

Exploring Protein Folding Landscapes: From Ensembles to Single Molecules

Sheena E. Radford

Astbury Centre for Structural Molecular Biology, Institute of Molecular and Cellular Biology, University of Leeds, LS2 9JT, UK
E-mail: s.e.radford@leeds.ac.uk

Introduction

An understanding of the folding energy landscape of a protein sequence requires all species formed along the reaction pathway to be characterised structurally and thermodynamically, from the initial unfolded state in denaturant to the final native structure. To understand how proteins fold studies have focused on the structural characterisation of accessible regions of the folding landscape, wherein intermediate, partially folded and unfolded states are populated and can be characterised directly using biophysical methods. In addition, kinetic analyses combined with site-directed mutagenesis, have been used to determine information about the interactions that are formed in transition state (TS) ensembles, using the so-called Φ -analysis [1] in which structural information is gained indirectly through interpretation of the effect of sequence alterations on the observed rate constants for folding and unfolding. In both scenarios, however, the challenge in understanding folding at a structural level is in how to interpret experimental parameters (most often chemical shifts, hydrogen exchange protection factors, inter-residue or inter-atom distance information determined using spin-labelling or fluorescence resonance energy transfer (FRET) methods, or Φ -values) that are necessarily ensemble averaged, in terms of an atomistic view of the species in question. One route towards achieving this, is to use the experimental parameters as restraints in molecular dynamics (MD) simulations, in order to generate all-atom structural models of these ensembles that then can be further tested and refined using new experiments [2]. Together, these approaches are beginning to shed light on the conformational properties of non-native species formed during folding.

Whilst Φ -value analysis and other approaches provide structural information about intermediate species and TS ensembles, several important questions remain unresolved, including the influence of non-native interactions on the folding landscape, the role of populated intermediates in defining the search to the native state and the nature of the very earliest events in folding. In addition, the heterogeneity of both the intermediate and TS ensembles is difficult to discern, the partial Φ -values that usually characterise these states most commonly being interpreted as resulting from heterogeneity of conformations within a single predominant folding route, although models involving folding on multiple micropathways to the native structure would also give rise to partial Φ -values.

To address the above issues and with the ultimate goal of discerning an atomistic view of how a protein folds, we have performed an extensive series of experiments on the

folding on one family of proteins, the four helical bacterial immunity proteins (see, for example, [3-5]). These studies have focused on two bacterial immunity family members, Im7 and Im9. These proteins are highly tractable experimentally and share 60% sequence identity, yet fold with mechanisms of different kinetic complexity: Im9 showing two-state kinetics whilst Im7 folds *via* a populated intermediate under the same experimental conditions [6]. Extensive studies using ultra-rapid mixing and stopped-flow fluorescence (at pH 7.0 and 10°C) have shown the kinetic intermediate observed during Im7 folding is a productive, on-pathway species that contains three of the four native helices [3];[7]. Perhaps most intriguingly, Φ -value analysis has suggested that this species is stabilised by both native and non-native interactions, suggesting that misfolding is an inherent and natural part of most folding processes [3].

In order to develop a more detailed picture of the folding of Im7, we have also investigated the structural properties of the initial denatured state of the protein in 6M urea [8]. The resulting data have shown that the initial denatured state of Im7 in 6M urea lacks helical structure, but contains hydrophobically collapsed regions that involve residues that ultimately form the native helices. How the protein folds to its three helical intermediate, however, remained unresolved. To address this issue necessitated developing a new ultra-rapid mixing device, capable of measuring folding on a sub millisecond timescale for a large number of variants of Im7, so that a Φ -value analysis of the first TS ensemble encountered during folding could be determined. This we ultimately achieved using a microfabricated T-mixer etched in stainless steel and coupled to a flow cell with a CCD detector, which allows the folding of a protein to be measured highly reproducibly from 100 μ sec to \sim 2msec [9]. Using this device, we have now re-measured the folding kinetics of an array of mutants of Im7 and have been able to construct a Φ -analysis of the entire folding reaction coordinate, from the early TS, through the populated intermediate, through the subsequent rate-limiting TS, to the native state. Moreover, in collaboration with Joerg Gsponer and Michele Vendruscolo (University of Cambridge) we have used these data to restrain MD simulations, providing an atomistic view of the entire folding landscape and allowing the contacts made by every residue at every stage of folding to be revealed. The results provide a fascinating view of how this small helical protein folds and pose challenges for the future in designing experiments to test these models further so that the resulting ensembles can be refined and the methods further developed. Such experiments are currently on-going. As part of this work

we are currently developing single molecule methods that will hopefully allow the results of the simulations to be directly tested. The future looks exciting, therefore, with new experimental methods coupled with new and refined methods of simulation promising to yield detailed information about the structural properties of non-native states that play an important part of folding, but also play an essential role in an array of biological events, including protein translocation, protein assembly and, of course, in protein misfolding diseases.

Acknowledgments

The above work was performed over the last eight years by many extremely talented researchers, all of whom it was a privilege to work with. These included Andrew Capaldi, Neil Ferguson, Stan Gorski, Graham Spence, Claire Friel, Sergui Masca, Inigo Rodriguez-Mendieta, Vicky Morton, Stuart Knowling, Alice Bartlett and Gareth Morgan. Thanks also to Alastair Smith, Heiner Roder, Colin Kleanthous, Emanuele Paci, Joerg Gsponer and Michele Vendruscolo for wonderful collaborations. Our work was funded by the BBSRC, the MRC, The Wellcome Trust and The University of Leeds.

References

1. Fersht, A. R. & Daggett, V. (2002). *Cell* 108, 573-582.
2. Gsponer, J., Hopearuoho, H., Whittaker, S. B. M., Spence, G. R., Moore, G. R., Paci, E., Radford, S. E. & Vendruscolo, M. (2006). *Proc. Natl. Acad. Sci. USA* 103, 99-104.
3. Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2002). *Nat. Struct. Biol.* 9, 209-216.
4. Spence, G. R., Capaldi, A.P. & Radford, S.E. (2004). *J. Mol. Biol.* 341, 215-226.
5. Friel, C. T., Beddard, G. S. & Radford, S. E. (2004). *J. Mol. Biol.* 342, 261-273.
6. Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C. & Radford, S.E. (1999). *J. Mol. Biol.* 286, 1597-1608.
7. Capaldi, A. P., Shastry, M. C. R., Kleanthous, C., Roder, H. & Radford, S. E. (2001). *Nat. Struct. Biol.* 8, 68-72.
8. S.B.M., W., Spence, G. R., Grossmann, J. G., Radford, S.E. & Moore, G. R. (2007). *J. Mol. Biol.* 366, 1001-1015.
9. Masca, S. I., Rodriguez-Mendieta, I. R., Friel, C. T., Radford, S. E. & Smith, D. A. (2006). *Rev. Sci. Instruments* 77, 055105.