Introducing the Levinthal's Protein Folding Paradox and Its Solution

ARTICLE *in* JOURNAL OF CHEMICAL EDUCATION · NOVEMBER 2014

Impact Factor: 1.11 · DOI: 10.1021/ed300302h

1 AUTHOR:

Leandro Martínez

University of Campinas

45 PUBLICATIONS **1,224** CITATIONS

JOURNAL OF CHEMICAL EDUCATION

Article

pubs.acs.org/jchemeduc

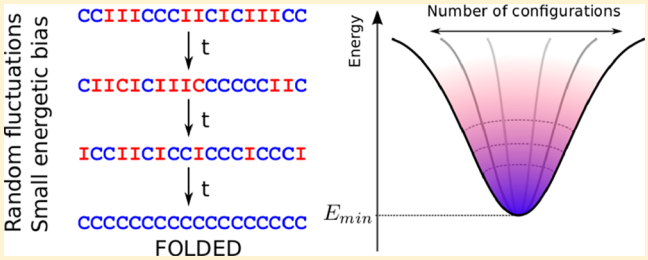
Introducing the Levinthal's Protein Folding Paradox and Its Solution

Leandro Martínez*

Department of Physical Chemistry, Institute of Chemistry, University of Campinas, Campinas, São Paulo 13083-970, Brazil

Supporting Information

ABSTRACT: The protein folding (Levinthal's) paradox states that it would not be possible in a physically meaningful time to a protein to reach the native (functional) conformation by a random search of the enormously large number of possible structures. This paradox has been solved: it was shown that small biases toward the native conformation result in realistic folding times of realistic-length sequences. This solution of the paradox is, however, not amenable to most chemistry or biology students due to the demanding mathematics. Here, a simplification of the study of the paradox and its solution is provided so that it is accessible to chemists and biologists at an undergraduate or graduate level. Despite its simplicity, the model captures some fundamental aspects of the protein folding mechanism and allows students to grasp the actual significance of the popular folding funnel representation of the protein energy landscape. The analysis of the folding model provides a rich basis for a discussion of the relationships between kinetics and thermodynamics on a fundamental level and on student perception of the time-scales of molecular phenomena.



KEYWORDS: *Upper-Division Undergraduate, Graduate Education, Biochemistry, Biophysical Chemistry, Molecular Biology, Proteins/Peptides, Theoretical Chemistry*

Understanding and controlling protein folding is arguably the most important challenge in structural biology. The ability to predict, and thus model, protein structure and function from sequence may, in the future, transform completely the molecular biology, synthetic chemistry, and medicinal chemistry fields. Achieving this goal requires both the comprehension of the general physicochemical aspects of folding and of the specificities leading each protein, with its particular sequence, to adopt a specific functional structure. Presently, there have been great advances on the comprehension of the physics of folding mechanisms in general, although the actual predictive ability of current models for particular proteins is small. Thus, the protein folding problem can be regarded at the same time solved as its physics is understood by simple models, or completely open, as methods for prediction of structure from sequence are still very limited.[1]

Protein structures are complex enough to stimulate educational efforts in many areas. For example, a continuous area of educational research is the proposition of mechanical models of representation of protein structures[2,3] and the classification of folds in periodic-table-like charts.[4] Experiments and macroscopic models for specific unfolding events have been proposed for undergraduate students.[5,6] Another amazing tool with both educational and scientific relevance is FoldIt,[7] which provides an interactive game for people with no particular scientific background to play with protein folding by manipulating structures, with remarkable achievements in solving and designing protein folds.[7,8] Nevertheless, the general theory of protein folding has not been published in simplified forms accessible for a more general audience.

From a general physicochemical point of view, the protein folding problem is stated in Levinthal's paradox:[9] How can proteins assume their functional structures given the enormous amount of unfolded structures possible? Levinthal's paradox was introduced from the supposition that the folding mechanisms of proteins occur by random search. Proteins, as linear chains of amino acids, should randomly search configurations until the native, functional arrangement is reached. Such a search can be shown[9,10] to be inconsistent with experimental folding rates in such a way that proteins with a few hundred residues would require astronomical scales of time to reach, in average, the folded state.

Of course, real proteins fold in time-scales ranging from microseconds to minutes,[11] and therefore, only a minimal fraction of the configurational space is sampled. The general requirements for folding mechanisms that solve this problem have been proposed; Bagchi and co-workers[12,13] have shown that small energetic biases toward the native state reduce the conformational search to the point that realistic folding times can be obtained. The analysis presented here is based on this simple argument but avoids the use of mathematical concepts not familiar to most chemistry or biology students. This analysis naturally leads to the "folding funnel" picture, which is widespread in both scientific and didactic books, many times with incomplete explanations or misinterpreted implications.[10]

Published: August 4, 2014

AUDIENCE

This analysis was presented during a graduate course on "Thermodynamics of Biological Systems". Most of the students who were exposed to the present arguments, particularly experimental molecular biology students, did not have strong mathematical backgrounds, a reason why the present derivation is intended to be very simple. At the same time, this analysis addresses some fundamental concepts behind the proposal of a physical theory of chemical reactions that graduates in physics also appreciate. In particular, it stimulates discussions about (1) the differences between thermodynamically and kinetically controlled reactions and their intricate correlation in the protein folding mechanism, (2) the validation of the simple models by intuitive guesses on the nature of molecular movements and energetics, and (3) validation or refusal of a theoretical model using experimental data.

Therefore, the theory presented here can be included in different undergraduate or graduate courses, for example, (1) as a illustrative example of chemical kinetics in a physical chemistry undergraduate course, (2) as an additional topic in any molecular biology, structural biology or biophysical course dealing with protein folding or protein structures, and (3) an extra topic in a general chemistry course that focuses on biochemistry or biophysics, after the introduction of basic chemical kinetics and thermodynamics.

Some suggested questions are available as Supporting Information as a follow-up assignment, an assessment for learning or to stimulate classroom discussions.

KINETIC DEFINITION OF THE PARADOX

The analysis of the protein folding problem is simplified by representing the sequence of a protein in a unidimensional way. The "protein" is defined by a sequence of letters

cicicicicicicic...

where c corresponds to a letter that corresponds to an "amino acid residue in the correct conformation", that is, in the conformation in which it should be in the native fold, and i is an "amino acid residue in the incorrect conformation", that is, one different from that of the native fold.

The first argument toward the comprehension of the Levinthal paradox is that the "correct fold" is characterized by a *unique* conformation, in this case the sequence

ccccccccccccccc....

If the two states c and i are equally probable, the probability is 0.5^N of such conformation in an evenly distributed set of sequences of N elements. If $N = 100$, this is $0.5^{100} \approx 10^{-30}$, which is a very small number. In one mole of protein, not a single folded structure would be generally observed if the folding occurred by chance. Thus, even for a small protein of 100 residues, a homogeneous population cannot explain function.

Therefore, there must be a bias toward the native fold so that the population of the native fold can increase and justify the activity of the protein solution.

ASSUMING THE STABILITY OF THE FOLDED STATE

This brief analysis is to introduce the first assumption: from experience, it is known that the population of proteins assuming the native fold relative to unfolded protein in functional conditions is very large (close to unity). In accordance with this experimental observation, the first assumption is

Once a structure reaches the folded state, it remains folded.

Therefore, regardless of the protein folding path, reaching the folded state would give rise to an increase of the folded population, thus explaining function.

At this point, the relationships between kinetics and thermodynamics are discussed; the assumption above implies, thermodynamically, that the equilibrium population of the folded state is unitary. Therefore, the equilibrium population cannot provide information on the folding mechanism and the folding kinetics must be studied. As with the determination of mechanisms in physical chemistry, one may guess a particular mechanism and deduce its kinetics, to be validated by comparison with experimental data.

Similar to using this assumption is the calculation of the first passage time of the structure through the folded state. Here, however, the stability assumption is used because it clearly identifies the folding paradox as a kinetic rather than thermodynamic problem, thus implying that the discussion leading to its solution will be based on an assumption of its mechanism. A fruitful discussion on the nature of protein native structure stability is introduced by this assumption: most students with an experimental background know, from experience, that the activities of their purified proteins are rarely preserved for more than a few days. Thus, one may argue that many proteins may be kinetically—but not actually thermodynamically—stable, thus justifying either the use of the first passage time or the assumption above to analyze folding rates without additional complications related to unfolding reactions.

ELAPSED TIME TO THE FOLDED STATE

At this point, the correct question is posed: given the first assumption on the stability of the folded state, how much time is needed for a protein to fold? The following model for the protein fold mechanism is proposed:

(1) A protein starts with any unfolded structure. An unfolded structure, in the representation used here, is any structure for which at least one residue is in the incorrect, i, conformation.
(2) At every time, τ, each residue (each letter) suffers a perturbation.
(3) The perturbation leads each residue to c or i with equal probabilities, independently of the previous state.

Within this simple kinetic model, at every τ, each residue is perturbed, so that for a sequence of N residues, the probability of obtaining the correct fold is 0.5^N. Thus, the probability for the sequence to remain unfolded at every τ is $1-0.5^N$, such that after t/τ units of time the probability of remaining unfolded is (eq 1)

$$P_{\text{remain unfolded}}(t) = (1 - 0.5^N)^{t/\tau} \quad (1)$$

Folding times clearly increase as a function of increasing sequence. (Students may benefit from explaining this last formula using the simple example where $N = 1$ and the analogy with the probability of throwing a coin at every τ seconds and systematically not obtaining heads—heads being the folded state.)

From eq 1, the probability for a particular sequence to be folded at time t is (eq 2)

dx.doi.org/10.1021/ed300302h | J. Chem. Educ. 2014, 91, 1918−1923

$$P_{\text{fold}}(t) = 1 - (1 - 0.5^N)^{t/\tau} \quad (2)$$

thus, the half-time, $t_{1/2}$, to sequence folding (the time necessary for 50% of the sequences to fold) is given by (eq 3)

$$1 - (1 - 0.5^N)^{t/\tau} = 0.5 \quad (3)$$

or (eq 4)

$$t_{1/2} = \tau \frac{\ln 0.5}{\ln(1 - 0.5^N)} \quad (4)$$

Do these time-scales make sense? This depends on the relationship between the folding time, $t_{1/2}$, and the characteristic time of the perturbations τ, which must be deduced, or supported, by real data.

By direct computation, for a sequence of 50 residues, $t_{1/2} \approx 10^{16}\tau$. Thus, for a typical protein folding time of one second, conformational changes in residues would need to be sampled at rates on the order of 10^{-16} s^{-1} = 0.1 fs. This does not make sense, as 0.1 fs is a time-scale faster than even the vibration of covalent bonds. No significant structural perturbation could occur that fast so that the peptide would fold in one second.

At the same time, for a sequence of 30 residues (N = 30), τ = 1 ns. Protein conformational changes *can* occur at the nanosecond time-scale, and therefore, a peptide of such size could fold by a random search with equal probabilities in a second. However, it is clear that for natural size proteins (i.e., 100 to 500 residues), this mechanism of a random search with equal probabilities cannot account for protein folding rates.

This is the actual nature of "Levinthal's paradox": How can a protein reach its folded state while avoiding a random search through all possible conformations *in realistic time*?

SOLUTION TO THE LEVINTHAL'S PARADOX

Bagchi and co-workers[12] have shown that under a similar scheme as the one described above, the introduction of a small energetic bias toward the native state leads to decreases in folding rates such that they can be compared with experiment. Their analysis involved the use of advanced calculus and thermodynamics, such that it is not amenable to most chemistry or biology students. Here, a simplified use of their arguments within the model above leads to the same semiquantitative analysis. A simple thermodynamic interpretation of the result confirms that it is a realistic explanation for the observed folding rates.

Kinetics of Protein Folding with Unequal Probabilities

Following the guidelines of Bagchi and co-workers,[12] a bias is introduced toward the native state. The simple way to do so in the context of the protein model of the previous section is to attribute unequal probabilities for a correct, c, or incorrect, i, fold for each residue at each perturbation time τ.

Thus, if α is the probability of the residue to turn into c at a given time, $1 - \alpha$ is the probability of i. Only cases where $\alpha > 0.5$ are of concern, introducing a bias toward correctly folded residues.

Following the same arguments above, but generalizing the probability of the folded residue at instant t, the probability for a sequence of N residues to remain unfolded is (eq 5)

$$P_{\text{remain unfolded}}(t) = (1 - \alpha^N)^{t/\tau} \quad (5)$$

and the probability of the sequence to fold is (eq 6)

$$P_{\text{fold}}(t) = 1 - (1 - \alpha^N)^{t/\tau} \quad (6)$$

The mean time to protein folding, $t_{1/2}$, becomes (eq 7)

$$t_{1/2} = \tau \frac{\ln 0.5}{\ln(1 - \alpha^N)} \quad (7)$$

Equation 7 shows that $t_{1/2}$ has a dramatic dependence on α, which is illustrated in Figure 1. Increasing α, that is, favoring at each element of the sequence the folded conformation, increases significantly the minimum size of the structure for which the folding rate becomes prohibitive. For α = 0.95, for instance, every sequence of less than 500 residues folds within $t_{1/2}$ = 60 s for a perturbation rate of 1 ns (used to build the figure). Thus, by favoring in the random search the folded conformation of each residue, the folding times can be driven to realistic times.

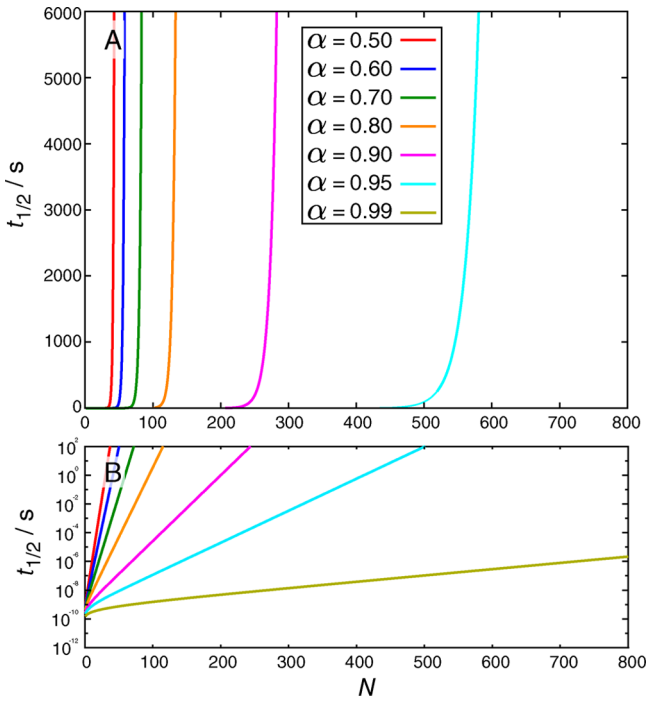


Figure 1. Dependence of $t_{1/2}$ on the sequence length, N, and on the probability α: (A) linear-scale plot and (B) logarithmic sale plot. For the preparation of this figure, τ = 1 ns.

Thermodynamic Interpretation and Validation

The above interpretation is reasonable whenever the favored transition toward the folded state of each residue can be physically explained. The energetic requirement for the differential probabilities is considered now. This connects the above solution of the paradox to the thermodynamics of protein folding.

In the model here, the relative probability of observing configurations c or i for a given residue at any time is (eq 8)

$$\frac{\alpha}{1 - \alpha} \quad (8)$$

This probability must be connected with the relative energies of states c and i. According to the Boltzmann distribution, the probability of finding configurations of energy E in a system is (eq 9)

$$P(E) = \frac{g e^{-E/(RT)}}{Z} \quad (9)$$

dx.doi.org/10.1021/ed300302h | J. Chem. Educ. 2014, 91, 1918−1923

where g is the degeneracy (the number of configurations with energy E) and Z is the partition function.

Therefore, the relative probability of two configurations of energies E_c and E_i, which are associated with states c and i, respectively, is (eq 10)

$$\frac{P(c)}{P(i)} = \frac{g_c e^{-E_c/(RT)}}{g_i e^{-E_i/(RT)}} \tag{10}$$

Using $\gamma = g_i/g_c$ as the relative degeneracies of incorrect and correct configurations (it is expected that $\gamma > 1$, as it is reasonable to suppose that the number of incorrect configurations is greater), the relative probabilities of configurations c and i are (eq 11)

$$\frac{P(c)}{P(i)} = \frac{e^{-E_c/(RT)}}{\gamma e^{-E_i/(RT)}} = \frac{\alpha}{1-\alpha} \tag{11}$$

or (eq 12)

$$E_c - E_i = -RT\ln\left(\gamma\frac{\alpha}{1-\alpha}\right) \tag{12}$$

Thus, this equation connects the relative probability of observing the correct configuration for a particular residue with the energy difference between the correct and incorrect states. It must be determined if realistic energy differences can explain statistical biases toward the correct configuration for each residue greater than 0.90, which could explain the folding of proteins larger than 250 residues (Figure 1).

Before actually computing the values, a crude approximation is needed of what one may consider "reasonable" from the point of view of an energetic bias toward the correct state for each residue. One can arrive at a reasonable estimate of a few kcal mol^{-1} as an acceptable energy bias toward the correct conformation of each residue. This is because the energy scales of intermolecular interactions are within 1 kcal mol^{-1} for van der Waals or weak dipolar interactions and about 10 kcal mol^{-1} for the formation of one hydrogen bond. Thus, if the energy bias required for a statistical drive toward the correct states is less than 10 kcal mol^{-1}, the current kinetic model is physically acceptable.

Table 1 displays the $E_c - E_i$ values, computed from eq 12 for various values of α and γ, for a system at T = 309 K (36 °C, body temperature). The energy differences of correct and incorrect configurations is obviously zero for zero degeneracy ($\gamma = 1$) and equal probabilities ($\alpha = 0.5$) at any temperature. For configurations with zero degeneracy, the energy bias toward the correct configuration rises from −1.3 kcal mol^{-1} for $\alpha = 0.9$ to −2.8 kcal mol^{-1} for $\alpha = 0.99$. Even if the degeneracy of the incorrect state is much larger than that of the correct state, the energy bias toward the correct state does not exceed 10 kcal mol^{-1}. As an illustration of reasonable degeneracies, the number of residue conformers in databases extracted from protein structures is of the order of 10, with one of them generally being correct for each residue in each protein. This provides an estimate of the degeneracy of correct versus incorrect residue conformations, not implying a major limitation for the current model of protein folding.

Table 1. Energy Differences[a] Favoring the Correct Configuration per Residue for Different Statistical Bias, α, and Degeneracies, γ

α	γ	$E_c - E_i$ (kcal mol^{-1})
0.5	1	0
0.9	1	−1.3
0.95	1	−1.7
0.99	1	−2.8
0.95	2	−2.4
0.95	10	−3.2
0.95	100	−4.6
0.95	1000	−6.1
0.95	100 000	−8.9

[a]Here, T = 309 K. For a characteristic configuration transition time of 1 ns, an α = 0.95 folds a protein of 500 residues in nearly 1 min.

The present model is able to explain the folding kinetics in the order of seconds to minutes of large proteins. The only requirement is that the search of the correct conformations for each residue involves an energy gain of a few kilocalories per mole. Notably, the size of single-domain proteins in whole genomes does not exceed 500 residues;[1] thus, the present model is able to explain the size limits of real-life proteins from the possibility of folding though a locally biased search.

The limitations of this model[11] are put in evidence by the quantitative comparison of the required bias with the actual enthalpy of folding of some proteins. For example, the enthalpy of folding of barnase (N = 110 residues) is −12 kcal mol^{-1}, implying an energy gain per residue of roughly −0.1 kcal mol^{-1} per residue.[11] This is less than the most optimistic estimates obtained in the previous analysis. Similar conclusions can be obtained for other proteins, such as ribonuclease (N = 124, ΔH = −60.3 kcal mol^{-1} at 25 °C or −0.48 kcal mol^{-1} per residue) and myoglobin (N = 153, ΔH = −42 kcal mol^{-1} at 25 °C or −0.27 kcal mol^{-1} per residue).[14] Additional suppositions on protein folding mechanisms are, therefore, required for a quantitative comparison with experiment, most notably all kinds of cooperative mechanisms, and the entropic contribution of the hydrophobic collapse, which reduces the conformational space for folding by increasing the entropy of the water while burying the side chains of hydrophobic residues.[14]

Number of Conformations and the Protein Folding Funnel

Admitting that a major contribution for the folding mechanisms results from the incremental energy decrease as predicted by the present model, one can deduce the implied energy landscape profile.

The total energy of a protein of N residues and N_c residues in the correct configuration is (eq 13)

$$E_T = N_c E_c + (N - N_c)E_i \tag{13}$$

Using the sensible hypothesis that $E_c < E_i$, there exists only one configuration of minimal energy $E_{min} = NE_c$. Furthermore, the number of configurations with energy E_T, $n(E_T)$, depends on N_c according to the number of combinations of N_c elements in N positions, that is (eq 14)

$$n(E_T) = \frac{N!}{N_c!(N - N_c)!} \tag{14}$$

This is a binomial distribution, which peaks at $N_c = N/2$ (Figure 2A).

This distribution of number of states and energies can be represented qualitatively as an energy landscape. The landscape is represented by an energy ordinate and funnel-like curve whose width roughly represents the number of configurations

dx.doi.org/10.1021/ed300302h | J. Chem. Educ. 2014, 91, 1918−1923

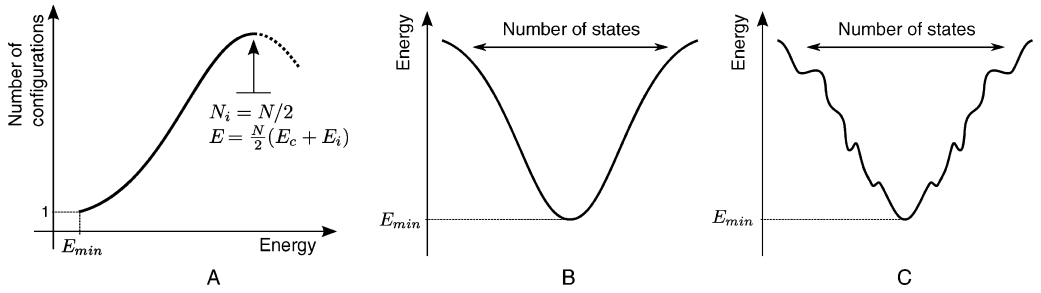


Figure 2. Visualizing the protein folding energy landscape. (A) The binomial model has a single configuration with minimum energy and peaks at $N/2$. (B) Schematic representation of the energy landscape, displaying an increased number of configurations with increasing energy. (C) General schematic representation of the protein folding energy landscape, or protein folding funnel, which is deduced from the present analysis and implies realistic protein folding kinetics (adapted from ref 15).

for each energy. Within the present model, this figure resembles an inverted binomial distribution (Figure 2B).

This funnel-like figure is the typical representation of the protein folding energy funnel. Without the simplifications introduced in the present two-state model (for which, for instance, every correct residue configuration has the same energy and every incorrect residue configuration has the same energy), the funnel would look like a roughed landscape with an average bias toward the minimum energy (Figure 2C). This figure, thus, means that the protein folding energy landscape, from which realistic folding times can be expected, is that for which small perturbations in the structure favor the native state by providing small but steady gains in energy every time a correct contact is formed. At the same time, the narrowing of the funnel upon folding represents an entropic penalty resulting from the small number of configurations that represent the functional form of the protein. For real proteins, for which the landscape is rugged (Figure 2C), there may be significant entropic barriers that might slow down the folding, as the protein might be trapped in local minima. Physically, these minima result from the formation of enthalpically favorable non-native ("wrong") contacts, such as non-native hydrophobic contacts or hydrogen bonds. Examples of such barriers to downhill folding exist even for small proteins.[15]

The entropy of the unfolded states can be quantified in this model. The calculation involves the statistical thermodynamic definition of entropy and might be introduced to advanced undergraduate or graduate students in chemistry. The number of conformations, for each total energy E_T, was already given by eq 14. The entropy of the folded state is, in this model, zero because it is represented by only one configuration. Therefore, the entropy of an unfolded state of N_c correct residues in a sequence of N residues, relative to the native state, is (eq 15)

$$S = R\ln \frac{N!}{N_c!(N - N_c)!} \tag{15}$$

Of course the entropy increases with the increasing number of incorrect configurations, to a maximum at $N/2$. This maximum entropy loss is computed from eq 15 to be of the order of 0.15 kcal mol^{-1} K^{-1} for a 100-residue sequence, which is of the order of, for example, the ribonuclease (124 residues) folding entropy loss of 0.20 kcal mol^{-1} K^{-1}.[14] Other proteins, such as myoglobin (154 residues) display much lower entropic penalties on folding (−0.095 kcal mol^{-1} K^{-1}),[14] such that the estimates from the model agree only very qualitatively with experimental data. The solvent entropy is certainly one of the important factors that should be included to reduce this discrepancy.

Additional Discussion

Part of the limitations of the model presented here results from the fact that there is no correlation between the current conformation of the protein (that is, the current sequence of correct and incorrect conformations) and the conformation in the proceeding step. Physically, this would mean that the structure jumps from one conformation to another in the energy landscape, only with greater probability for configurations with a greater number of residues in correct configurations. Of course, this is not very physically sensible, and it would be expected, in reality, that the conformation of a protein in a given time is correlated to the conformations assumed previously. Therefore, instead of jumping within uncorrelated conformations, the trajectory of configurations is continuous and can be imagined to be a path in the energy landscape. The process of folding, then, can be viewed as the diffusion of a particle in the energy landscape, following a trajectory that is biased toward the minimum energy.[16,17] To understand the kinetics of folding of a real protein completely, one must know the ensemble of trajectories that lead from denatured states to the folded state. This picture is different from what is usually imagined for simple chemical reactions, for which only a few possible reaction coordinates are relevant.

Following each of these trajectories to obtain the actual folding paths for any single protein is a major challenge, which has recently started to be accessed by computational techniques, such as molecular dynamics (MD) simulations. Some of the most promising simulations use massively parallel distributed computing, meaning that people around the world can contribute and interact with state-of-the art research in this area by running the simulations on personal computers (The Folding@Home project).[18] This strategy is particularly suited for the study of multiple folding paths, as discussed above. At the same time, the project provides an interesting platform for science education and dissemination. The results of these efforts support the energy landscape picture in the sense that even for small proteins with fast folding rates, the mechanisms of folding appear to be heterogeneous.[18,19]

■ CONCLUSIONS

Here, a simplified derivation of the solution of the paradox of protein folding was discussed in a form that can be easily introduced to chemistry and biology students with some background in thermodynamics. The derivation was based on simple, but sensible, assumptions on the stability of the

functional structure and allowed for a comprehensive discussion of the nature of the movements and interaction energies at a molecular level. This proved to be a fruitful introduction on the construction of a physical model of a biochemical phenomenon, particularly for undergraduate and graduate students in molecular biology. The validity of the model can be addressed by comparison with experimental thermodynamic data on protein folding, and the incompleteness of the model can also be discussed for the introduction of more sophisticated theories. Variations of the present model were provided by other authors in the scientific literature for the introduction of more advanced topics,[11] and these can follow the present discussion in advanced graduate courses, particularly for chemists and physicists.

■ ASSOCIATED CONTENT

Supporting Information

Questions that can be used for student assessment or to stimulate discussions in classroom. This material is available via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: leandro@iqm.unicamp.br.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The author thanks the financial support of FAPESP (Grants 2010/16947-9, 2011/51348-1, and 2013/05475-7) and CNPq (Grant 470374/2013-6).

■ REFERENCES

(1) Lin, M. M.; Zewail, A. H. Protein folding − simplicity in complexity. *Ann. Phys.* **2012**, *524*, 379−391.

(2) Chakraborty, P.; Zuckermann, R. N. Coarse-grained, foldable, physical model of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 13368−13373.

(3) Gurnon, D.; Voss-Andreae, J.; Stanley, J. Integrating Art and Science in Undergraduate Education. *PLoS Biol.* **2013**, *11*, e1001491.

(4) Garratt, R. C.; Orengo, C. A. *The Protein Chart*; Wiley-VCH: Weinheim, Germany, 2007.

(5) Jones, C. M. An Introduction to Research in Protein Folding for Undergraduates. *J. Chem. Educ.* **1997**, *74*, 1306−1310.

(6) Flores, R. V.; Sola, H. M.; Torres, J. C.; Torres, R. E.; Guzman, E. E. Effect of pH on the Heat-Induced Denaturation and Renaturation of Green Fluorescent Protein: A Laboratory Experiment. *J. Chem. Educ.* **2013**, *90*, 248−251.

(7) Cooper, S.; Khatib, F.; Treuille, A.; Barbero, J.; Lee, J.; Beenen, M.; Leaver-Fay, A.; Baker, D.; Popovic, Z.; Foldit players. Predicting protein structures with a multiplayer online game. *Nature* **2010**, *466*, 756−760.

(8) Eiben, C. B.; Siegel, J. B.; Bale, J. B.; Cooper, S.; Khatib, F.; Shen, B. W.; Stoddard, B. L.; Popovic, Z.; Baker, D.; Foldit players. Increased Diels-Alderase activity though backbone remodeling guided by Foldit players. *Nat. Biotechnol.* **2012**, *10*, 190−192.

(9) Levinthal, C. How to Fold Graciously. In *Mossbauer Spectroscopy in Biological Systems, Proceedings of a Meeting held at Allerton House, Monticello, Illinois*; Debrunner, P., Tsibris, J. C. M., Munck, E., Eds.; University of Illinois Press: Urbana, 1969.

(10) Karplus, M. Behind the folding funnel diagram. *Nat. Chem. Biol.* **2011**, *7*, 401−404.

(11) Plotkin, S. S.; Onuchic, J. N. Understanding Protein Folding with Energy Landscape Theory, Part I: Basic Concepts. *Q. Rev. Biophys.* **2002**, *35*, 111−167.

(12) Zwanzig, R.; Szabo, A.; Bagchi, B. Levinthal's paradox. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 20−22.

(13) Srinivas, G.; Bagchi, B. Study of the dynamics of protein folding through minimalist models. *Theor. Chem. Acc.* **2003**, *109*, 8−21.

(14) Cantor, C. R.; Schimmel, P. R. *Biophysical Chemistry - Part III*; W. H. Freeman: San Francisco, 1980; pp 1075−1104.

(15) Mukherjee, A.; Bagchi, B. Correlation between rate of folding, energy landscape, and topology in the folding of a model protein HP-36. *J. Chem. Phys.* **2003**, *118*, 4733−4747.

(16) Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Socci, N. D. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 3626−3630.

(17) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, Pathways, and the Energy Landscape of Protein Folding: A Synthesis. *Proteins* **1995**, *21*, 167−195.

(18) Snow, C. D.; Nguyen, N.; Pande, V. S.; Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **2002**, *420*, 102−106.

(19) Ensign, D. L.; Pande, V. S. The Fip35 WW domain folds with structural and mechanistic heterogeneity in molecular dynamics simulations. *Biophys. J.* **2009**, *96*, L53−L55.