ISSN 2321-807X

Myths and verities in protein folding theories

Part I: Anfinsen hypothesis and the search for the global minimum in the Gibbs energy landscape

Arieh Ben-Naim
Department of Physical Chemistry
The Hebrew University of Jerusalem
Edmond J. Safra Campus
Givat Ram, Jerusalem 91904
Israel
Email: arieh@fh.huji.ac.il

Abstract

Anfinsen's thermodynamic hypothesis may be interpreted in two ways: One, that the native 3D structure of the protein resides in the global minimum in the Gibbs energy landscape (GEL). The second, that the Gibbs energy functional, has a single (hence global) minimum at the equilibrium distribution of all accessible conformations of the protein. The second is equivalent to the Second Law of Thermodynamics. The first does not follow from the Second Law, and has no theoretical justification. Therefore, the search for a global minimum in the Gibbs energy landscape is not necessarily equivalent to a prediction of the native structure of the protein.

Keywords: Proteins, Second Law, Protein Folding, Native structure, Global minimum, Minimum Gibbs Energy, Anfinsen's Hypothesis.

Council for Innovative Research
Peer Review Research Publishing System

Journal: Journal of Advances in Chemistry

Vol 3, No. 2

editor@cirworld.com

www.cirworld.com, member.cirworld.com

ISSN 2321-807X

I. Introduction

Nowadays, the protein folding problem (PFP) is considered to be one of the most challenging problem in molecular biology. In 2006 the editors of "Science" listed the PFP as one of the 125 big questions of science.[1]

"*Can we predict how protein will fold? Out of a near infinitude of possible ways to fold, a protein picks one in just tens of microseconds. The same task takes 30 years of computer time."*

Actually, two different questions are contained in this quotation; the question of the "predictability" of the structure of the protein, and the question regarding the "speed" of the folding process.

This article is devoted to the "predictability" problem, and the methods developed by many scientists to solve this problem, i.e. to deciphering the so-called "folding code."[2]The idea of the existence of a "folding code," as well as a possible hint on how to find it are already contained in the writing of Anfinsen.[2,3]

"The studies on the renaturation of 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 inter-atomic interactions and hence by the amino acid sequence, in a given environment. In terms of natural selection through the 'design' of the macromolecules during evolution, this idea emphasized the fact that a protein molecule only makes stable, structural sense when it exists under condition similar to those for which it was selected, the so-called physiological value."

From the statement that the native conformation is determined by the sequence of amino acid, people concluded that there should be a "code" that translates from a sequence of amino acids to a 3D structure.

It was recently argued[2] that there cannot be such a "code" in the usual sense of the term "code." However, many scientists have used a more expanded meaning of the term "code," i.e. a computer algorithm which when fed with a sequence of amino acids as input will provide a 3D structure as an output. During the years various sophisticated algorithms were developed to predict the 3D structure of proteins. In a recent article we find the following statements.[4]

"It is well known that the folded state of a protein is the conformation of the lowest free energy…At sufficiently low temperature the structure of the minimal free energy corresponds to the Global minimum of its force field – also known as the potential energy surface (PES).

The protein structure prediction problem is the problem of finding the global minimum of the PES, while the potential has a huge number of local minima."

Then the authors continue to describe the methodology of finding the global minimum of the PES, and comparing their approach with other approaches in the literature. If one examines all the references provided in the article[4-15] one will find statements similar to the above quoted one citing references to older articles, and eventually leading to Anfinsen's thermodynamic hypothesis.[3] Since Anfinsen is considered to be the ultimate authority on this subject not too many authors questioned the validity of his hypothesis.

The purpose of this article is not to criticize any of the algorithms developed in order to find the global minimum in the Gibbs energy landscape (GEL) in the hope of finding where the native structure of the protein resides, hence deciphering the "folding code," Instead, we shall examine a more fundamental question. Does the native conformation of the protein really reside at the minimum of the Gibbs energy landscape (GEL)? In other words to the statement "it is well known," but it this statement also true?

We shall see in the following sections that the search for the global minimum in the GEL is a result of misinterpretation of Anfinsen's thermodynamic hypothesis. Furthermore, we shall see that most people who search for the global minimum in the GEL, are actually searching for a minimum in another function which is not even related to the GEL.

II. What can we, and what can't we infer from Anfinsen's hypothesis?

Reading carefully through the quotation from Anfinsen's article we find words and phrases that "belong" to two different statements:

(i) The 3D structure of the native protein is at its lowest minimum in the GEL.
(ii) The free energy of the whole system is lowest.

The second (ii) is simply a statement of the Second Law of Thermodynamics. If that is the meaning of Anfinsen's hypothesis then it is trivially true. However, if the first (i) meaning is accepted, then it is at best a speculation, and most likely to be wrong.

What did Anfinsen himself mean by his thermodynamic statement? My guess is that Anfinsen used the Second Law (ii), which is true, to imply (i). This seems to be the implication adopted by most researchers following Anfinsen leading to the strong motivation for searching for a global minimum in the GEL. In this view, a method of finding the global minimum would be tantamount to predicting the 3D structure of proteins, hence a solution to the protein folding problem. In my view there exists no solution to the prediction problem, and searching for such a solution in the "global minimum" is wishful thinking.

ISSN 2321-807X

The origin of the confusion between the two statements is that both statements (i) and (ii) refer to the "lowest free energy." Unfortunately, both statements do not specify the variable with respect to which the Gibbs energy function has a minimum. This is a very subtle point which needs clarification.

To do so we need to define the various functions involved. Any polypeptide of M amino acids has at least 2M internal rotational angles. The specification of all these angles defines a conformation of the protein. We shall use the notation $\psi = \psi_1, \cdots, \psi_n$ to describe the entire conformation of the protein, n includes all rotational angles along the backbone, as well as any other angles of internal rotation in the side chains. For each conformation ψ we can define the potential energy function, or the energy landscape (EL),

$$E(\psi) = U(\psi) - U(\psi_0) \qquad (2.1)$$

Thus, $E(\psi)$ is the potential energy of rotation about all the angles $\psi_1, \cdots, \psi_n$. ψ is measured with respect to a chosen reference conformation ψ_0, which may be chosen as the zero of the potential energy.

The EL is relevant to the internal energy of the single protein molecule in vacuum. In solution we need the GEL. This function may be obtained by adding to EL the solvation Gibbs energies[16] of the protein at the two conformations ψ and ψ_0. See Figure 1.

$$G(\psi) = E(\psi) + \Delta G^*(\psi) - \Delta G^*(\psi_0) \qquad (2.2)$$

ΔG^* is the solvation Gibbs energy from the gaseous phase (g) to the liquid phase (l).

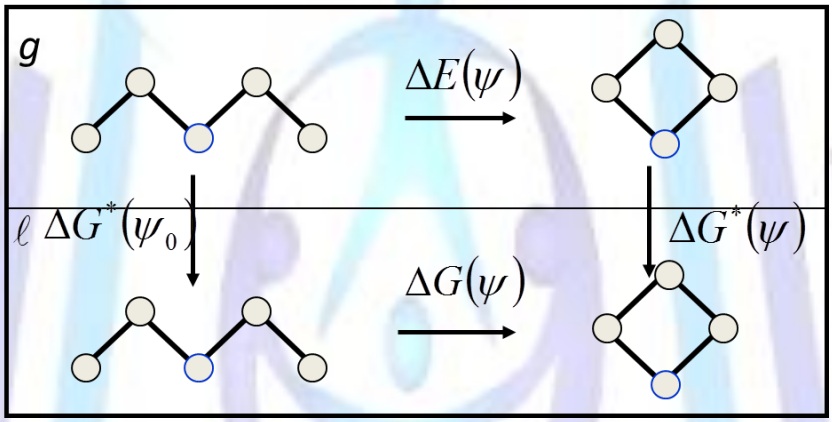

Figure 1 A schematic process of conformational change showing the relationships between the change in Gibbs energy, energy and the corresponding solvation Gibbs energies.

Since we are only interested in the shape of the GEL we may choose $\Delta G^*(\psi_0)$ as zero, much as we choose the potential energy at ψ_0 as our zero reference of the potential energy. Thus, we define the GEL up to an additive constant as the function

$$G = f(T, P, N_w; \psi) = E(\psi) + \Delta G^*(\psi) \qquad (2.3)$$

We use the notation $f(\psi)$ to define the Gibbs energy of the system of one protein at a specific conformation ψ, in a solution of composition $N_w = N_1, \ldots, N_c$, at some specific temperature T and pressure P.

When we talk about the GEL we refer to the function $G = f(T, P, N_w; \psi_1, \cdots, \psi_n)$ at a given T, P, N_w. A minimum in the GEL is a minimum of the function f with respect to the variables $\psi_1, \cdots, \psi_n$. We discussed specific examples of a GEL for a molecule having only one angle of internal rotation in reference 2.

The Second Law of thermodynamics also makes a statement about the minimum of the Gibbs energy which is different from the minimum in the GEL. For any given environment T, P, N_w there exists an equilibrium probability distribution of conformations $Pr^{eq}(\psi_1, \cdots, \psi_n)$. The Second Law states that the Gibbs energy functional $G = F(T, P, N_w; Pr(\psi))$ has a single (hence absolute) minimum over all possible distribution functions $Pr(\psi)$, at the "point" $Pr^{eq}(\psi)$. We use the letter F to distinguish the Gibbs energy functional from the GEL which we denoted by f. Thus, at equilibrium the Gibbs energy functional $G = F(T, P, N_w; Pr(\psi))$ has a single (absolute) minimum at the equilibrium distribution $Pr^{eq}(\psi)$. This is statement (ii) as referred to above. (This is also equivalent to the Second Law stated in terms of a maximum of the entropy in an isolated system[17]).

The GEL can have zero, one or many minima and maxima.[2] The Second Law does not state anything about the existence, nor about the number of minima or maxima of the GEL.

Thus, both statements (i) and (ii) refer to the "Gibbs energy minimum." However, one is a minimum of the GEL, i.e. of the function $G = f(T, P, N_w; \psi)$. The second is a minimum of the functional $G = F(T, P, N_w; Pr(\psi))$.In both T, P, N_w are constants.

It is unfortunate that a great amount of effort, as well as time and money has been expended in searching for the absolute minimum in the GEL. This is an extremely difficult mathematical, and computational problem. Unfortunately, even if such an absolute minimum in the GEL is found, it would not necessarily correspond to the conformation of the native structure of the protein. In the next sections we shall examine a few simple versions of the GEL, and the corresponding Gibbs energy functional. It will be shown that one can infer from Anfinsen what everyone already knows, that statement (ii) is true. On the other hand, one cannot in general infer that statement (i) is true, i.e. that the native conformation of the protein resides at the absolute minimum of the GEL.

III. The existence of a single minimum of the Gibbs energy functional

In this and in the following sections, we shall consider a system of N_s solute molecules (protein) in a solvent consisting of N_w solvent molecules (water) at a given temperature T and pressure P. We shall assume that the solute is much diluted in the solvent ($N_s \ll N_w$). If the solvent consists of several components, then we reinterpret N_w to be the composition vector of all solvent molecules. The triplet of variables T, P, N_w will be referred to as the "environment."

Each solute molecule is characterized by its conformation which is determined by specifying all the angles of internal rotations. For simplicity of notation we assume that there is only one angle of internal rotation denoted ϕ. Initially, we assume that there is a finite number of conformations characterized by the angles $\phi_1, \cdots, \phi_m$. Later, we take the limit of a continuous variation in ϕ.

Note carefully that in this section we characterize the conformation of the solute by one rotational angle ϕ, attaining m different (discrete) values $\phi_1, \cdots, \phi_m$. This is different from the notation in the previous section where the conformation of the protein is characterized by n angles $\psi_1, \cdots, \psi_n$, each of these can attain m different values.

Thus, our system consists of N_s solute molecules distributed in m different conformations; $\phi_1, \cdots, \phi_m$. We denote by $N_1, \ldots, N_m$ the composition of the solute molecules where N_i is the average number of solute molecules in the conformation ϕ_i. We also assume that all of the m conformations are accessible at the given environment, T, P, N_w.

The system can be viewed in two equivalent ways. Either as a two component system, w and s. In this view the Gibbs energy of the system is given by

$$G = N_w\mu_w + N_s\mu_s \qquad (3.1)$$

In the second, we view the system as $m + 1$ components, with composition $N_w, N_1, \ldots, N_m$. In this view we express the Gibbs energy as

$$G = N_w\mu_w + \sum_{i=1}^{m} N_i\mu_s(i) \qquad (3.2)$$

where $\mu_s(i)$ is the chemical potential of the solute being at a specific conformation ϕ_i

At equilibrium the Gibbs energy must have a single minimum with respect to the variables $N_1, N_2, \ldots, N_m$ with the constraint $\sum_{i=1}^{m} N_i = N_s$.

The condition for the minimum is easily obtained by using the method of Lagrange multiplier. We define the function

$$K(N_1, N_2, \ldots, N_m) = G(N_1, \ldots, N_m) - \lambda(\sum_{i=1}^{m} N_i - N_s) \qquad (3.3)$$

$$\frac{\partial K}{\partial N_i} = \frac{\partial G}{\partial N_i} - \lambda = \mu_s(i) - \lambda = 0 \qquad (3.4)$$

Hence

$$\mu_s(i) = \lambda \quad \text{for each i} \qquad (3.5)$$

At equilibrium we must have the equality

$$G = N_w\mu_w + N_s\mu_s = N_w\mu_w + \sum_{i=1}^{m} N_i\mu_s(i) \qquad (3.6)$$

Hence, we arrive at the general condition of equilibrium

$$\mu_s = \mu_s(i) = \mu_s(2) = \cdots \mu_s(m) \qquad (3.7)$$

This simply means that the chemical potential of the solute is equal to the chemical potential of any of its component (i). We define the mole fractions of the species i

$$x_i = N_i/N_s \qquad (3.8)$$

Thus, we showed that for a given environment the function $G(T, P, N_w; x_1, \cdots, x_m)$ has a minimum at equilibrium. The equilibrium distribution $x_1^{eq}, \ldots, x_m^{eq}$ is derived from Eq. (3.7), and given below in Eq. (3.12). Equivalently, we can say that the function $G(T, P, N_w; x_1, \cdots, x_m)$ has a minimum at equilibrium with the constraint $\sum_{i=1}^{m} x_i = 1$.

Next, we show that this minimum is a unique one, i.e. that there exists a unique distribution of species $x_1, \cdots, x_m$ which is referred to as the equilibrium distribution at which the Gibbs energy has a single absolute minimum.

To do this, suppose we start with any arbitrary initial distribution of species, denoted by $x_1^{in}, x_2^{in}, \ldots, x_m^{in}$. We shall show that starting with any arbitrary initial distribution $x_1^{in}, \ldots, x_m^{in}$ when releasing the constraint on the fixed distribution, the

system will evolve to a unique distribution $x_1^{eq}, \dots, x_m^{eq}$ for which the Gibbs energy is a minimum, i.e. we shall show that the Gibbs energy change for the process

$$(x_1^{in}, \dots, x_m^{in}) \rightarrow (x_1^{eq}, \dots, x_m^{eq}) \qquad (3.9)$$

Is always negative for any initial distribution. This is true provided all the states, i.e. angles $\phi_1, \cdots, \phi_m$ are accessible.

We write the chemical potential of the solute s as [16]

$$\mu_s = \mu_s^* + k_B T \ln \rho_s \Lambda_s^3 \qquad (3.10)$$

where Λ_s^3 is the momentum partition function of the solute s, $\rho_s = N_s/V$ is the number density of s, μ_s^* is the pseudo chemical potential defined in (3.10).[5] For very dilute solutions μ_s^* is independent of ρ_s. The pseudo chemical potential is the Gibbs energy change for inserting a solute at a fixed position in the solvent.

Also, for each species i of the solute we write its chemical potential as:

$$\mu_s(i) = \mu_s^*(i) + k_B T \ln \rho_i \Lambda_i^3 \qquad (3.11)$$

where $\rho_i = N_i/V$ and $\Lambda_i^3 = \Lambda_s^3$ for all i. We have denoted by $\mu_s(i)$ the chemical potential of the solute at a specific conformation ϕ_i.

From (3.10) and (3.11) we get the equilibrium distribution of the species

$$x_i^{eq} = \frac{\rho_i}{\rho_s} = \exp\left[\beta\left(\mu_s^* - \mu_s^*(i)\right)\right] \qquad (3.12)$$

Summing over all the species i we get from (3.12)

$$1 = \sum x_i^{eq} = \exp[\beta\mu_s^*] \sum_{i=1}^{m} \exp[-\beta\mu_s^*(i)] \qquad (3.13)$$

or equivalently

$$\exp[-\beta\mu_s^*] = \sum_{i=1}^{m} \exp[-\beta\mu_s^*(i)] \qquad (3.14)$$

which is the relationship between the pseudo chemical potential of s and the pseudo chemical potentials of all of the species.[16]

Since each of the terms on the rhs of (3.14) is positive we must have the inequality

$$\exp[-\beta\mu_s^*] \geq \exp[-\beta\mu_s^*(i)] \quad \text{for each i} \qquad (3.15)$$

or equivalently

$$\mu_s^* \leq \mu_s^*(i) \qquad (3.16)$$

The last inequality means that if we start with a system of N_s solute molecules, each at a fixed position, and fixed conformation ϕ_i, then releasing the constraint on a fixed conformation will always result in lowering the Gibbs energy of the system, i.e.

$$N_s\mu_s^* - N_s\mu_s^*(i) \leq 0 \quad \text{for each i} \qquad (3.17)$$

We now generalize these results in two steps. First, suppose we start with N_s solute molecules each at a fixed position (but far from each other) with composition $x_1^{in}, x_2^{in}, \dots, x_m^{in}$, and relax the constraint on the fixed composition. The Gibbs energy change of the system must be negative. To show that we multiply (3.17) by x_i^{in} and sum over i to obtain

$$\sum_{i=1}^{m} x_i^{in}[\mu_s^* - \mu_s^*(i)] = \mu_s^* - \sum_{i=1}^{m} x_i^{in}\mu_s^*(i)$$

$$= \sum_{i=1}^{m} x_i^{eq}\mu_s^* - \sum_{i=1}^{m} x_i^{in}\mu_s^*(i) \leq 0 \qquad (3.18)$$

where $\sum_{i=1}^{m} x_i^{eq} = 1$.

This is an important inequality. It states that the change in Gibbs energy of a system of solute particles at fixed positions (but far from each other) from any initial distribution x^{in} to the final distribution x^{eq} is always negative. This quantity may be referred to as the intrinsic change in the Gibbs energy, due to the release of the constraint only on a fixed distribution of conformations. This quantity does not involve the "liberation" Gibbs energies of the particles.[16] It should be noted that an inequality of the type (3.16) does not hold in general for the chemical potentials of s and of the species i. In general we have

$$\mu_s - \mu_s(i) = \mu_s^* - \mu_s^*(i) + k_B T \ln \rho_s \Lambda_s^3 - k_B T \ln \rho_i \Lambda_i^3$$

$$= [\mu_s^* - \mu_s^*(i)] + [-k_B T \ln x_i] \qquad (3.19)$$

The first term in square brackets on the rhs of (3.19) is negative by (3.16). The second term is always positive. Therefore, the sign of the difference $\mu_s - \mu_s(i)$ depends on whether x_i is larger or smaller than x_i^{eq} (Eq. 3.12). On the other hand, the average of $\mu_s - \mu_s(i)$ with the distribution x^{in} is always negative, see below Equations (3.20) and (3.21).

Next, we show the stronger statement. We start with any initial composition of the solute molecules $x^{in} = (x_1^{in}, \cdots, x_m^{in})$. The solute particles are not restricted to fixed positions. They are free to wander in the system. We relax the constraint on the fixed conformations and show that the Gibbs energy change must be negative.

The total change in the Gibbs energy for the process $x^{in} \rightarrow x^{eq}$ is

$$\Delta G = [N_w\mu_w + N_s\mu_s] - \left[N_w\mu_w - \sum_{i=1}^{m} N_1^{in}\,\mu_s(i)\right]$$

The change per solute molecule is

$$\Delta g = \frac{\Delta G}{N_s} = \mu_s - \sum_{i=1}^{m} x_i^{in}\mu_s(i)$$

$$= \left[\mu_s^* - \sum_{i=1}^{m} x_i^{in}\mu_s^*(i)\right] + k_B T \ln\rho_s\Lambda_s^3 - \sum_{i=1}^{m} x_i^{in}\, k_B T \ln\rho_i\Lambda_i^3 \qquad (3.20)$$

The three terms on the rhs of (3.20) correspond to the three processes: (a) the intrinsic change in Gibbs energy due to the change in conformation (see 3.18). (b) the "liberation" Gibbs energy of a single solute s. This is the change in Gibbs energy due to the release of a solute particle from a fixed position in the solvent w. (c) the loss of the "liberation" Gibbs energy of all the solute species due to fixing their locations.[16]

Using the (3.12) we can rewrite (3.20) as

$$\Delta g = k_B T \sum_{i=1}^{m} x_i^{in} \ln x_i^{eq} - k_B T \sum_{i=1}^{m} x_i^{in} \ln x_i^{in}$$

$$= k_B T \ln \sum_{i=1}^{m} x_i^{in} \ln \frac{x_i^{eq}}{x_i^{in}} \leq 0 \qquad (3.21)$$

The last inequality in (3.21) is valid for any two distributions.[17, 18] The equality holds if and only if $x_i^{in} = x_i^{eq}$ for all i.(This inequality, with a minus sign, is also known as the Kullback-Leibler distance between the two distributions).[17,18]

Thus, we also showed that there exists a specific distribution x^{eq}(given by 3.12) at which the Gibbs energy has a single minimum over all possible distributions x^{in}. We write this function as

$$G = F[T, P, N_w; x(\phi)] \qquad (3.22)$$

where $x(\phi) = (x_1(\phi_1), x_2(\phi_2), \cdots, x_n(\phi_n))$. This is the composition vector of the solute particles.

We have shown that the function $F(T, P, N_w; x(\phi))$ has a minimum for any discrete vector $x(\phi)$. However, we can always go to the limit of a continuous distribution function, and interpret the vector $x(\phi)$ as a function $x(\phi)$. In which case $F(T, P, N_w; x(\phi))$ is a functional of the continuous distribution $x(\phi)$.

Next, we show that $x(\phi)$ is related to the Gibbs energy landscape (GEL).

In the same system characterized by (T, P, N_w, N_s) with $N_s \ll N_w$ we can ask for the probability of finding a single solute particle at any specific angle ϕ.

The probability density is given by

$$Pr(\phi) = \frac{\exp[-\beta f(\phi)]}{\int_0^{2\pi} \exp[-\beta f(\phi)]\, d\phi} \qquad (3.23)$$

where $f(\phi)$ is the Gibbs energy of a system at a given T, P, N_w, and one solute particle at a specific angle ϕ. This is the GEL of the solute having a single rotational degree of freedom. Thus, the GEL in this system is

$$G = f(T, P, N_w; \phi) \qquad (3.24)$$

In a system characterized by T, P, N_w, N_s with $N_s \ll N_w$, the probability density $Pr(\phi)$ is the same as the function $x(\phi)$, where $x(\phi)d\phi$ is the mole fraction of solute molecules being at conformations between ϕ and $\phi + d\phi$.

Since $Pr(\phi)$, (or $x(\phi)$) is uniquely determined by the function $f(\phi)$ we can conclude that there exists an equilibrium GEL, denoted by $f^{eq}(\phi)$ which minimizes the functional $G = F(T, P, N_w, N_s; f(\phi))$, and the value of the Gibbs energy G at the minimum is the same as the one obtained for the equilibrium distribution $x^{eq}(\phi)$. (We use F to distinguish this functional from F defined in (3.22).

We have used the notation $f(\phi)$ and $F[x(\phi)]$ to distinguish between two different functions. Both are Gibbs energies. However, the first is a GEL, i.e. the Gibbs energy as a function of the angle ϕ. The second is a functional of the function $x(\phi)$.

The Second Law of thermodynamics states that at given environment (T, P, N_w, N_s), the functional F has a single (absolute) minimum at an equilibrium distribution $x^{eq}(\phi)$, or equivalently at the equilibrium GEL, $f^{eq}(\phi)$. The Second Law is a statement of an existence of an equilibrium GEL. It does not make any statement regarding the number of minima or maxima in the GEL, nor about the relative depths of the minima in the GEL. We shall discuss a few examples in the following sections.

Before generalizing for proteins we repeat the conclusion of this section. The Second Law of thermodynamics states that at a given T, P, N_w, N_s, when we release the constraint on the distribution of species $x^{in}(\phi)$, the Gibbs energy functional $G = F[x(\phi)]$ will reach a single absolute minimum at the equilibrium distribution $x^{eq}(\phi)$. The equilibrium distribution $x^{eq}(\phi)$ is related to the Gibbs energy landscape $G = f(\phi)$ by equation (3.23). Therefore, we can also say that the Gibbs energy functional $G = F[f(\phi)]$ has a unique minimum at the equilibrium GEL, corresponding to the equilibrium distribution $x^{eq}(\phi)$. It is now clear why we have used the different notation for the value of the Gibbs energy; the functional F and the GEL f. Using the same letter for the three, we would have the equation $G = G[T, P, N; G(\phi)]$ which could be potentially misleading since each of the three Gs has a different meaning; the name of the function f, the name of the functional F, and the value of the Gibbs energy G.

The generalization to proteins is quite straightforward. Simply replace the single internal rotational angles by any number of rotational angles. Alternatively, reinterpret ϕ in the above discussion as the vector of all rotational angles ψ as we have done in section I.

IV. Some concluding remarks

We have shown that the Second Law of thermodynamics is equivalent to the statement that the Gibbs energy has a single minimum with respect to all possible probability distributions of the conformations of the protein. The distribution that minimizes the Gibbs energy functional is related to the Gibbs energy landscape (GEL). The Second Law does not state anything about the shape of the GEL, nor on the number of minima and maxima of the GEL. The Second Law does not state that the native structure of the protein must be at the global minimum of the GEL.

Thus, the endeavor in searching for a global minimum in the GEL is not only unjustified theoretically, but even if one finds the global minimum, it is extremely unlikely that this minimum will coincide with the native conformation.

The unfortunate conclusion is that all the efforts expended in the search of the global minimum of the GEL are futile.

In most of the literatures we find that people looked at global minimum in the EL rather than in the GEL. This is sometimes referred to as the potential energy surface (PES). Note that the EL is relevant to a protein in vacuum.[4,19,20] It is sometimes argued that at $T \to 0$ the ground state must be occupied by all the molecules. This is true in the case that all the conformations are accessible. On the other hand, in solutions the GEL is the relevant quantity. In this case the argument that at $T \to 0$ the protein must attain the absolute minimum does not apply. Besides, the usage of the EL (or PES) cannot explain both heat and cold denaturation.[2]

There exists also confusion between the energy landscape, EL, and the thermodynamic internal energy of the system. These correspond to two different splitting of the Gibbs energy changes. Suppose we perform the process of transferring from one (initial) conformation at a fixed position to another (final) conformation, as in Figure 1. The Gibbs energy change for the process is

$$\Delta G(i \to f) = \Delta E(i \to f) + \Delta G_f^* - \Delta G_i^* \qquad (4.1)$$

On the other hand, we have the thermodynamic identity

$$\Delta G(i \to f) = \Delta H(i \to f) - T\Delta S(i \to f) \qquad (4.2)$$

Where ΔH and ΔS are the enthalpy and the entropy changes of the process $(i \to f)$ in the solution. Each of these can be written as

$$\Delta H(i \to f) = \Delta E(i \to f) + \Delta H_f^* - \Delta H_i^* \qquad (4.3)$$

$$\Delta S(i \to f) = \Delta S_f^* - \Delta S_i^* \qquad (4.4)$$

Since the process $(i \to f)$ is from a fixed conformation (i) to a fixed conformation (f), the change in the enthalpy in vacuum is $\Delta E(i \to f)$, and the change in entropy in vacuum is zero.

Therefore, from (6.5), (6.6) and (6.7) it follows that $T \to 0$.

$$\Delta G(i \to f) \to \Delta E(i \to f) + \Delta H_f^* - \Delta H_i^* \qquad (4.5)$$

Thus, it is not true that as $T \to 0$, the GEL tends to $\Delta E(i \to f)$. Even at $T \to 0$, the GEL differs from the EL by the solvation enthalpies of the conformers i and f.

In some cases one starts with the EL and "corrects" it to obtain an approximation to the GEL. Thus, the approximate GEL is written as

$$GEL \approx EL + \sum_{i,j} \delta G_{ij}^{H\phi O} \qquad (4.6)$$

where $\delta G_{ij}^{H\phi O}$ is pairwise hydrophobic (HϕO) interactions between groups i and j. Unfortunately, such a corrected EL cannot be used as an approximated GEL. First, because in (4.6) one takes only HϕO interactions, and neglect the more important (both stronger and more abundant) hydrophilic interactions. Second, and not less important is that the solvent-induced part of the GEL is not pair-wise additive.[16] Thus, in effect one ends up with searching for an unnecessary absolute minimum in a very complex function which is neither an EL, nor a GEL, and in effect not relevant to protein folding at all.

The analysis made in this article has also some relevance to simulation of the folding process of either a protein of known structure, or a completely new polypeptide. In the first case, the result of the simulated experiment would depend on the initial conditions. If one starts from any arbitrary initial conformation, it is extremely unlikely that the resulting equilibrium distribution will contain the native structure. On the other hand, for a new polypeptide we cannot know in advance in which T, P, N this polypeptide will fold, if it folds, to any stable 3D structure.

The search for the 3D structure of the native protein in the absolute minimum in the Gibbs energy landscape (GEL) is oftentimes described metaphorically as “looking for a needle in a haystack.” The conclusion of this article is better described as “looking for a needle in one out of many haystacks, in which the needle might, or might not reside.” On the other hand, the search for the 3D structure in the potential energy surface (PES) can be likened to a search in the wrong haystack.

References

1. Kennedy, D. and Norman, C., Science, 309, 75 (2005)
2. Ben-Naim, A., The Protein Folding Problem and Its Solutions, World Scientific, Singapore (2013)
3. Anfinsen, C. B., Science, 181, 223-230 (1973)
4. Goldstein, M., Fredj, E., and Gerber, R. B., J. of Computational Chemistry, 32, 1785 (2011)
5. Leach, A., Molecular Modeling – Principles and Applications, 2nd edition, Prentice Hall, England (2001)
6. Schlick, T., Modeling and Simulation – An Interdisciplinary Guide, Springer-Verlag, USA (2002)
7. Finkelstein, A. V., Curr Opin Struct Biol, 7, 60-71 (1997)
8. Nolting, B., Protein Folding Kinetics – Biophysical Methods, 2nd ed., Springer-Verlag, Germany, 2006
9. Prentiss, M. C., Hardin, C., Eastwood, M. P., Zong, Ch., Wolynes, P. G., J. Chem Theory Comput, 2, 705-716 (2006)
10. Finkelstein A. V., Galzitskaya, O. V., Phys. Life Rev., 1, 23-56 (2004)
11. Helling, R., Li, H., Melin, R., Miller, J., Wingree, N., Zheng, Ch. and Tang, Ch., J. Mol. Graph Model, 19, 157-167(2001)
12. Honig, B., J. Mol. Biol. 293, 283-293 (1999)
13. Karplus, M., Fold Des, 2, S69-S75 (1997)
14. Dobson, C. M. and Karplus, M., Curr Opin Struct Biol, 9, 91-101 (1999)
15. Dinner, A. R., Sali, A., Smith, L. J., Dobson, Ch. M. and Karplus, M., Trends Biochem Sci, 25, 331-339 (2000)
16. Ben-Naim, A., Molecular Theory of Solutions, Oxford University Press, Oxford (2006)
17. Ben-Naim, A., A Farewell to Entropy. Statistical Thermodynamics Based on Information, World Scientific, Singapore (2008)
18. Papoulis, A. and Pillain S. U., Probability Random Variables and Stochastic Processes, McGraw Hill, Boston (2002)
19. Ben-Naim, A., J. Chem. Phys. 135, 085104 (2011)
20. Wales, D. J. and Doye, J. P. K., J. Phys. Chem. A, 101, 5111-5116 (1997)