Scientific
Research
Publishing

# Single Molecule Thermodynamics Hypothesis of Protein Folding and Drug Design

**Fang Yi**

The Australian National University, Canberra, Australia
Email: yi.fang3@gmail.com

## Abstract

A single molecule view of protein folding leads to the concept of conformational Gibbs free energy function $G(\mathbf{X}; \mathfrak{E}, \mathfrak{U})$, where the variable $\mathbf{X}$ is a conformation of a protein $\mathfrak{U}$. $\mathfrak{U}$ and environments $\mathfrak{E}$ (solvent, temperature, pressure, other particles in solvent, etc.) in which $\mathfrak{U}$ folds appear here as parameters. Changing environment $\mathfrak{E}$ to $\mathfrak{E}'$ changes $G(\mathbf{X}; \mathfrak{E}, \mathfrak{U})$ to $G(\mathbf{X}; \mathfrak{E}', \mathfrak{U})$, this change can explain phenomena such as folding, unfolding, docking, etc., since this indicates that in different environments $\mathfrak{U}$ has different stable conformations $\mathbf{X}_{\mathfrak{E}}$ and $\mathbf{X}_{\mathfrak{E}'}$, satisfying $\nabla G(\mathbf{X}_{\mathfrak{E}}; \mathfrak{E}, \mathfrak{U}) = \mathbf{0}$ and $\nabla G(\mathbf{X}_{\mathfrak{E}'}, \mathfrak{E}', \mathfrak{U}) = \mathbf{0}$ respectively. The conformational Gibbs free energy function $G(\mathbf{X}; \mathfrak{E}, \mathfrak{U})$ also gives the deterministic part of folding force acting on an atom $\mathbf{a}_i$ of $\mathfrak{U}$, $\mathbf{F}_i = -\nabla_{\mathbf{x}_i} G(\mathbf{X}; \mathfrak{E}, \mathfrak{U})$. The equation of motion of folding is a Langevin equation. We suggest two verifiable predictions of the single molecule thermodynamic hypothesis that may confirm or refute the hypothesis: *ab initio* predictions of native structures of globular proteins and their folding paths. An application is suggested to drug design.

## Keywords

Protein Folding, Conformation, Thermodynamic System, Gibbs Free Energy, Single Molecule, Macroscopic, Quantum Statistics

## 1. Introduction: A Single Molecule View of Protein Folding

To resolve the protein folding problem, that is: predicting the native structure and describing the folding dynamics, we must work with the fundamental physical law that directly governs protein folding process. That law is the Thermodynamic Principle of Protein Folding [1], it is just the Second Law of Thermodynamics since Anfinsen and others already shown that the folding process is spontaneous. In the protein folding case, the second law is that the Gibbs free

energy achieves a minimum at the native structure.

Therefore, we have to figure out what is the Gibbs free energy. The question is, is there a Gibbs free energy function whose variables are all possible conformations of a given protein molecule? Or is it only a Gibbs free energy difference between the folded ensemble of protein molecules and its counterpart, the unfolded ensemble? The former is a single molecule view coming from contemplating a protein molecule comes out of ribosome and changes its conformations until it achieves its native structure; the latter is an ensemble view coming from staring at a tube of purified protein solution and trying to figure out the collective behaviours of the protein molecules in the solution while the solution is going towards equilibrium.

Via quantum statistics applied to a tiny thermodynamic system $\mathfrak{S}_\mathbf{X}$ which is tailor made for the conformation $\mathbf{X}$ and its immediate physiological environment $\mathfrak{E}_\mathfrak{U}$, we have derived the conformational Gibbs free energy (CGFE) function $G(\mathbf{X}; \mathfrak{E}_\mathfrak{U}, \mathfrak{U})$ for globular proteins $\mathfrak{U}$ [2] [3] [4].

Applying CGFE function we translate the thermodynamic principle of protein folding into the Single Molecule Thermodynamic Hypothesis (SMTH) of Protein Folding: putting a protein molecule $\mathfrak{U}$ in an environment $\mathfrak{E}$, a stable conformation $\mathbf{X}_\mathfrak{E}$ (may not be unique) of $\mathfrak{U}$ must be a minimizer (local or global) of the CGFE function $G(\mathbf{X}; \mathfrak{E}, \mathfrak{U})$. In particular, the gradient vanishes at $\mathbf{X}_\mathfrak{E}$, $\nabla G(\mathbf{X}_\mathfrak{E}, \mathfrak{E}, \mathfrak{U}) = \mathbf{0}$.

Another big question in protein folding is that is there a folding force? Leventhal in 1969 [5] has shown by contradiction that there must be a folding force, otherwise if the folding process were only random, it would have taken a time span longer than the Earth's age.

The CGFE function also gives us the deterministic part of the folding force $\mathbf{F}_i$ acting on an atom $\mathbf{a}_i$ of $\mathfrak{U}$. It is $\mathbf{F}_i(\mathbf{X}) = -\nabla_{\mathbf{x}_i} G(\mathbf{X}; \mathfrak{E}_\mathfrak{U}, \mathfrak{U})$.

A scientific hypothesis has to give verifiable predictions to let people confirm or refute it. We suggest two verifiable predictions of the single molecule thermodynamic hypothesis.

1) *ab initio* predictions of native structures of globular proteins: the native structure $\mathbf{X}_\mathfrak{U}$ of the protein $\mathfrak{U}$ is a (local or global) minimizer of the CGFE function:

$$G(\mathbf{X}_\mathfrak{U}; \mathfrak{E}_\mathfrak{U}, \mathfrak{U}) = \min_{\mathbf{X} \in U \subset \mathfrak{X}_\mathfrak{U}} G(\mathbf{X}; \mathfrak{E}_\mathfrak{U}, \mathfrak{U}), \quad \text{for some neighbourhood } U \text{ of } \mathbf{X}_\mathfrak{U}. \quad (1)$$

Especially, since $G(\mathbf{X}; \mathfrak{E}_\mathfrak{U}, \mathfrak{U})$ is smooth, $\nabla G(\mathbf{X}_\mathfrak{U}; \mathfrak{E}_\mathfrak{U}, \mathfrak{U}) = \mathbf{0}$. In case $\mathbf{X}_\mathfrak{U}$ is a global minimizer, then

$$G(\mathbf{X}_\mathfrak{U}; \mathfrak{E}_\mathfrak{U}, \mathfrak{U}) = \min_{\mathbf{X} \in \mathfrak{X}_\mathfrak{U}} G(\mathbf{X}; \mathfrak{E}_\mathfrak{U}, \mathfrak{U}). \quad (2)$$

Here $\mathbf{X}$ denotes a conformation and $\mathfrak{X}_\mathfrak{U}$ is the set of all possible conformations of $\mathfrak{U}$.

Therefore, for a globular protein $\mathfrak{U}$ (we know the formula of $G(\mathbf{X}; \mathfrak{E}_\mathfrak{U}, \mathfrak{U})$), the prediction of native structure is reduced to a pure mathematical problem, the minimization problem of a known smooth function. A mathematical theorem

guarantees that this problem has a solution, *i.e.*, mimimizers always exist, the real task is to find them and determine which is the native structure, a hard programming problem.

2) For a globular protein $\mathfrak{U}$, starting from any initial conformation $\mathbf{X}_0$, there is a folding path $\mathbf{X}(t) = \mathbf{X}(t; \mathbf{X}_0)$ satisfying $\mathbf{X}(t_0; \mathbf{X}_0) = \mathbf{X}_0$, and for $t \geq t_0$, the following Langevin equation:

$$m_i \frac{\mathrm{d}^2 \mathbf{x}_i(t)}{\mathrm{d}t^2} = \mathbf{F}_{\text{total}} = -\nabla_{\mathbf{x}_i} G(\mathbf{X}(t); \mathfrak{E}_{\mathfrak{U}}, \mathfrak{U}) - \eta_i \frac{\mathrm{d}\mathbf{x}_i(t)}{\mathrm{d}t} + \mathbf{F}_i(t), \quad i = 1, \cdots, n. \quad (3)$$

Here $\eta_i$ is the solvent friction. The random force $\mathbf{F}_i(t)$ is caused by occasionally bumping into another non-solvent molecule. Because of it, we do not have a completely deterministic folding path. Again, mathematical theorems guarantee that such folding path $\mathbf{X}(t; \mathbf{X}_0)$ exists. Moreover, mathematics also tells us that it is highly depending on its initial conformation $\mathbf{X}_0$, so an important issue is to know the protein's initial conformation as it is out of ribosome.

If the two predictions are positively verified, then we can say that theoretically the protein folding problem is resolved, at least for globular proteins.

## 2. The CGFE Function for Globular Proteins

To explain our CGFE function, we start with a conformation $\mathbf{X}$. A protein $\mathfrak{U}$ consists of $n$ atoms $(\mathbf{a}_1, \cdots, \mathbf{a}_n)$, a conformation of $\mathfrak{U}$ can be expressed as a point of 3n-dimensional space $\mathbb{R}^{3n}$, $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$, where $\mathbf{x}_i \in \mathbb{R}^3$ is the nuclear position of $\mathbf{a}_i$ in the 3-dimensional space $\mathbb{R}^3$. Not all points in $\mathbb{R}^{3n}$ are conformations of $\mathfrak{U}$, bond lengths, bond angles, and van der Waals distances in general, are natural constraints. Denote $\mathfrak{X}_{\mathfrak{U}}$ as the set of all possible conformations of $\mathfrak{U}$. A conformational function of $\mathfrak{U}$ is a function $f : \mathfrak{X}_{\mathfrak{U}} \to \mathbb{R}$. For example, all force fields used in molecular dynamics simulations are conformational functions.

The 3-dimensional conformation $P_{\mathbf{X}}$ of $\mathfrak{U}$ is $P_{\mathbf{X}} = \bigcup_{i=1}^n \mathbf{C}_i(\mathbf{x}_i) \subset \mathbb{R}^3$, where $\mathbf{C}_i \subset \mathbb{R}^3$ is the shape of the atom $\mathbf{a}_i$ in $\mathfrak{U}$, $\mathbf{C}_i(\mathbf{x}_i)$ indicates it has been congruently moved to the nuclear position $\mathbf{x}_i$. $\mathbf{C}_i$ exists, will change with $\mathbf{X}$, see [6]. A very good approximation to $\mathbf{C}_i(\mathbf{x}_i)$ is a solid ball $B(\mathbf{x}_i, r_i)$, centred at $\mathbf{x}_i \in \mathbb{R}^3$ with radius $r_i$ that is the $\mathbf{a}_i$'s van der Waals radius. For simplicity, we adopt that $\mathbf{C}_i(\mathbf{x}_i) = B(\mathbf{x}_i, r_i)$.

In natural, and even in most of artificial environments of protein folding, the immediate environment of a globular protein $\mathfrak{U}$ is just one layer of water molecules surrounding the conformation $P_{\mathbf{X}} = \bigcup_{i=1}^n B(\mathbf{x}_i, r_i)$. This is true even the protein molecule is inside a crystal [7].

$P_{\mathbf{X}}$ plus the one layer water molecules consists of a tiny thermodynamic system $\mathfrak{S}_{\mathbf{X}}$, tailor made for $P_{\mathbf{X}}$. As an open thermodynamics system, $\mathfrak{S}_{\mathbf{X}}$ has a Gibbs free energy $G(\mathfrak{S}_{\mathbf{X}})$. The CGFE function $G(\cdot; \mathfrak{E}_{\mathfrak{U}}, \mathfrak{U}) : \mathfrak{X}_{\mathfrak{U}} \to \mathbb{R}$ then is defined as $G(\mathbf{X}; \mathfrak{E}_{\mathfrak{U}}, \mathfrak{U}) = G(\mathfrak{S}_{\mathbf{X}})$.

Between $P_{\mathbf{X}}$ and the layer of water molecules, is an interface $M_{\mathbf{X}}$, for example, the solvent accessible surface $\partial P_{\mathbf{X}}$. The expression of $G(\mathfrak{S}_{\mathbf{X}})$ is via global

geometric features of $M_{\mathbf{X}}$ and its surface chemical potentials. A protein molecule has many moieties or atom groups, some are charged, some are polar, others are non-polar. They can be classified into hydrophobicity classes $H_i$, $1 \le i \le H$, $H > 1$, from most hydrophobic (non-polar) to the most hydrophilic (polar or charged). An atom $\mathbf{a}_i \in H_j$ if it belongs to a moiety of class $H_j$. Define $P_{\mathbf{X},i} = \bigcup_{\mathbf{a}_j \in H_i} B(\mathbf{x}_j, r_j)$, then $P_{\mathbf{X}} = \bigcup_{i=1}^{n} P_{\mathbf{X},i}$. The space containing water molecules in $\mathfrak{S}_{\mathbf{X}}$, the ring $\mathfrak{R}_{\mathbf{X}} = \overline{\mathfrak{S}_{\mathbf{X}} \setminus P_{\mathbf{X}}}$, is decomposed into $H$ parts (not necessarily connected) via the distance function $dist(\mathbf{x}, P_{\mathbf{X},i}) = \min_{\mathbf{y} \in P_{\mathbf{X},i}} |\mathbf{x} - \mathbf{y}|$, see **Figure 1**.

$$\mathfrak{R}_{\mathbf{X},i} = \{\mathbf{x} \in \mathfrak{R}_{\mathbf{X}} : dist(\mathbf{x}, P_{\mathbf{X},i}) \le dist(\mathbf{x}, P_{\mathbf{X},j}), \text{for any } j \ne i\}, i = 1, \cdots, H. \quad (4)$$

The interface $M_{\mathbf{X}}$ then is decomposed accordingly

$$M_{\mathbf{X}} = \bigcup_{i=1}^{H} M_{\mathbf{X},i}, \quad M_{\mathbf{X},i} = M_{\mathbf{X}} \bigcap \mathfrak{R}_{\mathbf{X},i}. \quad (5)$$

A water molecule in $\mathfrak{R}_{\mathbf{X},i}$ will touch $M_{\mathbf{X},i}$, so it will be attracted ($H_i$ is charged or polar) or repulsed ($H_i$ is non-polar) by $P_{\mathbf{X}}$. So the same water molecule in different $\mathfrak{R}_{\mathbf{X},i}$ has different chemical potentials $\mu_i$. For non-polar $H_i$, $\mu_i > 0$, for charged or polar $H_i$, $\mu_i < 0$. Thus there is a $1 < k < H$, such that

$$\mu_1 > \mu_2 > \cdots > \mu_k > 0 > \mu_{k+1} > \cdots > \mu_H. \quad (6)$$

$\bigcup_{i=1}^{k} M_{\mathbf{X},i}$ is called hydrophobic surface, $\bigcup_{i=k+1}^{H} M_{\mathbf{X},i}$ is called hydrophilic surface.

Since water molecules and electrons can enter or leave $\mathfrak{S}_{\mathbf{X}}$, $\mathfrak{S}_{\mathbf{X}}$ is an open thermodynamic system. Hence, the ensemble used here is the grand canonical ensemble. Applying quantum statistics to the open system $\mathfrak{S}_{\mathbf{X}}$, the number $N_i$ of water molecules in $\mathfrak{R}_{\mathbf{X},i}$ turns to be a Hermitian operator $\hat{N}_i$ with mean value $\langle \hat{N}_i \rangle(\mathbf{X})$. The same is true to the number of electrons $N_e$. The Gibbs free energy of $G(\mathfrak{S}_{\mathbf{X}}) = G(\mathbf{X}; \mathfrak{E}_{\mathfrak{U}}, \mathfrak{U})$ then is

$$G(\mathbf{X}; \mathfrak{E}_{\mathfrak{U}}, \mathfrak{U}) = \mu_e \langle \hat{N}_e \rangle(\mathbf{X}) + \sum_{i=1}^{H} \mu_i \langle \hat{N}_i \rangle(\mathbf{X}) + \langle \hat{T}_{\mathbf{X}} \rangle. \quad (7)$$

Since $P_{\mathbf{X}}$ is fixed, here the kinetic energy operator $\hat{T}_{\mathbf{X}}$ vanishes.
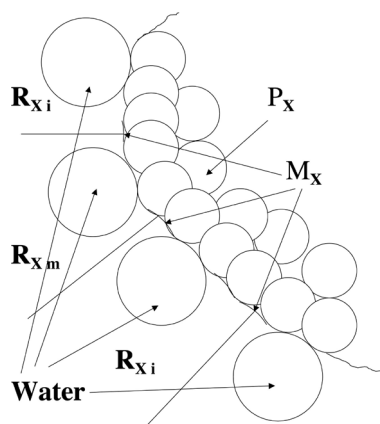


**Figure 1.** Water molecules are contained inside $\mathfrak{R}_{\mathbf{X},i}$, note it is not necessarily connected. $M_{\mathbf{X}}$ is an interface.

Since every water molecule in $\mathfrak{R}_{X,i}$ has contact with the surface $M_{X,i}$, $\langle \hat{N}_i \rangle(\mathbf{X})$ is proportional to the area $A(M_{X,i})$. Therefore, there are $\nu_i > 0$, such that

$$\nu_i A(M_{X,i}) = \langle \hat{N}_i \rangle(\mathbf{X}), \quad 1 \leq i \leq H. \tag{8}$$

Let $\mathbf{r}$ be the coordinates of one electron in $\mathfrak{S}_{\mathbf{X}}$, $\mathbf{W}$ the coordinates of $N = \sum_{i=1}^{H} N_i$ water molecules in $\mathfrak{R}_{\mathbf{X}}$. Then the electronic density distribution function $\rho_N(\mathbf{r}; \mathbf{X}, \mathbf{W})$ ((1.3) of ([6], page 6)) gives

$$N_e = \int_{\mathfrak{S}_{\mathbf{X}}} \rho_N(\mathbf{r}; \mathbf{X}, \mathbf{W}) d\mathbf{r} = \nu_{\mathbf{X},N} \int_{\mathfrak{S}_{\mathbf{X}}} d\mathbf{r} = \nu_{\mathbf{X},N} V(\mathfrak{S}_{\mathbf{X}}), \quad \nu_{\mathbf{X},N} > 0. \tag{9}$$

The $\nu_{\mathbf{X},N} > 0$ is obtained by mean value theorem of integrals.

Let $\Omega_{M_{\mathbf{X}}}$ be the domain enclosed by $M_{\mathbf{X}}$ ( $P_{\mathbf{X}} \subset \overline{\Omega_{M_{\mathbf{X}}}}$ ), we have roughly the volume $V(\mathfrak{S}_{\mathbf{X}} \setminus \Omega_{M_{\mathbf{X}}}) = d_w A(M_{\mathbf{X}})$, where $d_w$ is the diameter of a water molecule and $A(M_{\mathbf{X}})$ is the area of $M_{\mathbf{X}}$. Then taking the mean in (9) we have

$$\begin{aligned} \langle \hat{N}_e \rangle &= \langle \hat{\nu}_{\mathbf{X},N} V(\mathfrak{S}_{\mathbf{X}}) \rangle = \langle \hat{\nu}_{\mathbf{X},N} \rangle V(\mathfrak{S}_{\mathbf{X}}) = \nu_e V(\mathfrak{S}_{\mathbf{X}}) \\ &= \nu_e [V(\Omega_{M_{\mathbf{X}}}) + V(\mathfrak{S}_{\mathbf{X}} \setminus \Omega_{M_{\mathbf{X}}})] \\ &= \nu_e V(\Omega_{M_{\mathbf{X}}}) + \nu_e d_w A(M_{\mathbf{X}}). \end{aligned} \tag{10}$$

Let $\omega_e = \nu_e \mu_e$ and $\omega_i = \nu_i \mu_i$. Substitute (8) and (10) into (7), the conformational Gibbs free energy function of a globular protein $\mathfrak{U}$ in its physiological environment $\mathfrak{E}_{\mathfrak{U}}$ is

$$G(\mathbf{X}; \mathfrak{E}_{\mathfrak{U}}, \mathfrak{U}) = \omega_e V(\Omega_{M_{\mathbf{X}}}) + \omega_e d_w A(M_{\mathbf{X}}) + \sum_{i=1}^{H} \omega_i A(M_{\mathbf{X},i}). \tag{11}$$

The function is smooth, *i.e.*, the first and second derivatives exist in $\mathbb{R}^{3n}$ except at points $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ such that $\mathbf{x}_i = \mathbf{x}_j$ for some $i \neq j$. But such case cannot happen for a conformation, because of the van der Waals distances must be positive. Therefore, on $\mathfrak{X}_{\mathfrak{U}}$, $G(\mathbf{X}; \mathfrak{E}_{\mathfrak{U}}, \mathfrak{U})$ is smooth.

## 3. Explanations of Protein Folding

A scientific hypothesis has to be able to explain natural and artificial phenomena. We will explain several phenomena of protein folding, unfolding, and docking, and suggest an application to drug design, according to the SMTH.

### 3.1. What the CGFE Function in (11) Reveal?

It is well known ([8] and [9]) that native structures of globular proteins have three important global geometric features. Comparing to unfolded conformations, they have: 1) Smaller volume $V(\Omega_{M_{\mathbf{X}}})$; 2) Smaller surface area $A(M_{\mathbf{X}})$; and 3) Compactly packed hydrophobic cores.

Hence folding towards native structure, the volume and area are going to shrink. Then the first two terms in (11) must have positive coefficients, *i.e.*, $\mu_e > 0$, otherwise shrinking volume and area would have enlarged the Gibbs free energy.

As the hydrophobic core, look at (11), the hydrophobic surfaces $\bigcup_{i=1}^{k} M_{\mathbf{X},i}$ have positive chemical potentials $\mu_i$, thus shrinking $A(\bigcup_{i=1}^{k} M_{\mathbf{X},i}) = \sum_{i=1}^{k} A(M_{\mathbf{X},i})$ will reduce the Gibbs free energy. On the other hand, for hydrophilic surface $\bigcup_{i=k+1}^{H} M_{\mathbf{X},j}$, $\mu_i < 0$, therefore, enlarging $A(\bigcup_{i=k+1}^{H} M_{\mathbf{X},j})$ will reduce the Gibbs free energy. Enlarging of hydrophilic surface area $A(\bigcup_{j=k+1}^{H} M_{\mathbf{X},j})$ is equivalent to shrinking the hydrophobic surface area $A(\bigcup_{i=1}^{k} M_{\mathbf{X},i})$, since

$A(M_{\mathbf{X}}) = A(\bigcup_{i=1}^{k} M_{\mathbf{X},i}) + A(\bigcup_{j=k+1}^{H} M_{\mathbf{X},j})$ is also shrinking towards the native structure.

According to the SMTH, predicting native structure of a globular protein is to minimize the CGFE function (11) as in (1) and (2). As analysed above, it is essentially making an ever better hydrophobic core by shrinking volume, area, and hydrophobic area simultaneously and cohesively. This is the "cooperativity" searched in [10], of "the concurrent participation of different regions of the biomolecule to promote and sustain intramolecular or intermolecular interactions".

One may ask that where are hydrogen bonds in (11)? The answer is that secondary structures and hydrogen bonds are products of minimizing the CGFE function to find the native structures. A judicious examination of an exhaustive PDB sample of small soluble globular proteins of moderate size (residues $N < 102$) showed that the hydrophobic collapsing is coupled with backbone hydrogen-bond formation [11]. In [12], we neglected the volume and area, only shrank the hydrophobic surface area (equivalent to make better hydrophobic core), hydrogen bonds, secondary structures such as $\alpha$ helices, $\beta$ strands, and $\beta$ turns, duly appeared with statistical significance.

The explanation is: proteins in their physiological environment are special among polymers. Polymers do not have specified structures, proteins have native structures in their physiological environment. Why? Globular proteins' peptide chains are special, folding in their physiological environment, while collapsing to hydrophobic cores the residues are putting just in places to be able to form secondary structures and hydrogen bonds simultaneously. Evolution selects the very few peptide chains to be foldable globular proteins. In fact, randomly picking a 400 residue peptide chain, the probability that it is a protein's peptide chain is at most $10^{-460}$ [2]. In any computer, $10^{-460}$ is zero.

## 3.2. Explanation of Denaturation

According to the pioneer research of denaturation [13], all known denature phenomena are caused by environment change from $\mathfrak{E}_{\mathfrak{U}}$ to some $\mathfrak{E}$. Denaturation, or unfolding, is the same as folding, only in a different environment $\mathfrak{E}$ and with a different CGFE function $G(\mathbf{X};\mathfrak{E},\mathfrak{U})$. According to the SMTH, the unfolding will end at minimizers $\mathbf{X}_{\mathfrak{E}}$'s (may not be unique, local or global), of $G(\mathbf{X};\mathfrak{E},\mathfrak{U})$.

Experiments show that the difference between folding and unfolding is that folding leads to a unique native structure, unfolding leads to many different sta-

ble conformations [1].

Either in folding or unfolding, a conformation moves along a folding (unfolding) path which satisfies an equation of motion, the Langevin Equation (3), with deterministic forces $\nabla_{x_i} G(\mathbf{X}; \mathfrak{E}_{\mathfrak{U}}, \mathfrak{U})$ and $\nabla_{x_i} G(\mathbf{X}; \mathfrak{E}, \mathfrak{U})$. This may explain why folding leads to a unique native structure and unfolding leads to many stable conformations. The random forces $\mathbf{F}_i(t)$ in the Langevin equation for folding and unfolding may be also different, in a denatured environment randomly bumping upon some other molecules will happen more often.

Moreover, the initial conformation of the folding or unfolding path also determines where the path ends. Any local minimizer $\mathbf{Y}$ has a domain of attractive basin $A_{\mathbf{Y}}$ ($U$ in (1)) such that any initial conformation in $A_{\mathbf{Y}}$ will fold to $\mathbf{Y}$. Because of evolutionary selection for a protein $\mathfrak{U}$'s peptide chain made it fit in the protein's physiological environment $\mathfrak{E}_{\mathfrak{U}}$, the attractive basin of $\mathbf{X}_{\mathfrak{U}}$ is large enough to contain all the initial conformations freshly come out of ribosome, thus even $G(\mathbf{X}; \mathfrak{E}_{\mathfrak{U}}, \mathfrak{U})$ has more than one minimizer, the native structure is the unique folding result.

But is not in denaturation, the initial conformation is the unique native structure $\mathbf{X}_{\mathfrak{U}}$? Why it unfolds to many stable denatured conformations? It is because catastrophe, a phenomenon often happens in nature, a description of it is given in [14]: "Catastrophe theory is concerned with the mathematical modelling of sudden changes—so called 'catastrophes'—in the behaviour of natural systems, which can appear as a consequence of continuous changes of the system parameters".

Actually, from physiological environment $\mathfrak{E}_{\mathfrak{U}}$ to the denaturation environment $\mathfrak{E}$, there must be a family of environments $\mathfrak{E}_t$ connecting them. Except $\mathfrak{E}_{\mathfrak{U}}$ and $\mathfrak{E}$, these environments are not in equilibrium or quasi-static so the function $G(\mathbf{X}; \mathfrak{E}_t, \mathfrak{U})$ is not well defined. Hence, although the parameter $t$ varies continuously, catastrophe does happen so that various copies of the same native structure $\mathbf{X}_{\mathfrak{U}}$ suddenly changed to different structures and when the environment finally changes to the denaturation environment $\mathfrak{E}$, these changed structures became different initial conformations of denaturation paths under $G(\mathbf{X}; \mathfrak{E}, \mathfrak{U})$.

### 3.3. Explanation of Docking

Docking is trying to bound two molecules to form a stable complex.

Let $\mathfrak{U}$ and $\mathfrak{V}$ be two molecules (proteins or others), and $\mathbf{X} \in \mathfrak{X}_{\mathfrak{U}}$ and $\mathbf{Y} \in \mathfrak{X}_{\mathfrak{V}}$. The 3-dimensional conformations $P_{\mathbf{X}}$ and $P_{\mathbf{Y}}$ are contained in their tailor made thermodynamic systems $\mathfrak{S}_{\mathbf{X}}$ and $\mathfrak{S}_{\mathbf{Y}}$ respectively in their common physiological environment $\mathfrak{E}_{\mathfrak{U}}$. In **Figure 2**, in the beginning, $\mathfrak{S}_{\mathbf{X}} \cap \mathfrak{S}_{\mathbf{Y}} = \varnothing$. Now suppose that there is a congruence $C: \mathbb{R}^3 \to \mathbb{R}^3$ brings $\mathfrak{S}_{\mathbf{X}}$ to $\mathfrak{S}_{\mathbf{Y}}$, *i.e.*, $C(\mathfrak{S}_{\mathbf{X}}) \cap \mathfrak{S}_{\mathbf{Y}} \neq \varnothing$. Van der Waals repulse tells us that it is always 1) $C(P_{\mathbf{X}}) \cap P_{\mathbf{Y}} = \varnothing$ and 2) $C(\mathfrak{R}_{\mathbf{X}}) \cap \mathfrak{R}_{\mathbf{Y}} \neq \varnothing$. Thus, we define $\mathbf{X}$ and $\mathbf{Y}$
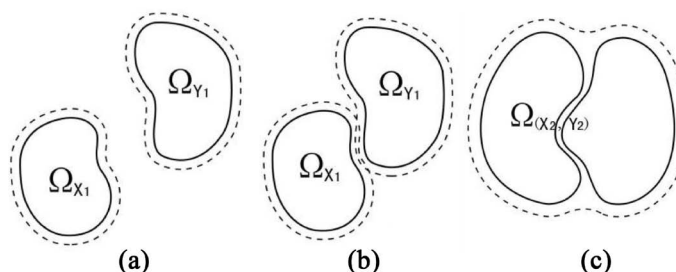
**Figure 2.** (a) Two independent molecules; (b) Invading each other; (c) Forming new minimiser $(\mathbf{X}, \mathbf{Y}) \in \mathfrak{X}_{(\mathfrak{U}, \mathfrak{V})}$.

binding if and only if 1) $P_{\mathbf{X}} \bigcap P_{\mathbf{Y}} = \varnothing$ and 2) $\mathfrak{R}_{\mathbf{X}} \bigcap \mathfrak{R}_{\mathbf{Y}} \neq \varnothing$.

If $\mathbf{X} \in \mathfrak{X}_{\mathfrak{U}}$ and $\mathbf{Y} \in \mathfrak{X}_{\mathfrak{V}}$ binding, there is a binding energy $E(\mathbf{X}, \mathbf{Y})$ depending on the net effect of the way of bring $P_{\mathbf{X}}$ to close to $P_{\mathbf{Y}}$. Neglecting $E(\mathbf{X}, \mathbf{Y})$, consider $\hat{G}((\mathbf{X}, \mathbf{Y}); \mathfrak{E}_{\mathfrak{U}}, (\mathfrak{U}, \mathfrak{V}))$ the Gibbs free energy for the conformation $(\mathbf{X}, \mathbf{Y})$ of the "molecule" $(\mathfrak{U}, \mathfrak{V})$), then

$$G((\mathbf{X}, \mathbf{Y}); \mathfrak{E}_{\mathfrak{U}}, (\mathfrak{U}, \mathfrak{V})) = \hat{G}((\mathbf{X}, \mathbf{Y}); \mathfrak{E}_{\mathfrak{U}}, (\mathfrak{U}, \mathfrak{V})) + E(\mathbf{X}, \mathbf{Y}). \tag{12}$$

If $\mathfrak{E}_{\mathfrak{U}}$ is just water, then $\hat{G}((\mathbf{X}, \mathbf{Y}); \mathfrak{E}_{\mathfrak{U}}, (\mathfrak{U}, \mathfrak{V}))$ is given by (11) for the conformation $(\mathbf{X}, \mathbf{Y})$ of the "molecule" $(\mathfrak{U}, \mathfrak{V})$. If $\mathbf{X}$ and $\mathbf{Y}$ are not binding, then $E(\mathbf{X}, \mathbf{Y}) = 0$, thus (12) is true for any $(\mathbf{X}, \mathbf{Y}) \in \mathfrak{X}_{(\mathfrak{U}, \mathfrak{V})}$.

In general, even $\mathbf{X}_{\mathfrak{U}}$ and $\mathbf{Y}_{\mathfrak{U}}$ are minimizers of $G(\mathbf{X}; \mathfrak{E}_{\mathfrak{U}}, \mathfrak{U})$ and $G(\mathbf{Y}, \mathfrak{E}_{\mathfrak{U}}, \mathfrak{V})$ respectively, $(\mathbf{X}_{\mathfrak{U}}, \mathbf{Y}_{\mathfrak{U}})$ will not be a minimizer of $G((\mathbf{X}, \mathbf{Y}); \mathfrak{E}_{\mathfrak{U}}, (\mathfrak{U}, \mathfrak{V}))$.

## 3.4. Drug Design

We address a question asked in [10], "the rational drug designer faces a many-body problem: the interactions between the protein target and the drug/ligand involve more than groups matched up in a pairwise fashion at the target-ligand interface... what sort of many-body problem is the drug designer facing and how can this knowledge play advantageously to address the major therapeutic imperatives of today and tomorrow"?

In fact, a drug $\mathfrak{U}$ is much smaller than its target, a globular protein $\mathfrak{V}$. Thus, the binding energy $E(\mathbf{X}, \mathbf{Y})$ will be easier to figure out than that between two large proteins. Then the stable structures of the drug/target complex will be (local or global) minimizers of $G((\mathbf{X}; \mathbf{Y}); \mathfrak{E}_{\mathfrak{U}}, (\mathfrak{U}, \mathfrak{V}))$ in (12). Thus the drug's efficiency and safety depend on the properties of minimizers of $G((\mathbf{X}, \mathbf{Y}); \mathfrak{E}_{\mathfrak{U}}, (\mathfrak{U}, \mathfrak{V}))$. This might be an answer to what is the "many-body problem" in [10] and also suggests a particular solution.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

[1] Anfinsen, C.B. (1973) Principles That Govern the Folding of Protein Chains.

*Science*, **181**, 223-230. https://doi.org/10.1126/science.181.4096.223

[2] Fang, Y. Gibbs Free Energy Formula for Protein Folding. In: Morales-Rodriguez, R., Ed., *Thermodynamics—Fundamentals and Its Application in Science*, 47-82. http://www.intechopen.com/books/thermodynamics-fundamentals-and-its-application-in-science

[3] Fang, Y. (2014) A Gibbs Free Energy Formula for Protein Folding Derived from Quantum Statistics. *Science China, Physics, Mechanics & Astronomy*, **57**, 1547-1551. https://doi.org/10.1007/s11433-013-5288-x

[4] Fang, Y. (2015) Thermodynamic Principle Revisited: Theory of Protein Folding. *Advances in Bioscience and Biotechnology*, **6**, 37-48. https://doi.org/10.4236/abb.2015.61005

[5] Levinthal, C. (1969) How to Fold Graciously. Mössbauer Spectroscopy in Biological Systems Proceedings. Proceedings of a Meeting Held at Allerton House, Monticello, Illinois. *University of Illinois Bulletin*, **67,** 22-26. https://web.archive.org/web/20110523080407/http://www-miller.ch.cam.ac.uk/levinthal/levinthal.html

[6] Bader, R.F.W. (1990) Atoms in Molecules: A Quantum Theory. Clarendon Press, Oxford.

[7] Lattman, E.E. and Loll, P.J. (2008) Protein Crystallography: A Concise Guide. The Johns Hopkins University Press, Baltimore.

[8] Novotny, J., Bruccoleri, R. and Karplus, M. (1984) An Analysis of Incorrectly Folded Protein Models. Implications for Structure Predictions. *J. Mol. Biol.*, **177**, 787-818. https://doi.org/10.1016/0022-2836(84)90049-4

[9] Novotny, J., Rashin, A.A. and Bruccoleri, R. (1988) Criteria That Discriminate between Native Proteins and Incorrectly Folded Models. *Proteins*, **4**, 19-30. https://doi.org/10.1002/prot.340040105

[10] Fernádez Stigliano, A. (2015) Biomolecular Interfaces: Interactions, Functions and Drug Design. Springer International Publishing.

[11] Fernandez, A., Kardos, J. and Goto, Y. (2003) Protein Folding: Could Hydrophobic Collapse Be Coupled with Hydrogen-Bond Formation? *FEBS Letters*, **536**, 187-192. https://doi.org/10.1016/S0014-5793(03)00056-5

[12] Fang, Y. and Jing, J. (2010) Geometry, Thermodynamics, and Protein. *Journal of Theoretical Biology*, **262**, 382-390. https://doi.org/10.1016/j.jtbi.2009.09.013

[13] Wu, H. (1931) Studies of Denaturation of Proteins XIII. A Theory of Denaturation. *Chinese J. Physiol.*, **5**, 321-344.

[14] Sanns, W. (2009) Catastrophe Theory. In: Meyers, R.A., Ed., *Encyclopedia of Complexity and System Science*, Springer, Vol. 4, 703-719. https://doi.org/10.1007/978-0-387-30440-3_47