JPP JOURNAL OF PROTEINS AND PROTEOMICS
4(3), December 2013, pp. 165-174

Review Article

MORPHOGENESIS OF PROTEINS: CLASSIC VIEWS AND CHALLENGES AHEAD

Rajanish Giri[1], Carlo Travaglini-Allocatelli[1], Stefano Gianni[1,2], Giorgio Giardina[1], and Maurizio Brunori[1,*]

[1] *Istituto Pasteur-Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR Dipartimento di Scienze Biochimiche "A. Rossi Fanelli". Sapienza Università di Roma, P.le A. Moro 5, 00185, Rome, Italy*

[2] *Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom*

Abstract: Protein folding has stimulated since decades a large number of experimental and theoretical studies. Given the complexity of the problem, efforts have been generally directed towards the understanding of the fundamental rules governing the morphogenesis of a few paradigmatic globular proteins. A classic approach has been to tackle the folding mechanism of proteins belonging to the same "family" and thereby characterized by the same overall fold (and function) in spite of substantial differences in primary structure. Recently we have investigated the mirror situation i.e. that of "heteromorphic" proteins characterised by substantially different fold in spite of very similar or nearly identical amino acid sequence. In this paper we shall present and discuss some of the more stimulating aspects of the folding problem based on our results on small domain proteins investigated by transient kinetics, mutagenesis and MD simulations. Finally, we highlight some of the recent findings on the so called instrinsically unstructured proteins, unfolded polypeptides that acquire a three-dimensional structure upon binding to their target/ligand. These apparently numerous protein domains represent an interesting experimental system to explore the binding induced folding reaction and to investigate the role of disorder in cellular functions.

Keywords: Protein folding; mechanism; dynamics; denatured state; folding simulations

Introduction

A central concept in protein science, based on extensive theoretical and experimental work, implies that the "native" three dimensional (3D) structure of a globular protein is, by definition, the only functionally competent state. The famous experiment of Christian Anfinsen showed that in water, a fully denatured inactive polypeptide acquires spontaneously the native 3D structure and thereby its specific function (Anfinsen *et al.*, 1961). Thus it was concluded that in water the amino acid sequence contains all the information necessary to achieve such a complex improbable task, often very rapidly. This proved to be a fundamental progress in the life sciences, and stimulated a huge number of productive and conceptually original work to tackle what is generally called the *protein folding problem*.

It is nowadays agreed that small globular proteins (roughly around 100 amino acids) conform to a reversible folding and unfolding transition (Jackson, 1998). Thus the general rules of chemical equilibrium can be applied to assess the thermodynamic stability of a protein and quantitatively evaluate the effect of different denaturants. In order to obtain information on the pathway followed by the polypeptide in the transition from a high entropy disordered state to

Corresponding Author: **Maurizio Brunori**
E-mail: maurizio.brunori@uniroma1.it

Received: November 24, 2013
Accepted: December 19, 2013
Published: December 31, 2013

a low entropy native state, folding and unfolding events have to be characterized by rapid reaction techniques, given that generally folding speed ranges from seconds down to microseconds.

In this review we shall summarize some of the basic concepts and experimental approaches developed to unveil the mechanism whereby a functionally competent form is emerging from a myriad of disordered structures, i.e. the process of *morphogenesis of a protein* and the fundamental role of the primary structure in dictating the pathway for this complex event. Moreover we deal very briefly with some of the basic features of the coupling between binding and folding, a theme that has acquired centre stage after the discovery that a significant fraction of proteins is "natively unfolded" (Uversky and Dunker, 2010; Tompa, 2011). Given that in the cell an unfolded polypeptide is generally more vulnerable than the native state, it would be instructive to find out why a significant fraction of proteins or protein domains is natively unfolded.

Equilibrium and Kinetic Behaviour

Equilibrium experiments provide information about the reversible folding process and thereby allow estimating the thermodynamic stability of a protein and the cooperativity of the unfolding reaction. Generally, the protein is unfolded with a chemical denaturant and a spectroscopic signal, such as fluorescence emission or circular dichroism, is recorded as a function of denaturant concentration; typically equilibrium transitions display a sigmoidal shape (Figure 1). By applying simplifying assumptions on the effect of a denaturant (Tanford, 1964), protein stability under native conditions in water is obtained (Myers and Oas, 1999).

As shown if Figure 1, the transition from native to denatured (or unfolded) can be analyzed using the linear correlation between free energy and denaturant concentration (see inset panel) (Fersht, 1999). The thermodynamic stability, expressed by the change in free energy between the native (N) and the denatured (D) state ($\Delta G_{D\text{-}N}$), can be estimated by extrapolation. From the slope of the linear dependence of $\Delta G_{D\text{-}N}$ on denaturant concentration it is possible to obtain a parameter, called $m_{D\text{-}N}$ value, which is correlated

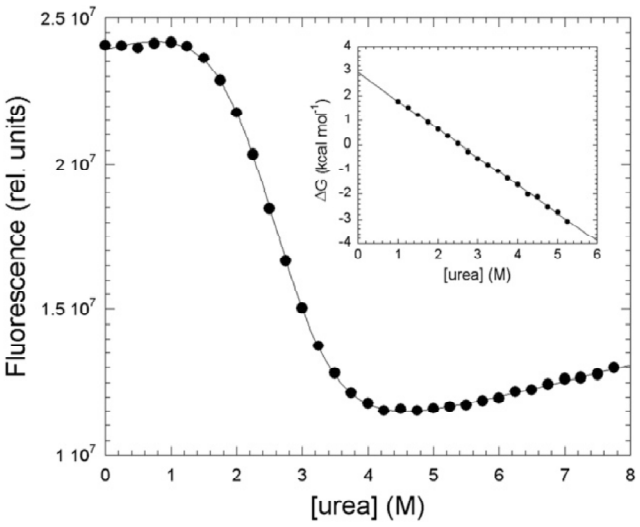

Figure 1: **Equilibrium denaturation profile. The unfolding of the second PDZ domain from PTP-BL (PDZ2) monitored by fluorescence is presented. It may be observed that at low denaturant concentrations the protein resists denaturation but at higher concentrations denaturation occurs *via* a concerted unfolding event. A quantitative analysis of the observed spectroscopic signal based on a two-state mechanism allows calculation of the free energy for unfolding at different denaturant concentrations (inset Panel). The slope of the observed linear dependence of the unfolding free energy, classically named as the $m_{D\text{-}N}$ value, which describes the co-operativity of the process, is correlated with the exposure of accessible surface area upon unfolding**

to the surface area exposed to solvent during unfolding (Tanford, 1964).

Some general conclusions emerge from this simple analysis. First of all the thermodynamic stability of proteins is generally quite small, ranging from a few to approx 10 kcal/mole; thus these macromolecules are marginally stable, stability being just sufficient for the job but not excessive as to hinder functionality at physiological temperatures. Second, the reversible transition is cooperative, mimicking in some cases a real "phase transition"; of course, cooperativity implies that intermediates are not populated to any significant extent along the transition and therefore the overall process conforms to a simple two-state model. Finally, the estimate of the buried surface area that becomes solvent exposed upon unfolding is an important parameter to define the progress of the reaction, and it is often reported as the β_T to honor Charles Tanford who contributed fundamental insight into the role of the hydrophobic effect in protein folding and stability (Tanford, 1968).

Kinetics is the ideal approach to provide, in principle, information about the folding pathway of a protein and the reaction dynamics. Given that small globular proteins fold and unfold quite rapidly (from sec to μsec), rapid reaction techniques have to be employed; most frequently either rapid (or super-rapid) mixing or temperature jump. Kinetic studies are based on imposing a rapid perturbation of the equilibrium to change the relative population of the relevant states; by rapid mixing this can be obtained by dilution of a denaturant to induce refolding. The perturbation is followed by a relaxation of the system to a new equilibrium, and the time course allows estimating the relaxation time of the process.

Folding/unfolding kinetics is often monitored by following the change in Trp fluorescence emission or far UV circular dichroism during the relaxation. Folding and unfolding rate constants at a variety of denaturant concentrations are determined by fitting the time course to exponential equations. In the case of two-state folding, only a single exponential folding phase is generally observed (Jackson, 1991), whereas multi-state systems may lead complex kinetics *via* population of intermediates that may be either on-pathway or off-pathway (Travaglini-Allocatelli *et al.*, 2003).

As stated above, if the folding reaction is consistent with a two-state model, only the denatured and native states are significantly populated:

$$D \underset{k_U}{\overset{k_F}{\rightleftharpoons}} N \qquad \text{(Scheme 1.1)}$$

When a perturbation is imposed on this system, the observed rate constant k_{obs} is contributed (as usual) by the forward and reverse rate constants, according to:

$$k_{obs} = k_F + k_U \qquad \text{(Eq. 1.1)}$$

where k_U and k_F represent the unfolding and refolding rate constants respectively. In a two-state reaction, the log of the observed rate constant is dependent on denaturant concentration following a characteristic V-shape with limiting values at low and high denaturant concentrations that are linear. Figure 2 depicts the

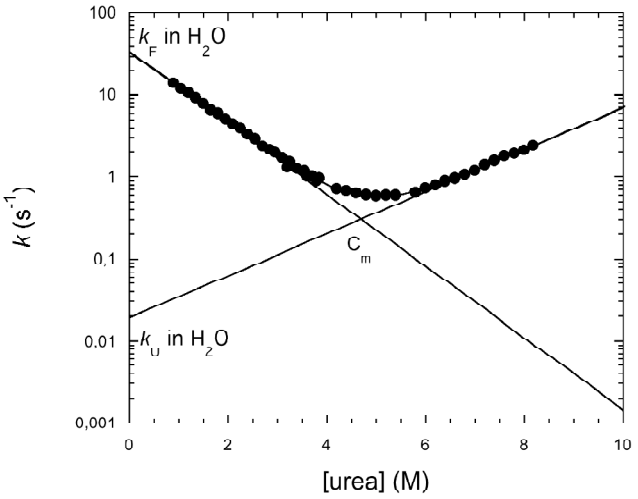

Figure 2: **Semilogarithmic plot of the observed rate constant as a function of denaturant concentration (chevron plot); the data refer to the chevron plot ot PDZ2 measured in the presence of 0.4 M sodium sulfate and 50 mM phosphate buffer pH 7.0. As described in the text, for a two-state system, the observed rate constant is the sum of the folding and unfolding microscopic rate constants, which are reported as lines. The extrapolated folding and unfolding rate constants in the absence of denaturant, as well as the midpoint of the unfolding transition (commonly called the C_m), are highlighted in the figure**

folding/unfolding rate constants observed for a two-state system. Because of its classical V-shaped appearance, this is called a "chevron" plot by the protein folding community. Chevron plot analysis allows obtaining by extrapolation the folding and unfolding rate constants k_F and k_U in the absence of denaturant (see legend to Fig. 2). The slopes m_F and m_U, calculated from the linear dependence on denaturant concentration, correlate with the change in accessible surface area between the two ground states and the transition state in between.

Although quite often small domain proteins follow a two state mechanism and thereby conform to a V-shaped chevron plot, this is by no means general (Brockwell and Radford, 2007; Capaldi *et al.*, 2002; Ferguson *et al.*, 1999; Jemth, 2004; Khorasanizadeh *et al.*, 1996; Parker *et al.*, 1995; Travaglini-Allocatelli *et al.*, 2003). Deviations from a simple behaviour are very informative since may reflect the population of productive folding intermediates, the accumulation of misfolded states or, sometimes, transient aggregation occurring upon dilution of the denaturant. Discrimination between these

possible mechanisms is a pre-requisite for any further analysis of the significance of intermediate species.

Folding Transition State(s): The Φ-value analysis

Protein folding might be described as a unimolecular chemical reaction, in which the "reactant" (an unfolded protein) is converted to a "product" (the folded state). Unimolecular chemical reactions imply the system going over a free-energy barrier *via* a transition state (TS). The observation that protein folding kinetics often conforms to a single exponential time course allows to postulate the presence of an energy barrier in between the two equilibrium ground states.

According to transition state theory, the progress of a reaction can be described by a reaction coordinate that may be any degree of freedom that connects reactant(s) and product(s) along the lowest free energy continuous path; the highest free energy point along the path being the transition state. Given the complex nature of protein folding (with many weak bonds involved), the transition state may correspond to a number of high free energy conformations; and therefore it is generally referred to as the transition state ensemble (Onuchic, et al. 1996; Onuchic and Wolynes, 2004).

To unveil the reaction pathway in the folding of a protein (no matter how simple!), knowledge of structural and energetic properties of the transition state would be very valuable. In principle we should acquire information nor only on the height of the barrier and its position along the reaction coordinate, but also on the three-dimensional structure of the transition state; a rather formidable goal. To unveil the structure of the folding transition state(s), a protein engineering approach, called Φ value analysis, was developed by Alan Fersht and collaborators (Fersht *et al.*, 1992; Matouschek, 1990) and validated by many groups working on protein folding (see for example (Chiti *et al.*, 1999; Fersht and Sato, 2004; Friel *et al.*, 2003; Gianni *et al.*, 2007; Itzhaki *et al.*, 1995; Lindberg and Oliveberg, 2007; Riddle *et al.*, 1999; Zarrine-Afsar *et al.*, 2010). By this approach the degree of structure formation of individual residues in the transition state is actually inferred from analysis of the effect of single-site mutations on folding rates and stability (see legend to Figure 3).

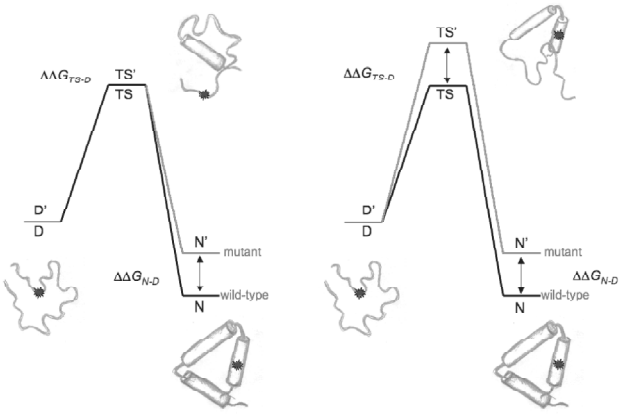

Figure 3: **Φ-value analysis. The depicted energy profiles represent two different scenarios in which a hypothetical probed residue (shown in red on the structures) gives rise to different effects in the folding activation energy and in the stability of the protein, depending on the structure of the transition state. Left panel: the site of mutation is in an unstructured region of the transition state (Φ = 0). Right panel: the probed residue is highly native-like in the transition state (Φ = 1)**

The Φ value is calculated as the ratio of the energetic perturbation induced on the transition state versus that induced in the native folded state, introducing a non-disruptive mutation intended to cause a small perturbation:

$$\Phi = \frac{\Delta\Delta G^{*}_{TS-D}}{\Delta\Delta G_{N-D}} \qquad \text{(Eq. 1.3)}$$

where ΔG_{N-D} is the change induced by a particular mutation on the energetics of the folded state, and ΔG^{*}_{TS-D} that induced on the activation energy for folding. Φ values are expected to range between 0 and 1. When Φ is near unity, it is assumed that the mutation is perturbing the TS as much as the native state; and thus in the TS the mutated residue is engaged in (quasi) native contacts. On the other hand, a Φ-value near zero is taken as evidence that the TS is not energetically perturbed by the mutation, while the native state is (i.e. the mutated residue is as unstructured in the TS as in the denatured ensemble).

Indeed, experimentally determined Φ values are generally in between 0 and 1. The traditional

interpretation of fractional Φ values is, however, not straightforward as they might indicate the existence of multiple folding pathways or a unique transition state ensemble with genuinely weakened interactions (Oliveberg and Fersht, 1996). Also the interpretation of the so-called non-classical or unusual Φ-values (Φ>1 and Φ<0), that are seldom observed, is not straightforward. In some cases this anomaly may be due to an incorrect mutation, while in others they appear genuine. Negative Φ-values can be observed when the native state is destabilized while the transition state is stabilized and vice-versa. Φ-values higher than 1 may be detected when the transition state is affected by the mutation more than the native state. Because unusual Φ-values are often indicative of non-native contact formation (Ozkan *et al.*, 2001), they can be used to detect local misfolding "spots" in transient intermediates either *en-route* to the productive folding pathway, as in the case of the Im7 protein (Capaldi *et al.*, 2002), or off-pathway kinetic traps, as in the case of the circularly permuted PDZ domain of D1p protein (Gianni *et al.*, 2010).

The impact of the Φ-value analysis made a significant step forward with the introduction of Restrained Molecular Dynamics (MD) simulations (Vendruscolo *et al.*, 2001), that helped in identifying or predicting the 3D structure of transition and/or intermediate states along the folding pathway. Conceptually the method introduces into the calculation of a folding trajectory for a given protein, a series of restrains represented by the experimentally determined Φ-values. This leads to a dramatic drop in the number of degrees of freedom and allows computing structures of the folding TS.

In the case of GB1 (Morrone *et al.*, 2011a), protein MD simulations have endorsed the presence of an intermediate identified by kinetic experiments and Φ-value analysis. To identify the transition and intermediate states in GB1, a multidimensional property space derived from 15 physical properties of the protein was prepared, and then a one-dimensional reaction coordinate based on these 15 properties was calculated (Toofanny *et al.*, 2010). The mean distance in property space was calculated for each time point in a simulation of interest to the native state reference, which contained all the structures of the native state simulation excluding the first nsec. The 1D reaction coordinate was created from a histogram of the mean distance to reference for all structures. To compare with experimental Tanford β-values, the average buried surface area for the TS ensemble was compared to that calculated for the native state. The interesting result of this combined approach will be illustrated below when dealing with the folding mechanism of the heteromorphic pair.

A Novel Perspective: Comparing Heteromorphic Proteins

As reported above, our understanding of the protein folding reaction has greatly advanced by extensively characterizing the folding mechanisms of simple protein systems. The classical experimental strategy to elucidate the relationships between sequence information and folding mechanism, is the analysis of the folding kinetics of homologous proteins (Nickson *et al.*, 2008). By comparing the folding processes for different members of a given protein family, such a strategy assumes that general correlations between primary structure and folding pathway(s) may be extrapolated. The results obtained by many groups on different protein families highlighted that the overall folding mechanism is generally conserved within a given fold-family, and hidden common features may be unveiled even when apparently different folding mechanisms were observed (Clarke *et al.*, 1999; Ferguson *et al.*, 1999; Gianni *et al.*, 2001; Martínez and Serrano, 1999). The results obtained by our group in the case of the cytochrome c fold family are paradigmatic. By studying the folding mechanism of different prokaryotic c-type cytochromes, and by comparing our results with those obtained by others on eukaryotic cytochromes c, we were able to show that members of this family, despite differences in physico-chemical properties and thermodynamic stabilities, fold *via* a consensus mechanism (Travaglini-Allocatelli, 2004). Such a mechanism involves an obligatory intermediate that may represent either a low- or a high-energy state, but in all cases is characterized by conserved structural properties. These observations suggest that protein topology (i.e. the architectural

organization of the elements of secondary structure in the native state framework) plays a crucial role in determining folding pathways; according to this interpretation, much experimental work is still needed to address the relationships between sequence composition and protein topology.

In the last few years, a growing number of naturally evolved protein sequences characterized by the surprising ability to fold into different 3D structures have been identified. These heteromorphic proteins now offer an opportunity to study the protein folding problem from an innovative and complementary perspective, whereby the "signal" (the protein topology) is not disturbed by the "noise" (the sequence variations among protein homologues).

Taking advantage of the availability of a set of artificially evolved protein pairs with a high degree of sequence identity but different folds (Alexander *et al.*, 2007; Alexander *et al.*, 2009), we carried out a research project on the folding of the heteromorphic G_A and G_B protein pairs (sharing from 30% to an extraordinary 98% sequence identity but completely different folds). We expected to identify hot spot residues for the commitment of either one of the two different topologies and to unveil hypothetical relationships (if any) between protein topology and folding mechanism. By using a combination of biophysical and computational techniques, we have shown that, contrary to expectation, G_B88, characterized by a native a+b fold (see Figure 4), in the denatured state displays a content of native-like helical structure greater than that of G_A88, which is all a in its native state (Morrone et al., 2011b). Thus, despite the high sequence identity, the folding pathways for these two proteins appear to diverge as early as in the denatured state. These results suggested that native state topology is committed very early along the folding pathway, as it is per-sculped in the flickering residual structure of the denatured state. Although these results may seem at odd with those obtained on the PDZ family where a strong native bias is seen only at the late stages of folding (Calosci *et al.*, 2008), we have proposed that these apparently contrasting results can be reconciled by taking into account a major

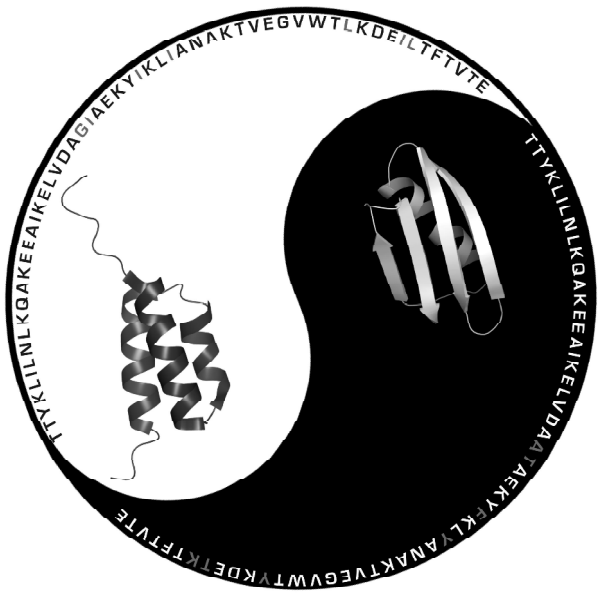

Figure 4: **Pictorial description of heteromorphic pairs. One of the most popular and recognizable images in the world, the yin and yang symbols represent two fundamental forces of nature, which complement yet oppose each other. By simplifying these relationships, ancient Chinese people tried to explain complicated phenomena in the universe; likewise the design of two proteins, sharing 88% sequence identity (49 out of 56 amino acids), offers a unique opportunity of studying the complex problem of protein folding in a simplified way. The G_A88 structure is represented in black on the white background, whereas the G_B88 is represented in white on the black background. The primary structures of the two proteins are also reported; highlighted in red are the residues that are different in G_A and G_B**

difference between the two protein systems. In particular, our results suggested that when alternative nucleation motifs, are present (where a nucleation motif is defined as a cluster of residues displaying high Φ-values), the native state fold is committed only at the latest stages of folding, such as in the case of the PDZ family. On the contrary, when such alternative nuclei are absent, the coarse topology of the native state, being pre-sculpted in the denatured state, is committed in the earliest stages of folding.

In an effort to obtain a quasi-atomic description of the transition state(s) for folding, we have carried out an extensive Φ-value analysis of several heteromorphic pairs (i.e. G_A 30/ G_B 30, G_A 77/ G_B 77, G_A 88/ G_B 88), characterizing the thermodynamic and kinetic parameters of nearly 150 site-directed mutants (Giri *et al.*, 2012). Overall, our results indicate that, despite the very high level of sequence identity, G_A and G_B fold

via completely independent pathways, quantitatively described by a simple two-state model for the G_A variants and by a three-state model for the G_B variants. Moreover, by analyzing the structural distribution of measured Φ-values (Figure 5), we were able to show that in the case of G_B the folding process is quite dependent on sequence composition, as alternative folding nuclei may be selectively (de)stabilized; on the contrary, in the case of G_A the mechanism appears to resist to differences in primary structure.

This finding is of general significance, as it has been anticipated that the number of accessible pathways for folding depends on the presence of different nucleation motifs contained within a given native topology (Haglund *et al.*, 2008). Overall, these results indicate that the malleability of folding pathways is determined by the occurrence of multiple nuclei; eventually, the stabilization of such nuclei, experimentally mirrored by the stabilization of folding intermediates, leads to multi-state folding whereas the presence of a unique diffused nucleus is featured by two-state systems, generally characterized by robust folding pathway.

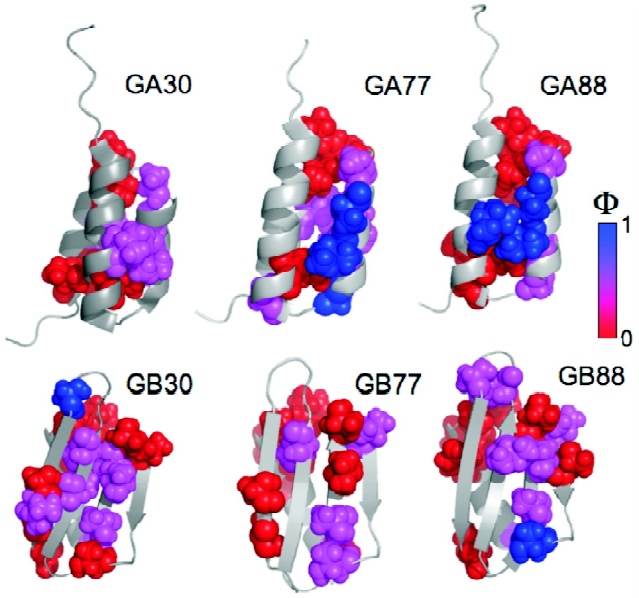

Figure 5: **Structural distribution of the measured Φ-values for the heteromorphic pairs. The experimentally determined Φ-values were divided into three categories and reported on the structure of the proteins using the following colour code: red, 0<Φ<0.30; magenta, 0.30<Φ<0.70; blue, 0.70<Φ<1. Whilst a conserved nucleus between the helices α1 and α2 is clearly evident in all the G_A variants, in the case of the G_B variants there is a clear change in pathway when the sequence composition is altered from the G_B30 (with high Φ-values in the second β hairpin) to G_B88 (with high Φ-values in the first β hairpin)**

Intrinsically Denatured Proteins

The explosion of the protein data bank (PDB), as well as the increase of power in computers, has suggested that employing an approach based on bioinformatics, rather than using *in vitro* studies on the thermodynamics of folding and unfolding, may provide novel complementary information. At the same time, the interest of the protein folding community focussed on the study of the mechanisms of protein folding, with particular attention for misfolding events in view of their significance for the pathogenic events leading to neurodegenerative diseases. The bioinformatics approach, coupled with the ever increasing availability of 3D structures, led to a new perspective once it was discovered that several proteins or protein domains perform their biological function while lacking a well-defined 3D structure (Dunker *et al.*, 2008; Dyson and Wright, 2002; Tompa, 2005, 2011; Tompa and Fuxreiter, 2008; Uversky and Dunker, 2010). These proteins, nowadays called Intrinsically Denatured Proteins (IDPs), generally undergo a structural transition to a folded state upon recognition and binding to their physiological partner. It is not surprising that the amino acid composition of IDPs has peculiar features, such as a lower content of hydrophobic amino acid residues as well as a higher content of charged residues, which allows to predict disorder from amino acid sequence.

The scientific community has been interested in attempting to reconcile the unexpectedly high fraction of IDPs with the structure-function dogma, posing some simple yet unanswered questions (Dunker *et al.*, 2008; Dyson and Wright, 2002; Tompa, 2005, 2011; Tompa and Fuxreiter, 2008; Uversky and Dunker, 2010). Is there any potential value for a protein to display a flexible structure, sometimes so extreme to compromise the acquisition of the native state? Is the expected increase in the *repertoire* of states characteristic for IDPs associated with some peculiar "survival value", given that in the cell disorder is usually associated to higher vulnerability?

Several (untested) hypotheses have been put forward to rationalize why so many proteins seem

to be intrinsically unstructured, such as: (i) being a mechanism to decouple affinity and specificity (Spolar and Record, 1994); (ii) increasing the association rate with the target/ligand (Shoemaker *et al.*, 2000); (iii) providing an increased plasticity with regard to the target/ligand (Wells *et al.*, 2008); (iv) having a more rapid turnover in the cell (Uversky and Dunker, 2010). The first four possibilities are of high interest from a biophysical perspective and they could be tackled *via* experimental and theoretical approaches. In fact, it has been suggested that a potential value for a protein to be disordered would arise from an increased capture radius (Shoemaker *et al.*, 2000). This fascinating hypothesis implies that natively-unfolded proteins should bind their partner according to an "induced-fit" model (i.e. binding precedes folding), and that the association rate constants should be inversely related to protein stability.

A valuable system to test the fly-casting hypothesis as summarized below has been recently described (Gianni *et al.*, 2012). The CREB-binding protein (CBP) is a co-activator that modulates the interaction between DNA-bound activator proteins and the components of the basal transcription complex; a globular domain of CBP, called KIX, is the principal mediator of such interactions (Radhakrishnan *et al.*, 1997; Zor *et al.*, 2004). Despite its small size (87 amino acids) and a relatively simple fold, the KIX domain binds different IDP systems *via* two distinct, but energetically connected, binding sites, known as c-Myb and MLL sites (named after two characteristic ligands of each site, i.e. the transactivation domain of the protein c-Myb and the mixed lineage leukaemia MLL protein) (Radhakrishnan *et al.*, 1997). In a recent study (Gianni *et al.*, 2012), we tested the applicability of the fly-casting mechanism proposed by Wolynes and coworkers (Shoemaker *et al.*, 2000) to c-Myb, a prototype of an IDP, which assumes helical structure upon binding to KIX (Dyson and Wrigth, 2002). A kinetic analysis of the recognition mechanism by which the KIX domain binds c-Myb, based on temperature jump and stopped flow experiments in the presence and absence of 1,1,1-tri-fluoro-ethanol (Gianni *et al.*, 2012), revealed that this IDP recognizes KIX by following a folding-after-binding scenario; accordingly, it binds to its target in a relatively unstructured conformation, with the locking of the hydrogen bonds of the helical structure occurring only downhill the primary rate limiting step. Surprisingly, however, the transition state for the folding-after-binding reaction revealed that c-Myb is well ordered in the rate limiting barrier and indicated that the interaction between KIX and c-Myb is characterized by a considerable geometrical fit (Giri *et al.*, 2013). On the basis of these findings, and by considering our earlier observation that the association rate constant is insensitive to the stability of c-Myb, we suggested that in this case disorder by itself did not accelerate the recognition events between the IDP and its partner and therefore, the potential advantages in protein disorder were not related to the kinetics of initial recognition. Thus, given that in the cellular system disordered polypeptides may potentially be more vulnerable than compact domains, the selective advantage in being natively unfolded remains an interesting *conundrum* for the future that demands additional work.

Acknowledgements

Work partly supported by a grant from the Italian Ministero dell'Istruzione dell'Universita' e della Ricerca (Progetto di Interesse 'Invecchiamento' to S.G.).

Abbreviations

MD, molecular dynamics; PDZ, postsynaptic density protein (PSD-95), discs-large tumor suppressor (dlg) and zonula occludens protein 1 (ZO-1); IDP, Intrinsically Denatured Proteins; CBP, CREB Binding Protein

References

Alexander, P.A., He, Y., Chen, Y., Orban, J., and Bryan, P.N. (2007). The design and characterization of two proteins with 88% sequence identity but different structure and function. Proc Natl Acad Sci U S A *104*, 11963-11968.

Alexander, P.A., He, Y., Chen, Y., Orban, J., and Bryan, P.N. (2009). A minimal sequence code for switching protein structure and function. Proc Natl Acad Sci U S A *106*, 21149-21154.

Anfinsen, C.B., Haber, E., Sela, M., and White, F.H., Jr. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci U S A *47*, 1309-1314.

Brockwell, D.J., and Radford, S.E. (2007). Intermediates: ubiquitous species on folding energy landscapes? Curr Opin Struct Biol *17*, 30-37.

Calosci, N., Chi, C.N., Richter, B., Camilloni, C., Engstrom, A., Eklund, L., Travaglini-Allocatelli, C., Gianni, S., Vendruscolo, M., and Jemth, P. (2008). Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. Proc Natl Acad Sci U S A *105*, 19241-19246.

Capaldi, A.P., Kleanthous, C., and Radford, S.E. (2002). Im7 folding mechanism: misfolding on a path to the native state. Nat Struct Biol *9*, 209-216.

Chiti, F., Taddei, N., White, P.M., Bucciantini, M., Magherini, F., Stefani, M., and Dobson, C.M. (1999). Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. Nat Struct Biol *6*, 1005-1009.

Clarke, J., Cota, E., Fowler, S.B., and Hamill, S.J. (1999). Folding studies of Ig-like beta-sandwich proteins suggest they share a common folding pathway. Structure *7*, 1145-1153.

Dunker, A.K., Silman, I., Uversky, V.N., and Sussman, J.L. (2008). Function and structure of inherently disordered proteins. Curr Opin Struct Biol *18*, 756-764.

Dyson, H.J., and Wright, P.E. (2002). Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol *12*, 54-60.

Ferguson, N., Capaldi, A.P., James, R., Kleanthous, C., and Radford, S.E. (1999). Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. J Mol Biol *286*, 1597-1608.

Fersht, A. (1999). Structure and Mechanism in Protein Science. Freeman WH and Co New York.

Fersht, A.R., Matouschek, A., and Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol *224*, 771-782.

Fersht, A.R., and Sato, S. (2004). Phi-value analysis and the nature of protein-folding transition states. Proc Natl Acad Sci U S A *101*, 7976-7981.

Friel, C.T., Capaldi, A.P., and Radford, S.E. (2003). Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: similarities and differences in the folding of homologous proteins. J Mol Biol, 326 293-305.

Gianni, S., Geierhaas, C.D., Calosci, N., Jemth, P., Vuister, G.W., Travaglini-Allocatelli, C., Vendruscolo, M., and Brunori, M. (2007). A PDZ domain recapitulates a unifying mechanism for protein folding. Proc Natl Acad Sci U S A *104*, 128-133.

Gianni, S., Ivarsson, Y., De Simone, A., Travaglini-Allocatelli, C., Brunori, M., and Vendruscolo, M. (2010). Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. Nat Struct Mol Biol *17*, 1431-1437.

Gianni, S., Morrone, A., Giri, R., and Brunori, M. (2012). A folding-after-binding mechanism describes the recognition between the transactivation domain of c-Myb and the KIX domain of the CREB-binding protein. Biochem Biophys Res Comm *428*, 205-209.

Gianni, S., Travaglini-Allocatelli, C., Cutruzzolà, F., Bigotti, M.G., and Brunori, M. (2001). Snapshots of protein folding. A study on the multiple transition state pathway of cytochrome c(551) from Pseudomonas aeruginosa. J Mol Biol *309*, 1177-1187.

Giri, R., Morrone, A., Toto, A., Brunori, M., and Gianni, S. (2013). Structure of the transition state for the binding of c-Myb and KIX highlights an unexpected order for a disordered system. Proc Natl Acad Sci U S A *110*, 14942-14947.

Giri, R., Morrone, A., Travaglini-Allocatelli, C., Jemth, P., Brunori, M., and Gianni, S. (2012). Folding pathways of proteins with increasing degree of sequence identities but different structure and function. Proc Natl Acad Sci U S A *109*, 17772-17776.

Haglund, E., Lindberg, M.O., and Oliveberg, M. (2008). Changes of protein folding pathways by circular permutation. Overlapping nuclei promote global cooperativity. J Biol Chem *283*, 27904-27915.

Itzhaki, L.S., Otzen, D.E., and Fersht, A.R. (1995). 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*, 260-288.

Jackson, S.E. (1998). How do small single-domain proteins fold? Fold Des *3*, R81-91.

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

Jemth, P., Gianni, S., Day, R., Li, B., Johnson, C.M., Daggett, V., Fersht, A.R. (2004). 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*, 6450-6455.

Khorasanizadeh, S., Peters, I.D., and Roder, H. (1996). Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. Nat Struct Biol *3*, 193-205.

Lindberg, M.O., and Oliveberg, M. (2007). Malleability of protein folding pathways: a simple reason for complex behaviour. Curr Opin Struct Biol *17*, 21-29.

Martínez, J.C., and Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. Nat Struct Biol *6*, 1010-1016.

Matouschek, A., Kellis, J.T. Jr, Serrano, L., Bycroft, M., Fersht, A.R. (1990). Transient folding intermediates characterized by protein engineering. Nature *346*, 440-445.

Morrone, A., Giri, R., Toofanny, R.D., Travaglini-Allocatelli, C. Brunori, M. Daggett, and V. Gianni, S. (2011a). Gb1 Is not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate. Biophys J *101*, 1-8.

Morrone, A., McCully, M.E., Bryan, P.N., Brunori, M., Daggett, V., Gianni, S., and Travaglini-Allocatelli, C.

(2011b). The denatured state dictates the topology of two proteins with almost identical sequence but different native structure and function. J Biol Chem *286*, 3863-3872.

Myers, J.K., and Oas, T.G. (1999). Contribution of a buried hydrogen bond to lambda repressor folding kinetics. Biochemistry *38*, 6761-6768.

Nickson, A.A., Stoll, K.E., and Clarke, J. (2008). Folding of a LysM domain: entropy-enthalpy compensation in the transition state of an ideal two-state folder. J Mol Biol *380*, 557-569.

Oliveberg, M., and Fersht, A.R. (1996). Formation of electrostatic interactions on the protein-folding pathway. Biochemistry *35*, 2726-2737.

Ozkan, S.B., Bahar, I., and Dill, K.A. (2001). Transition states and the meaning of Phi-values in protein folding kinetics. Nat Struct Biol *8*, 765-769.

Parker, M.J., Spencer, J., and Clarke, A.R. (1995). An integrated kinetic analysis of intermediates and transition states in protein folding reactions. J Mol Biol *253*, 771-786.

Radhakrishnan, I., Pérez-Alvarado, G.C., Parker, D., Dyson, H.J., Montminy, M.R., and Wright, P.E. (1997). Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. Cell *91*, 741-752.

Riddle, D.S., Grantcharova, V.P., Santiago, J.V., Alm, E., Ruczinski, I., and Baker, D. (1999). Experiment and theory highlight role of native state topology in SH3 folding. Nat Struct Biol *6*, 1016-1024.

Shoemaker, B.A., Portman, J.J., and Wolynes, P.G. (2000). Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci USA *97*, 8868-8873.

Spolar, R.S., and Record, M.T.J. (1994). Coupling of local folding to site-specific binding of proteins to DNA. Science *263*, 777-784.

Tanford, C. (1964). Isothermal unfolding of globular proteins in aqueous urea solutions. . J Am Chem Soc *86*, 2050-2059.

Tanford, C. (1968). Protein denaturation. Adv Protein Chem *23*, 121-282.

Tompa, P. (2005). The interplay between structure and function in intrinsically unstructured proteins. FEBS Letters *579*, 3346-3354.

Tompa, P. (2011). Unstructural biology coming of age. Curr Opin Struct Biol *21*, 419-425.

Tompa, P., and Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Bioch Sci *33*, 2-8.

Toofanny, R.D., Jonsson, A.L., and Daggett, V. (2010). A Comprehensive Multidimensional-Embedded, One-Dimensional Reaction Coordinate for Protein Unfolding/Folding. . Biophys J *98*, 2671-2681.

Travaglini-Allocatelli, C., Gianni, S., Morea, V., Tramontano, A., Soulimane, T., and Brunori, M. (2003). Exploring the cytochrome c folding mechanism: cytochrome c552 from thermus thermophilus folds through an on-pathway intermediate. J Biol Chem *278*, 41136-41140.

Travaglini-Allocatelli, C., Gianni, S., Brunori, M. (2004). A common folding mechanism in the cytochrome c family. Trends Biochem Sci *29*, 535-541.

Uversky, V.N., and Dunker, A.K. (2010). Understanding protein non-folding. Biochim Biophys Acta *1804*, 1231-1264.

Vendruscolo, M., Paci, E., Dobson, C.M., and Karplus, M. (2001). Three key residues form a critical contact network in a protein folding transition state. Nature *409*, 641-645.

Wells, M., Tidow, H., Rutherford, T.J., Markwick, P., Jensen, M.R., Mylonas, E., Svergun, D.I., Blackledge, M., and Fersht, A.R. (2008). Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci U S A *105*, 5762-5767.

Zarrine-Afsar, A., Lin, S.L., and Neudecker, P. (2010). Mutational investigation of protein folding transition states by Phi-value analysis and beyond: lessons from SH3 domain folding. Biochem Cell Biol *88*, 231-238.

Zor, T., De Guzman, R.N., Dyson, H.J., and Wright, P.E. (2004). Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. J Mol Biol *337*, 521-534.