Equations of Life: Revealing Physical Laws of Protein Folding

Yi Fang

Australian National University

Abstract:

To resolve the protein folding problem, that is: predicting the native structure and describe 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**, it is just the **Second Law of Thermodynamics**. 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 microscopic view; the latter is a macroscopic view.

A microscopic 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 $\mathfrak{U}$. $\mathfrak{U}$ appears here as a parameter to indicate that for different proteins, we should have different Gibbs free energy functions. Other very important parameters are the environment $\mathfrak{E}$ (solvent, temperature, pressure, other particles in solvent, etc.) in which $\mathfrak{U}$ folds.

Conformational Gibbs free energy function derived from first principle is the core of the **Microscopic Thermodynamic Hypothesis for 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 conformational Gibbs free energy function $G(\mathbf{X}; \mathfrak{E}, \mathfrak{U})$, therefore, $\nabla G(\mathbf{X}_{\mathfrak{E}}, \mathfrak{E}, \mathfrak{U}) = \mathbf{0}$.**

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 environment in $\mathfrak{E}_N$, we have derived by first principle the conformational Gibbs free energy function $G(\mathbf{X}; \mathfrak{E}_N, \mathfrak{U})$ for globular proteins $\mathfrak{U}$. It is an analytic function expressed by global geometric features of the conformation $\mathbf{X}$. For

non-globular proteins, the derivation idea is the same but are much more mathematically difficult, due to their complicated native environments.

A scientific hypothesis has to be able to explain nature phenomena. In this talk, we will explain why an over-simplified statement that a protein's amino acid sequence totally determines its native structure is wrong. The parameters $\mathfrak{E}$ in $G(\mathbf{X};\mathfrak{E},\mathfrak{U})$ shows, for most non-native environment $\mathfrak{E}$, not only the native structure $\mathbf{X}_{\mathfrak{U}}$ is not a local minimizer of $G(\mathbf{X};\mathfrak{E},\mathfrak{U})$, but also in general there will be many many stable conformations, all of them are local minimizers of $G(\mathbf{X};\mathfrak{E},\mathfrak{U})$. This can explain the unfolding phenomenon. Therefore, environment is an essential part of the protein folding process, the same amino acid sequence will turn out dramatically different shapes in different environment. Indeed, for a given protein $\mathfrak{U}$, we have infinite Gibbs free energy functions according to different environments. In natural, proteins fold in their native environments $\mathfrak{E}_N$; In unfolding, the environment is different from $\mathfrak{E}_N$; when two protein molecules dock, the native environment $\mathfrak{E}_N$ changes to include other proteins, etc. Environments changes leading to different conformational Gibbs free energy functions in control of folding process, can explain natural phenomena such as folding, unfolding, docking, allosteric change, amyloid disease, etc.

.

Another big question is that is there a folding force? Leventhal in 1960's 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.

By the microscopic thermodynamic hypothesis, the conformational Gibbs free energy function gives a deterministic fording force acting on each atom of the protein $\mathfrak{U}$, it is,

$$\mathbf{F}_i = -\nabla G_{\mathbf{x}_i}(\mathbf{X};\mathfrak{E}_N,\mathfrak{U}), \quad i=1,\cdots,n, \quad \mathbf{X}=(\mathbf{x}_1,\cdots,\mathbf{x}_n)\in\mathfrak{X}_{\mathfrak{U}}\subset\mathbb{R}^{3n}.$$

A scientific hypothesis has to give verifiable predictions to let people confirm or refute it. We suggest two verifiable predictions of the microscopic thermodynamic hypothesis: *ab initio* predictions of native structures of globular proteins and equations of motion of their folding dynamics.

Under the microscopic thermodynamic hypothesis, the native structure $\mathbf{X}_{\mathfrak{U}}$ of a protein molecule $\mathfrak{U}$ is a (local) minimiser,

$$G(\mathbf{X}_{\mathfrak{U}};\mathfrak{E}_N,\mathfrak{U})\leq G(\mathbf{X};\mathfrak{E}_N,\mathfrak{U});\quad \mathbf{X}\in U\subset\mathfrak{X}_{\mathfrak{U}}.$$

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

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

Therefore, for a globular protein $\mathfrak{U}$ (we know $G(\mathbf{X};\mathfrak{E}_N,\mathfrak{U})$), 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.

Another verifiable prediction of the microscopic thermodynamic hypothesis is the the folding dynamics which needs solving an equation of motion. That is, there is a folding path $\mathbf{X}(t)$ from an initial conformation to the native structure conformation, part deterministic and part random, the equation of motion is a Langevin equation

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

Here η_i is the solvent friction. The random force $\mathbf{F}_i(t)$ is caused by probably 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 exists. Moreover, mathematics also tells us that a folding path is highly depending on its initial conformation, so an important issue is to know the shape of protein conformation just out of ribosome.

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

All the discussions depend on the conformational Gibbs free energy function $G(\mathbf{X};\mathfrak{E},\mathfrak{U})$ and its mathematical property. Let me write down the formula of $G(\mathbf{X};\mathfrak{E}_N,\mathfrak{U})$ of a globular protein $\mathfrak{U}$ in its native environment $\mathfrak{E}_N$:

$$G(\mathbf{X};\mathfrak{E}_N,\mathfrak{U})=\omega_eV(\Omega_{S_{\mathbf{X}}})+\omega_ed_wA(S_{\mathbf{X}})+\sum_{i=1}^{H}\omega_iA(S_{\mathbf{X},i})+\sum_{1\leq i<j\leq n}\frac{Z_iZ_je^2}{4\pi\epsilon_0|\mathbf{x}_i-\mathbf{x}_j|}. \tag{2}$$

Because of the extremely importance of protein folding for life process, equations (1) and (2) should be included among the Equations of Life.

References

[1] Fang, Y. Mathematical protein folding problem. In: Hoffman, D. (ed.), *Global Theory of Minimal Surfaces. Proceedings of the Clay Mathematical Proceedings*, Vol. **2**, pp. 611-622, 2005.

[2] Fang, Y. and Jing, J. Implementation of a mathematical protein folding model. *International Journal of Pure and Applied Mathematics* **42**(4), 481-488, 2008.

[3] Fang, Y. and Jing, J. Geometry, thermodynamics, and protein. *Journal of Theoretical Biology*, **262**, 382-390, 2010.

[4] Fang, Y. Gibbs free energy formula for protein folding. In: Morales-Rodriguez, R. (ed.), *Thermodynamics - Fundamentals and Its Application in Science*. pp. 47-82. DOI: 10.5772/2615. Open source: http://www.intechopen.com/books/thermodynamics-fundamentals-and-its-application-in-science

[5] Fang, Y. Ben-Naim's "pitfall": Don Quixote's windmill. *Open Jouranl of Biophysics*, **3**:13-21, 2013. doi:10.4236/ojbiphy.2013.31002 Published Online January 2013 (Open source: http://www.scirp.org/journal/ojbiphy)

[6] Fang, Y. The second law, Gibbs free energy, geometry, and protein folding. *Journal of Advances in Physics* **3**(3): 278-285, 2014.

[7] Fang, Y. A Gibbs free energy formula for protein folding derived from quantum statistics. *Science China, Physics, Mechanics & Astronomy* **57**(8), 1547-1551, 2014.

[8] Fang, Y. Thermodynamic principle revisited: Theory of protein folding. *Advances in Bioscience and Biotechnology* **6**: 37-48, 2015.

[9] Fang, Y. Why Ben-Naim's deepest pitfall does not exist. *Open Journal of Biophysics* **5**: 45-49, 2015.

[10] Fang, Y., Protein Folding: Towards a single molecule theory. In Hale, M. (ed.), *Advances in Protein Folding Research*, pp. 65-126, 2015. Nova Publishers, New York, 2015.

Biography:

My last 18 years of research can be summarized as: A mathematician trying to solve a biological problem by revealing how fundamental physical laws govern protein folding. I first heard protein folding from Jiri Novotny. He told me that a folded globular protein has smaller volume, surface area, and better hydrophobic core than unfolded one. As a geometer used to calculus of variations, my instinct is that why not minimize volume, area,

and hydrophobic area (the smaller it is, the better the hydrophobic core) simultaneously and cohesively to get the native structure? A mathematical model was formulated during the conversation and was published [1]. In 2006 Junmei Jing and I participated CASP7, submitted our models [3], which were results of shrinking the hydrophobic area. Neglecting volume and area certainly will not give us good 3-dimensional models. But our models contain a lot of secondary structures and hydrogen bonds, even though we only calculated the gradient of the hydrophobic area function. This result confirms that I was on the right track. But we cannot just publish a mathematical model without computer simulation. Most of my papers were rejected without going to referees.

Realzing that protein folding is a physics problem, I determined to study physics to prove my mathematical model has a firm physics foundation. I am used to teach myself. I was born in 1953, finished primary school in 1966, missed secondary education because all schools in China were closed down in 1966. After wasting 4 years, from 1970 to 1978, I worked as a farm labor. I tried to teach myself by reading textbooks. In 1997, I passed university admission examination of 1977, the first one in 12 years since 1966, entered Jilin University to learn mathematics. In 1985, I went to University of Massachusetts and got my PhD in 1990.

Broad reading and mathematical skills enable me to learn biology, chemistry, thermodynamics, statistical and quantum mechanics, etc. After years of study, I created a new physical method to derive conformational Gibbs free energy function $G(\mathbf{X}; \mathfrak{E}_N, \mathfrak{U})$. As anticipated, it is a refined version of my original mathematical model. With more learning of physics, the microscopic thermodynamic hypothesis for protein folding follows suit.