The protein folding problem (PFP) has been revisited in a recent paper (Ben-Naim, 2012). This author has addressed the Levinthal’s question and mainly focused on how a protein folds to a unique structure (functionally active) from the information that is coded in its sequence of amino acids. Ben-Naim provided his answers through several algorithms and discussions. We propose additional discussions with respect to this author and others (as Anfinsen, 1973; Brygelson & Wolynes, 1987; Dill & Chan, 1997; Rose, Fleming, Banavar, & Maritan, 2006; Zwanzig, Szabo, & Bagchi, 1992). First, we discuss on the Anfinsen’s thermodynamic hypothesis. This hypothesis, derived from the Second Law, is not sufficient to affirm that the native functional conformation always coincides with the lowest minimum of the Gibbs free energy. We also emphasize on the role of entropy that was often underestimated in current molecular simulations. Entropy however plays an important role to maintain a moderate free energy difference and accessibility between intermediate states and native conformation. We are cautious with the so-called hydrophobic paradigm. Indeed, for proteins larger than ~100 amino acids, the three effects of hydrophobicity, hydrophilicity and flexibility participate in folding and balance between each others. Evolution determines which among newly-synthesized and already-available proteins have the best functional conformation to perform desired cellular tasks. There is no evidence of correct and incorrect amino-acid binding in an evolutionist perspective. Our comments extend to large proteins.

**Anfinsen’s thermodynamic hypothesis and PFP**

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 $G$. 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 $\Delta G_{\text{relax}}^{\text{relax}} = G_N - \tilde{G}_N < 0$ 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 $\tilde{G}_X = \tilde{G}(C_X, Q_N)$ and $G_N = \tilde{G}(C_N, Q_N)$. $C_N$ 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 $\Delta G_{\text{relax}} = G_N - \tilde{G}_X$, where $G_X = \tilde{G}(C_X, Q_N)$ is the equilibrium free energy for $Q_X$. The Anfinsen’s thermodynamic hypothesis can, therefore, only hold with a good likeliness if $\tilde{G}_X \approx G_N$. However, cases where $G_X$ is a deep minimum with $G_X < \tilde{G}_X$ cannot be excluded, which may lead to $\Delta G_{\text{relax}} = G_N - \tilde{G}_X > 0$. Inclusion bodies with lower free energies than that of the native state could, therefore, be obtained after folding, as it can be observed in large proteins. However, if such inclusion bodies were formed in the cell, they would be eliminated. The cell only circulates the native conformation of a protein corresponding to the desired function (Chen, Retzlaff, Roos, & Frydman, 2011; Vabulas, Raychaudhuri, Hayer-Hartl, & Hartl, 2010). $G_N$ is related to the minimal free energy out of an ensemble of possible conformations but does not always correspond to the global minimum of the full Gibbs free energy landscape (GFL). We, therefore, support cautionary note of Ben-Naim (2012) that the

**COMMENT**

**Concepts on the protein folding problem**

Jean-Numa Gillet and Indira Ghosh*

School of Computational and Integrative Sciences, Jawaharlal Nehru University (JNU), New Delhi 110067, India

*Corresponding author. Email: indirag@mail.jnu.ac.in

Copyright © 2012 Taylor & Francis
Anfinsen’s thermodynamic hypotheses should be considered with care. This hypothesis and fact that the native conformation corresponds to the functional structure have been often oversimplified. These oversimplifications led to the Levinthal’s paradox (Levinthal, 1968) and idea of existence of a unique folding pathway, which was explained as a “target-based folding,” and consequently, the funnel concept was introduced in the GEL (Dill & Chan, 1997).

Role of entropy in the PFP

The physical quantity to be minimized is \( G = H - TS \), but neither enthalpy \( H \) nor internal energy \( E \). We feel a lack of discussions about the entropy \( S \) in the paper presented by Ben-Naim (2012). However, \( S \) is an important quantity for the PFP that has been unfortunately not considered in the current molecular dynamics (MD) models. For example, when inclusion bodies are found, even if its free energy \( G_N \) is the lowest, the native conformation of a protein could present a higher \( H = H_N \) than the \( H \) values of inclusion bodies, which seems unfavorable for the native conformation. Nevertheless, \( G_N = H_N - TS_N \) could remain the lowest \( G \) due to a higher entropic contribution \( TS_N \). It has been indeed observed that the relative contact order of various folded proteins remains moderate (below \( \approx 20\% \)) even for large proteins with an amino acid number \( \geq 100 \) (Galzitskaya, Garbuzynskiy, Ivanov, & Finkelstein, 2003; Plaxco, Simons, Ruczinski, & Baker, 2000). Hence, it may be concluded that not all possible long-distance amino acid favorable contacts are always made in the native state (as in large proteins). Our remark gets supported when one measures the compactness of folded structures in terms of hydrophobicity, mass, or polarizability using fractal dimensions. For instance, the all-\( \beta \) proteins were identified with a maximum amount of “unused hydrophobicity” compared to the all-\( \alpha \) proteins (Banerji & Ghosh, 2011). This may result in quite high remaining conformational entropy in the functional structure. The available \( S \) would keep a fast folding kinetics by increasing the degree of accessibility (and escape) of intermediate states.

Hydrophobic, hydrophilic, or flexibility paradigm?

Predominance of the hydrophobic, hydrophilic, or flexibility effect is dependent on the protein type. All the three effects are intimately related to obtain folding of the protein to its native state in an aqueous solvent. In any general model of the PFP, there should be no bias with respect to one of these effects taken alone (see Supplementary Material). To confirm absence of paradigm, we computed the effective hydrophobicities, hydrophilicities, and side-chain flexibilities of the proteins contained in a PDB database (see Supplementary Material). The calculated quantities, from protein amino acid sequences, present almost constant values with low standard deviations for proteins larger than \( \sim 50-100 \) residues. Hence, considering only average values per amino acid of these three characteristics gives no information on the degrees of hydrophobicity, hydrophilicity, or flexibility in proteins. We, therefore, agree with Ben-Naim (2012) that considering an excessive importance of the hydrophobic forces to lead the protein to folding is not an acceptable paradigm in the PFP. Differently, we are skeptical regarding some of his discussions suggesting that hydrophilicity would give a predominant force fortification over hydrophobicity. A too high fortification of the hydrophilic properties of exposed amino acids can also make less significant the entropic advantage of providing a better GEL model, due to worse amino acids or drive of the unfolded conformation away from the functional minimum structure in the GEL. Balancing between the three parameters, hydrophobicity, hydrophilicity, and flexibility, should drive folding to the protein’s functional form. However, all the three effects may appear to act only locally while protein folding takes place through many microstates.

Large proteins also fold

While does it wrongly seem that a number of large proteins fold too slowly or does not fold at all? Inhomogeneities through the overall volume \( \tilde{V} \) of a large interacting system can occur, especially in inhomogeneous quite-diluted systems as the bacterial medium. It is always possible to envision that, in a smaller system subvolume \( \tilde{V}_I < \tilde{V} \), the equilibrium composition \( Q_I \) is not the same than those in other subvolumes. In \( \tilde{V}_I \), a significant concentration of the protein in an intermediate state \( I \) with the non-native conformation \( C_I \) can, therefore, be found. The equilibrium \( G \) in the subset \( \tilde{V}_I \) is given by \( G_I = \tilde{G}(C_I, Q_I) \). The local equilibrium with \( Q_I \) is not the same than the native state equilibrium with \( Q_N \) that can exist in other system subsets. Therefore, state \( I \) could coexist in the cellular medium during a given lifetime. However, the local equilibrium conditions in \( \tilde{V}_I \) can be changed (or even forced) to the native equilibrium. The latter can be imposed after activation of a network \( \Pi_N \) of cellular regulation processes. Due to introduction of the new native conditions in \( \tilde{V}_I \), state \( I \) is out of equilibrium with the non-equilibrium \( \tilde{G}_I = \tilde{G}(C_I, Q_N) \). Since \( \tilde{G}_I \) is normally higher than \( G_N \), a spontaneous process can thus be started (according to the Anfinsen’s thermodynamic hypothesis) and the protein folds until the native state with the free energy minimum \( G_N \) and native conformation \( C_N \) is reached because they correspond to the new equilibrium with \( Q_N \). For a normally activated \( \Pi_N \), proteins in intermediate states should, therefore, fold.
anyway to the minimum $G_N$ of the native state (possible inclusion bodies are removed by the cell). These developments about the cellular networks (which were missing in Ben-Naim, 2012) as well as those proposed in the previous paragraphs apply to large proteins. In addition, due to entropy–enthalpy compensations, the amplitude of the free energy change $AG$ should remain moderate to keep reversibility between intermediate and native states, which is another help for large proteins to fold by escaping intermediate states. After interpolations of experimental databases where the logarithm of the folding rate $k_f$ with respect to the protein length $L$ is given for various proteins (Bogatyreva, Osypov, & Ivankov, 2009; Ivankov, Bogatyreva, Lobanov, & Galzitskaya, 2009), we observed no clear decrease relationship between $k_f$ and $L$ (see Supplementary Material), in agreement with the precedent.

**Evolution is cellular: proteins cannot evolve alone**

Proteins are pure physicochemical molecular objects. If classical mechanics holds, we see no possibility for a protein to memorize a narrow range of folding pathways leading to the native conformation. Therefore, proteins taken alone cannot be randomly selected by evolution, and neither “correct” nor “incorrect” non-covalent bonds can be defined by evolution for the PFP in contrast to Zwanzig et al. (1992) but in agreement with Ben-Naim (2012). A cell during evolution of its related organism genetic code can use and synthesize a quite large number of newly available protein types to generate novel and improved cellular functions. The choice of new protein types can be targeted to the adaptation of the cell processes to new thermodynamic equilibrium conditions $\Omega (N')$ that have changed from the previous ones $\Omega (N)$. The functional state $N$ is now out of equilibrium and presents a non-equilibrium free energy $\tilde{G}_N = \tilde{G} (C_N; \Omega (N'))$ with respect to the modified equilibrium native state $N'$ with $\Omega (N')$. However, if $G_N < G_{N'}$ or states $N$ and $N'$ are too distant from each other in the GEL, it is not sure if refolding of the same protein type to $C_N$, which would correspond to a more functional conformation for $\Omega (N')$, is likely to happen after simple relaxation to equilibrium (Klein-Seetharaman et al., 2002; Shortle & Ackerman, 2001). Moreover, new protein types could fold to conformations that may show even better functionalities than that of $C_N$ related to the older protein. The best cells can, therefore, use other protein types for the survival of the fittest organisms to a new $\Omega (N')$ equilibrium.

**Our answer to the Levinthal’s paradox**

The pitfall is that this paradox avoids considering significant aspects of the PFP that we, however, discussed throughout this comment. First, the Levinthal’s paradox was only defined at the microscopic conformation scale of an individual protein. It does not include the macroscopic kinetic equilibrium effects between large ensembles of microstates in the cell. Moreover, the number of intermediate micro state types present in a cell is finite and usually small (even for large proteins). Therefore, in contrast to the Levinthal’s paradox, the global minimum of the entire GEL does not have to be discovered from a random walk that can be much longer than the biological time scale for protein folding. The useful part of the GEL to be explored, therefore, shrinks to the only conformations in the vicinities of those of the intermediate and native functional states present in the cell. Second, balancing between the hydrophobic, hydrophilic, and flexibility decreases the number of dimensions of the useful part of the GEL due to reduction in the number of important reaction coordinates for protein folding. Hence, the functional conformation corresponds to the search of the deepest minimum of $\tilde{G}$ in a much smaller zone with fewer dimensions than the entire GEL. Third, the Levinthal’s paradox does not consider importance of entropic effects for protein folding. In this paradox, the search of the global minimum is biased to that of the internal energy landscape (EL) instead of that of the GEL. Due to many balance effects of entropy–enthalpy (or entropy-internal energy) compensations, the natural GEL should be much smoother than the EL presenting many artifacts as too high energy barriers and too deep local minima (with locations that could be incorrect in the GEL). From the precedent, one can easily envision that protein folding to the functional structure is indeed as fast as in the cell, with no exception for many large proteins. Therefore, the Levinthal’s paradox is an artifact but not a nature’s paradox, as also noted in Ben-Naim (2012).

The presented discussions hold for proteins that fold with the help of chaperones and enzymes for faster kinetics. Indeed, the protein coordinate set $C_X$ can be included to the enzyme coordinate set $E_X$ to form an extended conformational ensemble ($C_X, E_X$). Possible covalent binding between the protein and enzyme may also be introduced from the binding relative affinities $\Delta \Delta G_{\text{bind}}$ inside the overall system $\Delta G$, if one would like to expand the concept.

**Acknowledgement**

For the Supplementary Material, we thank Mr. Taushif Khan for help in achievement of the data analyses presented in Figs. S1 and S2. JNG acknowledges the J. Nehru Institute of Advanced Study (JNIAS) in JNU for its support as well as (both authors) the Department of Biotechnology – Centre of Excellence (DBT – COE) funds for this publication.

**Supplementary Material**

The supplementary material for this paper is available online at http://dx.doi.org/10.1080/07391102.2012.748546.
References
Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science, 181*, 223–230.
Banerji, A., & Ghosh, I. (2011). Fractal symmetry of protein interior: What have we learned? *Cellular and Molecular Life Sciences, 68*, 2711–2737.
Ben-Naim, A. (2012). Levinthal’s question revisited, and answered. *Journal of Biomolecular Structure and Dynamics, 30*, 113–134.
Bogatyreva, N. S., Osypov, A. A., & Ivankov, D. N. (2009). KineticDB: A database of protein folding kinetics. *Nucleic Acid Research, 37*, D342–D346.
Brygelson, J. D., & Wolynes, P. G. (1987). Spin glasses and the statistical mechanics of protein folding. *Proceedings of the National Academy of Sciences of the USA, 84*, 7524–7528.
Chen, B., Retzlaff, M., Roos, T., & Frydman, J. (2011). Cellular strategies of protein quality control. *Cold Spring Harbor Perspectives in Biology, 3*, a004374, 1–14.
Dill, K. A., & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Structural Biology, 4*, 10–19.
Galzitskaya, O. V., Garbuzynskiy, S. O., Ivankov, D. M., & Finkelstein, A. V. (2003). Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics. *Proteins: Structure, Function, and Genetics, 51*, 162–166.
Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Deubert, E., Ueda, T., … Schwalbe, H. (2002). Long-range interactions within a nonnative protein. *Science, 293*, 1719–1722.
Ivankov, D. N., Bogatyreva, N. S., Lobanov, M. Y., & Galzitskaya, O. V. (2009). Coupling between properties of the protein shape and the rate of protein folding. *PLoS ONE, 4*, e6476, 1–10.
Levinthal, C. (1968). Are there pathways for protein folding. *Journal de Chimie Physique, 65*, 44.
Plaxco, K. W., Simons, K. T., Ruczinski, I., & Baker, D. (2000). Topology, stability, sequence, and length: Defining the determinants of two-state protein folding kinetics. *Biochemistry, 39*, 11177–11183.
Rose, G. D., Fleming, P. J., Banavar, J. R., & Maritan, A. (2006). A backbone-based theory of protein folding. *Proceedings of the National Academy of Sciences of the USA, 103*, 16623–16633.
Shortle, D., & Ackerman, M. S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. *Science, 293*, 487–489.
Vabulas, R. M., Raychaudhuri, S., Hayer-Hartl, M., & Hartl, F. U. (2010). Protein folding in the cytoplasm and the heat shock response. *Cold Spring Harbor Perspectives in Biology, 2*, a004390, 1–18.
Zwanzig, R., Szabo, A., & Bagchi, B. (1992). Levinthal’s paradox. *Proceedings of the National Academy of Sciences of the USA, 89*, 20–22.