Protein Folding

Deutsche Ausgabe: DOI: 10.1002/ange.201504835
Internationale Ausgabe: DOI: 10.1002/anie.201504835

Frustration Sculpts the Early Stages of Protein Folding**

*Eva Di Silvio, Maurizio Brunori, and Stefano Gianni**

Abstract: *The funneled energy landscape theory implies that protein structures are minimally frustrated. Yet, because of the divergent demands between folding and function, regions of frustrated patterns are present at the active site of proteins. To understand the effects of such local frustration in dictating the energy landscape of proteins, here we compare the folding mechanisms of the two alternative spliced forms of a PDZ domain (PDZ2 and PDZ2as) that share a nearly identical sequence and structure, while displaying different frustration patterns. The analysis, based on the kinetic characterization of a large number of site-directed mutants, reveals that although the late stages for folding are very robust and biased by native topology, the early stages are more malleable and dominated by local frustration. The results are briefly discussed in the context of the energy-landscape theory.*

The Levinthal paradox highlights how folding of a protein to the native state cannot occur through a random search between all possible states.[1] One of the intriguing approaches to solve the paradox is to assume that proteins are minimally frustrated, implying a funneled energy landscape.[2] But, proteins are optimized not only to fold but also to function. Therefore, because the evolutionary constrains that select for function contrast with those optimizing folding, it is observed that local frustration of non-optimized patterns is overlapped with the active site.[3] The effects of such frustration, which is a signature of the conflicting demands between folding and function, are still poorly understood and demand experimental characterization.

The second PDZ domain from protein tyrosine phosphatase-bas like (PDZ2) is a globular protein–protein recognition domain, containing two α helices and six β strands. The binding pocket of PDZ2 is located, like in all PDZ domains, between helix α2 and strand β2, but its function is also regulated by a sparse energetic network within the domain.[4] An alternative spliced isoform of this PDZ domain (PDZ2as), displays an insertion of five residues (VLFDK) at the beginning of the β2–β3 loop. Such an insertion effectively abrogates binding, due to a small but detectable conformational change observed at the level of the binding site.[5] By comparing the three-dimensional structures of PDZ2 and PDZ2as by using a methodology aimed at identifying frustration in protein structures,[6] we observed a change in the frustration pattern, which appears consistent with the perturbed binding properties of PDZ2as (Figure 1). In fact, as expected from an allosterically regulated protein,[3] PDZ2 displays signatures of local frustration in different parts of its structure. On the other hand, such frustration is lower in PDZ2as, a feature that is consistent with its lower binding capability.

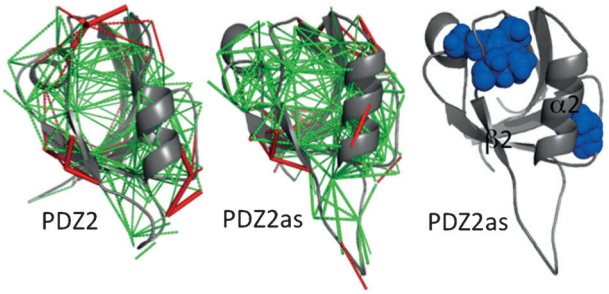

Figure 1. Frustration and structure of PDZ2 and PDZ2as. The frustration profiles, reported for both PDZ2 and PDZ2as were calculated using the frustratometer server.[6] Frustrated patterns are highlighted in red, whereas the minimally frustrated interactions are reported in green. The right panel highlights in blue spheres the positions of the two domains that are different in structure in the case of TS1. It is evident that these positions map on the structural regions (up-left and down-right) that are frustrated in PDZ2, although being minimally frustrated in PDZ2as.

The PDZ domain family is an ideal candidate for protein folding studies. In fact, because their folding pathways involve the presence of two sequential transition states, the early and late events are both experimentally accessible.[7] To obtain a glimpse on how frustration affects the different stages of the folding reaction, we report in this communication the comparison between the folding pathways of PDZ2 and PDZ2as by the so-called Φ-value analysis, a powerful experimental method to obtain structural information about transition states.

Because of their nature, transition states never accumulate; thus, their structure can be inferred only indirectly. By systematically mutating amino acid side chains and probing the effect of this perturbation on the activation and ground states' free energies, it is possible to map interaction pattern(s) in the transition state and to unveil their structure.[8] In practice, the analysis demands the determination of the so-called Φ value, which is formalized by normalizing the changes in activation free energy to the changes in native

[*] Dr. E. Di Silvio, Prof. M. Brunori, Prof. S. Gianni
Istituto Pasteur—Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli" Sapienza, University of Rome
00185 Rome (Italy)
E-mail: stefano.gianni@uniroma1.it
Prof. S. Gianni
Department of Chemistry, University of Cambridge
Cambridge CB2 1EW (UK)

[**] This work was partly supported by grants from the Italian Ministero dell'Istruzione dell'Universitáe della Ricerca (Progetto di Interesse "Invecchiamento" to S.G.).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201504835.

stability upon each specific non-disruptive mutation. If some critical experimental caveats are taken into account,[9] the Φ value returns a robust index that reports on the extent of native-like structure in the transition state, approaching 1 and 0 for native- and denatured-like structures, respectively.

Twenty-five conservative mutants of PDZ2as were produced, expressed, purified, and characterized (Table S1). Five of these mutants expressed poorly and could not be analyzed (V65A, L66A, V68A, L94A, L96A); thus, we performed urea-induced kinetic denaturation experiments on the remaining 20 mutants. The semi-logarithmic plots of the observed unfolding and refolding rate constant versus denaturant concentration (chevron plot) are shown in Figure S1 (see the Supporting Information). In analogy with previous work on PDZ domains,[7] analysis of the observed rate constants revealed the presence of two successive transition states in the folding reaction, namely a denatured like TS1 and a native like TS2. Thus, the chevron plots for all mutants (Figure S1) were fitted globally to a three-state model with shared overall structural features of the transition states in terms of their accessible surface area (see the Experimental Section). We were able to calculate reliable values of Φ for 13 mutants, whereas the remaining 7 mutants displayed a thermodynamic stability too similar to wild-type PDZ2as, preventing an accurate calculation of their Φ values. In summary, two sets of Φ values were obtained for all 13 mutants, reflecting the index of native-like structure for the early transition state TS1 and the late transition state TS2 (Table S1).

It is of interest to compare the structural features of TS1 and TS2 obtained for PDZ2as with those previously defined for PDZ2, in the light of their differences in frustration patterns. In fact, whereas the two proteins share a nearly identical structure and amino acid sequence, a Φ versus Φ plot for TS1 and TS2 (Figure 2) highlights that although the late stages for folding are extremely robust, there is a weaker native bias at the early stages, with TS1 being much more malleable in structure than TS2. In particular, mutations at positions L18A, L25A, A52G, and L73A, although displaying a similar behavior in the two proteins in determining the structure of TS2, represent clear outliers in the case of TS1. These positions, which are highlighted as spheres in Figure 1, are located close to the active site of the domain, in the regions that display a clearly different frustration pattern in PDZ2 compared to PDZ2as. On the basis of these observations we conclude that, in the case of this PDZ domain, local frustration sculpts the early stages of folding, whereas it has little effects on the late stages of the reaction.

The principle of minimal frustration and the funneled energy landscape represent the milestones of protein-folding theory since decades.[2,10] Nevertheless, the effect of local frustration on the folding mechanisms of proteins has been relatively poorly addressed experimentally and still represents an open challenge. The comparison of the folding pathways of two alternative spliced forms of a PDZ domain, which display a nearly identical sequence and structure but a different frustration pattern, provides important insight on how frustration determines the folding mechanism. In fact, we show that, in agreement with what is predicted by the

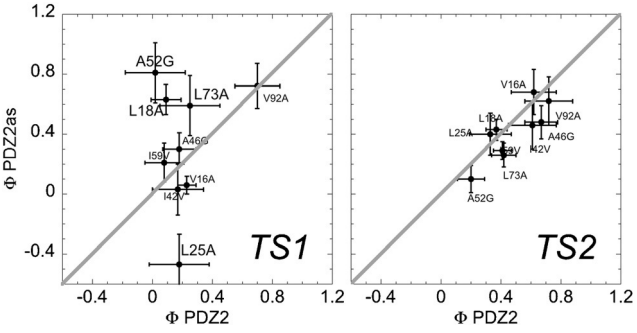

Figure 2. Comparison of the folding mechanisms of PDZ2 and PDZ2as. The calculated Φ values obtained for the two proteins for the denatured-like TS1 (left panel) and native-like TS2 (right panel) are reported with their respective errors. A line passing through the origin with slope one has been drawn to guide the eye. Whilst a linear fit of the data for TS2 returns a good correlation with an *r* value of 0.86, data for TS1 display no correlation ($r = 0.18$). As described in the text, by considering that the two proteins display a nearly identical sequence and structure, in the light of the structural localization of the clear outliers in TS1 (i.e., mutations L18A, L25A, A52G, and L73A) we explain these effects by comparing the different frustration patterns of the domains.

funneled energy landscape theory, frustration is critical in shaping the early stages of folding, which are more malleable and heterogeneous, whereas the late stages appear more robust and biased by native topology. In this perspective, the results presented in this work represent an experimental validation of the funneled energy landscape in protein folding.

Experimental Section

Site-Directed Mutagenesis and Protein Expression and Purification: PDZ2as was expressed and purified as described.[11] Site-directed mutants were produced by using a quick-change site-directed mutagenesis kit (Stratagene), according to the manufacturer's Instructions and purified as PDZ2as. Site-directed variants were confirmed by DNA sequencing.

Stopped-flow measurements: Single mixing kinetic folding experiments were carried out on an SX-18 stopped-flow instrument (Applied Photophysics, Leatherhead, UK); the excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter. The experiments were performed at 25 °C and all experiments were carried out in the presence of 50 mM sodium phosphate pH 7. In all experiments, refolding and unfolding were initiated by an 11-fold dilution of the denatured or the native protein, in the appropriate buffer, with a typically final protein concentration of 5 μM.

Kinetic data analysis: The folding chevron plots of PDZ2as and its site-directed mutants were fitted to a model implying a high-energy intermediate, intervening between two consecutive transition states, namely a more denatured like TS1 and a more native-like TS2, as previously described.[7] The presence of such intermediate reflects a deviation from linearity in the unfolding arm of the chevron plot and data was analyzed according to a three-state model, with the observed rate constants equal to:

$$k_{obs} = k_F + k_{U2}/(1 + K_{part}) \quad (1)$$

in which k_F is the folding-rate constant, k_{U2} is the unfolding-rate constant from the native to the second transition state TS2, and K_{part} is the partition factor between the two transition states. For the early

transition state, ϕ values were calculated as

$$\phi_{TS1} = \Delta\Delta G_{D-TS1}/\Delta\Delta G_{D-N} \quad (2)$$

where $\Delta\Delta G_{D-TS1}$ is the difference in the free energy between the denatured state and TS1 upon mutation

$$\Delta\Delta G_{D-TS1} = R\,T \ln ({k_f}^{WT}/{k_f}^{mut}) \quad (3)$$

and

$$\Delta\Delta G_{D-N} = \Delta G_{D-N}{}^{WT} - \Delta G_{D-N}{}^{mut} \quad (4)$$

For the late transition state, ϕ values were calculated as

$$\phi_{TS2} = \Delta\Delta G_{D-TS2}/\Delta\Delta G_{D-N} = 1 - \Delta\Delta G_{TS2-N}/\Delta\Delta G_{D-N} \quad (5)$$

in which

$$\Delta\Delta G_{TS2-N} = R\,T \ln ({k_{U2}}^{mut}/{k_{U2}}^{WT}) \quad (6)$$

Data analysis was performed by Prism (Graphpad) and Kaleidagraph (Synergy Software) software packages.

Keywords: frustrated proteins · kinetics · mutagenesis · protein folding

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 10867–10869
Angew. Chem. **2015**, *127*, 11017–11019

[1] C. Levinthal, *J. Chem. Phys.* **1968**, *48-49*, 44–45.
[2] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, *Proteins Struct. Funct. Bioinf.* **1995**, *21*, 167–195.
[3] a) M. Jager, Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, M. E. Bowman, J. P. Noel, M. Gruebele, J. W. Kelly, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10648–10653; b) D. U. Ferreiro, J. A. Hegler, E. A. Komives, P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19819–19824; c) D. U. Ferreiro, J. A. Hegler, E. A. Komives, P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 3499–3503.
[4] a) T. Walma, C. A. Spronk, M. Tessari, J. Aelen, J. Schepens, W. Hendriks, G. W. Vuister, *J. Mol. Biol.* **2002**, *316*, 1101–1110; b) S. Gianni, S. R. Haq, L. C. Montemiglio, M. C. Jurgens, A. Engstrom, C. N. Chi, M. Brunori, P. Jemth, *J. Biol. Chem.* **2011**, *286*, 27167–27175; c) G. Hultqvist, S. R. Haq, A. S. Punekar, C. N. Chi, A. Engstrom, A. Bach, K. Stromgaard, M. Selmer, S. Gianni, P. Jemth, *Structure* **2013**, *21*, 1193–1202.
[5] a) T. Walma, J. Aelen, S. B. Nabuurs, M. Oostendorp, L. van den Berk, W. Hendriks, G. W. Vuister, *Structure* **2004**, *12*, 11–20; b) E. Di Silvio, A. Toto, D. Bonetti, A. Morrone, S. Gianni, *Sci. Rep.* **2015**, *5*, 9299.
[6] M. Jenik, R. G. Parra, L. G. Radusky, A. Turjanski, P. G. Wolynes, D. U. Ferreiro, *Nucleic Acids Res.* **2011**, *39*, W348–351.
[7] a) N. Calosci, C. N. Chi, B. Richter, C. Camilloni, Å. Engstrom, L. Eklund, C. Travaglini-Allocatelli, S. Gianni, M. Vendruscolo, P. Jemth, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 19241–19246; b) S. Gianni, C. D. Geierhaas, N. Calosci, P. Jemth, G. W. Vuister, C. Travaglini-Allocatelli, M. Vendruscolo, M. Brunori, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 128–133; c) Y. Ivarsson, C. Travaglini-Allocatelli, M. Brunori, S. Gianni, *J. Am. Chem. Soc.* **2009**, *131*, 11727–11733; d) Y. Ivarsson, C. Travaglini-Allocatelli, V. Morea, M. Brunori, S. Gianni, *Protein Eng. Des. Sel.* **2008**, *21*, 155–160; e) P. Jemth, S. Gianni, *Biochemistry* **2007**, *46*, 8701–8708.
[8] A. R. Fersht, A. Matouschek, L. Serrano, *J. Mol. Biol.* **1992**, *224*, 771–782.
[9] a) A. R. Fersht, S. Sato, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7976–7981; b) M. A. De Los Rios, B. K. Muralidhara, D. Wildes, T. R. Sosnick, S. Marqusee, P. Wittung-Stafshede, K. W. Plaxco, I. Ruczinski, *Protein Sci.* **2006**, *15*, 553–563; c) S. Gianni, P. Jemth, *IUBMB Life* **2014**, *66*, 449–452.
[10] a) J. N. Onuchic, N. D. Socci, Z. Luthey-Schulten, P. G. Wolynes, *Folding Des.* **1996**, *1*, 441–450; b) P. G. Wolynes, *Philos. Trans. R. Soc. London Ser. A* **2005**, *363*, 453–464; c) K. A. Dill, J. L. MacCallum, *Science* **2012**, *338*, 1042–1046.
[11] S. Gianni, A. Engström, M. Larsson, N. Calosci, F. Malatesta, L. Eklund, C. C. Ngang, J. P. Travaglini-Allocatelli, *J. Biol. Chem.* **2005**, *280*, 34805–34812.

Received: May 28, 2015
Published online: July 23, 2015

