

Protein Folding: An Antagonistic Reaction of Spontaneous Folding and Diffusion Limited Aggregation in Nature

Chia-Ching Chang^{1,2,3,*} and Lou-Sing Kan^{4,†}

¹*Department of Biological Science and Technology,
National Chiao Tung University, HsinChu, Taiwan 30050*

²*National Nano Device Laboratory, HsinChu, Taiwan 30078*

³*Institute of Physics, Academia Sinica, Taipei, Taiwan 11529*

⁴*Institute of Chemistry, Academia Sinica, Taipei, Taiwan 11529*

(Received Jan 9, 2007)

Protein folding may follow a spontaneous process or a reaction-path directed process, determined by various folding transition boundaries due to intrinsic properties of the protein. A general first-order-like state-transition model predicts that a protein might be trapped in an aggregated state when the folding path crosses the transition boundary. Both experimental study and molecular simulation indicated that protein within this particular transition region might involve intermolecular interactions. Therefore, a direct folding process may have been a combination of an antagonistic reaction of spontaneous folding and a diffusion limited aggregation. In this paper, the protein folding mechanism and the theoretical basis of time limited diffusion/aggregation process are elaborated. The application of auto-correlation function in time dependent biological studies is also discussed.

PACS numbers: 87.15.Cc, 87.15.hg, 87.15.km

I. PROTEIN FOLDING: A FIRST-ORDER-LIKE STATE-TRANSITION REACTION

Protein folding may follow a spontaneous process [1] or a reaction-path directed process [2] *in vitro*. A choice between the two may be determined by the intrinsic properties of proteins, e.g. the varying folding transition boundaries. However, a general model, named “first-order-like state transition model”, in which aggregated proteins exist within finite boundaries, can encompass both processes without any conflicts [3~7]. According to this model, the folding path of the protein may not be unique. It can be folded, without being trapped in an aggregated state, via a carefully designed refolding path circumventing the transition boundary, i.e., via an overcritical path [3~7]. The intermediates, following an overcritical path, are in a molten globular state [8], and their behavior is consistent with both a sequential [9] and a collapse model [10]. However, both soluble (folded) and precipitated (unfolded) proteins can be observed in direct folding reaction path *in vitro*. In terms of the “first-order-like state transition model” language, this can be described as stepping across the state transition line in the protein folding reaction phase diagram [3,4]. Since in protein refolding it is important to prevent protein aggregation *in vitro*, similarly, in bio-medical applications, the revelation of the mechanism of the formation of the two

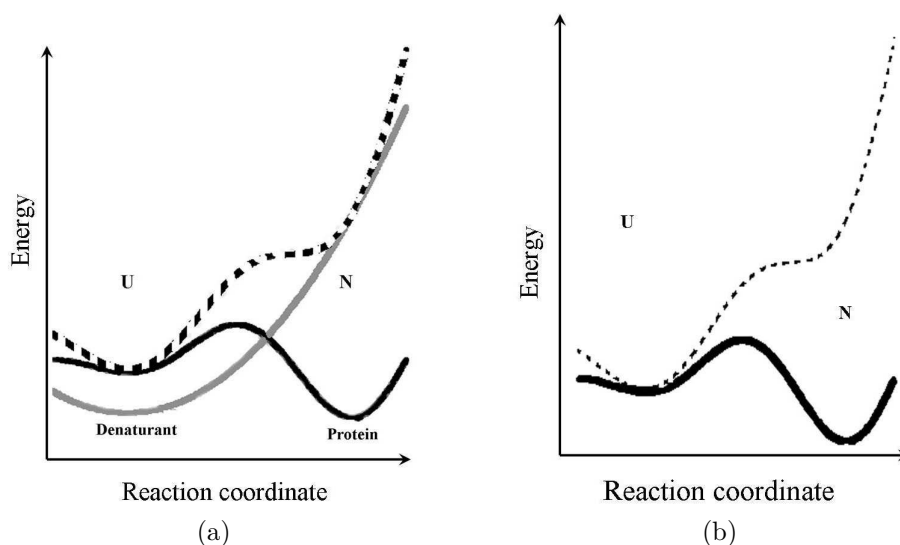


FIG. 1: (a) Protein folding two-state (dual well) transition model. The gray line denoted the proposed potential well of denaturant. The dot line denoted the hybrid protein-denaturant complex potential wells. The black line is the reaction energy from unfolded to native state. (b) The model of direct folding reaction.

states (unfolded and folded) becomes significant.

Previous studies have indicated that chemical environment [11], temperature [12], pH [13], ionic strength [14], dielectric constant [15], and pressure [16] - considered as solvent effects collectively - could affect the fundamental structure, thermodynamics and dynamics of polypeptides/proteins. The reaction ground state can be expressed as a dual-well potential according to the two-state transition model. The conformational energy, in general, of unfolded state is relatively higher than that of the native state (Figure 1). When the system reaches thermal equilibrium most protein molecules are found in their native state. No unfolded or intermediate states are observable. However, as describe previously [3~7], if a denaturant is added the reaction potential may change accordingly as indicated in Figure 1a. Then, an unfolded protein may be stable, as it is now at the lowest energy under the newly established equilibrium. The energy of the system can be expressed as follows:

$$H_T = H_p + \lambda H_s \quad (1)$$

Where H_T , H_p , H_s denote the potential energy of the interacting protein-solvent total system, the protein, and the solvent, respectively. λ is a weighting factor of the solvent environment ($0 \leq \lambda \leq 1$), approaching unity when a denaturant is present as pure solvent; and decreases in value as the concentration of the denaturant is reduced.

When λ of the system is changed drastically, direct folding ensues, leading to release of some of the bound denaturant (Figure 1b). According to the Donnan effect in a macromolecule-counter ions interactive system, the diffusion of the bound denaturant can

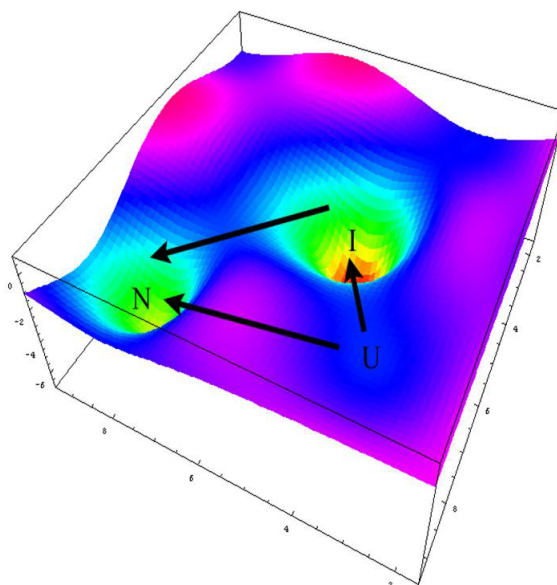


FIG. 2: Three-well model of multi-protein molecules folding reaction. The U denoted the unfolded state. N denoted the native state and I denoted the protein-protein complex (aggresome) intermediate. Where the I may be deeper than native state N.

be expressed by Fick's first law:

$$\vec{J} = -D\nabla n \quad (2)$$

Where n denotes the concentration of the denaturant that is dissociated from the protein. D denotes the diffusive constant and vector J denotes the flux, respectively, of the solute. According to the Einstein relation:

$$D = \frac{kT}{6\pi\eta R_H} \quad (3)$$

Where k is the Boltzmann constant; T is the temperature in Kelvin; η is the viscosity of the solvent, and R_H is the hydration radius of the solutes. Due to the intrinsic diffusion process, the solute exchange processes are not synchronous for all protein molecules. Therefore, the folding rate of protein may not be measured directly by simple spectral technique, i.e., the stopped-flow CD [17], continue-flow CD [18] or fluorescence [19]. However, the reaction interval of protein folding can be revealed by auto-correlation of reaction time from these direct measurements. The detail mechanism and example will be discussed later.

If we look at the energy landscape funnel model of protein folding [20], it appears that proteins can be trapped in a multitude of local minima of the potential well in a complicated protein system. The native state, though, is at the lowest energy level. When thermal

equilibrium is reached, most of the protein molecules are located in the lowest energy state, with a population ratio as low as $e^{-\frac{\Delta E}{kT}}$, according to the Maxwell-Boltzmann distribution in thermodynamics. The ΔE denotes the energy difference between native state and a local minimum; k and T denote the Boltzmann constant and temperature in Kelvin, respectively. At high concentration (≥ 0.1 mg/ml), however, considerable amount of insoluble protein has been observed in protein folding [5,6,21], indicating that insoluble proteins are at an even lower energy state than the native protein. Therefore, by considering the inter-molecular interactions during the protein folding process the reaction energy landscape may be expressed as a three-well model (Figure 2). As shown in Figure 2, the unfolded protein (U) is in the highest energy state; the native protein (N) is in the lower energy state. However, the intermediate (I) that may cause further protein aggregation/precipitation is in the lowest energy state. Although the energy state of the intermediate/aggregosome is the lowest energy state, the conformational energy of the individual proteins composing the aggregosome may not be lower than the native protein. Namely, in single molecular simulation this extra potential well of intermediate (I) is non-existent. Therefore, in the conventional energy landscape model (the single molecule simulation model), the lowest energy state “I” cannot be observed. According to the Zwanzig’s definition of state, the protein molecules in the intermediate (I) belong to an unfolded state [22]. Hence, in a direct folding reaction, the soluble (N) and the insoluble parts (U) can coexist and be observed simultaneously, similar to the situation where the phase transition line is crossed in reactions congruent to the “first-order phase transition” model. Therefore, we named the protein folding reaction as “first-order like state transition model” (Figure 3).

The $\Phi(n_1, n_2, \dots)$ in Figure 3 denotes the folding status of protein, where n_1, n_2, \dots represent the variables affecting the folding status, such as, temperature, concentration of denaturants, and etc. The reaction curve indicates an over-critical reaction path of a quasi-static folding reaction. The gray area in Figure 3 indicates the state transition boundary of protein folding. The gray line and dash line indicate the reaction path of direct folding. By combining the three-well model (Figure 2) and direct folding reaction of the “first-order like state transition model” (Figure 3), we realized that the protein molecules those folded along the direct folding path might fold spontaneously or form aggregate. The detailed mechanisms of these two reaction fates of protein will be discussed below.

II. SPONTANEOUS FOLDING MAY BE DRIVEN BY ENTHALPY-ENTROPY COMPENSATION

As indicated previously, the conformation of protein changed with changes of the solvent environment. It seems that the protein may fold spontaneously, such as in Anfinsen’s experiment [1] and direct folding reactions. The protein folding reaction, similar to all chemical reactions, reaches its equilibrium by following the fundamental laws of thermodynamics. Although protein folding has been studied extensively in certain model systems for over forty years, the driving force at the molecular level remains unclear until recently.

It is known that polymers and macromolecules may self-assemble/self-organize into

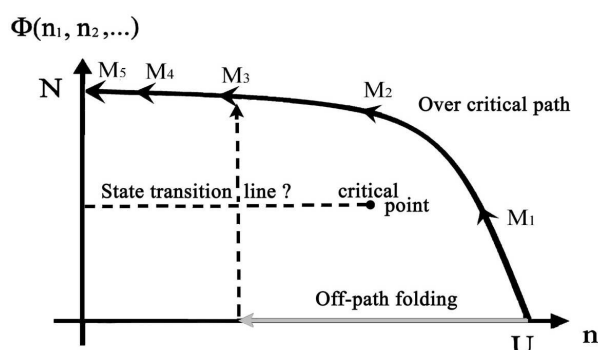


FIG. 3: The protein folding diagram of “First-order like state transition model”. Where the $\Phi (n_1, n_2, \dots)$ denoted the folding status (the order parameter) of protein. The n, n_1, n_2, \dots denoted the variables of which affect the folding status such as temperature, concentration of denaturants, and etc. The horizontal dash-line denoted the proposed state transition line. The vertical dash line denoted the direct folding (off-path) reaction pathway. Where the unfolded protein in denature solvent environment which was diluted into native solvent environment rapidly and the folding reaction may happen spontaneously and the crosses the state transition line and may cause protein aggregation/precipitation

a wide range of highly ordered phase/state at thermal equilibrium [23~26]. In a condensed solvent environment, large molecules may self-organize to reduce their effective volume. Meanwhile, the number of the allowed states (Ω) of small molecules, such as buffer salt and other counter-ions in solution, increases considerably. Therefore, the entropy of the system, $\Delta S = R \ln (\Omega_f / \Omega_i)$, becomes large, where i and f denote initial and final state, respectively. Meanwhile, the enthalpy change (ΔH) between unfolded and native protein is around hundreds Kcal/mol [27]. Therefore, the Gibbs free energy of the system, $\Delta G = \Delta H - T\Delta S$, becomes more negative in this system when the large molecules self-organize [28]. Similar entropy-enthalpy compensation mechanism has been used to solve the reaction of colloidal crystals that self-assemble spontaneously [29,30].

According to our studies [3~7], the effective diameter of unfolded protein is about 1.7 to 2.5 fold larger than the folded protein. Therefore, with the same mechanism, those macromolecules (proteins) may tend to reduce their effective volumes and increase the system entropy when thermal equilibrium is reached. The increase in entropy may compensate for the change of the enthalpy of the system and enable the reaction to take place spontaneously. This may be the reaction molecular mechanism of protein spontaneous folding reactions.

III. INTER-MOLECULAR INTERACTION OF UNFOLDED PROTEIN LEADS TO PROTEIN AGGREGATION

From the energy landscape model the native state is at the lowest energy of all conformations [31,32]. There are local minima on the potential energy landscape that may trap proteins into a glassy or misfolded state, forming aggregates or precipitates. However, most protein molecules are found in the lowest energy state, i.e. the native state when the reaction reaches a thermal equilibrium according to the Maxwell-Boltzmann distribution. In the direct folding experiment, most of the proteins formed aggregates or precipitates, with only a minute portion folded into the native state. This, however, is in conflict with the model of energy landscape [31,32]. We found that the aggregates may be caused by the inter-molecular interaction microscopically, via the MALDI-TOF mass spectroscopy [33]. Therefore, the intermolecular interactions should be considered in the molecular simulation as well.

By combining molecular simulation and experimental approaches [33] we found that protein folding and diffusion limited aggregation processes are mutually antagonistic. The unfolded protein, such as lysozyme, can be refolded by direct dilution into 100-fold native buffer by way of a so-called direct folding process path or off-path folding process under an intrinsic entropy-driven force. The folding reaction may take a path across the state transition line/boundary, as in a first order like transition model resulting in both folded and aggregated proteins. By using Raman spectroscopy, we found that the protein aggregates contained both disulfide bridges and reduced thiol groups, which may have been caused, partly, by intermolecular mis-linkages. These stable aggresomes were dissolved in acetonitrile and then analyzed by MALDI-TOF mass spectroscopy. For example, stable aggresomes such as tri-mer, hepta-mer and nona-mers of lysozyme were detected in acetonitrile in a previous study [33]. The dielectric constant of acetonitrile is around 37.5 at 21.1 ° C, in the realm of a low hydrophilic solvent, and is a better solvent than water for hydrophobic aggregates. Therefore, the aggresomes precipitated in aqua solution may have formed through hydrophobic interactions besides mis-linkages. However, a native aggresome, such as inclusion body of *E. coli*, did not show to contain oligomers of protein by using the same procedure. Comparing both *in vitro* and *in vivo* experiments we conclude that the hydrophobic interaction is the major factor for aggresome formation [33]. Therefore, suitable protection chaperons and enzymes preventing hydrophobic interactions and mis-linked disulfide bonds formation are necessary *in vivo*. However, the proteins do not aggregate in the native state. This may indicate the aggregation process only occurs during the folding process.

We performed a simulation study in which the protein particles were randomly distributed in a defined three-dimensional box with periodic boundaries. The size of the simulation box is determined by the cubic root of the particle numbers multiplied by the mean distance between protein particles. In order to reduce the calculation time, we used a periodical boundary for the simulation space. Meanwhile, the optimal particle number for molecular simulation is 216 and the mean distance between each protein can be determined from the initial concentration of the protein [33]. The particles are allowed to move ran-

domly with fixed step size, and the relative walking time between collisions can be derived from Equation 6 in the next section. The gyration radius of the particle is shrunk from non-fold to its native radius by a function of temperature dependent-exponential decay. As shown in gyration radii shrink, the aggregation activation factor was set as exponential decay. The rest of the protein particles that do not interact with each other will be treated as the ones fold spontaneously. By comparing the initial number of particles and the number of particles that fold spontaneously, the protein aggregation fraction can be obtained.

By comparing the aggregation experimental and the simulation results, we established that model simulation is a feasible way of exploring the molecular mechanism of protein aggregation. The fundamental basis of this simulation is identical, in concept, to the diffusion-limited aggregation (DLA). In summary, the protein folding in a cell-free system composes of antagonistic reactions of both spontaneous folding and DLA.

IV. AUTO-CORRELATION FUNCTION REVEALED THE PROTEIN FOLDING TIME

As mentioned in the previous section, aggregations occur when proteins do not fold properly into the stable state. Therefore, by measuring the ratio of folded and aggregated proteins, we can estimate the time scale for the protein to fold into a stable state. The stable folding time can be calculated as follows:

The concentrations of the protein, diluted by the native buffer, can be converted to the mean-distance (\bar{X}) between the proteins. According to a three dimensional random-walk model, the mean distance (\bar{X}) is equal to the root mean square distance (X_{rms}); and the mean collision time for the protein diffusing across this distance can be determined by the following equation:

$$\bar{X} \approx X_{rms} = \langle X^2 \rangle^{\frac{1}{2}} = (6Dt)^{\frac{1}{2}}, \quad (4)$$

where D and t denote the diffusion constant of protein and time, respectively, as defined in Eq. 3. η can be replaced by the viscosity of water (η 1 centipoises (cp)) since it did not change much in dilute solutions [3, 33]. R_H is the effective hydration radius of the protein. According to ours and other studies, the unfolded protein is about 1.5~2 times larger than its native protein [3~7]; and the radius of protein, R_H , at time t, may be expressed as an exponential decaying function [32]:

$$R_H(t) = R_0(1 + ne^{-ct}) \quad (5)$$

where R_0 is the hydration radius in its native form; cgs the hydration radius collapse factor and the optimal value is 0.001 as described in the simulation section; n is the radius expansion ratio of the unfolded protein.

The general form of mean collision time during protein diffusive process can be expressed as follows:

$$nX_{rms}^2 \times e^{-ct} - 6D_0t + X_{rms}^2 = 0 \quad (6)$$

$$t = \frac{cX_{rms}^2 + 6D_0 \text{ProductLog}\left[\frac{ce^{-\frac{cX_{rms}^2}{6D_0}} nX_{rms}^2}{6D_0}\right]}{6cD_0}, \quad (7)$$

where the ProductLog can be expanded as follows,

$$\text{ProductLog}\left[\frac{cne^{-\frac{cX_{rms}^2}{6D_0}} X_{rms}^2}{6D_0}\right] = \frac{cnX_{rms}^2}{6D_0} - \frac{(cn)^2 X_{rms}^4}{18D_0^2} + \frac{(cn)^3 X_{rms}^6}{54D_0^3} - \frac{7(cn)^4 X_{rms}^8}{972D_0^4} + \frac{(cn)^5 X_{rms}^{10}}{324D_0^5} - \dots \quad (8)$$

Therefore, the mean collision/diffusion time of each initial concentration of unfolded protein can be determined. In the case of hen egg white lysozyme, their mean collision/diffusion times are 8.2 μs for 200 $\mu\text{g/mL}$ and 101.3 μs for 5 $\mu\text{g/mL}$ at 295 K [33]. The mean collision time is also an upper limit for the protein to reach a stable state in the diffusive processes. Namely, if proteins do not fold but settle into a stable state during this period, they may aggregate and precipitate.

However, macroscopically speaking, as that conducted in an experiment, it is impossible to synchronize all particles of proteins to fold from unfolded state to their native state. The relationship between aggregation and reaction time can be revealed by the autocorrelation function (ACF) of the protein aggregation fraction. Therefore, the ACF function can be expressed as follows:

$$\Gamma(\Delta t) = \langle C(t) \cdot C(t - \Delta t) \rangle, \quad (9)$$

where $C[t]$ denotes the aggregation fraction at time t . Δt denotes the time lapse between the observations. The value of $\Gamma(\tau_c)$ is the half-width of the ACF [34] according to Equation 9. The stable-state folding time of lysozyme is about 25.5 to 27.5 μs at 295 and 279 K, respectively [33]. This correlation time indicated that the lysozyme in such solvent environment may need 25.5 to 27.5 μs to reach the stable state. However, if the molecules of protein collide and interact with each other during this time period, they may have the chance to form aggregates, and precipitate as a result [33].

By comparison with the model of the first-order-like state transition, the state transition line/region of direct folding reaction may have been caused by the interactions among unfolded proteins. An over critical reaction regulated by solvent environment and denaturant might play a role to avoid effective collision between proteins and protein folds spontaneously during this period [3]. These may explain the possible mechanism and function of molecular chaperons [3~7] those facilitate the protein folding.

V. PROTEIN FOLDING REACTION FOLLOWS THE FIRST-ORDER-LIKE STATE TRANSITION IN A CELL

Cells avoid accumulating potentially toxic aggregates by mechanisms, such as, the suppression of aggregate formation by molecular chaperones, and the degradation of misfolded proteins by proteasomes [35]. In eukaryotic systems, HSP70 serves as molecular chaperone, which directly interacts with unfolded proteins, and HSP60-HSP10 complex provides a hydrophobic cave in facilitating protein folding [36]. In prokaryotic systems, GroEL and GroES play roles similar to those of HSP60-HSP10 complexes [37]. These direct or indirect interactions with unfolded proteins will change the reaction potential in a way similar to the behavior of urea and solvent acidity (or pH) in this study. Moreover, under physiological conditions, proteins such as growth hormones will not aggregate. These observations indicated that chaperone helps the protein fold while preventing the state transition from taking place. Our studies with chemical molecules mimic those effects of molecular chaperone. In fact, those *in vivo* reactions are more efficient than those *in vitro*. Therefore, we suspect that for those proteins that have the ability to fold may follow the over critical reaction path and fold continuously without being trapped in the precipitate/aggregate state *in vivo*, and the first-order like state transition reaction model may represent the true reaction model of protein folding. The model describes a non-unique reaction path for proteins such that they can fold continuously when the reaction path does not cross the state transition line [3,4].

VI. CONCLUSION

In summary, the folding reaction or kinetics always follows thermal dynamic rules. Even though protein is a mesoscopic system, it still has to follow physical laws. In our opinion, the first order like state transition model may be a universal model for protein folding reactions, although approaches from all possible angles are needed to reveal a complete picture of the true protein folding kinetics and mechanism.

VII. ACKNOWLEDGMENTS

This study was supported in part by grants NSC 95-2112-M-009-019 (to CCC), NSC 94-2120-M-001-008 (to LSK) and Academia Sinica, Taipei, Taiwan.

References

- * Electronic address: ccchang01@faculty.edu.tw
- † Electronic address: lskan@chem.sinica.edu.tw
- [1] C. B. Anfinsen, E. Haber, M. Sela, and F. H. White Jr., Proc. Natl. Acad. Sci. **47**, 1309 (1961).
- [2] C. Levinthal, J. Chem. Phys. **65**, 44 (1968).

- [3] C.-C. Chang, X.-C. Yeh, H.-T. Lee, P.-Y. Lin, and L. S. Kan, *Physical Review E* **70**, 011904 (2004)
- [4] C.-C. Chang, Y.-C. Su, M.-S. Cheng, and L. S. Kan, *Physical Review E* **66**, 021903 (2002)
- [5] C.-C. Chang, C. T. Tsai, and C. Y. Chang, *Protein Eng.* **5**, 437 (2002)
- [6] C.-C. Chang, Y.-C. Su, M.-S. Cheng, and L. S. Kan, *J. Biomol. Struct. Dyn.* **21**, 247 (2003).
- [7] Y.-L. Liu, H.-T. Lee, C.-C. Chang, and L. S. Kan, *Biochem. Biophys. Res. Commun.* **306**, 59 (2003).
- [8] P. S. Kim, and R.L. Baldwin, *Annu. Rev. Biochem.* **51**, 459 (1982).
- [9] T. Y. Tsong, R. L. Baldwin, *J. Mol. Biol.* **63**, 453 (1972).
- [10] H. S. Chan, and K. A. Dill, *J. Chem. Phys.* **100**, 9238 (1994)
- [11] Y. Levy, J. Jortner, and O. Becker, *Proc. Natl. Acad. Sci.* **98**, 2188 (2001)
- [12] T. E. Creighton, in *Proteins*, W. H. Freeman ed., New York. (1993)
- [13] A. H. Elcock, *J. Mol. Biol.* **294**, 1051 (1999).
- [14] N. P. Ulrikh, G. Anderluh, P. Macek, and T. V. Chalikian, *Biochemistry.* **43**, 9536 (2004).
- [15] F. Despa, A. Fernandez, and R. S. Berry, *Phys. Rev. Lett.* **93**, 228104 (2004).
- [16] M. B. Seefeldt, J. Ouyang, W. A. Froland, J. F. Carpenter, and T. W. Randolph, *Protein Sci.*, **13**, 2639 (2004).
- [17] Z. D. Su, M. T. Arooz, H. M. Chen, C. J. Gross, T. Y. Tsong, *Proc Natl. Acad. Sci* **93**, 2539 (1996).
- [18] H. Roder, K. Maki, H. Cheng, M. C. Shastry, *Methods.* **34**, 15 (2004).
- [19] C. A. Royer, *Methods Mol. Biol.* **40**, 65 (1995).
- [20] P. G. Wolynes, *Proc. Natl. Acad. Sci.* **94**, 6170 (1997).
- [21] S. Misawa, and I. Kumagai, *Biopolymers* **51**, 297 (1999).
- [22] R. Zwanzig, *Proc. Natl. Acad. Sci.* **94**, 148 (1997).
- [23] A. D. Dinsmore, J. C. Crocker, and A. G. Yodh, *Curr. Opin. Colloid Interface Sci.* **3**, 5 (1998).
- [24] A. P. Gast, and W. B. Russel, *Phys. Today* **51**, 24 (1998).
- [25] P. N. Pusey, and W. van Megen, *Nature* **320**, 340 (1986)
- [26] B. J. Ackerson, and P. N. Pusey, *Phys. Rev. Lett.* **61**, 1033 (1988).
- [27] T. Y. Tsong, R. P. Hearn, D. P. Warthal, and J. M. Sturtevant, *Biochemistry*, **9**, 2666 (1970)
- [28] N. Tokuriki, M. Kinjo, S. Negi, M. Hoshino, Y. Goto, I. Urabe, T. Yomo, *Protein Sci.* **13**, 125 (2004)
- [29] K.-H. Lin, J. C. Crocker, V. Prasad, A. Schofield, D. A. Weitz, T. C. Lubensky, and A. G. Yodh, *Phys. Rev. Lett.* **85**, 1770 (2000)
- [30] A. W. C. Lau, K.-H. Lin, and A. G. Yodh, *Phys. Rev. E.* **66**, 020401 (2002).
- [31] K.A. Dill, *Protein Sci.* **8**, 1166 (1999).
- [32] K. A. Dill, and H. S. Chan, *Nat. Struct. Biol.* **4**, 10 (1997).
- [33] C.-C. Chang, P.-Y. Lin, X.-C. Yeh, K.-H. Deng, Y. P. Ho, and L. S. Kan, *Biochem. Biophys. Res. Commun.* **328**, 845 (2005).
- [34] L. E. Reichl, *A modern Course in Statistical Physics*. Wiley-Interscience publication. New York (1998).
- [35] L. E. Reichl, *A modern Course in Statistical Physics*. Wiley-Interscience publication. New York (1998).
- [36] J. Frydman, *Annu. Rev. Biochem.* **70**, 603 (2001).
- [37] D. Thirumalai, and G. H. Lorimer, *Annu. Rev. Biophys. Biomol. Struct.* **30**, 245 (2001).