases. Newly derived Gibbs free energy function $G(\mathbf{X}; \mathfrak{U}, \mathfrak{E}_N)$ is in-

roduced, its roles in protein structure prediction and in explaining folding process are discussed. The derivation of $G(\mathbf{X}; \mathcal{U}, \mathcal{E}_N)$ is applying quantum statistics to thermodynamic systems \mathcal{T}_X tailor made for a single conformation \mathbf{X} and its immediate environment, following Anfinsen's original single molecule orientation.

References

- [1] Ben-Naim, A. (2011) Pitfalls in Anfinsen's Thermodynamic Hypothesis. *Chemical Physics Letters*, **511**, 126-128. <http://dx.doi.org/10.1016/j.cplett.2011.05.049>
- [2] Müller, I. (2010) Miscellanea about Entropy, Energy, and Available Free Energy. *Symmetry*, **2**, 916-934. <http://dx.doi.org/10.3390/sym2020916>
- [3] Anfinsen, C.B. (1973) Principles that Govern the Folding of Protein Chains. *Science*, **181**, 223-230. <http://dx.doi.org/10.1126/science.181.4096.223>
- [4] Wang, Y., Zhang, H., Li, W. and Scott, R.A. (1995) Discriminating Compact Nonnative Structure from the Native Structure of Globular Proteins. *PNAS*, **92**, 709-713.
- [5] Kaushik, A. and Gupta, E. (2013) Protein Folding grand Challenge: Hydrophobic vs. Hydrophilic Forces. *Journal of Biomolecular Structure and Dynamics*, **31**, 1011-1012.
- [6] Ben-Naim, A. (2012) Levinthal's Question Revisited, and Answered. *Journal of Biomolecular Structure and Dynamics*, **30**, 113-124. <http://dx.doi.org/10.1080/07391102.2012.674286>
- [7] Lazaridis, T. and Karplus, M. (2003) Thermodynamics of Protein Folding: A Microscopic View. *Biophysical Chemistry*, **100**, 367-395. [http://dx.doi.org/10.1016/S0301-4622\(02\)00293-4](http://dx.doi.org/10.1016/S0301-4622(02)00293-4)
- [8] Ben-Naim, A. (2013) Response to Comments on My Article: Levinthal's Question Revisited and Answered. Ben-Naim A. (2012) *Journal of Biomolecular Structure and Dynamics*, **30**, 113-124. *Journal of Biomolecular Structure and Dynamics*, **31**, 1028-1033. <http://dx.doi.org/10.1080/07391102.2012.748548>
- [9] Schafer, N.P., Kim, B.L., Zhang, Q. and Wolynes, P.G. (2013) Learning to Fold Protein Using Energy Landscape Theory. arXiv:1312.7283v1 [q-bio.BM] <http://arxiv.org/pdf/1312.7283.pdf>
- [10] Liu, S.Q., Ji, X.L., Tao, Y., Tan, D.Y., Zhang, K.Q. and Fu, Y.X. (2012) Protein Folding, Binding and Energy Landscape: A Synthesis. In: Kaumaya, P.T.P., Ed., *Protein Engineering*, Intech, Rijeka, 207-253.
- [11] Fang, Y. (2015) Why Ben-Naim's Deepest Pitfall Does Not Exist. To Appear in *Open Journal of Biophysics*.
- [12] Fang, Y. (2012) Gibbs Free Energy Formula for Protein Folding. In: Morales-Rodriguez, R., Ed., *Thermodynamics—Fundamentals and Its Application in Science*, Intech, Rijeka, 47-82. <http://www.intechopen.com/books/thermodynamics-fundamentals-and-its-application-in-science>
- [13] Fang, Y. (2013) Ben-Naim's Pitfalls: Don Quixote's Windmill. *Open Journal of Biophysics*, **3**, 13-21.
- [14] Fang, Y. (2014) The Second Law, Gibbs Free Energy, Geometry, and Protein Folding. *Journal of Advances in Physics*, **3**, 278-285.
- [15] Fang, Y. (2014) A Gibbs Free Energy Formula for Protein Folding Derived from Quantum Statistics. *Science China Physics, Mechanics & Astronomy*, **57**, 1547-1551. <http://dx.doi.org/10.1007/s11433-013-5288-x>
- [16] Gillet, J. and Ghosh, I. (2013) Concepts on the Protein Folding Problem. *Journal of Biomolecular Structure and Dynamics*, **31**, 1020-1023. <http://dx.doi.org/10.1080/07391102.2012.748546>
- [17] Bader, R.F.W. (1990) *Atoms in Molecules: A Quantum Theory*. Clarendon Press, Oxford.
- [18] Pippard, A.A. (1957) *The Elements of Classical Thermodynamics*. Cambridge University Press, Cambridge.
- [19] Richards, F.M. (1977) Areas, Volumes, Packing, and Protein Structure. *Annual Review of Biophysics and Bioengineering*, **6**, 151-176. <http://dx.doi.org/10.1146/annurev.bb.06.060177.001055>
- [20] Tuñón, I., Silla, E. and Pascual-Ahuir, J.L. (1992) Molecular Surface Area and Hydrophobic Effect. *Protein Engineering, Design and Selection*, **5**, 715-716. <http://dx.doi.org/10.1093/protein/5.8.715>
- [21] Jackson, R.M. and Sternberg, M.J.E. (1993) Protein Surface Area Defined. *Nature*, **366**, 638. <http://dx.doi.org/10.1038/366638b0>
- [22] Eisenberg, D. and McLachlan, A.D. (1986) Solvation Energy in Protein Folding and Binding. *Nature*, **319**, 199-203. <http://dx.doi.org/10.1038/319199a0>
- [23] Fang, Y. and Jing, J. (2008) Implementation of a Mathematical Protein Folding Model. *International Journal of Pure and Applied Mathematics*, **42**, 481-488.
- [24] Ben-Naim, A. (2013) Comment on a Paper: "Ben-Naim's 'Pitfalls': Don Quixote's Windmill" by Y. Fang, *Open Journal of Biophysics*, 2013, 3, 13-21. *Open Journal of Biophysics*, **3**, 275-276.

- [25] Fang, Y. and Jing, J. (2010) Geometry, Thermodynamics, and Protein. *Journal of Theoretical Biology*, **262**, 383-390. <http://dx.doi.org/10.1016/j.jtbi.2009.09.013>
- [26] Hubner, I.A. and Shakhnovic, E.I. (2005) Geometric and Physical Considerations for Realistic Protein Models. *Physical Review E*, **72**, Article ID: 022901. <http://dx.doi.org/10.1103/PhysRevE.72.022901>
- [27] Lee, B. and Richards, F.M. (1971) The Interpretation of Protein Structures: Estimation of Static Accessibility. *Journal of Molecular Biology*, **55**, 379-400. [http://dx.doi.org/10.1016/0022-2836\(71\)90324-X](http://dx.doi.org/10.1016/0022-2836(71)90324-X)
- [28] Janin, J. (1976) Surface Area of Globular Proteins. *Journal of Molecular Biology*, **105**, 13-14. [http://dx.doi.org/10.1016/0022-2836\(76\)90192-3](http://dx.doi.org/10.1016/0022-2836(76)90192-3)
- [29] 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 Research Communications*, **44**, 47-63. <http://dx.doi.org/10.1007/BF02906521>
- [30] Novotny, J., Brucoleri, R. and Karplus, M. (1984) An Analysis of Incorrectly Folded Protein Models: Implications for Structure Predictions. *Journal of Molecular Biology*, **177**, 787-818. [http://dx.doi.org/10.1016/0022-2836\(84\)90049-4](http://dx.doi.org/10.1016/0022-2836(84)90049-4)
- [31] Novotny, J., Rashin, A.A. and Brucoleri, R. (1986) Criteria that Discriminate between Native Proteins and Incorrectly Folded Models. *Proteins*, **4**, 19-30. <http://dx.doi.org/10.1002/prot.340040105>
- [32] Fang, Y. (2005) Mathematical Protein Folding Problem. In: Hoffman, D., Ed., *Global Theory of Minimal Surfaces. Proceedings of the Clay Mathematical Proceedings*, Vol. 2, American Mathematical Society, Clay Mathematics Institute, 611-622.