Review

The Molten Globule, and Two-State *vs.* Non-Two-State Folding of Globular Proteins

Kunihiro Kuwajima,*

Department of Physics, School of Science, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan, and School of Computational Sciences, Korea Institute for Advanced Study (KIAS), Seoul 02455, Korea ; kuwajima@ims.ac.jp

* Correspondence: kuwajima@ims.ac.jp

Abstract: From experimental studies of protein folding, it is now clear that there are two types of folding behavior, i.e., two-state folding and non-two-state folding, and understanding the relationships between these apparently different folding behaviors is essential for fully elucidating the molecular mechanisms of protein folding. This article describes how the presence of the two types of folding behavior has been confirmed experimentally, and discusses the relationships between the two-state and the non-two-state folding reactions, on the basis of available data on the correlations of the folding rate constant with various structure-based properties, which are determined primarily by the backbone topology of proteins. Finally, a two-stage hierarchical model is proposed as a general mechanism of protein folding. In this model, protein folding occurs in a hierarchical manner, reflecting the hierarchy of the native three-dimensional structure, as embodied in the case of non-two-state folding with an accumulation of the molten globule state as a folding intermediate. The two-state folding is thus merely a simplified version of the hierarchical folding caused either by an alteration in the rate-limiting step of folding or by destabilization of the intermediate.

Keywords: protein folding; molten globule state; two-state proteins; non-two-state proteins

1. Introduction

Elucidation of the molecular mechanisms of protein folding is a fundamental problem in biophysics and molecular biology, although almost 60 years have elapsed since Anfinsen and his coworkers discovered the genetic control of the native tertiary structure of proteins [1]. In the interim, there have been many experimental studies on protein folding (refs. [2, 3] and references cited therein). From these studies, it is now clear that there are two types of folding behavior, i.e., non-two-state folding, involving at least one folding intermediate, and two-state folding without any detectable intermediate during kinetic refolding from the fully unfolded (U) state to the native (N) state [4, 5]. Understanding the relationships between the two-state and non-two-state folding may be essential in order to fully elucidate the molecular mechanisms of protein folding. This article will discuss these relationships, on the basis of available data on the correlations of the rate constant of folding with various structure-based properties, which are determined primarily by the backbone structure (topology) of proteins (see section 4 below).

Traditionally, the presence of a specific pathway of folding was thought to be a unique solution of Levinthal's paradox [6], and hence, detection and characterization of specific intermediates, present along the pathway of folding, were basic approaches for elucidating the mechanisms of folding. The molten globule (MG) state was proposed as such a specific intermediate of folding [7],

and experimentally the MG state was observed as an equilibrium unfolding intermediate in certain proteins, and also as a transient folding intermediate, which accumulates at an early stage of kinetic refolding [8, 9]. The transient MG state was observed not only in the proteins that show the equilibrium MG state but also in the proteins that show the two-state unfolding transition at equilibrium. Oleg B. Ptitsyn, I and our colleagues thus conducted a cooperative research project supported by the Human Frontier Science Program (HFSP) for three years from 1993. The title of the project was "Role of the Molten Globule State in the Folding of Globular Protein." In the next section (section 2) of this article, the MG state in protein folding is thus described.

The two-state folding behavior, in which both equilibria and kinetics of folding/unfolding occur only between N and U, was discovered for a number of small globular proteins, usually with less than 100 amino acid residues [10, 11], and this discovery changed the focus of many researchers in regard to the two-state proteins. These initial efforts suggested that the two-state folding, which required no stable intermediates for complete and successful folding to the N state, was simpler than the non-two-state folding. Theoretical and computational studies of protein folding have proposed an energy landscape theory (a funnel model) of folding, and the two-state folding behavior has been interpreted reasonably well by this model [12]. In section 3, the two-state *vs*. non-two-state folding of globular proteins will be described. This section mainly deals with the significance of a burst-phase intermediate, which is accumulated within the dead time of kinetic folding experiments, because the burst-phase intermediate has often been observed in non-two-state proteins [9]. Sub-millisecond continuous-flow techniques [13] and more recently in-gel folding techniques [14] have revealed that the burst-phase intermediate is a real folding intermediate separated from the U state by a free-energy barrier, so that there are two-types of globular proteins in terms of the folding behavior, i.e., two-state proteins and non-two-state proteins.

In section 4, the relationships between the two-state and non-two-state folding reactions will be discussed. The logarithmic rate constant of folding, $\ln(k_f)$, shows significant correlations with structure-based properties, which are determined by the backbone structure. Recently, various structure-based properties that correlate with $\ln(k_f)$ were reported, and the correlations of $\ln(k_f)$ with the structure-based properties were found to be quite similar between two-state and non-two-state proteins, indicating that the physical mechanisms behind two-state and non-two-state folding are essentially identical (see section 4). From these results, it is concluded that the general mechanism of protein folding can be described using a two-stage hierarchical model, with the two-state folding being merely a simplified version of the hierarchical folding.

2. Role of the Molten Globule State in Protein Folding

The molten globule (MG) state is an intermediate between the N and the U states of globular proteins. It has the following characteristics: (1) the presence of a substantial amount of native-like secondary structure; (2) the absence of most of the specific tertiary structure associated with the tight packing of side chains; (3) the presence of a native-like backbone fold (topology), and the compactness of the overall shape of the molecule, with a radius only 10–30% larger than that in the N state; and (4) the presence of a loosely organized hydrophobic core, and the heterogeneity of the three-dimensional structure, in which certain subdomains of the molecule are more organized than others [8, 9]. For more recent and comprehensive information on the MG state, see [15]. The loosely organized hydrophobic core is often characterized by binding experiments using a hydrophobic fluorescent probe, 1-anilino-8-naphthalene sulfonate (ANS) [16]. The equilibrium MG state has been observed for a number of globular proteins under mildly denaturing conditions (e.g., at acidic pH or in the presence of an intermediate concentration of a strong denaturant such as guanidinium chloride (GdmCl) or urea). The role of the MG state in kinetic folding of globular proteins was an intriguing issue when our project started [17], and Oleg proposed that the MG state is a general intermediate in protein folding [7]. In fact, transient kinetic intermediates were detected and characterized for a number of globular proteins, including α-lactalbumin [18], lysozyme [18, 19], cytochrome *c* [20], ribonuclease A [21] and apo-myoglobin [22], by a kinetic circular dichroism (CD) technique and a

pulsed hydrogen/deuterium (H/D)-exchange method combined with two dimensional (2D) NMR spectroscopy. Close similarity between the equilibrium MG state and the kinetic folding intermediate thus characterized was demonstrated for certain proteins by coincidence of the equilibrium unfolding transition curve of the MG state and the pre-equilibrium unfolding transition curve of the kinetic intermediate [23] and by close similarity of the H/D-exchange protection profile between the equilibrium and kinetic intermediates [22].
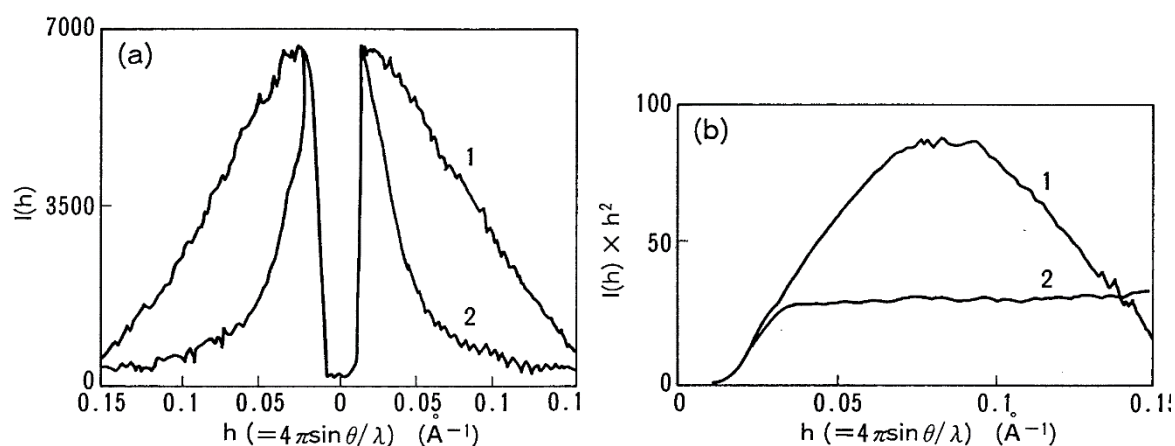

Figure 1. Synchrotron SAXS patterns (a) and Kratky plots (b) for bovine carbonic anhydrase (1) in the N state (0.05 M Tris-HCl (pH 8)) and (2) in the U state (0.05 M Tris-HCl (pH 8), 8.5 M urea). Reproduced with permission from [24].

To further characterize the equilibrium and kinetic MG states, Ptitsyn's group and the Japanese group carried out joint experiments, and Gennady V. Semisotnov (Institute of Protein Research, Russia) often visited Japan to participate in these efforts. Hiroshi Kihara (Kansai Medical Univ.), Yoshiyuki Amemiya (Univ. Tokyo), Kazumoto Kimura (Dokkyo Univ.), and students of our laboratories at that time were also involved in these experiments. We utilized a synchrotron radiation facility at the High Energy Accelerator Research Organization, Tsukuba, Japan to characterize the MG state of proteins by a small angle X-ray scattering (SAXS) technique [24] (Fig. 1). In particular, we were interested in a time-resolved SAXS study. If the early kinetic folding intermediate is identical to the MG state, it would be possible to directly observe a compact shape of the protein molecule, an important characteristic of the MG state, at an early stage of kinetic folding. We investigated the kinetic refolding reactions of β-lactoglobulin and α-lactalbumin induced by a denaturant (urea or GdmCl) concentration jump using a stopped-flow SAXS apparatus [25, 26]. Both the proteins formed a compact globular structure with a radius of gyration (R_g) identical to that of the MG state, although the intermediate of β-lactoglobulin exhibited a non-native α-helical structure [25, 27].

Figure 2 shows time-dependent changes in the R_g value during refolding of α-lactalbumin and the Kratky plots of the protein in the early stages of refolding [26]. Changes in the R_g^2 values were fitted to a single-exponential function, and the rate constant thus obtained was 0.49 ($\pm$0.07) s^{-1}, which is in agreement with the rate measured by other spectroscopic techniques. The R_g value after complete refolding was 15.5 ($\pm$0.1) Å, and the R_g value at zero time of refolding obtained by extrapolation of the fitting curve to zero time was 17.5 ($\pm$0.2) Å. Because this protein forms the transient folding intermediate within the dead time (~7 ms) of the stopped-flow measurement, the folding intermediate has an R_g of 17.5 ($\pm$0.2) Å, which is in good agreement with the known R_g value of the equilibrium MG state [26]. The Kratky plot obtained by averaging the scattering pattern within 10–20 ms of refolding is coincident with that of the equilibrium MG state, indicating that both the size and shape of the transient folding intermediate were identical to those of the equilibrium MG state.
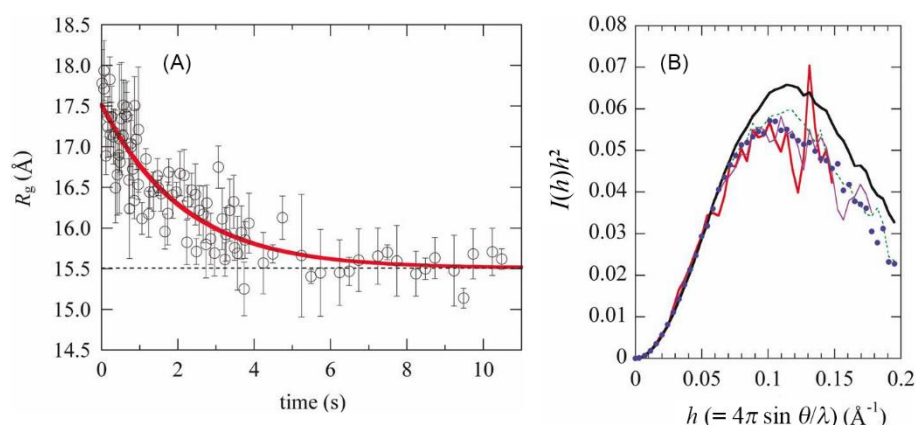
Figure 2. Time-dependent changes in the R_g during the refolding of α-lactalbumin (A), and the Kratky plots of the protein during the refolding (B). The refolding was induced by a concentration jump of GdmCl from 4 M to 0.77 M in the presence of 1 mM CaCl$_2$ (pH 7.0 and 4.5°C). (A) The time-dependent changes in R_g^2 were fitted to a single-exponential function, and the fitting curve of R_g thus obtained is shown as a thick red line. The dotted line shows the R_g value after the refolding is completed, and errors are the standard errors in fitting of Guinier plots. (B) The Kratky plots averaged between 10 and 20 ms after the initiation of refolding (red thick line), between 10 and 110 ms (purple line), between 10 and 960 ms (green dotted line), and between 8 and 11 s (black thick line). Blue filled circles show the Kratky plot of the equilibrium MG state at pH 2. The scattering intensities of each plot are normalized with the respective zero-angle intensities. Reproduced with permission from [26].

In the SAXS experiments of Fig. 2, we used a 2D charge-coupled device (CCD)-based X-ray detector with a beryllium-windowed X-ray image intensifier, by which the S/N ratio of the scattering data was dramatically improved as compared with the data obtained by a one-dimensional detector [26]. This 2D CCD-based detector was developed by Amemiya and coworkers [28], and detailed procedures for data correction and the analysis software were described by Ito *et al*. [29]. The 2D CCD-based detector was used for nearly a decade to characterize compact folding intermediates of other globular proteins [30-36], but at present, a more sensitive PILATUS pixel detector is widely used for the SAXS experiments [37-39].

3. Two-state *vs*. Non-two-state Folding of Globular Proteins

In 1991, just two years before the start of the above-mentioned project supported by HFSP, Jackson and Fersht [10] in Cambridge reported that barley chymotrypsin inhibitor 2 (CI2), a small monomeric protein of 83 residues (the first 18 residues are unstructured), appears to be a rare example in which both equilibria and kinetics are described by a two-state model. As far as I know, this was the first report on observation of the two-state folding of globular proteins. The criteria for the two-state folding are (i) observation of a single equilibrium unfolding transition as a function of an external parameter such as denaturant (GdmCl or urea) concentration (*c*) or temperature; (ii) linear free-energy relationships of the equilibrium unfolding free energy and the activation free energies for kinetic folding and unfolding with respect to *c*; (iii) single-exponential kinetics for both folding and unfolding with the observation of the full change of a structural probe, fluorescence, CD ellipticity or any other index used to monitor the conformational transition, between U and N (allowing for exceptions of complex refolding kinetics due to proline isomerization in the U state); (iv) consistency between the equilibrium and the kinetic parameters of folding/unfolding; and (v) agreement between the van't Hoff enthalpy and the calorimetric enthalpy of equilibrium thermal unfolding. Any deviations from the above criteria suggest the presence of an intermediate between N and U [40].

Although Jackson and Fersht [10] reported that CI2 was a rare example, later many small proteins, usually with less than 100 amino-acid residues, were shown to fold with a similar simple two-state mechanism [11]. My colleagues in KIAS and I recently constructed a standardized protein folding database (PFDB) with temperature correction, in which the ln(k_f) values of all listed proteins were

calculated at the standard temperature (25°C) [41]. Among the total 141 proteins registered in PFDB, 89 are two-state proteins, and hence two-state proteins are no longer rare examples. However, this does not necessarily mean that a majority of naturally occurring proteins are of the two-state type. Because two-state proteins are apparently simpler than non-two-state proteins, researchers in the protein folding field tend to investigate two-state proteins, which would account for the large number of two-state proteins registered in PFDB. Among the 89 two-state proteins registered in PFDB, 53 are composed of an isolated domain of a multi-domain protein. The average chain length of the two-state proteins registered in PFDB is 92.6, while the average chain length of the non-two-state proteins is 139.4 [41].

The folding intermediate appears not to be required for the complete folding of two-state proteins, which has raised questions as to the significance of the folding intermediates previously detected and characterized in non-two-state proteins [42, 43]. The formation of an early kinetic intermediate often occurs too quickly to measure directly by a conventional rapid reaction technique such as the stopped-flow method, and the process occurs in a burst phase within the dead time of the measurement [9, 44, 45]. This situation raises a question in regard to the burst-phase—namely, whether an observed burst-phase signal reflects a real folding event. To answer this question, the burst-phase fluorescence or CD signals at an early stage of kinetic refolding of cytochrome *c* and ribonuclease A were thus investigated by Sosnick *et al*. [46-49]. The refolding reactions were induced by stopped-flow concentration jumps of GdmCl, and the burst-phase signals thus obtained were compared with the signals of non-folding polypeptides: fragments 1–80 and 1–65 for cytochrome *c* and a disulfide broken form for ribonuclease A. Rather surprisingly, the signals of the intact proteins and the non-folding polypeptides were coincident with each other. It was thus concluded that the sub-millisecond burst-phase signals may not reflect the fast formation of the folding intermediate, but rather may reflect some solvent-dependent contraction/collapse of the U state, since the water solution at a low concentration of GdmCl is a poor solvent for the polypeptides [46-48], and the initial barrier mechanism in protein folding was proposed [49].

Roder's group [50, 51] and Takahashi's group [33, 52] thus investigated the sub-millisecond processes of folding reactions of cytochrome *c* and ribonuclease A by continuous-flow techniques, which have a higher time resolution than the conventional stopped-flow method. Shastry and Roder [50] found that the kinetics of the cytochrome *c* folding measured by tryptophan (Trp 59) fluorescence exhibits a major exponential process with a time constant of ~50 μs, independent of initial conditions and heme ligation state, indicating that a common free energy barrier is encountered during the formation of an early compact intermediate. Therefore, there are at least two distinct stages in the cytochrome *c* folding, with the compact folding intermediate accumulating in the first stage, followed by the rate-limiting formation of the N state. Interestingly, Qui *et al*. [53] found by the laser temperature-jump method that not only full-length cytochrome *c* but also non-folding peptides 1–65 and 1–80 displayed similar exponential decays, indicating that the free-energy barrier to collapse was similar in the folding and non-folding sequences. Because the non-folding peptides contain significant portions of the natural cytochrome *c* sequence, it might be possible that such sequences can form a similar compact structure [54]. For ribonuclease A, Welker *et al*. [51] reported that the very fast folding U state (U_{vf}), which contains only native proline isomers, exhibits an exponential decay in fluorescence with a time constant of ~80 μs, whereas the equilibrium U state (U_{eq}), which contains nonnative proline isomers, did not show this exponential decay, indicating that the isomerization state of proline peptide bonds can have a major impact on the early stages of the kinetic folding. It may be noted that the burst phase previously observed in ribonuclease A by Houry *et al*. [45] using a stopped-flow double-jump apparatus was for U_{vf} and not for U_{eq}. Kimura *et al*. [33] have shown that the sub-millisecond folding intermediate of intact ribonuclease A strongly binds to ANS and has a compact structure, characteristics of the MG state, but that disulfide-reduced ribonuclease A shows neither any significant binding to ANS nor such compaction of the structure, indicating the difference in the sub-millisecond conformations of ribonuclease A and its reduced form.
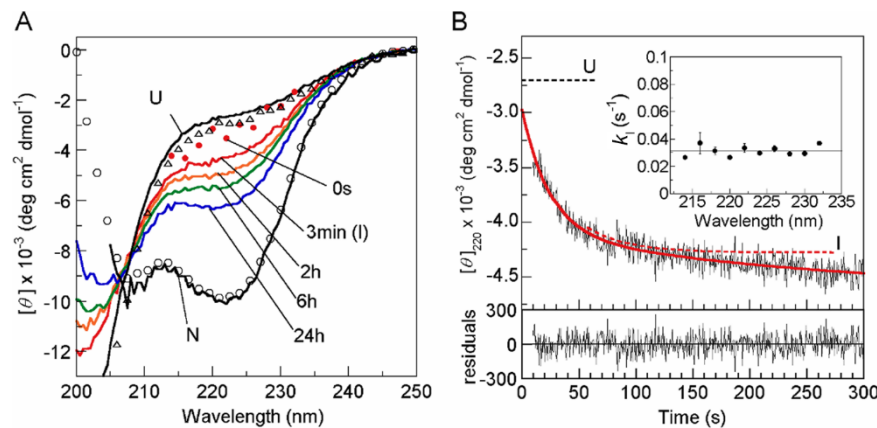
Figure 3. Representative far-UV CD spectra recorded during the folding of cytochrome *c* at 25°C and pH 4.5 in a silica gel containing 80% (w/w) water. The times indicated are those after initiation of the folding reaction. The red circles correspond to the mean residue ellipticity [θ] obtained by the extrapolation of kinetic curves to time zero. The solid black lines represent the spectra of the N and U states in gel. The open circles and triangles are the [θ] values of the spectra of the N and U states, respectively, in solution. Note that the spectra of the solution N and U states were recorded using the same folding and unfolding conditions as were used for the gel measurements. (B) A representative kinetic curve plotted using the [θ] at 220 nm for the initial resolvable cytochrome *c* in-gel folding phase at 25°C. The times are those after initiation of the folding reaction. The thick red line is the fitted curve drawn using a multi-exponential decay function. The dotted red curve is the single exponential curve. The dotted black line in the upper panel is the [θ] at 220 nm for the U state. The residuals are shown in the lower panel. The inset shows the rate constants of the initial phase calculated for different wavelengths. Reproduced with permission from [14].

Recently, Okabe *et al.* [14] employed an elegant technique to delineate the solution burst-phase events of folding by encapsulating the proteins in silica gels. Within a porous silica gel with a large water content, large scale protein motions are dramatically slowed, but the in-gel N state can retain its solution properties [55-58]. In particular, the folding of a protein in such a gel is dramatically slowed, and this makes it possible to directly observe the entire folding process without substantially perturbing the folding pathway [59-62]. Figure 3 shows (A) representative far-UV CD spectra of cytochrome *c* during the refolding and (B) a typical kinetic refolding curve measured by the CD ellipticity at 220 nm for the initial resolvable phase of the in-gel refolding; the protein that was initially acid unfolded at pH 1.8 was refolded at pH 4.5 and 25°C in a silica gel containing 80% (w/w) water [14]. The CD spectra in the N and U states in the gel are nearly identical to the corresponding spectra in solution, indicating that the conformations in the N and U states are not influenced by the gel matrix (Fig. 3(A)). As shown in Fig. 3 (B), the initial phase of the in-gel folding exhibits an exponential decay with a time constant of ~30 s. Because the fastest phase of the protein in solution at pH 4.5 and 22°C has a time constant of 57 μs [50], the refolding rate was at least 5.3×10^5 times reduced in the in-gel folding compared with the refolding in solution. The CD spectrum of the early kinetic intermediate observed at 3 min, 6 times the time constant (~30 s), in the in-gel folding (Fig. 3(A)) is very similar to the CD spectrum at 400 μs, 7 times the time constant (57 μs), in the free solution refolding reported by Akiyama *et al.* [52], suggesting that the same transient folding intermediate accumulates at the initial stage in the in-gel and the free solution refolding reactions. It may also be noted that the acid compact state of cytochrome *c* stabilized by KCl, known also as the MG state [63, 64], is an equilibrium counterpart of a late folding intermediate separated from U by the rate-limiting free-energy barrier [65], and hence it is different from the intermediate shown in Fig. 3 and the MG state in this article

Figure 4 shows kinetic refolding curves measured by the CD ellipticity at 220 nm for the initial resolvable phase of the in-gel refolding reactions of four proteins, equine β-lactoglobulin, human tear lipocalin, bovine α-lactalbumin, and hen-egg white lysozyme [14]. The proteins were initially unfolded in the silica gel immersed into 6 M GdmCl, and the refolding was initiated in the gel by a GdmCl concentration jump produced by washing the protein-embedded gel with a refolding buffer solution.

From Fig. 4, all four proteins exhibit an exponential decay and accumulate a transient folding intermediate. The rate constants measured at different wavelengths were coincident with each other (insets of Fig. 4), and the CD spectrum of the transient intermediate for each protein was obtained from the kinetic refolding curves measured at different wavelengths. From comparisons of the transient CD spectrum thus obtained with the CD spectrum of the burst-phase intermediate in free refolding in solution and also with the CD spectrum of the equilibrium MG state, if available, for each protein, it is concluded that the folding intermediate formed at the fastest initial stage in the in-gel refolding is equivalent to the intermediate formed in the burst phase in the free refolding in solution. Figure 4 also shows that the CD ellipticity value for each protein obtained by extrapolation of the exponential decay curve to zero time does not coincide with the value in the U state, indicating the presence of a burst phase before the exponential decay. However, the U state values shown in Fig. 4 are those at 6 M GdmCl, and when we assume a linear dependence of the U state values on GdmCl concentration, the zero-time ellipticity becomes much closer to the values in the U state under the refolding condition. The burst-phase signals in Fig. 4 may represent the rapid collapse of the protein molecules caused by changing the solvent environment from a good solvent (6 M GdmCl) to a poor solvent (~0 M GdmCl).
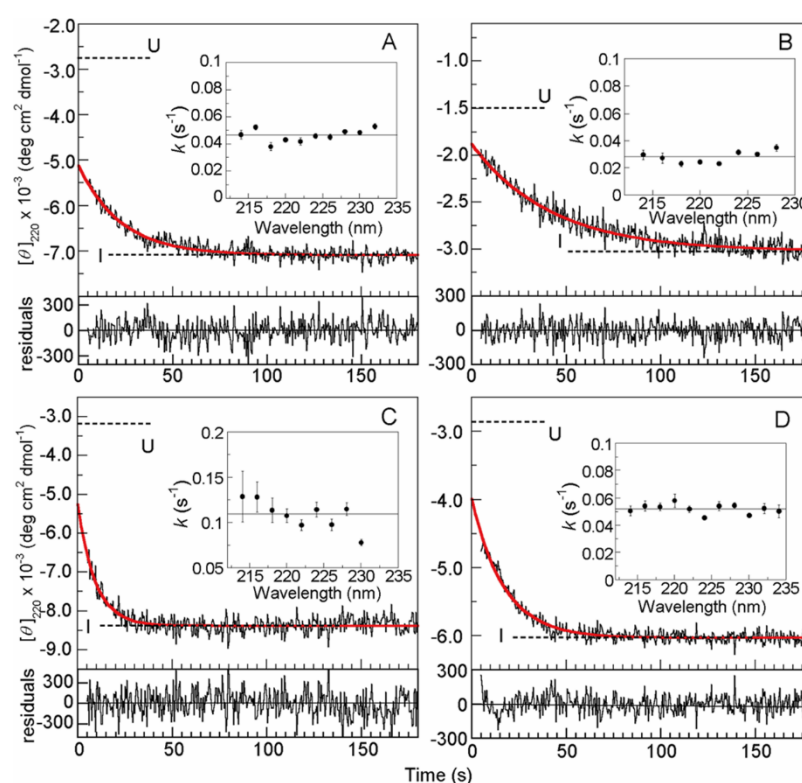

Figure 4. Representative plots of the $[\theta]$ at 220 nm during the U $\rightarrow$ I transitions at 25 °C for (A) equine β-lactoglobulin, (B) human tear lipocalin, (C) bovine α-lactalbumin, and (D) hen egg-white lysozyme in the gels. The times are those after initiation of the folding reaction. The thick red lines are each a fit to a single exponential. The dotted lines are the $[\theta]$ values for the U and I states at 220 nm. The residuals are shown under each kinetic plot. The insets show the rate constants for the U $\rightarrow$ I phase calculated for each wavelength, with the horizontal lines being the average values. Reproduced with permission from [14].

4. The Relationships between the Two-state and Non-two-state Folding Reactions

From the above results, it now appears that there are two types of globular proteins in terms of the folding behavior, i.e., two-state proteins and non-two-state proteins. The two-state folding reactions under native conditions are simply represented by equation (1), where the backward reaction is neglected because it is negligible as compared with the forward reaction:

$$U \xrightarrow{k_f} N \tag{1}$$

On the other hand, the non-two-state proteins often accumulate the MG state at an early stage in kinetic refolding from the U state [9]. The simplest reaction scheme of the non-two-state folding is thus given by:

$$U \xrightarrow{k_I} I \xrightarrow{k_f} N \tag{2}$$

where I is the folding intermediate that has the characteristics of the MG state, k_I is the rate constant of the formation of I, and the backward reactions are again neglected because they are sufficiently small under native folding conditions. Understanding the relationship between the two types of proteins may be important for fully elucidating the molecular mechanisms of protein folding.

For two-state proteins, significant correlations between $\ln(k_f)$ and various structure-based properties have been well documented. The structure-based properties include the relative contact order (RCO) [66-68], the long-range order (LRO) [69, 70], the number of sequence-distant native pairs (Q_d) [71, 72], the chain topology parameter [73], the total contact distance [74], the cliquishness [75] and the relative logarithmic contact order [76]. All these structure-based properties are determined by the number of contacts between two residues in close proximity, with their C_α–C_α distance less than 6–8 Å, and by the separation lengths between the contacting residues along the primary sequence. These structure-based properties thus depend on the backbone structure (topology) of a protein, with a more complex backbone topology corresponding to a slower the folding rate. Originally, it was proposed that the correlations of $\ln(k_f)$ with the above structure-based properties (except cliquishness) are present only for two-state proteins, and for non-two-state proteins, whose folding reactions may be complicated by the rate of escape from the folding intermediates trapped on a ragged energy landscape of folding, such simple correlations of $\ln(k_f)$ with the structure-based properties may be absent [72].

However, later it became clear that non-two-state proteins also show similar correlations with the structure-based properties. The $\ln(k_f)$ values of non-two-state proteins show a significant correlation with the chain length (L), i.e., the number of amino acid residues, and also with the absolute contact order (ACO), which is given by ACO = RCO × L [77-79]. Similar correlations of $\ln(k_f)$ with ACO were observed for two-state and non-two-state proteins, but the correlation with L was worse for two-state proteins than for non-two-state proteins [77, 79]; see also [80]. Among the above-mentioned structure-based properties for two-state proteins, three properties, LRO, Q_d and the cliquishness, were also reported to exhibit significant correlations for non-two-state proteins [75, 79, 81]; LRO and Q_d are closely related as LRO ≈ Q_d/L. More sophisticated structure-based properties, well correlating with $\ln(k_f)$ for both two-state and non-two-state proteins, have also been proposed [82-90]. These structure-based properties include the n-order contact distance [83], the geometric contact number, which is the number of nonlocal contacts well packed by a Voronoi criterion [84], the inter-residue interaction parameter, which considers the distances between all residue pairs in a protein [87], the average topological information, which is derived from probability densities of all residue pairs based on a self-avoiding random walk model [89], the entanglement of the native backbone structure [88], the logarithmic ACO, computed on shortcut networks of the native structure [90], the prediction by an algorithm for the PREdiction of Folding and Unfolding Rates (PREFUR) using L and the structural class as only protein-specific input [86], and the effective chain length, evaluated from the number of helical residues and the number of helices [82] or from the amino acid composition of a protein [85]. More strict physical models to evaluate $\ln(k_f)$ for two-state and non-two-state proteins from the free-energy profiles obtained by simple statistical thermodynamic calculations were also reported [91-94]. Garbusynskiy *et al.* [95] have shown that physical theory, which describes scaling relationships between L and the activation free energy of folding, together with biological constraints outlines a "golden triangle" limiting the possible range of $\ln(k_f)$ for single-domain globular proteins. The $\ln(k_f)$ values of most of the globular proteins, including both two-state and non-two-state folders, are distributed within this narrow triangle.
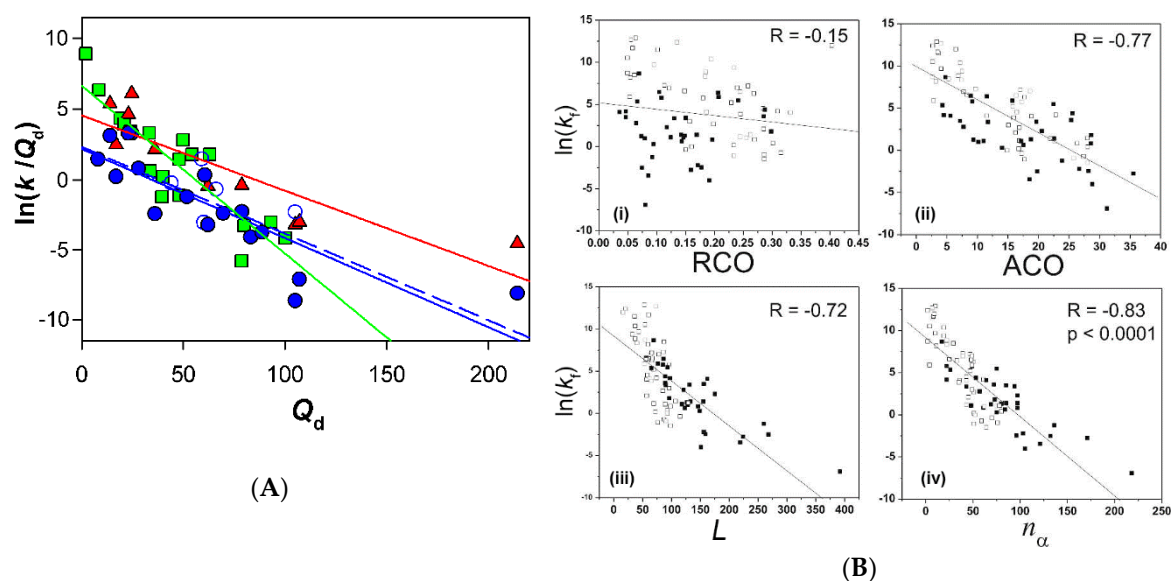
(A)

(B)

Figure 5. A comparison of the folding rates of two-state and non-two-state proteins. (A) The $\ln(k_f/Q_d)$ and $\ln(k_I/Q_d)$ values are plotted against Q_d. Filled green squares (■) represent the $\ln(k_f/Q_d)$ values of two-state proteins. Filled red triangles (▲) represent the $\ln(k_I/Q_d)$ values of 10 non-two-state proteins for which the formation of the intermediate was kinetically observed. Filled circles (●) and open circles (○) represent the $\ln(k_f/Q_d)$ values for the 17 non-two-state proteins, i.e., the above 10 proteins plus the 7 proteins that show a burst-phase intermediate, and the 22 non-two-state proteins that additionally include the 5 proteins that show only a roll-over in the chevron plot, respectively. The green, red, and blue continuous lines represent the best linear fit for $\ln(k_f/Q_d)$ for the 18 two-state proteins, $\ln(k_I/Q_d)$ for the 10 non-two-state proteins, and $\ln(k_f/Q_d)$ for the 17 non-two-state proteins, respectively. The broken blue line represents the best linear fit for $\ln(k_f/Q_d)$ for the 22 non-two-state proteins. Reproduced with permission from [79]. (B) Relationship between different structure-based properties and $\ln(k_f)$ of two-state (open squares) and non-two-state (filled squares) proteins. (i) Relative contact order, RCO ($R = -0.15$); (ii) absolute contact order, ACO ($R = -0.77$); (iii) chain length, L ($R = -0.72$); and (iv) geometric contact number, n_α ($R = -0.83$), where R is the correlation coefficient. Reproduced with permission from [84].

Figure 5(A) shows the dependence of the logarithmic rate constant of folding on Q_d for two-state and non-two-state proteins, taken from Kamagata *et al.* [79]. The Q_d in Fig. 5(A) is defined as:

$$Q_d = \sum_{i=2}^{L} \sum_{j=1}^{i-1} \Delta_{ij} \tag{3}$$

where i and j are the residue numbers of two residues for which the C_α–C_α distance in space is within 6 Å in the PDB structure, and $\Delta_{ij} = 1$ if $i - j > 12$, and otherwise $\Delta_{ij} = 0$. Q_d thus represents the number of sequence-distant native pairs. In Fig. 5(A), the values of $\ln(k_f/Q_d)$ are shown for the two-state and the non-two-state proteins, and for the 10 non-two-state proteins whose k_I values were determined experimentally, the values of $\ln(k_I/Q_d)$ are also presented. From Fig. 5(A), both the value of $\ln(k_f/Q_d)$ and its dependence on Q_d in the case of the two-state proteins are very similar to those found in the plots of $\ln(k_I/Q_d)$ and $\ln(k_f/Q_d)$ against Q_d for the non-two-state proteins. Such similarities between two-state and non-two-state proteins were also found rather generally in the relationships between $\ln(k_f)$ and structure-based properties (see Fig. 5(B)), which include ACO [77-79], LRO [81], the cliquishness [75], the *n*-order contact distance [83], the geometric contact number [84], the inter-residue interaction parameter [87], the average topological information [89], the entanglement of the native backbone structure [88], and the other structure-based properties [82, 85, 86, 90]. As to the golden triangle for scaling of protein folding rates, proposed by Garbuzynskiy *et al.* [95], we also find similar distributions of $\ln(k_f)$ between the two-state and the non-two-state proteins. These similarities thus clearly demonstrate that the physical mechanisms of folding are not essentially different between the two types of proteins.

If the mechanisms behind two-state and non-two-state folding are essentially identical, what accounts for the differences between the two-state and the non-two-state proteins? Figure 5(A), which includes the data for both k_f and k_I for non-two-state proteins, may provide a reasonable answer to this question. From Fig. 5(A), the slope of the plot of $\ln(k_f/Q_d)$ against Q_d is steeper (i.e., more negative) for the two-state proteins than for the non-two-state proteins. This trend can also be found in the plots of $\ln(k_f)$ against the other structure-based properties, which usually represent the backbone topological complexity of a protein (Fig. 5(B)) [83, 84, 87]. The steeper dependence of the plots against Q_d and the other structure-based properties for the two-state proteins may be interpreted in terms of a shift in the rate-limiting step, caused by a change in the value of each structure-based property. From Fig. 5(A), the k_f values of the two-state proteins are larger than those of the non-two-state proteins at Q_d values less than 25, but rather coincide with the k_I values of the non-two-state proteins. This apparently suggests that the $U \rightarrow I$ process of equation (2) has become rate-limiting in these two-state proteins. At Q_d values between 25 and 70, the two-state k_f shows variations in either the non-two-state k_I or k_f, and at Q_d values larger than 70, the two-state k_f shows a coincidence with the non-two-state k_f. It is thus suggested that the $I \rightarrow N$ process of equation (2) is now rate-limiting in the two-state proteins with a Q_d larger than 70.
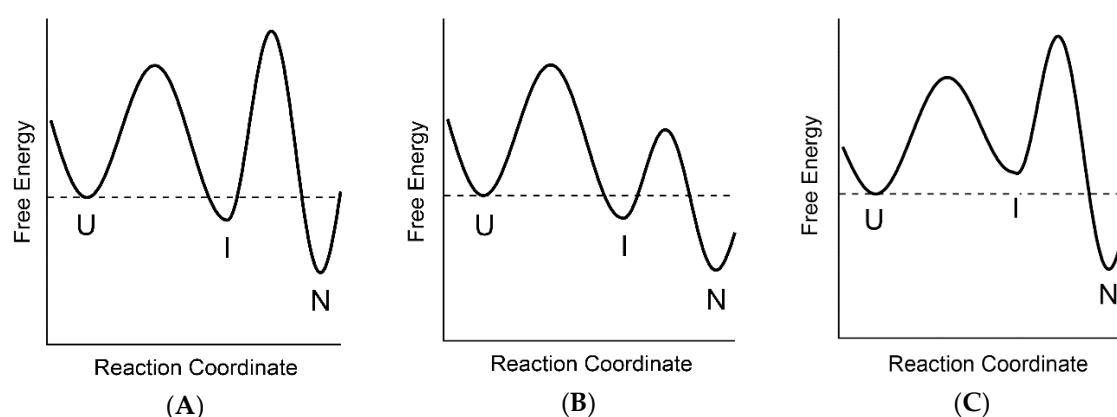

Figure 6. The free-energy profiles of the non-two-state and the two-state folding reactions. (A) Hierarchical three-state folding, (B) two-state folding in the first scenario, and (C) two-state folding in the second scenario. The horizontal broken line in each panel shows the free-energy level of U.

From the above results, a two-stage hierarchical model is proposed as a general mechanism of protein folding. In this model, protein folding occurs in a hierarchical manner, reflecting the hierarchy of the native three-dimensional structure, as embodied in the case of non-two-state folding with an accumulation of the MG state as a folding intermediate. In this model, the protein folding process is divided into two stages: the first stage is formation of the MG state ($U \rightarrow I$ in equation (2)); and the second stage is formation of the N state from the MG state ($I \rightarrow N$). The two-state folding is thus merely a simplified version of the hierarchical folding caused either by an alteration in the rate-limiting step of folding or by destabilization of the intermediate. Therefore, there are two scenarios for interpreting the two-state folding behavior. First, when the backbone topology is sufficiently simple and organized by short sequence-distant contacts, it is easy to determine the specific conformation of side-chain packing, because the number of possible specific interactions is so small that the formation of the backbone structure rather uniquely determine the side-chain packing interactions. In this case, the first stage of the hierarchical folding becomes rate-limiting, leading to the two-state kinetic behavior of the protein. This first scenario is thus consistent with the initial barrier mechanism proposed by Krantz *et al.* [49]. In the second scenario, the backbone structure of the protein is organized by long sequence-distant contacts, and hence there are too many ways to make the side-chain packing interactions quickly, keeping the second stage of the hierarchical folding rate-limiting, but destabilization of the intermediate leads to the two-state kinetic behavior. The free-energy profiles of the non-two-state and the two-state folding reactions are shown in Fig. 6; panels (A), (B), and (C) represent the hierarchical three-state folding, the two-state folding in the first scenario, and the two-state folding in the second scenario,

respectively. The existence of an unstable high-energy intermediate was expected experimentally for certain two-state proteins from the unfolding-limb or the refolding-limb curvature of the chevron plot [96], and this is consistent with the above-described two-state behaviors in the second scenario.

5. Conclusions

1. In classic experimental studies on protein folding, detection and characterization of kinetic folding intermediates were basic approaches to elucidate the molecular mechanisms of folding. The MG state was found and characterized as such a folding intermediate. The MG state in protein folding was thus briefly described. The characterization of the folding intermediate by time-resolved SAXS techniques, which was originally started by a collaboration with Ptitsyn's group, led us to conclude that the intermediate thus characterized is equivalent to the MG state.

2. Due to the observation of the two-state folding behavior for a number of small globular proteins, the two-state folding engendered much interest among researchers in the field of protein folding. Apparently, the folding intermediate is not required for complete folding of two-state proteins, and this raised questions as to the significance of the folding intermediates, such as the previously observed and characterized MG state. This was especially the case for the burst-phase intermediate that accumulates within the dead time of conventional rapid reaction techniques such as the stopped-flow method.

3. The use of the continuous-flow techniques and more recently the in-gel folding techniques made it possible to directly observe the previously unobservable solution burst-phase process of folding. The solution burst-phase intermediate was shown to be a real folding intermediate separated from the U state by a free-energy barrier. Therefore, it is now clear that there are two types of globular proteins in terms of the folding behavior, i.e., two-state proteins and non-two-state proteins.

4. There are significant correlations between the logarithmic rate constant of folding, $\ln(k_f)$, and various structure-based properties such as RCO, LRO, Q_d and so on. Although it was originally proposed that the correlation of $\ln(k_f)$ with the structure-based properties exists only for two-state proteins, later it became clear that non-two-state proteins also show similar correlations with the structure-based properties, and hence the physical mechanisms behind two-state and non-two-state folding are essentially identical.

5. Finally, the two-stage hierarchical model was proposed as a general mechanism of protein folding. In this model, protein folding occurs in a hierarchical manner, reflecting the hierarchy of the native three-dimensional structure, as embodied in the case of non-two-state folding with an accumulation of the MG state as a folding intermediate. The two-state folding is thus merely a simplified version of the hierarchical folding caused either by an alteration in the rate-limiting step of folding or by destabilization of the intermediate.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Epstein, C. J., Goldberger, R. F. & Anfinsen, C. B. (1963) The genetic control of tertiary protein structure: Studies with model systems, *Cold Spring Harb Symp Quant Biol.* **28**, 439-449.

2. Sinha, K. K. & Udgaonkar, J. B. (2009) Early events in protein folding, *Curr Sci.* **96**, 1053-1070.

3. Bergasa-Caceres, F., Haas, E. & Rabitz, H. A. (2019) Nature's Shortcut to Protein Folding, *J Phys Chem B.* **123**, 4463-4476.

4. Barrick, D. (2009) What have we learned from the studies of two-state folders, and what are the unanswered questions about two-state protein folding?, *Phys Biol.* **6**, 9.

5. Tsytlonok, M. & Itzhaki, L. S. (2013) The how's and why's of protein folding intermediates, *Arch Biochem Biophys.* **531**, 14-23.

6. Levinthal, C. (1968) Are there pathways for protein folding?, *J Chim Phys.* **65**, 44-45.

7. Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E. & Razgulyaev, O. I. (1990) Evidence for a molten globule state as a general intermediate in protein folding, *FEBS Lett.* **262**, 20-24.

8. Ptitsyn, O. B. (1995) Molten globule and protein folding, *Adv Protein Chem.* **47**, 83-229.

9. Arai, M. & Kuwajima, K. (2000) Role of the molten globule state in protein folding, *Adv Protein Chem.* **53**, 209-282.

10. Jackson, S. E. & Fersht, A. R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition, *Biochemistry.* **30**, 10428-10435.

11. Jackson, S. E. (1998) How do small single-domain proteins fold?, *Fold Des.* **3**, R81-r91.

12. Onuchic, J. N., Nymeyer, H., García, A. E., Chahine, J. & Socci, N. D. (2000) The energy landscape theory of protein folding: insights into folding mechanisms and scenarios, *Adv Protein Chem.* **53**, 87-152.

13. Roder, H., Maki, K. & Cheng, H. (2006) Early events in protein folding explored by rapid mixing methods, *Chem Rev.* **106**, 1836-1861.

14. Okabe, T., Tsukamoto, S., Fujiwara, K., Shibayama, N. & Ikeguchi, M. (2014) Delineation of solution burst-phase protein folding events by encapsulating the proteins in silica gels, *Biochemistry.* **53**, 3858-3866.

15. Bychkova, V. E., Semisotnov, G. V., Balobanov, V. A. & Finkelstein, A. V. (2018) The Molten Globule Concept: 45 Years Later, *Biochem-Moscow.* **83**, S33-S47.

16. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F. & Gilmanshin, R. I. (1991) Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe, *Biopolymers.* **31**, 119-128.

17. Kuwajima, K. (1989) The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure, *Proteins.* **6**, 87-103.

18. Kuwajima, K., Hiraoka, Y., Ikeguchi, M. & Sugai, S. (1985) Comparison of the transient folding intermediates in lysozyme and alpha-lactalbumin, *Biochemistry.* **24**, 874-881.

19. Radford, S. E., Dobson, C. M. & Evans, P. A. (1992) The folding of hen lysozyme involves partially structured intermediates and multiple pathways, *Nature.* **358**, 302-307.

20. Roder, H., Elöve, G. A. & Englander, S. W. (1988) Structural characterization of folding intermediates in cytochrome *c* by H-exchange labeling and proton NMR, *Nature.* **335**, 700-704.

21. Udgaonkar, J. B. & Baldwin, R. L. (1988) NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A, *Nature.* **335**, 694-699.

22. Jennings, P. A. & Wright, P. E. (1993) Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin, *Science.* **262**, 892-896.

23. Ikeguchi, M., Kuwajima, K., Mitani, M. & Sugai, S. (1986) Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: A comparative-study of the folding reactions of α-lactalbumin and lysozyme, *Biochemistry.* **25**, 6965-6972.

24. Semisotnov, G. V., Kihara, H., Kotova, N. V., Kimura, K., Amemiya, Y., Wakabayashi, K., Serdyuk, I. N., Timchenko, A. A., Chiba, K., Nikaido, K., Ikura, T. & Kuwajima, K. (1996) Protein globularization during folding. A study by synchrotron small-angle X-ray scattering, *J Mol Biol.* **262**, 559-574.

25. Arai, M., Ikura, T., Semisotnov, G. V., Kihara, H., Amemiya, Y. & Kuwajima, K. (1998) Kinetic refolding of β-lactoglobulin. Studies by synchrotron X-ray scattering, and circular dichroism, absorption and fluorescence spectroscopy, *J Mol Biol.* **275**, 149-162.

26. Arai, M., Ito, K., Inobe, T., Nakao, M., Maki, K., Kamagata, K., Kihara, H., Amemiya, Y. & Kuwajima, K. (2002) Fast compaction of α-lactalbumin during folding studied by stopped-flow X-ray scattering, *J Mol Biol.* **321**, 121-132.

27. Kuwajima, K., Yamaya, H. & Sugai, S. (1996) The burst-phase intermediate in the refolding of β-lactoglobulin studied by stopped-flow circular dichroism and absorption spectroscopy, *J Mol Biol.* **264**, 806-822.

28. Amemiya, Y., Ito, K., Yagi, N., Asano, Y., Wakabayashi, K., Ueki, T. & Endo, T. (1995) Large-aperture TV detector with a beryllium-windowed image intensifier for X-ray diffraction, *Rev Sci Instrum.* **66**, 2290-2294.

29. Ito, K., Kamikubo, H., Yagi, N. & Amemiya, Y. (2005) Correction method and software for image distortion and nonuniform response in charge-coupled device-based X-ray detectors utilizing X-ray image intensifier, *Jpn J Appl Phys.* **44**, 8684-8691.

30. Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y. & Fujisawa, T. (2002) Conformational landscape of cytochrome *c* folding studied by microsecond-resolved small-angle x-ray scattering, *Proc Natl Acad Sci U S A.* **99**, 1329-1334.

31. Uzawa, T., Akiyama, S., Kimura, T., Takahashi, S., Ishimori, K., Morishima, I. & Fujisawa, T. (2004) Collapse and search dynamics of apomyoglobin folding revealed by submillisecond observations of α-helical content and compactness, *Proc Natl Acad Sci U S A.* **101**, 1171-1176.

32. Larios, E., Li, J. S., Schulten, K., Kihara, H. & Gruebele, M. (2004) Multiple probes reveal a native-like intermediate during low-temperature refolding of ubiquitin, *J Mol Biol.* **340**, 115-125.

33. Kimura, T., Akiyama, S., Uzawa, T., Ishimori, K., Morishima, I., Fujisawa, T. & Takahashi, S. (2005) Specifically collapsed intermediate in the early stage of the folding of ribonuclease A, *J Mol Biol.* **350**, 349-362.

34. Kimura, T., Uzawa, T., Ishimori, K., Morishima, I., Takahashi, S., Konno, T., Akiyama, S. & Fujisawa, T. (2005) Specific collapse followed by slow hydrogen-bond formation of β-sheet in the folding of single-chain monellin, *Proc Natl Acad Sci U S A.* **102**, 2748-2753.

35. Arai, M., Kondrashkina, E., Kayatekin, C., Matthews, C. R., Iwakura, M. & Bilsel, O. (2007) Microsecond hydrophobic collapse in the folding of *Escherichia coli* dihydrofolate reductase, an α/β-type protein, *J Mol Biol.* **368**, 219-229.

36. Konuma, T., Kimura, T., Matsumoto, S., Got, Y., Fujisawa, T., Fersht, A. R. & Takahashi, S. (2011) Time-resolved small-angle X-ray scattering study of the folding dynamics of barnase, *J Mol Biol.* **405**, 1284-1294.

37. Broennimann, C., Eikenberry, E. F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H. & Wagner, A. (2006) The PILATUS 1M detector, *J Synchrot Radiat.* **13**, 120-130.

38. Blanchet, C. E., Spilotros, A., Schwemmer, F., Graewert, M. A., Kikhney, A., Jeffries, C. M., Franke, D., Mark, D., Zengerle, R., Cipriani, F., Fiedler, S., Roessle, M. & Svergun, D. I. (2015) Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY), *J Appl Crystallogr.* **48**, 431-443.

39. Osaka, K., Matsumoto, T., Taniguchi, Y., Inoue, D., Sato, M. & Sano, N. (2016) High-Throughput and Automated SAXS/USAXS Experiment for Industrial Use at BL19B2 in SPring-8 in *Proceedings of the 12th*

International Conference on Synchrotron Radiation Instrumentation (Shen, Q. & Nelson, C., eds), Amer Inst Physics, Melville.

40. Morris, E. R. & Searle, M. S. (2012) Overview of protein folding mechanisms: experimental and theoretical approaches to probing energy landscapes, *Curr Protoc Protein Sci.* **Chapter 28**, Unit 28.2.1-22.

41. Manavalan, B., Kuwajima, K. & Lee, J. (2019) PFDB: A standardized protein folding database with temperature correction, *Sci Rep.* **9**, 1588.

42. Silow, M. & Oliveberg, M. (1997) Transient aggregates in protein folding are easily mistaken for folding intermediates, *Proc Natl Acad Sci U S A.* **94**, 6084-6086.

43. Went, H. M., Benitez-Cardoza, C. G. & Jackson, S. E. (2004) Is an intermediate state populated on the folding pathway of ubiquitin?, *FEBS Lett.* **567**, 333-8.

44. Colón, W., Elöve, G. A., Wakem, L. P., Sherman, F. & Roder, H. (1996) Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome *c* folding, *Biochemistry.* **35**, 5538-5549.

45. Houry, W. A., Rothwarf, D. M. & Scheraga, H. A. (1996) Circular dichroism evidence for the presence of burst-phase intermediates on the conformational folding pathway of ribonuclease A, *Biochemistry.* **35**, 10125-10133.

46. Sosnick, T. R., Mayne, L. & Englander, S. W. (1996) Molecular collapse: The rate-limiting step in two-state cytochrome *c* folding, *Proteins.* **24**, 413-426.

47. Sosnick, T. R., Shtilerman, M. D., Mayne, L. & Englander, S. W. (1997) Ultrafast signals in protein folding and the polypeptide contracted state, *Proc Natl Acad Sci U S A.* **94**, 8545-8550.

48. Qi, P. X., Sosnick, T. R. & Englander, S. W. (1998) The burst phase in ribonuclease A folding and solvent dependence of the unfolded state, *Nat Struct Biol.* **5**, 882-884.

49. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W. & Sosnick, T. R. (2002) Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding, *J Mol Biol.* **324**, 359-71.

50. Shastry, M. C. R. & Roder, H. (1998) Evidence for barrier-limited protein folding kinetics on the microsecond time scale, *Nat Struct Biol.* **5**, 385-392.

51. Welker, E., Maki, K., Shastry, M. C., Juminaga, D., Bhat, R., Scheraga, H. A. & Roder, H. (2004) Ultrarapid mixing experiments shed new light on the characteristics of the initial conformational ensemble during the folding of ribonuclease A, *Proc Natl Acad Sci U S A.* **101**, 17681-17686.

52. Akiyama, S., Takahashi, S., Ishimori, K. & Morishima, I. (2000) Stepwise formation of α-helices during cytochrome *c* folding, *Nat Struct Biol.* **7**, 514-520.

53. Qiu, L., Zachariah, C. & Hagen, S. J. (2003) Fast chain contraction during protein folding: "Foldability" and collapse dynamics, *Phys Rev Lett.* **90**, 168103.

54. Goldbeck, R. A., Chen, E. & Kliger, D. S. (2009) Early events, kinetic intermediates and the mechanism of protein folding in cytochrome *c*, *Int J Mol Sci.* **10**, 1476-1499.

55. Ellerby, L. M., Nishida, C. R., Nishida, F., Yamanaka, S. A., Dunn, B., Valentine, J. S. & Zink, J. I. (1992) Encapsulation of proteins in transparent porous silicate glasses prepared by the sol-gel method, *Science.* **255**, 1113-1115.

56. Shibayama, N. & Saigo, S. (1995) Fixation of the quaternary structures of human adult haemoglobin by encapsulation in transparent porous silica gels, *J Mol Biol.* **251**, 203-209.

57. Shibayama, N. & Saigo, S. (1999) Kinetics of the allosteric transition in hemoglobin within silicate sol-gels, *J Am Chem Soc.* **121**, 444-445.

58. Ronda, L., Bruno, S., Campanini, B., Mozzarelli, A., Abbruzzetti, S., Viappiani, C., Cupane, A., Levantino, M. & Bettati, S. (2015) Immobilization of proteins in silica gel: Biochemical and biophysical properties, *Curr Org Chem.* **19**, 1653-1668.

59. Samuni, U., Navati, M. S., Juszczak, L. J., Dantsker, D., Yang, M. & Friedman, J. M. (2000) Unfolding and refolding of sol-gel encapsulated carbonmonoxymyoglobin: An orchestrated spectroscopic study of intermediates and kinetics?, *J Phys Chem B.* **104**, 10802-10813.

60. Peterson, E. S., Leonard, E. F., Foulke, J. A., Oliff, M. C., Salisbury, R. D. & Kim, D. Y. (2008) Folding myoglobin within a sol-gel glass: Protein folding constrained to a small volume, *Biophys J.* **95**, 322-332.

61. Shibayama, N. (2008) Slow motion analysis of protein folding intermediates within wet silica gels, *Biochemistry.* **47**, 5784-94.

62. Shibayama, N. (2008) Circular dichroism study on the early folding events of beta-lactoglobulin entrapped in wet silica gels, *FEBS Lett.* **582**, 2668-2672.

63. Ohgushi, M. & Wada, A. (1983) 'Molten-globule state': a compact form of globular proteins with mobile side-chains, *FEBS Lett.* **164**, 21-24.

64. Nakamura, S., Seki, Y., Katoh, E. & Kidokoro, S. (2011) Thermodynamic and structural properties of the acid molten globule state of horse cytochrome *c*, *Biochemistry.* **50**, 3116-3126.

65. Colón, W. & Roder, H. (1996) Kinetic intermediates in the formation of the cytochrome *c* molten globule, *Nat Struct Biol.* **3**, 1019-1025.

66. Plaxco, K. W., Simons, K. T. & Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins, *J Mol Biol.* **277**, 985-994.

67. Baker, D. (2000) A surprising simplicity to protein folding, *Nature.* **405**, 39-42.

68. 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.

69. Gromiha, M. M. & Selvaraj, S. (2001) Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction, *J Mol Biol.* **310**, 27-32.

70. Harihar, B. & Selvaraj, S. (2009) Refinement of the long-range order parameter in predicting folding rates of two-state proteins, *Biopolymers.* **91**, 928-935.

71. Makarov, D. E., Keller, C. A., Plaxco, K. W. & Metiu, H. (2002) How the folding rate constant of simple, single-domain proteins depends on the number of native contacts, *Proc Natl Acad Sci U S A.* **99**, 3535-3539.

72. Makarov, D. E. & Plaxco, K. W. (2003) The topomer search model: A simple, quantitative theory of two-state protein folding kinetics, *Protein Sci.* **12**, 17-26.

73. Nölting, B., Schälike, W., Hampel, P., Grundig, F., Gantert, S., Sips, N., Bandlow, W. & Qi, P. X. (2003) Structural determinants of the rate of protein folding, *J Theor Biol.* **223**, 299-307.

74. Zhou, H. & Zhou, Y. (2002) Folding rate prediction using total contact distance, *Biophys J.* **82**, 458-463.

75. Micheletti, C. (2003) Prediction of folding rates and transition-state placement from native-state geometry, *Proteins.* **51**, 74-84.

76. Dixit, P. D. & Weikl, T. R. (2006) A simple measure of native-state topology and chain connectivity predicts the folding rates of two-state proteins with and without crosslinks, *Proteins.* **64**, 193-197.

77. Galzitskaya, O. V., Garbuzynskiy, S. O., Ivankov, D. N. & Finkelstein, A. V. (2003) Chain length is the main determinant of the folding rate for proteins with three-state folding kinetics, *Proteins.* **51**, 162-166.

78. Ivankov, D. N., Garbuzynskiy, S. O., Alm, E., Plaxco, K. W., Baker, D. & Finkelstein, A. V. (2003) Contact order revisited: Influence of protein size on the folding rate, *Protein Sci.* **12**, 2057-2062.

79. Kamagata, K., Arai, M. & Kuwajima, K. (2004) Unification of the folding mechanisms of non-two-state and two-state proteins, *J Mol Biol.* **339**, 951-965.

80. Istomin, A. Y., Jacobs, D. J. & Livesay, D. R. (2007) On the role of structural class of a protein with two-state folding kinetics in determining correlations between its size, topology, and folding rate, *Protein Sci.* **16**, 2564-2569.

81. Harihar, B. & Selvaraj, S. (2011) Analysis of rate-limiting long-range contacts in the folding rate of three-state and two-state Proteins, *Protein Pept Lett.* **18**, 1042-1052.

82. Ivankov, D. N. & Finkelstein, A. V. (2004) Prediction of protein folding rates from the amino acid sequence-predicted secondary structure, *Proc Natl Acad Sci U S A.* **101**, 8942-8944.

83. Zhang, L. & Sun, T. (2005) Folding rate prediction using *n*-order contact distance for proteins with two- and three-state folding kinetics, *Biophys Chem.* **113**, 9-16.

84. Ouyang, Z. & Liang, J. (2008) Predicting protein folding rates from geometric contact and amino acid sequence, *Protein Sci.* **17**, 1256-63.

85. Chang, L., Wang, J. & Wang, W. (2010) Composition-based effective chain length for prediction of protein folding rates, *Phys Rev E.* **82**, 12.

86. De Sancho, D. & Muñoz, V. (2011) Integrated prediction of protein folding and unfolding rates from only size and structural class, *Phys Chem Chem Phys.* **13**, 17030-17043.

87. Huang, S. R. & Huang, J. T. T. (2013) Inter-residue interaction is a determinant of protein folding kinetics, *J Theor Biol.* **317**, 224-228.

88. Baiesi, M., Orlandini, E., Seno, F. & Trovato, A. (2017) Exploring the correlation between the folding rates of proteins and the entanglement of their native states, *J Phys A-Math Theor.* **50**, 16.

89. Censoni, L. & Martinez, L. (2018) Prediction of kinetics of protein folding with non-redundant contact information, *Bioinformatics.* **34**, 4034-4038.

90. Khor, S. (2018) Folding with a protein's native shortcut network, *Proteins.* **86**, 924-934.

91. Muñoz, V. & Eaton, W. A. (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures, *Proc Natl Acad Sci U S A.* **96**, 11311-11316.

92. Alm, E., Morozov, A. V., Kortemme, T. & Baker, D. (2002) Simple physical models connect theory and experiment in protein folding kinetics, *J Mol Biol.* **322**, 463-476.

93. Garbuzynskiy, S. O., Finkelstein, A. V. & Galzitskaya, O. V. (2004) Outlining folding nuclei in globular proteins, *J Mol Biol.* **336**, 509-525.

94. Galzitskaya, O. V. & Glyakina, A. V. (2012) Nucleation-based prediction of the protein folding rate and its correlation with the folding nucleus size, *Proteins.* **80**, 2711-2727.

95. Garbuzynskiy, S. O., Ivankov, D. N., Bogatyreva, N. S. & Finkelstein, A. V. (2013) Golden triangle for folding rates of globular proteins, *Proc Natl Acad Sci U S A.* **110**, 147-150.

96. Sanchez, I. E. & Kiefhaber, T. (2003) Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding, *J Mol Biol.* **325**, 367-376.