



Article

The Finite Size Effects and Two-State Paradigm of Protein Folding

Artem Badasyan ^{1,*} , Matjaz Valant ^{1,2}, Jože Grdadolnik ³ and Vladimir N. Uversky ⁴

¹ Materials Research Laboratory, University of Nova Gorica, Vipavska 13, SI-5000 Nova Gorica, Slovenia; matjaz.valant@ung.si

² Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu 610054, China

³ National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia; joze.grdadolnik@ki.si

⁴ Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, 12901 Bruce B. Downs Blvd., MDC07, Tampa, FL 33612, USA; vuversky@health.usf.edu

* Correspondence: artem.badasyan@ung.si; abadasyan@gmail.com

Abstract: The coil to globule transition of the polypeptide chain is the physical phenomenon behind the folding of globular proteins. Globular proteins with a single domain usually consist of about 30 to 100 amino acid residues, and this finite size extends the transition interval of the coil-globule phase transition. Based on the pedantic derivation of the two-state model, we introduce the number of amino acid residues of a polypeptide chain as a parameter in the expressions for two cooperativity measures and reveal their physical significance. We conclude that the k_2 measure, defined as the ratio of van 't Hoff and calorimetric enthalpy is related to the degeneracy of the denatured state and describes the number of cooperative units involved in the transition; additionally, it is found that the widely discussed $k_2 = 1$ is just the necessary condition to classify the protein as the two-state folder. We also find that Ω_c , a quantity not limited from above and growing with system size, is simply proportional to the square of the transition interval. This fact allows us to perform the classical size scaling analysis of the coil-globule phase transition. Moreover, these two measures are shown to describe different characteristics of protein folding.

Keywords: protein folding; two-state model; size scaling; thermodynamic cooperativity



Citation: Badasyan, A.; Valant, M.; Grdadolnik, J.; Uversky, V.N. The Finite Size Effects and Two-State Paradigm of Protein Folding. *Int. J. Mol. Sci.* **2021**, *22*, 2184. <https://doi.org/10.3390/ijms22042184>

Academic Editor: Shin-ichi Tate

Received: 27 January 2021

Accepted: 15 February 2021

Published: 22 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

From the point of view of polymer physics, the folding of a protein is similar to the coil-globule transition of a short polypeptide chain [1]. The coil-globule transition is known as the phase transition of first (in rigid) or second order (in flexible chains) [1,2]. By following the behavior of the order parameter (degree of “nativeness”) $f_N(T) \in [1, 0]$ or its counterpart (“denaturation” degree) $f_D(T) = 1 - f_N(T)$, it is possible to describe the phenomenon; the condition:

$$f_N(T_D) = f_D(T_D) = 0.5 \quad (1)$$

defines the folding temperature T_D . If there are no finite size effects or heterogeneity (the account of heteropolymeric effects in the coil-globule transition is outside the scope of the current study), the order parameter at the transition point undergoes an abrupt all-or-none transformation. Responsible for this coil-globule phase transition are strong correlations between repeat unit conformations, which occur due to the van der Waals interactions between the remote repeat units [1]. Changes in external conditions (temperature, pressure, pH, solution composition, etc.) shift the equilibrium in these effective interactions from repulsion (good solvent regime) via neutral conditions (ideal or theta conditions) to attraction (poor solvent regime), which forces the protein to fold. The hydrogen bonds, which

are responsible for the formation of secondary structures, have a shorter span and influence the conformations locally. In α -helices, one hydrogen bond fixes the conformations of three subsequent residues [3]. Although the hydrogen bond loops in β -hairpins are roughly twice as long and typically span over five to seven residues [4], the interaction is still local. Therefore, according to the Landau–Peierls theorem [5], such hydrogen bonds cannot per se lead to the coil-globule (phase) transition [1,6]. However, if there are long-range interactions present in the system, the formation of secondary structures can change the effective stiffness of the polypeptide chain, increase stability, and thus, promote the coil-globule transition at equal external conditions. Indirect support for such a mechanism arises from the fact that both the coil-helix transition and protein folding occur at the same interval of external parameters [7]. Thermodynamic cooperativity as a concept is often attributed to the sharpness of the phase transition, which results from the spatially correlated behavior of the particles (in this case, repeating units). The situation of the idealized first order phase transition with correlations that extend throughout the system and lead to the discontinuity of the order parameter corresponds to infinite cooperativity and the zero transition interval. When it comes to the folding of single domain globular proteins of just $N < 100$ repeating units long, the limited system sizes impose constraints onto otherwise infinite correlations at the transition point. Consequently, the folding happens over some small temperature interval $\Delta T (\neq 0)$, which needs to be estimated. Using the Taylor expansion cut at first order, it is possible to approximate the order parameter as f_D^{appr} with the help of the tangent at the transition point:

$$f_D(T) \approx f_D^{appr}(T) = f_D(T_D) + f'_D(T)|_{T_D}(T - T_D). \quad (2)$$

From the definitions of initial and final temperatures as $f_D^{appr}(T_1) = 0$ and $f_D^{appr}(T_2) = 1$, one can define the transition interval (see, e.g., [8,9]) as:

$$T_2 - T_1 = \Delta T = f'_D(T)|_{T_D}^{-1}. \quad (3)$$

The derivative of the order parameter at the transition point is the experimentally measurable quantity that provides access to information on the system's cooperativity. The temperature is not the only possible external parameter that can induce the transition. The experiments are often set by changing the concentration of the denaturant such as urea or guanidinium chloride (GdmCl). After repeating the steps behind Equation (2), the resulting expression for the change in the number of bound denaturant molecules during the transition is:

$$n_2 - n_1 = \Delta n = f'_D(n)|_{n_D}^{-1}, \quad (4)$$

so that the thermodynamic cooperativity of the transition can be still estimated by the measured slope of the transition curve at its middle point.

In this paper, we introduce the protein chain length as a parameter into the two-state model, perform the finite-size scaling of protein folding, and compare the two famous criteria of cooperativity.

2. Materials and Methods

The two-state model is the simplest among the folding models, yet is very general and fruitful and therefore deserves a detailed, even pedantic derivation of its well-known formulas, enabling us to trace their origins and limitations. Within the two-state paradigm, the presence of just two possible macroscopic states is assumed: the native globular state with the energy value E_N and the denatured coil one with the energy E_D . To reflect the uniqueness of the native state, a degeneracy $g_N = 1$ is attributed; a $g_D \gg 1$ degeneracy is set for the denatured state to reflect its greater conformational entropic freedom. Without loss of generality, one can assume $E_N = 0$, $E_D \neq 0$ and write down the density of states for the two-state model:

$$g(E) = \delta(E) + g_D \delta(E - E_D), \quad (5)$$

where $\delta(x)$ is the Dirac delta function, resulting in the partition function:

$$Z(\beta) = \int_0^\infty dE g(E) e^{-\beta E} = 1 + g_D e^{-\beta E_D} = [N] + [D], \quad (6)$$

where [...] is the number of repeat units in the native or denatured state and $\beta = 1/T$ is the inverse temperature. The average energy is just the internal energy of the system and follows directly as:

$$\langle E(\beta) \rangle = \frac{\int_0^\infty dE g(E) e^{-\beta E} E}{\int_0^\infty dE g(E) e^{-\beta E}} = -\frac{d \log Z(\beta)}{d\beta} = \frac{g_D e^{-\beta E_D}}{1 + g_D e^{-\beta E_D}} E_D, \quad (7)$$

leading to the heat capacity:

$$C_V(\beta) = -\beta^2 \frac{d \langle E \rangle}{d\beta} = (\beta E_D)^2 \frac{g_D e^{-\beta E_D}}{(1 + g_D e^{-\beta E_D})^2}. \quad (8)$$

The denaturation degree reads:

$$f_D(\beta) = \frac{[D]}{[N] + [D]} = \frac{g_D e^{-\beta E_D}}{1 + g_D e^{-\beta E_D}} = \frac{\langle E(\beta) \rangle}{E_D} = -\frac{1}{E_D} \frac{d \log Z(\beta)}{d\beta}, \quad (9)$$

and the equilibrium constant:

$$K_{eq}(\beta) = \frac{[D]}{[N]} = \frac{f_D(\beta)}{1 - f_D(\beta)} = g_D e^{-\beta E_D}. \quad (10)$$

At the transition point, the numbers of repeat units in the native N or the denatured D state are equal, and with the help of Equation (9), we can express the transition temperature Equation (1) and interval Equation (3) in terms of the two-state model parameters as:

$$T_D = \frac{E_D}{\log g_D}; \quad \Delta T = \frac{4E_D}{\log^2 g_D}. \quad (11)$$

The last expression can be rewritten as:

$$E_D = \frac{4T_D^2}{\Delta T}, \quad (12)$$

resulting in the famous expression for the energetic price of the transition between the two states. Privalov and Kheshinashvili [10] referred to Equation (12) as an approximation, but as we showed above, it is indeed exact within the two-state picture. Since all the above formulae are derived under the assumption of the existence of strictly two states, the results can only be attributed to one cooperative unit, i.e., a part of a molecule that undergoes the transition from N to D as a whole. Microcalorimetry allows the simultaneous measurement of the transition enthalpies for the whole protein molecule and for the cooperative unit [11]. Potentiometric titration also allows the difference in the degree of ionization to be measured for the entire molecule and compared with the value for the cooperative unit [12].

The order of a conformational transition can be evaluated by analyzing the dependence of the slope of the transition on the molecular weight of the protein (M), which is linearly proportional to the degree of polymerization N . It is clear that the slope of the phase transition in small systems depends on the dimensions of this system [1,13]. In the case of the first order phase transition, the slope increases proportionally to the number of units in a system [13], while the slope for the second order phase transition is proportional to the square root of this number [1].

The system sizes can be introduced by the reasonable assumption that each repeating unit of the polypeptide chain can be found in one out of $Q > 2$ rotational isomeric states,

only one of which corresponds to the native state. Since there is N such repeating units, the number of possible states in the denatured conformation for the whole macromolecule and the additive energy of the system read:

$$g_D = (Q - 1)^N; E_D = \epsilon_D N. \quad (13)$$

In view of Equation (11), this means:

$$T_D(N) = \frac{\epsilon_D}{\log(Q - 1)}; \Delta T(N) = \frac{4T_D}{\log(Q - 1)} \frac{1}{N}. \quad (14)$$

This is a very interesting result, which shows that within the two-state paradigm, the denaturation temperature does not depend on the system size. Instead, the transition interval is inversely proportional to N , which naturally leads to a zero interval at $N \rightarrow \infty$, just as it should in the case of the phase transition.

3. Results and Discussion

The criterion of the two-state cooperativity k_2 of protein folding has already been discussed in detail (see, e.g., [14,15] and the references therein). It is defined as the ratio of van 't Hoff and calorimetric enthalpy (energy):

$$k_2 = \frac{\Delta E_{vH}}{\Delta E_{cal}}, \quad (15)$$

where the van 't Hoff energy is:

$$\Delta E_{vH} = - \frac{d \log K_{eq}(\beta)}{d\beta} = E_D, \quad (16)$$

and the amount of heat exchanged during the transition is calculated as the integral under the heat capacity curve:

$$\Delta E_{cal} = \int_0^\infty dT C_V(T) = E_D \frac{g_D}{1 + g_D}. \quad (17)$$

According to Equations (13), (16), and (17), the resulting:

$$k_2 = \frac{\Delta E_{vH}}{\Delta E_{cal}} = 1 + 1/g_D = 1 + O(e^{-N \log(Q-1)}), \quad (18)$$

is an expression that asymptotically tends to one (from above) for large N . It can be concluded that the two-state ansatz, expressed in Equation (5), results in $k_2 = 1$, making it the necessary condition for the transition to be classified as a two-state one. Please note, strictly speaking, that it follows from noting that $k_2 = 1$ means that the transition is a two-state one. In a certain sense, the condition is negative: if k_2 is different from unity, the transition cannot be a two-state one, while if it is close to unity, it is not enough to conclude the two-state behavior. The folding cooperativity measure:

$$\Omega_c = \frac{T_D^2}{\Delta T} \frac{df_D}{dT} \Big|_{T=T_D} \quad (19)$$

was proposed by Klimov and Thirumalai [16] to compare the cooperativities of different proteins. Based on their collection of experimental and simulation data of protein folding, the same authors later suggested a size scaling law for the folding cooperativity measure $\Omega_c \propto N^\zeta$ [17], where $\zeta = 1 + \gamma$ and γ is a susceptibility exponent.

Using our Equation (3), we can significantly simplify the expressions for Ω_c . Li et al. defined the interval $\Delta T^* = T_2^* - T_1^*$ as the width at half-height of the differential curve [17]. One can approximate the peaked curve by a rectangle with sides at T_1^* and T_2^* and a height

$|f'_D(T)|_{T_D}$ in such a way, that $1 = \int_0^\infty f'_N(T)dT \approx |f'_D(T)|_{T_D}(T_2^* - T_1^*)$. With the account of Equation (3), this leads to the obvious $\Delta T^* = \Delta T$, proving that both definitions of the transition interval are equivalent, at least in the sense of asymptotic, size scaling relations. The same Equation (3), when inserted into the cooperativity measure Equation (19), simply results in:

$$\Omega_c = \left(\frac{T_D}{\Delta T} \right)^2. \quad (20)$$

The result is not surprising, since the $\frac{T_D}{\Delta T}$ ratio is common in the studies of finite size effects at phase transitions [18–20]. In view of Equation (14), valid for the two-state model, it simply means that:

$$\Omega_c = (\log g_D/4)^2 \propto N^2. \quad (21)$$

However, if not bound to the two-state paradigm, the more general and model-independent formula expressed with Equation (20) allows establishing direct links between the well-known size scaling relations and the cooperativity measure Ω_c . To take into account the possibility for both the first and second order mechanisms of the phase transition, $\frac{T_D}{\Delta T} \propto N^{1/d\nu}$ scaling should be considered [18–21] (instead of N^1 , used by Li et al.), where $d\nu$ is a critical exponent of the correlation length or radius of gyration; the $d\nu = 1$ and $d\nu = 2$ values would correspond to the first and second order phase transition, accordingly. From Equation (20), it immediately follows that:

$$\Omega_c \propto N^{2/d\nu}, \quad (22)$$

with $d\nu = 1$ for our case of exactly two-state folders. Klimov, Thirumalai, and Li [16,17] justified the necessity for the new critical exponent $\zeta = 1 + \gamma$ by fitting the points from their dataset reported in [17] to their expression Equation (19). However, our Equation (22) can be used instead, without invoking the new critical exponent ζ . In order to compare the two approaches, in Figure 1, we re-plot the data from [17] and compare them with our Equation (22). The data points for $\ln \Omega_c$ and $2 \ln \frac{T_D}{\Delta T}$ vs. $\ln N$ are superimposed, and the corresponding fitted straight lines are indistinguishable, thus validating Equation (22) over the set of data from [17]. The fit results in $d\nu_{exp} = 0.92$, which is close to, but not equal to one. The scaling on the basis of Equation (22) nicely fits the experimental trends and thus allows us to treat protein folding as a true phase transition in a finite system in the sense of Lifshitz–Grosberg–Khokhlov [2]. The fact that the transition interval has the same size scaling exponent as the correlation length is a nice example of the contribution of correlations in protein conformations to folding cooperativity.

There is further experimental evidence that supports our view. Ptitsyn and Uversky proposed the molten globule as the third thermodynamic state of protein molecules in a number of publications [22,23]. Based on the systematic analysis of data on urea and guanidinium chloride induced transition of globular proteins from the native to the unfolded state ($N \rightarrow U$), from the native to the molten globule ($N \rightarrow MG$) state, and from the molten globule to the unfolded state ($MG \rightarrow U$), it has been shown that in all these cases, the cooperativity of unfolding increases linearly with the increase in the molecular weight of the protein up to 25–30 kDa [22,23]. In fact, this cooperativity of all three transitions measured in terms of Δn (see Equation (4)) follows $\log \Delta n = d\nu \log M - b$, with $d\nu_{N-U} = 0.97$, $d\nu_{N-MG} = 1.02$, and $d\nu_{MG-U} = 0.89$, all close to the $d\nu = 0.92$ value, estimated from the temperature inspired set of data from [17]. This means that such a dependence of the cooperativity of urea-induced and guanidinium chloride-induced transitions in small proteins on their molecular weight suggests that all three types of transitions are all-or-none, indicating that the molten globule state is separate from the native and unfolded state by all-or-none transitions [22,23]. Thus, the experimental data on denaturant-induced unfolding of small globular proteins are consistent with the linear $\log \Omega_c$ vs. $\log N$ dependence described in [17].

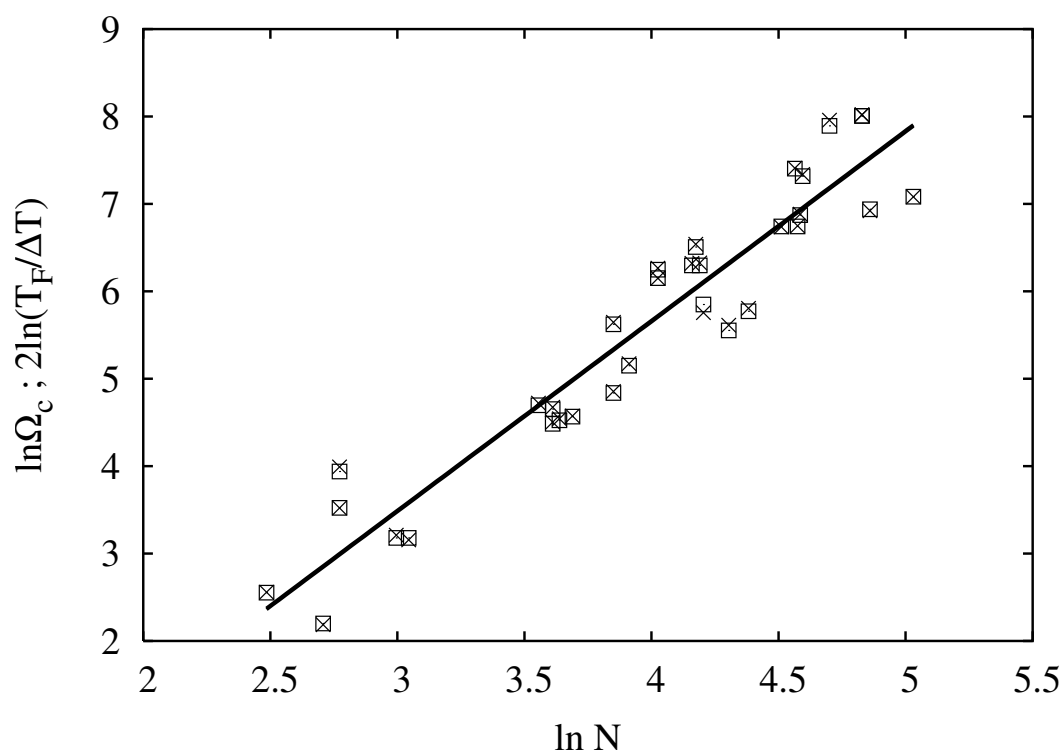


Figure 1. The dependence of $\ln \Omega_c$ ($\times$) and $2 \ln \left(\frac{T_D}{\Delta T} \right)$ ($\square$) vs. $\ln N$. Data taken from the database reported in [17]. ($\times$) correspond to the Klimov and Thirumalai Equation (19), while ($\square$) visualize our result Equation (20). The straight lines, corresponding to the linear fits for both data point collections, are indistinguishable on the graph.

The comparison of cooperativity measures shows that each of them has advantages and drawbacks. The strict two-state assumption, expressed in Equation (5) allows the derivation of $k_2 \approx 1$ at large N , which is therefore a necessary condition for the two-state folding. Independent of the chain length, k_2 allows the statement about which of the proteins under consideration comes closer to the ideal two-state behavior. Instead, in the same $N \rightarrow \infty$ limit, Ω_c tends to infinity, which means that under other equal conditions, longer chains have higher values of the cooperativity measure Ω_c . On the other hand, k_2 , as defined by Equation (15), contains both equilibrium and kinetic quantities, which are only equal when the system has reached equilibrium, and the deviation from unity can be attributed to kinetic traps (see also [15] for the definition and discussion about the kinetic cooperativity). Regarding Ω_c , once expressed through $\frac{T_D}{\Delta T}$, it becomes a criterion similar to those introduced in other areas of physics to deal with the effects of a finite size at phase transitions. The last fact puts it on very solid grounds.

4. Conclusions

In summary, we contribute to a better understanding of the physical basis of the two cooperation criteria under consideration. For the first time, the size scaling expressions for the cooperation criteria are derived and analyzed (Equations (18) and (20)). As a result, we concluded that k_2 can be conveniently used to compare cooperativity for individual proteins, while Ω_c is more useful for comparing protein folding data sets with respect to size scaling analysis.

Author Contributions: Conceptualization, A.B. and V.N.U.; writing—original draft preparation, A.B. and V.N.U.; writing—review and editing, A.B., M.V., and J.G.; funding acquisition, A.B., M.V., and J.G. All authors read and agreed to the published version of the manuscript.

Funding: A.B. acknowledges the partial financial support from Erasmus+ Project No. (2018-1-SI01-KA107-046966); A.B. and J.G. acknowledge the partial financial support from the Javna Agencija

za Raziskovalno Dejavnost RS through Project No. (J1-1705); A.B. and M.V. acknowledge the partial financial support from the Javna Agencija za Raziskovalno Dejavnost RS through Program No. (P2-0412).

Data Availability Statement: Data Availability Statement: Data sharing not applicable.

Acknowledgments: Discussions with Hue Sun Chan from the University of Toronto are thankfully acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Grosberg, A.Y.; Khokhlov, A.R. *Statistical Physics of Macromolecules*; AIP Press: New York, NY, USA, 1994.
2. Lifshitz, I.M.; Grosberg, A.Y.; Khokhlov, A.R. Some problems of the statistical physics of polymer chains with volume interaction. *Rev. Mod. Phys.* **1978**, *50*, 683.
3. Cantor, C.; Shimmel, T. *Biophysical Chemistry*; Freeman and Co.: San-Francisco, CA, USA, 1980.
4. Stanger, H.E.; Syud, F.A.; Espinosa, J.F.; Giriat, I.; Muir, T.; Gellman, S.H. Length-dependent stability and strand length limits in antiparallel β -sheet secondary structure. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 12015.
5. Landau, L.; Lifshits, E. *Statistical Physics*; Pergamon: Oxford, UK, 1988.
6. Chan, H.S.; Bromberg, S.; Dill, K.A. Models of cooperativity in protein folding. *Phil. Trans. R. Soc. Lond. B* **1995**, *348*, 61.
7. Uversky, V.N.; Fink, A.L. The chicken-egg scenario of protein folding revisited. *FEBS Lett.* **2002**, *515*, 79–83.
8. Lazurkin, Y.S.; Frank-Kamenetskii, M.D.; Trifonov, E.N. Perspectives report: Melting of DNA: Its study and application as a research method. *Biopolymers* **1970**, *9*, 1253–12306.
9. Mamasakhlisov, E.S.; Badasyan, A.V.; Tsarukyan, A.V.; Grigoryan, A.V.; Morozov, V.F. The generalized model of polypeptide chain describing the helix-coil transition in biopolymers. *arXiv* **2005**, arXiv:cond-mat/0506791.
10. Privalov, P.L.; Khechinashvili, N.N. A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study. *J. Mol. Biol.* **1974**, *86*, 665–684.
11. Privalov, P.L. Stability of Proteins Small Globular Proteins. *Adv. Prot. Chem.* **1979**, *33*, 167–241.
12. Hermans, J., Jr.; Acampora, G. Reversible denaturation of sperm whale myoglobin. I. Dependence on temperature, pH, and composition. *J. Am. Chem. Soc.* **1967**, *89*, 1543–1547.
13. Hill, T.L. *Thermodynamics of the Small Systems*; Wiley: New York, NY, USA, 1963.
14. Kaya, H.; Chan, H.S. Polymer principles of protein calorimetric two-state cooperativity. *Proteins* **2000**, *40*, 637–661.
15. Chan, H.S.; Shimizu, S.; Kaya, H. Cooperativity principles in protein folding. *Meth. Enzym.* **2004**, *380*, 350–379.
16. Klimov, D.K.; Thirumalai, D. Cooperativity in protein folding: from lattice models with sidechains to real proteins. *Fold. Des.* **1998**, *3*, 127–139.
17. Li, M.S.; Klimov, D.; Thirumalai, D. Finite size effects on thermal denaturation of globular proteins. *Phys. Rev. Lett.* **2004**, *93*, 268107.
18. Fisher, M.E.; Berker, A.N. Scaling for first-order phase transitions in thermodynamic and finite systems. *Phys. Rev. B* **1982**, *26*, 2507.
19. Challa, M.S.S.; Landau, D.P.; Binder, K. Finite-size effects at temperature-driven first-order transitions. *Phys. Rev. B* **1986**, *34*, 1841.
20. Alves, N.A.; Hansmann, U.H.E. Partition function zeros and finite size scaling of helix-coil transitions in a polypeptide. *Phys. Rev. Lett.* **2000**, *84*, 1836.
21. De Gennes, P.-G. *Scaling Concepts in Polymer Physics*; Cornell University Press: New York, NY, USA, 1985.
22. Uversky, V.N.; Ptitsyn, O.B. All-or-none solvent-induced transitions between native, molten globule and unfolded states in globular proteins. *Fold. Des.* **1996**, *1*, 117–222.
23. Ptitsyn, O.B.; Uversky, V.N. The molten globule is a third thermodynamical state of protein molecules. *FEBS Lett.* **1994**, *341*, 15–18.