ELSEVIER

Biophysical Chemistry 128 (2007) 105–113

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Review

Identification and characterization of protein folding intermediates

Stefano Gianni [a,*], Ylva Ivarsson [a], Per Jemth [b], Maurizio Brunori [a], Carlo Travaglini-Allocatelli [a]

[a] *Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Università di Roma "La Sapienza", Piazzale A. Moro 5, 00185 Rome, Italy*

[b] *Department of Medical Biochemistry and Microbiology, Uppsala University, BMC Box 582, SE-75123 Uppsala, Sweden*

Received 29 March 2007; received in revised form 16 April 2007; accepted 16 April 2007
Available online 24 April 2007

Abstract

In order to understand the mechanism by which a polypeptide chain folds into its functionally active native state it is necessary to characterize in detail all the species accumulated along the pathway. The elusive nature of protein folding intermediates poses their identification and characterization as an extremely difficult task in the protein folding field. In the case of small single domain proteins, the direct measurement of the thermodynamics and structural parameters of protein folding intermediates has provided new insights on the nature of the forces involved in the stabilization of nascent protein structures. Here we summarize some of the experimental approaches aimed at the detection and characterization of folding intermediates along with a discussion of some general structural features emerging from these studies.

Keywords: Protein folding; Intermediate; Transition state; Folding kinetics; Mutagenesis

Contents

1. Introduction

It dates back to the early 1960s when Anfinsen discovered that small proteins can adopt their biologically relevant structure by self-assembly [1]. This observation led to the conclusion that, at least for small globular proteins, the structure of the native state is determined only by the primary structure. More than 40 years later, efforts are still being made to understand how the information contained in the amino acid sequence is sufficient to define the adoption of a three-dimensional structure on a biologically relevant time scale.

Classically, protein folding was thought to proceed in a stepwise manner, via a series of partially structured intermediates.

* Corresponding author. Tel.: +39 0649910548; fax: +39 064440062.
E-mail address: stefano.gianni@uniroma1.it (S. Gianni).

0301-4622/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2007.04.008

All proteins were considered to fold through a framework model that postulated secondary structure to form before the tertiary structure was locked in place. This model was challenged in 1991 when chymotrypsin inhibitor 2 was shown to fold in a cooperative two-state transition without detectable kinetic intermediates [2]. Several other small single domain proteins were later shown to fold in a similar way [3]. The two state mechanism provided the basis for a nucleation-condensation mechanism, with folding through simultaneous formation of secondary and tertiary structure centered around a small folding nucleus [4,5]. Recent work on the homeodomain superfamily [6,7] and on a PDZ domain [8], led to the view that the framework and nucleation-condensation models represent extreme manifestation of an underlying common mechanism and that proteins may appear to fold by either the nucleation-condensation or framework mechanism depending on the inherent stability of their secondary structure elements [7,9].

The existence and importance of intermediates are fundamental issues for our understanding of protein folding [10]. It is generally accepted that the folding pathways of large proteins involve the population of partially structured species en route to the native state. In contrast, accumulation of intermediates in the folding of small protein domains (<100 amino acids) is still widely debated [11]. In this review we focus on the intermediates in folding of protein domains, how to identify them, how to characterize their role in the folding process and their structural features. These are particularly difficult tasks considering the fact that folding intermediates are often only transiently populated and play different mechanistic roles, being either on-pathway productive species, or off-pathway kinetic traps that have to unfold for proper folding to take place. Furthermore, it has been suggested that accumulation of folding intermediates may be a crucial step prior to protein aggregation and amyloid fiber formation [12], posing the characterization of these partially structured species as a central problem in structural biology.

2. Linear free-energy relationship

Early works on protein folding revealed a key characteristic of the solution behavior of many proteins, i.e. the co-operativity of protein folding and unfolding [13]. The study of protein folding requires the use of perturbations affecting the stability of the system under study; the parameters affecting protein stability typically being temperature, pH, ionic strength and solvent composition. Chaotropic agents (such as urea and guanidine) are generally used to affect the protein stability. Fig. 1 shows a representative unfolding transition of a PDZ domain [14], where the native state of a protein resists denaturation at low denaturant concentrations. Higher denaturant concentrations induce a monotonic unfolding transition that leads to the fully denatured state. This type of all-or-none denaturation is a general characteristic of protein folding/unfolding; a number of weak, but mutually dependent, non-covalent interactions stabilize the native state relative to other states. With a few exceptions (see for example [15,16]), small single domain proteins have been observed to populate only the native and the unfolded state at equilibrium. Thus, it is generally rare to observe folding intermediates at equilibrium and proteins tend to display a two-state transition as represented in Fig. 1.

It has been empirically determined that the stability of proteins can be expressed as a linear function of denaturant concentration [13,17]. Thus, the linear extrapolation method is routinely used in protein folding studies. Following this approach it is possible to estimate the protein stability in the absence of denaturant by monitoring the fraction of unfolded protein under different denaturant concentrations [17]. The inset panel in Fig. 1 shows an example of how such linear relationship can be employed using a classical co-operative unfolding profile. The slope of the line in the inset panel in Fig. 1 is a constant denoted m_{D-N} value (expressed in kcal mol^{-1} M^{-1}), a key parameter for experimental protein folding analysis. The m_{D-N} value defines the sensitivity of the system to solvent denaturation and has been shown empirically to be related to the change in solvent exposed surface area upon denaturation [18].

3. Folding kinetics: The chevron plot

While equilibrium studies provide useful information about folding in terms of stability and co-operativity, only few details can be determined about the reaction dynamics. Understanding the mechanism by which a polypeptide chain folds into a unique native protein requires the characterization of the free energy landscape of protein folding in detail, including partially (un) folded intermediates, transition states, and their order in the folding process. To achieve this aim it is necessary to perform time resolved protein folding studies. In such studies, a perturbation of the energetics of the system is imposed by changing, for example, the urea concentration, the pH or the temperature. The observed rate constant and its associated amplitude are measured as the system relaxes to the new equilibrium.

In order to study the protein folding kinetics, proteins are usually denatured (using either urea or guanidium hydrochloride)

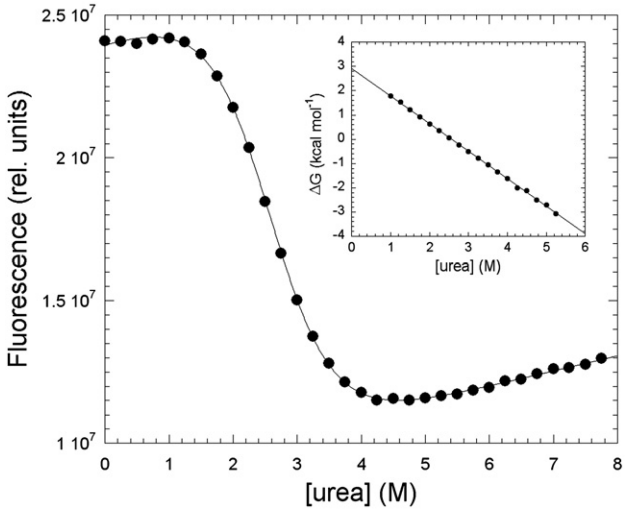


Fig. 1. Equilibrium unfolding of the second PDZ domain from PTP-BL monitored by fluorescence. Inset Panel. Linear free energy extrapolation. As discussed in the text a quantitative analysis of the observed spectroscopic signals as a function of denaturant allows the estimate of the stability of the protein in the absence of denaturant.

and then refolded through rapid dilution in a stopped-flow apparatus (dead time on the order of 1–2 ms). In a two-state model only the native (N) and the unfolded, or denatured state (D) are populated as depicted in Scheme 1

$$D \underset{k_U}{\overset{k_F}{\rightleftarrows}} N. \qquad \text{(Scheme 1)}$$

In a refolding experiment of a two state folding protein, the concentration of native state as a function of time follows Eq. (1):

$$[D]_t = \frac{D^0}{k_F + k_U} \cdot \{k_F \cdot \exp[-(k_F + k_U)t] + k_U\} \qquad (1)$$

$[D]_t$ is the concentration of denatured protein at time t, and D^0 is the initial concentration of (denatured) protein. The time course is thus described by an exponential function with an apparent rate constant of

$$k_{\text{obs}} = k_F + k_U. \qquad (2)$$

As shown in Fig. 1 the linear dependence of the free energy of unfolding is defined by the constant m_{D-N}. Denaturants affect the stability of the different conformational states of polypeptides by selectively stabilizing them according to their degree of solvent exposed surface area. For a protein which displays a two-state transition the activation energy for folding and unfolding reactions will therefore also be linearly dependent on denaturant concentration [2,19]. Such a relationship is formalized in Eq. (3), which defines the dependence of the observed (un)folding rate constant (k_{obs}) on denaturant concentration:

$$k_{\text{obs}} = k_F^0 \exp(-m_F[\text{urea}]/RT) + k_U^0 \exp(m_U[\text{urea}]/RT) \qquad (3)$$

where k_{F}^0 and k_{U}^0 are the folding and unfolding rate constants in the absence of denaturant and m_{F} and m_{U} reflect their dependence on denaturant concentration and correlate with the change in accessible surface area between the two ground states and the transition state between them. Fig. 2 shows the folding/unfolding rate constants expected for a two-state system together with the calculated profile using Eq. (3). Because of its classical V-shaped appearance, this kind of semilogaritmic plot is currently called "chevron" plot by the protein folding community, i.e. *chevron*, from Latin *caper*, and modern French *chevre*, goat. V-shaped chevron plots, indicative of two-state folding transitions, have been observed by stopped flow kinetics for a number of different structurally unrelated proteins (reviewed by [3]).

4. Identification and characterization of folding intermediates: Burst phase analysis

Under certain circumstances proteins that exhibit a co-operative, two-state equilibrium transition, may display more complex kinetics. Such proteins are generally interpreted to fold or unfold via transiently populated intermediates [20]. Folding intermediates generally form very fast (sub-millisecond time range) and this generally prevents the direct measurement of the rate constants of their formation [21]. However, despite the rapidity of their formation, the presence of intermediates can often be inferred indirectly from endpoint analysis. In such analysis, the initial signal (at time zero) and the final signal (at infinite time) of a reaction, determined from extrapolation of one or more exponential functions, are plotted as a function of denaturant concentration [22]. The dependence of the signal at infinite time, the endpoint of the reaction, on denaturant concentration should describe the equilibrium transition, as the signals of the native and denatured states in refolding and unfolding experiments are defined by the endpoint of the reaction kinetics. While in a two-state system, the initial signal in refolding experiments defines the signal of the denatured state under refolding conditions (i.e. the refolding amplitude reflects the difference between the native and denatured state at every denaturant concentrations), a significant deviation from the expected signal has been observed in many proteins [22,23]. In these cases, the observed initial signal reflects the endpoint of a faster transition of the unfolded state to a transient intermediate, implying that a rapid event is lost in the dead-time of the stopped-flow experiment (burst phase) [22]. The concept of "burst phase" was originally introduced in enzymology by Hartley and Kilby [24], for the reaction of chymotrypsin with excess of p-nitrophenyl acetate, and it is routinely used to detect protein folding intermediates escaping stopped-flow time range. However, in protein folding studies amplitude analyses are often complicated by the uncertainties of the expected signal of the denatured state under native conditions and the expected signal of the native state under denaturing conditions [11].

5. Identification and characterization of folding intermediates: The roll-over effect

Analysis of chevron plots is a common and powerful way to detect folding intermediates. In the chevron plot, the rate constants

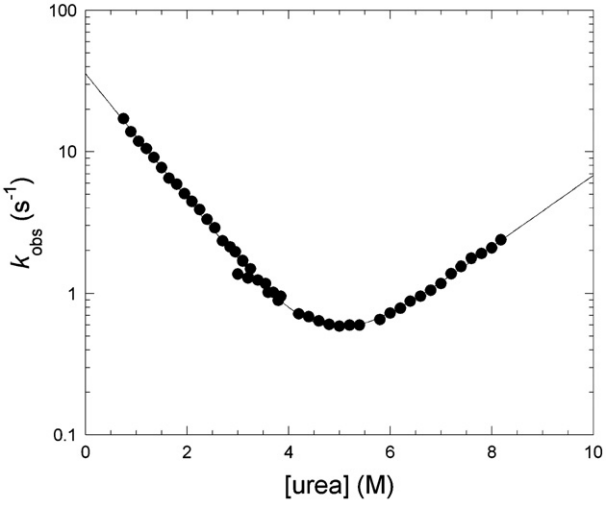


Fig. 2. Semilogarithmic plot of the folding rate constant as a function of urea (chevron plot) of the second PDZ domain from PTP-BL. The line represents the best fit to a two state model, as formalized in Eq. (3). As described in the text, in a two-state system, the observed rate constant is the sum of the folding and unfolding microscopic rate constants.

for folding and unfolding are plotted as a function of the denaturant concentration on a semilogarithmic scale. If there is only one rate-limiting energy barrier in the folding pathway a linear relation between the free energies of the different folding states and the denaturant concentration is expected, giving rise to the classical V-shaped chevron plot. Deviation from linearity in the (un)folding branch of chevron plots (so called rollover effects, or curved chevrons) have been taken as evidence for the accumulation of partially folded intermediates (either on- or off-the productive folding pathway) or for changes in rate limiting steps invoking sequential barriers on-the-path to the native state [25–27].

If a partially folded intermediate is present, the folding reaction can be described by a sequential mechanism:

$$D \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} I \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} N \qquad \text{(Scheme 2)}$$

The kinetics of a two step reaction should be fitted to the two roots of a quadratic equation, as previously shown for Im7 [28], TT cyt c_{552} [29]and the FF domain [30]. In many cases, however, only one relaxation rate constant can be experimentally observed, which jeopardizes a quantitative curve fitting. Two approximations have been introduced to describe the folding pathway of three-state systems. On one hand, the intermediate is assumed to be in a fast pre-equilibrium with one of the ground states [31]. Under such conditions:

$$k_{\text{obs}} = k_F + \frac{k_U}{1 + K_{IN}^{\text{eq}}} \qquad (4)$$

Alternatively, the intermediate is assumed to be at steady state and at low (approx. zero) concentration [26,27]. In this case, the observed kinetics may show a change in rate limiting step with increasing denaturant concentration:

$$k_{\text{obs}} = k_F + \frac{k_U}{1 + \frac{k_{IU}}{k_{IN}}} \qquad (5)$$

Note that both approximations lead to very similar solutions with a partition factor, describing alternatively either the $K_{\text{eqI-N}}$ or k_{IU}/k_{IN}, implying a deviation from the classical V-shaped chevron appearance (namely the roll-over effect).

6. Folding intermediates: On- or off- the path to the native state

It is of crucial importance to define whether a folding intermediate represents an on pathway productive species en-route to the native state or an unproductive kinetic trap. As described above, the existence of folding intermediates is often proposed from indirect observations, such as the presence of sub-millisecond burst phases or roll-over effects. However, the development of instrumentation for protein folding studies, for example ultra-rapid mixing devices [32], and temperature jump relaxation techniques [33], has allowed the direct characterization of events taking place in the dead-time of conventional stopped-flow instruments. These techniques have provided new insight in the role of intermediates in protein folding [10]. If a stable intermediate is formed before the rate-limiting transition in the folding process this will result in deviation from single exponential kinetics and at least two relaxation rate constants will be observed. It has been shown that quantitative analysis of both observed rate constants and amplitudes may be applied to determine unequivocally whether the identified folding intermediate is on- or off- the pathway to the native state, as depicted in the following reaction schemes:

$$D \underset{k_{ID}}{\overset{k_{DI}}{\rightleftarrows}} I \underset{k_{NI}}{\overset{k_{IN}}{\rightleftarrows}} N \qquad \text{On pathway} \qquad \text{(Scheme 3)}$$

$$I \underset{k_{DI}}{\overset{k_{ID}}{\rightleftarrows}} D \underset{k_{ND}}{\overset{k_{DN}}{\rightleftarrows}} N \qquad \text{Off pathway} \qquad \text{(Scheme 4)}$$

The solution of a given chemical network, involving interconnected monomolecular reaction pathways, involves simultaneous integration of a number of linear differential equations. It is often difficult to determine the analytical solution of chemical networks involving folding intermediates and approximations may be introduced (i.e. the pre-equilibrium or steady state assumptions). However, when and if the observed relaxation rate constants are close enough (<10-fold) to allow kinetic coupling, the exact analytical solution of the reaction network should be calculated. Following a stochastic approach to reaction kinetics [34], the exact analytical solution of any reaction system involving monomolecular reactions may be determined by calculating the eigenvalues of the square matrix of linear differential equations describing the reaction system. For example, for a two step reaction system:

$$A \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} B \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} C \qquad \text{(Scheme 5)}$$

the following differential equations and the associated square matrix may be derived:

$$\frac{\mathrm{d}[A]}{\mathrm{d}t} = -k_1[A] + k_{-1}[B] + 0[C] \qquad (6)$$

$$\frac{\mathrm{d}[B]}{\mathrm{d}t} = k_1[A] - (k_2 + k_{-1})[B] + k_{-2}[C] \qquad (7)$$

$$\frac{d[C]}{dt} = 0[\mathrm{A}] + k_2[\mathrm{B}] - k_{-2}[\mathrm{C}]$$

$$\Delta = \begin{vmatrix} -k_1 & k_{-1} & 0 \\ k_1 & -(k_2 + k_{-1}) & +k_{-2} \\ 0 & k_2 & -k_{-2} \end{vmatrix} \qquad (8)$$

the eigenvalues λ_1 and λ_2 of the Δ matrix will then define the two relaxation rate constants of the reaction network. For the two step mechanism the two relaxation rate constants are:

$$\lambda_1 = (p + q)/2 \qquad (9)$$

$$\lambda_2 = (p - q)/2 \qquad (10)$$

where $p = (k_1 + k_{-1} + k_2 + k_{-2})$ and $q = ((k_1 + k_{-1} + k_2 + k_{-2})^2 - 4(k_1k_2) - 4(k_1k_{-2}) - 4(k_{-1}k_{-2}))^{1/2}$.

It is obvious that both the on- (Scheme 3) and the off- (Scheme 4) pathway schemes are captured by the general two-step depicted in Scheme 5. Quantitative discrimination between the two models can be obtained by assuming an increase in native structure formation in the first step of the on-pathway model (Scheme 3) (with a decrease and increase of k_1 and k_{-1}, respectively, upon increasing denaturant concentration), or loss of structure in the first step of the off-pathway model (Scheme 4) (increase and decrease of k_1 and k_{-1}, respectively, upon increasing denaturant concentration). A crucial corollary of the analytical solution of Scheme 5 is that, under native conditions, the observed rate constant λ_2 can only be lower than or equal to (in the limit condition in which $k_1 < k_{-2}$) the microscopic rate constant k_1. Given that k_1 represents either the intermediate formation (k_{DI}) or its breakage (k_{ID}) in the on- and off-pathway scenarios respectively, it is possible to quantitatively test the two models: only the on-pathway model allows the microscopic rate constant for the unfolding of the intermediate k_{ID} to be lower than the observed rate constant λ_2 [35]. Such an observation, summarized in Fig. 3, represents a key test to distinguish between on- and off-pathway intermediates.

The folding of c_{552} from *Thermus thermophilus* [29] and *Hydrogenobacter thermophilus* [36], the FF domain [30] and Im7 protein [28] represent four examples where the role of the observed intermediate could be tested. In these four cases, folding displays double exponential kinetics and all the rate constants can be measured over a wide range of denaturant concentrations. In the case of the c-type cytochromes, in fact, the folding intermediate can be populated to a considerable extent by diluting rapidly the denaturant in a first mixing step, and then characterized in a second mixing step. For example, the accumulated intermediate can be challenged again with high denaturant concentrations to measure its rate of unfolding. On the other hand, in the case of the FF domain and Im7, a complete determination of the two observed relaxation rate constants called for two different mixing methods: stopped-flow and continuous-flow ultra rapid mixing techniques. In all cases, global analysis of observed kinetics led to the conclusion that the observed folding intermediate was an on-pathway species on the route between the denatured and native states.

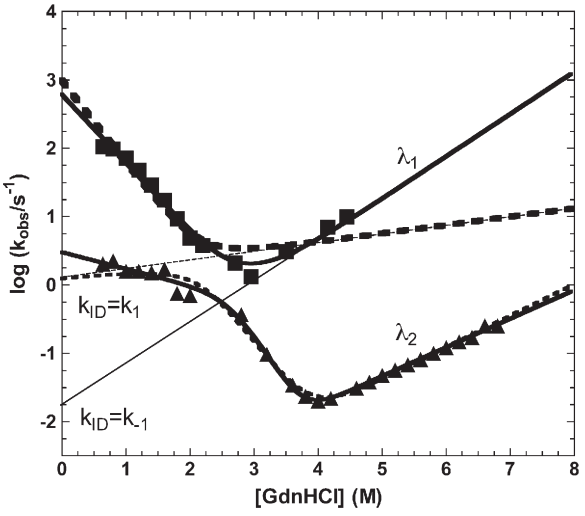


Fig. 3. On pathway versus off pathway folding intermediates. Chevron plot of HT cyt c_{552}, recorded at pH 4.7 10 °C, from fluorescence measurements. Continuous and broken lines represent the best global fit for an on- and off-pathway model respectively.

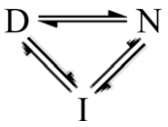
Recently, a novel type of kinetic test has been proposed to define the mechanistic role of folding intermediates [37]. In this experiment, a large excess of ligand is added to denatured protein, thereby trapping newly folded protein by ligand binding, posing the folding step quasi-irreversible (Scheme 6).

$$D \underset{k_U}{\overset{k_F}{\rightleftarrows}} N \overset{k_{on}[P]}{\longrightarrow} N \cdot P \qquad \text{(Scheme 6)}$$

It has been shown that when and if (i) the ligand interacts only with the fully native protein and (ii) the binding reaction is much faster than the folding, this allows determination of folding rates independently of the unfolding even under strongly denaturing conditions. Hence, the full folding limb can be added to the conventional chevron plot, which eliminates the need for long extrapolations in the curve fitting and allows discrimination between the two models described by Eqs. (1) and (2). Following this approach, the previously detected intermediate in the folding of PDZ2 was shown to be an on-pathway species to the native state [37]. Under the same token, Teilum et al. used NMR relaxation dispersion techniques to determine the (un)folding rate constants for bovine ACBP at several denaturant concentrations [38], thereby complementing the conventional chevron with an inverted chevron plot and identifying an unfolding intermediate bearing large resemblance to the native state.

7. Sequential versus parallel folding mechanisms

Lately, a new view of protein folding has emerged, namely that proteins fold *via* parallel routes, which may be depicted as a flux of molecules along different folding channels partitioned by the relative effect of denaturant on the relevant transition states [39]. According to this model, intermediates are not necessary steps during the folding but rather kinetic traps due to the ruggedness of the folding landscape. Since there might be many routes to the native state, the use of on- versus off-pathway intermediates may be sometimes simplistic. An accurate comprehension of a given folding process should therefore take into account the distinction between sequential (Scheme 5) and parallel folding mechanisms, as depicted below for a simple case:



(Scheme 7)

A powerful experimental approach to test the presence of parallel folding routes is afforded by double mixing interrupted refolding experiments. Such methodology, originally introduced in the protein folding field to detect prolyl *cis–trans*

isomerisation events in the folding of RNaseA [40], was later successfully exploited to monitor parallel folding routes involving an homogeneous denatured state [41]. The unfolded protein is first mixed against a refolding buffer (first mix) and after a controlled delay time, refolding is interrupted by rapid addition of high concentrations of the same denaturant (second mix). This approach makes it possible to distinguish partially folded intermediates from native molecules since these states are characterized by different unfolding rates. In particular, the native protein being separated from the unfolded one by the highest energy barrier should unfold more slowly than any partially folded intermediate. A plot of the amplitudes of the observed unfolding rate constants as a function of the delay time between the first and the second mix allows monitoring the fraction of native protein formed when refolding was interrupted. Moreover, this plot gives the time course for the formation of native and intermediate protein state(s).

Kiefhaber and coworkers exploited this technique to dissect the folding pathway of lysozyme showing that under strongly native conditions where partially folded states populate during refolding, this protein can attain the native state via two different kinetic tracks [41,42]. The results clearly demonstrated that, while most of native lysozyme is formed in a slow kinetic reaction, some of the molecules reach the native state on a fast refolding channel. In the case of c-type cytochromes, this technique permitted to distinguish between members characterized by obligatory intermediates on the path to the native state (Scheme 3; [36]) from those where a non-obligatory intermediate is present in a triangular folding mechanism (Scheme 7) characterized by parallel pathways [43,44].

8. Experimental methods to address structural features of protein folding intermediates

The practical problem in studying folding intermediates is that the folding reactions are generally highly co-operative, so that intermediates are not populated at equilibrium. However, as described above, even if intermediates cannot be found at equilibrium they may accumulate transiently in time-resolved folding experiments prior to the formation of the native state. Generally structural information about folding intermediates can only be inferred indirectly from analysis of the folding rates and protein engineering. In particular, by systematically mutating side chains while probing the effects on the kinetics, it is possible to map out detailed interaction patterns in folding intermediates and transition states. In fact, mutations that destabilize the folding intermediate target contact formed in its structure. The strength of the contacts is measured by the Φ value which normalizes the stability loss of the intermediate state to that of the native state. $\Phi=1$ indicates that the site of mutation is fully structured in the intermediate state, $\Phi=0$ indicates that such a site is as unfolded as the denatured state. This technique, Φ value analysis, has been successfully employed to determine the structure of the folding intermediates observed in the folding of barnase [20,45] and of the bacterial immunity protein Im7 [28,46]. The structural features arising from these studies will be discussed in the next section.

There is a controversy about the accuracy of experimentally determined Φ values. Since Φ values are ratios of differences between experimental observables, they may be sensitive to errors when the observed differences are relatively small. Three independent laboratories performed a blind replicate Φ value analysis on the FynSH3 domain [47]. In this study, it has been shown that small changes in free energy associated with the probing mutation may compromise the precision of Φ values. However, the same study showed that inter-laboratory consistence of experimentally determined Φ values may be greatly improved when the slopes of the chevron plots, namely the *m*-values, where constrained for the different mutants. This observation suggests that long extrapolations in chevron plot analysis represent the major source of errors in Φ value analysis. Fersht and Sato have recently highlighted the experimental power and limitations of the Φ value analysis clearly indicating that accurate Φ values may be calculated only when (i) non-disruptive deletion mutants are characterized and (ii) the change in free energy associated with the mutation is more than ≈ 0.6 kcal mol^{-1} [48].

A new method for measuring the extent of development of interactions in transition states has been recently introduced by Sosnick and co-workers, namely the Ψ value analysis [49]. This technique involves the substitution of two target surface amino acids by His residues that may be reversibly cross-linked in the native state by adding an ion such as Co^{2+}. The rational of such analysis proposes two parallel routes for folding and unfolding: one in which the His residues are in their native state proximity in the transition state ("present-route"), which can bind M^{2+}; and one in which they are apart ("absent-route"). In the absence of M^{2+}, the protein folds with a rate constant $k_{obs}=k_{abs}+k_{pres}$. On the addition of M^{2+}, the rate constant for the present-route (k_{pres}) is enhanced, but k_{abs} remains constant as M^{2+} increases. In analogy with the Φ value analysis, the kinetics and equilibria of folding and unfolding may be determined as a function of M^{2+} concentration, to extract structural information about folding intermediates and transition states. According to Sosnick and coworkers [50], a potential advantage of Ψ versus Φ value analysis lies in the more conservative nature of the mutation introduced in the Ψ analysis, the Φ value analysis always relying in post-mutation conditions for which the importance of an interaction may be reduced per se by the mutation. A surprising result from Ψ analysis was the presence of parallel folding pathways in all reported studies and a major discrepancy between Φ and Ψ values measured in the same protein [50]. Two independent studies have later explained such discrepancy by showing that Ψ values cannot be analyzed in the same way as other rate-equilibrium free energy relationships due to the involvement of bimolecular reactions [51,52]. In particular, since the native, unfolded and intermediate states may have different dissociation constants for the metal ion, Ψ values reflect the relative degree of structure of transition states and folding intermediates only for the extreme values of $\Psi=0$ or $\Psi=1$. In our opinion, this observation poses the Φ value analysis as the best experimental available technique to characterize the structural features of protein folding intermediates and transition states.

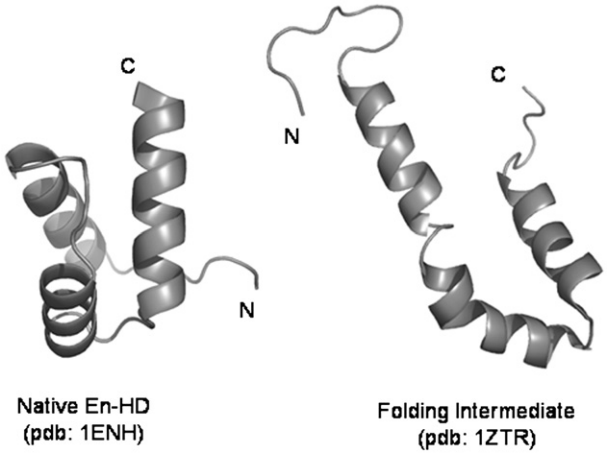


Fig. 4. Three-dimensional structure of native En-HD and its point mutant L16A. The structure of the L16A variant recorded by NMR at physiological conditions resembles the structure of the folding intermediate intervening in wild type En-HD folding pathway.

A direct method to extract information on the structural features of protein folding intermediate has been recently introduced by Fersht and co-workers. By specifically destabilizing the native state without altering the structure of intermediates intervening in folding pathways, it is in theory possible to populate such intermediate state(s) at equilibrium and physiological conditions. The structure of these intermediates may be then addressed by solution NMR methods. In such cases, however, a complete biophysical analysis of engineered mutants in comparison with the wild-type protein is required, in order to prove that structures resulting from mutagenesis arise from a destabilization of the native state rather than stabilization of mis-folded states which may be only peripheral to the folding reaction [33]. As described below, this approach has been successfully employed to determine the structure of the folding intermediate of a small alpha-helical protein, the Engrailed homeodomain (En-HD) (Fig. 4) [53].

9. Structural features of protein folding intermediates

Recent developments of powerful experimental methods as well as the synergy between experimental and theoretical techniques have contributed to define complex folding pathways at nearly atomic resolution. In this section, we will discuss some general structural aspects of folding intermediates emerging from these studies.

The folding pathway of barnase has been extensively characterized in the past two decades. A large number of experiments have shown that this protein folds via at least one intermediate. The first evidence for a folding intermediate came from a roll-over effect in the chevron plot [20]. More recently, analysis of pulse labeling hydrogen exchange protection factors has shown that the intermediate is an on-pathway species to the native state [54]. The structural features of the barnase folding intermediate has been addressed both experimentally (e.g. by protein engineering [20]) and theoretically (i.e. by molecular dynamics [45,55]). Overall, it appears that this protein folds by parts as discrete regions of the protein are highly structured whereas others are flexible and denatured-like. Importantly, the overall topology of the folding intermediate resembles that of the native structure. Several interactions between the strands β1, β2 and β3 appear to be consolidated together with helix 1; on the other hand, the second module of the protein, involving helix 2 and the loops appears to be unstructured.

The helical bacterial immunity proteins Im7 and Im9 have been shown by Radford and co-workers to fold via kinetic mechanisms of differing complexity, despite having 60% sequence identity [56,57]. At pH 7.0 and 10 °C, Im7 folds in a three-state mechanism involving an on-pathway intermediate, whereas Im9 folds in an apparent two-state transition. Kinetic modeling of the folding and unfolding data for Im7 between pH 5.0 and 8.0 shows that the on-pathway intermediate is stabilized by more acidic conditions, whilst the native state is destabilized. At pH 7.0, the folding and unfolding kinetics for Im9 can be fitted adequately by a two-state model, in accord with previous results. However, under acidic conditions there is a clear roll over in the chevron plot, consistent with the population of one or more intermediate(s) early during folding [58]. These observations indicate that, as observed in the case of cytochrome c [59], apo-myoglobin [16], the PDZ domain [60] and homeodomain-like superfamilies [6,7], intermediate formation is a general step in immunity protein folding and suggest that it is necessary to explore a wide range of refolding conditions in order to show that intermediates do or do not form.

It appears that the overall structure of the folding intermediate of Im7 is well-defined. Three of the four native helices are fully structured and docked around the hydrophobic core, which is largely formed but in the process of being consolidated [46]. Interestingly, many of the hydrophobic interactions that stabilise the intermediate become disrupted in the transition state ensemble. Moreover, several lines of evidence indicate that such an intermediate is stabilized by a network of hydrophobic non-native interactions around helix 4, suggesting that Im7 forms an on pathway [28] but mis-folded intermediate [46]. The results for Im7 suggest that mis-folding may be a natural consequence of hierarchical folding (diffusion-collision). The early stages of folding involve the formation of fully structured secondary elements prior to solvent exclusion; the protein then collapses to form a mis-folded on-pathway intermediate and then it folds to the native state.

As recalled above, a single point mutation in En-HD from *Drosophila melanogaster* specifically destabilizes the native state, leading to the population of an intermediate under native conditions [33]. Interestingly, the denatured state of En-HD seems compact and partly structured at conditions that favor folding but is disorganized under denaturing conditions. It has been proposed that, at physiological pH, the denatured state is the folding intermediate because it is the most stable of the denatured conformations [33]. Solving the structure of the intermediate/denatured state using standard NMR methods revealed that the conformation of helices 2 and 3 was highly native like, whereas helix 1 had secondary structure but no tertiary interactions (Fig. 4) [53].

Experimental studies on folding intermediates have contributed to provide invaluable insights in defining the forces driving

the early events in protein folding. Overall, it appears that protein folding intermediates are stabilized both by native and non-native interactions [10,46,53]. Recent observation that non-native interactions may be observed for productive on-pathway intermediates suggests that partial protein mis-folding might be an obligatory step preceding native state consolidation. These results parallel the observation that protein aggregation and amyloid fiber formation may arise from accumulation of partially folded intermediates [12] and poses the characterization of the structural features of protein folding intermediates as a central problem in structural biology.

Ackowledgements

Y.I. is supported by a fellowship from the Wenner-Gren Foundations (Sweden). Work partly supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (RBLA03B3KC_004 and 2005050270_004 to M.B. and 2005027330_005 to C.T.A.).

References

[1] C.B. Anfinsen, E. Haver, M. Sela, F.H.J. White, The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain, Proc. Natl. Acad. Sci. U. S. A. 47 (1961) 1309–1314.
[2] S.E. Jackson, A.R. Fersht, Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition, Biochemistry 30 (1991) 10428–10435.
[3] S.E. Jackson, How do small single-domain proteins fold? Fold. Des. 3 (1998) R81–R91.
[4] V.I. Abkevich, A.M. Gutin, E.I. Shakhnovich, Specific nucleus as the transition state for protein folding: evidence from the lattice model, Biochemistry 33 (1994) 10026–10036.
[5] L.S. Itzhaki, D.E. Otzen, A.R. Fersht, The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding, J. Mol. Biol. 254 (1995) 260–288.
[6] S. Gianni, N.R. Guydosh, F. Khan, T.D. Caldas, U. Mayor, G.W. White, M.L. DeMarco, V. Daggett, A.R. Fersht, Unifying features in protein-folding mechanisms, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 13286–13291.
[7] G.W. White, S. Gianni, J.G. Grossmann, P. Jemth, A.R. Fersht, V. Daggett, Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding, J. Mol. Biol. 350 (2005) 757–775.
[8] S. Gianni, C.D. Geierhaas, N. Calosci, P. Jemth, G.W. Vuister, C. Travaglini-Allocatelli, M. Vendruscolo, M. Brunori, A PDZ domain recapitulates a unifying mechanism for protein folding, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 128–133.
[9] V. Daggett, A.R. Fersht, Is there a unifying mechanism for protein folding? Trends Biochem. Sci. 28 (2003) 18–25.
[10] D.J. Brockwell, S.E. Radford, Intermediates: ubiquitous species on folding energy landscapes? Curr. Opin. Struck. Biol. 17 (2007) 30–37.
[11] B.A. Krantz, L. Mayne, J. Rumbley, S.W. Englander, T.R. Sosnick, Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding, J. Mol. Biol. 324 (2002) 359–371.
[12] T.R. Jahn, M.J. Parker, S.W. Homans, S.E. Radford, Amyloid formation under physiological conditions proceeds via a native-like folding intermediate, Nat. Struct. Mol. Biol. 13 (2006) 195–201.
[13] C. Tanford, Protein denaturation, Adv. Protein Chem. 24 (1968) 1–95.
[14] S. Gianni, N. Calosci, J.M. Aelen, G.W. Vuister, M. Brunori, C. Travaglini-Allocatelli, Kinetic folding mechanism of PDZ2 from PTP-BL, Protein Eng. Des. Sel. 18 (2005) 389–395.
[15] M.S. Kay, C.H.I. Ramos, R.L. Baldwin, Specificity of native-like interhelical hydrophobic contacts in the apomyoglobin intermediate, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 2007–2012.
[16] R.A. Staniforth, S. Giannini, M.G. Bigotti, F. Cutruzzola, C. Travaglini–Allocatelli, M. Brunori, A new folding intermediate of apomyoglobin from *Aplysia limacina*: stepwise formation of a molten globule, J. Mol. Biol. 297 (2000) 1231–1244.
[17] C.N. Pace, K.L. Shaw, Linear extrapolation method of analyzing solvent denaturation curves, Proteins (Suppl 4) (2000) 1–7.
[18] J.K. Myers, C.N. Pace, J.M. Scholtz, Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, Protein Sci. 4 (1995) 2138–2148.
[19] T. Schindler, M. Herrler, M.A. Marahiel, F.X. Schmid, Extremely rapid protein folding in the absence of intermediates, Nat. Struct. Biol. 2 (1995) 663–673.
[20] A. Matouschek, J.T. Kellis Jr., L. Serrano, M. Bycroft, A.R. Fersht, Transient folding intermediates characterized by protein engineering, Nature 346 (1990) 440–445.
[21] H. Roder, M.R. Shastry, Methods for exploring early events in protein folding, Curr. Opin. Struck. Biol. 9 (1999) 620–626.
[22] S. Khorasanizadeh, I.D. Peters, H. Roder, Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues, Nat. Struct. Biol. 3 (1996) 193–205.
[23] J.M. Sauder, N.E. MacKenzie, H. Roder, Kinetic mechanism of folding and unfolding of *Rhodobacter capsulatus* cytochrome c2, Biochemistry 35 (1996) 16852–16862.
[24] B.S. Hartlley, B.A. Kilby, The reaction of p-nitrophenyl esters with chymotrypsin and insulin, Biochem. J. 56 (1954) 288–297.
[25] M. Oliveberg, Alternative explanations for "multistate" kinetics in protein folding: transient aggregation and changing transition-state ensembles, Acc. Chem. Res. 31 (1998) 765–772.
[26] F. Khan, J.I. Chuang, S. Gianni, A.R. Fersht, The kinetic pathway of folding of barnase, J. Mol. Biol. 333 (2003) 169–186.
[27] I.E. Sanchez, T. Kiefhaber, Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding, J. Mol. Biol. 325 (2003) 367–376.
[28] A.P. Capaldi, M.C. Shastry, C. Kleanthous, H. Roder, S.E. Radford, Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate, Nat. Struct. Biol. 8 (2001) 68–72.
[29] C. Travaglini-Allocatelli, S. Gianni, V. Morea, A. Tramontano, T. Soulimane, M. Brunori, Exploring the cytochrome c folding mechanism: cytochrome c552 from thermus thermophilus folds through an on-pathway intermediate, J. Biol. Chem. 278 (2003) 41136–41140.
[30] P. Jemth, S. Gianni, R. Day, B. Li, C.M. Johnson, V. Daggett, A.R. Fersht, Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 6450–6455.
[31] M.J. Parker, J. Spencer, A.R. Clarke, An integrated kinetic analysis of intermediates and transition states in protein folding reactions, J. Mol. Biol. 253 (1995) 771–786.
[32] M.C. Shastry, S.D. Luck, H. Roder, A continuous-flow capillary mixing method to monitor reactions on the microsecond time scale, Biophys. J. 74 (1998) 2714–2721.
[33] U. Mayor, N.R. Guydosh, C.M. Johnson, J.G. Grossmann, S. Sato, G.S. Jas, S.M.V. Freund, D.O.V. Alonso, V. Daggett, A.R. Fersht, The complete folding pathway of a protein from nanoseconds to microseconds, Nature 421 (2003) 863–867.
[34] N. Piskunov, Differential and Integral Calculus, Mir Publisher, 1974.
[35] Y. Bai, Kinetic evidence for an on-pathway intermediate in the folding of cytochrome c, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 477–480.
[36] C. Travaglini-Allocatelli, S. Gianni, V.K. Dubey, A. Borgia, A. Di Matteo, D. Bonivento, F. Cutruzzola, K.L. Bren, M. Brunori, An obligatory intermediate in the folding pathway of cytochrome c552 from *Hydrogenobacter thermophilus*, J. Biol. Chem. 280 (2005) 25729–25734.
[37] Y. Ivarsson, C. Travaglini-Allocatelli, P. Jemth, F. Malatesta, M. Brunori, S. Gianni, An on-pathway intermediate in the folding of a PDZ domain, J. Biol. Chem. 282 (2007) 8568–8572.
[38] K. Teilum, F.M. Poulsen, M. Akke, The inverted chevron plot measured by NMR relaxation reveals a native-like unfolding intermediate in acyl-CoA binding protein, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 6877–6882.
[39] J.N. Onuchic, Z. Luthey-Schulten, P.G. Wolynes, Theory of protein folding: the energy landscape perspective, Annu. Rev. Phys. Chem. 48 (1997) 545–600.

[40] F.X. Schmid, Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction via structural intermediates, Biochemistry 22 (1983) 4690–4696.
[41] T. Kiefhaber, Kinetic traps in lysozyme folding, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 9029–9033.
[42] G. Wildegger, T. Kiefhaber, Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate, J. Mol. Biol. 270 (1997) 294–304.
[43] S. Gianni, C. Travaglini-Allocatelli, F. Cutruzzola, M. Brunori, M.C. Shastry, H. Roder, Parallel pathways in cytochrome c(551) folding, J. Mol. Biol. 330 (2003) 1145–1152.
[44] A. Borgia, D. Bonivento, C. Travaglini-Allocatelli, A. Di Matteo, M. Brunori, Unveiling a hidden folding intermediate in C-type cytochromes by protein engineering, J. Biol. Chem. 281 (2006) 9331–9336.
[45] X. Salvatella, C.M. Dobson, A.R. Fersht, M. Vendruscolo, Determination of the folding transition states of barnase by using PhiI-value-restrained simulations validated by double mutant PhiIJ-values, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 12389–12394.
[46] A.P. Capaldi, C. Kleanthous, S.E. Radford, Im7 folding mechanism: misfolding on a path to the native state, Nat. Struct. Biol. 9 (2002) 209–216.
[47] M.A. de los Rios, B.K. Muralidhara, D. Wildes, T.R. Sosnick, S. Marqusee, P. Wittung-Stafshede, K.W. Plaxco, I. Ruczinski, On the precision of experimentally determined protein folding rates and phi-values, Protein Sci. 15 (2006) 553–563.
[48] A.R. Fersht, S. Sato, Phi-value analysis and the nature of protein-folding transition states, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 7976–7981.
[49] B.A. Krantz, R.S. Dothager, T.R. Sosnick, Discerning the structure and energy of multiple transition states in protein folding using psi-analysis, J. Mol. Biol. 337 (2004) 463–475.
[50] T.R. Sosnick, R.S. Dothager, B.A. Krantz, Differences in the folding transition state of ubiquitin indicated by phi and psi analyses, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 17377–17382.
[51] C. Bodenreider, T. Kiefhaber, Interpretation of protein folding psi values, J. Mol. Biol. 351 (2005) 393–401.
[52] A.R. Fersht, Phi value versus psi analysis, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 17327–17328.
[53] T.L. Religa, J.S. Markson, U. Mayor, S.M. Freund, A.R. Fersht, Solution structure of a protein denatured state and folding intermediate, Nature 437 (2005) 1053–1056.
[54] A.R. Fersht, A kinetically significant intermediate in the folding of barnase, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 14121–14126.
[55] A. Li, V. Daggett, Molecular dynamics simulation of the unfolding of barnase: characterization of the major intermediate, J. Mol. Biol. 274 (1998) 677–694.
[56] N. Ferguson, A.P. Capaldi, R. James, C. Kleanthous, S.E. Radford, Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9, J. Mol. Biol. 286 (1999) 1597–1608.
[57] N. Ferguson, W. Li, A.P. Capaldi, C. Kleanthous, S.E. Radford, Using chimeric immunity proteins to explore the energy landscape for alpha-helical protein folding, J. Mol. Biol. 307 (2001) 393–405.
[58] S.A. Gorski, A.P. Capaldi, C. Kleanthous, S.E. Radford, Acidic conditions stabilise intermediates populated during the folding of Im7 and Im9, J. Mol. Biol. 312 (2001) 849–863.
[59] C. Travaglini-Allocatelli, S. Gianni, M. Brunori, A common folding mechanism in the cytochrome c family, Trends Biochem. Sci. 29 (2004) 535–541.
[60] C.N. Chi, S. Gianni, N. Calosci, C. Travaglini-Allocatelli, Å. Engstrom, P. Jemth, A conserved folding mechanism for PDZ domains, FEBS Lett. 581 (2007) 1109–1113.