Understanding how small helical proteins fold: conformational dynamics of Im proteins relevant to their folding landscapes

Angelo Miguel Figueiredo*, Geoffrey R. Moore*† and Sara B.-M. Whittaker*†

*Centre for Molecular and Structural Biochemistry, School of Chemistry, University of East Anglia, Norwich NR4 7TJ, U.K., and †School of Cancer Sciences, University of Birmingham, Vincent Drive, Edgbaston, Birmingham B15 2TT, U.K.

Abstract
Understanding the mechanism of folding of small proteins requires characterization of their starting unfolded states and any partially unfolded states populated during folding. Here, we review what is known from NMR about these states of Im7, a 4-helix bundle protein that folds via an on-pathway intermediate, and show that there is an alignment of non-native structure in urea-unfolded Im7 with the helices of native Im7 that is a consequence of hydrophobic helix-promoting residues also promoting cluster-formation in the unfolded protein. We suggest that this kind of alignment is present in other proteins and is relevant to how native state topology determines folding rates.

Introduction
Knowledge of protein structures is a central feature of biochemistry today. It is part of the curricula of all biochemistry degree courses, and probably all biochemical researchers have wondered about the structural basis of the particular biochemical phenomenon they are studying. Even before the first three-dimensional structure of a protein was reported there was considerable speculation about what the nature of protein structures would be. Now we can look at numerous protein structures in detail and tell our students that we understand the principles of their formation [1,2]; the balance of entropic and enthalpic forces that make most globular proteins only marginally stable under physiological conditions, the existence of secondary structure and recurrence of super-secondary structural motifs that contribute extensively to stability, and the distribution of charged residues that helps determine important properties of proteins. But we are not yet able to say precisely how these structures came to be; how a given amino acid sequence folded into its particular three-dimensional structure. Even for the relatively small proteins that we discuss below there is much that is unknown, but consider a complex multi-subunit protein with cofactors, such as the transmembrane mitochondrial cytochrome oxidase [3]. This has 13 subunits, a haem a1-CuA binuclear centre, another haem a3 and a binuclear CuA centre. How does it attain its native state? It is clear that complex proteins such as cytochrome oxidase [4] need an array of helper proteins for them to be constructed, and these lie outside the scope of this article. Simpler proteins do not require helper proteins to assist them in folding and, as Anfinsen was the first to show [5], the information required for them to fold is encoded in their amino acid sequences so that they will fold spontaneously under the right solution conditions.

In the present study, we describe key structural aspects, deduced from NMR measurements, of the folding of Im7, an 87 residue, single-domain helical protein [6], and consider the implications of these to the folding of other helical proteins.

Colicin Im proteins
The Im proteins are inhibitors of DNase bacteriocins and provide immunity (hence Im) to the producing cells [6]. They are homologous, with Im7 and Im9 having 57% sequence identity and Im2 having 68% sequence identity with Im9. They share a common distorted 4α-helical bundle structure, as illustrated by Cα RMSDs (root-mean-square deviation) between the NMR structure of Im9 (PDB code 1IMQ) [7] and the X-ray structure of Im7 (PDB code 1AYI) [8], which is 1.7 Å (excluding residues in termini and loops that were poorly defined in the structure of Im9) (1 Å = 0.1 nm). Despite the high structural similarity, Radford and co-workers have shown from stopped-flow and ultra-rapid mixing experiments that Im2 and Im9 fold in two-state transitions from the urea-unfolded state to the folded state [9,10], while Im7 folds in a three-state transition [10] (Figure 1) via an on-pathway KIS (kinetic intermediate state) [11] that Φ-analysis has revealed to be a compact structure with only three helices, equivalent to I, II and IV of the native state, arranged in a manner that allows both native and non-native inter-helical contacts [12].

Energy landscapes and folding funnels
The development of energy landscapes and folding funnels as conceptual models for describing protein folding [19–21], along with computational procedures for defining how proteins traverse such landscapes, has contributed to a

Key words: amino acid cluster, Im protein, kinetic intermediate state (KIS), NMR, protein folding.
Abbreviations used: KIS, kinetic intermediate state; R1, longitudinal relaxation rate; R2, transverse relaxation rate.
†To whom correspondence should be addressed (email g.moore@uea.ac.uk).
structural description of protein folding that complements the kinetic and energetic descriptions obtained by a variety of experimental procedures. The folding funnel in Figure 2 is smooth so that unfolded molecules descending to the native state do not have a preferred route. However, smooth funnels are not likely to be common for proteins. The occurrence of interatomic interactions as a molecule descends the funnel may cause the funnel surface to have a roughness to it. This roughness, which has been called frustration [19], can lead to different routes from the top to the bottom of the funnel having different characteristics. In some cases, the roughness can cause false local minima at energies above the global free energy minimum and, if these become populated during folding, kinetic intermediates may result [22] though non-native interactions do not always retard folding [23,24]. Although schematic, what Figure 2 illustrates is that the ensemble of unfolded states needs to be described as fully as possible in order to get a full structural description of folding [25,26], and, of relevance to the KIS of Im7, any populated false local minima also need to be characterized.

**NMR view of important states in folding of Im7**

Using NMR, we have studied urea-unfolded and native states of Im proteins, as well as various mutant forms of Im7. Backbone $^{15}$N relaxation time measurements show that Im7 and Im9 have significantly different conformational dynamics, as illustrated by their $R_2$ (transverse relaxation rate)/$R_1$ (longitudinal relaxation rate) ratios (Figure 3), that correlate with the difference in their kinetic folding pathway. While Im9 has a featureless sequence profile, a number of residues of Im7 have heightened $R_2$/$R_1$ ratios indicative of conformational exchange [18]. In fact, the native state of Im7, unlike that of Im9, is in equilibrium with a different conformation of the protein that has a higher energy and so is present in solution at low population (∼2% at 25°C) [15,18]. Relaxation-dispersion NMR [18] confirms that this low-populated state at equilibrium is the same as the KIS [20] detected by Radford in stopped-flow experiments [11,12]. Moreover, the NMR data provide unambiguous information about the structure of this form of Im7: while helices I, II and IV are formed, the residues that form the native helix III are in a disordered loop with backbone chemical shifts resembling those of urea-unfolded Im7, though the conformational differences between the native state of Im7 and its KIS are not confined to helix III but are spread throughout the structure. Furthermore, there is a remarkable correlation between the chemical shifts of the non-helix III residues of the KIS detected by RD-NMR (resistively detected NMR) with those of the 3-helix bundle L53A/I54A variant of Im7® [18,27], which Spence et al. [28] showed was a stabilized form of the KIS. What this means is that the highly dynamic 3-helix bundle of the KIS must be interacting with the side chain of both of, or one of, Leu$^{33}$ and Ile$^{44}$ else why would replacing these by alanine residues cause the population of the KIS to increase from ∼2% to approximately 100%? Such an interaction is not unexpected since previous
NMR work has shown that the side chains of one or both of Tyr55 and Tyr56 of the KIS interact with the 3-helix bundle of the KIS [27].

Im7 in 6 M urea is largely extended but it is not completely disordered, having four clusters of interacting side chains that restrict the motion of the backbone [16]. The clusters do not interact with each other, probably a consequence of their solvation by urea since similar clusters in intrinsically disordered proteins [29] and in acid-unfolded apomyoglobin [30] interact with other clusters in the absence of chaotropes. Each Im7 cluster occurs in a region of the protein that forms a helix in the native state, with the largest clusters being associated with the three helices that are formed in the KIS, and the smallest cluster involving residues that form helix III in the native state. This alignment of clusters in unfolded Im7 with helices in native Im7 is a consequence of hydrophobic helix-promoting residues also promoting cluster-formation in the unfolded protein. An important question is: is there any evidence that the clusters contain helical character? At the time the work on urea-unfolded Im7 was reported there was no clear cut evidence, but since then more sophisticated methods of chemical shift analysis have been developed that suggest [31] the larger clusters have 5–10% helical structure, which is likely to be transient. Even though 5–10% is low we believe it is significant since random polypeptide chains have a preference for populating the $\beta$-region of $(\phi, \psi)$ space [32] and the clusters are at least partially solvated by urea, which we envisage reduces the helix content.

**Implications for the folding of Im proteins and other small helical proteins**

It is well established for small proteins that the topology of their native states determines their folding rates [33], which raises the question of how early in the folding pathways native interatomic contacts appear. It is in this context that the clusters of urea-unfolded Im7 become significant.

Although they are probably largely, or entirely, formed from non-native contacts and at best have a low helical content, they are formed from residues with a propensity to form a helix, and the clusters do mirror the helices of the native protein. So, when folding is initiated by dilution of the urea we envisage clusters I, II and IV coming together in the initial stage of hydrophobic collapse. Since clustering of hydrophobic groups to form a compact globule in the absence of hydrogen bond formation is energetically unfavourable [34], the peptide bonds buried in the globule will be driven to form hydrogen bonds, which may lead to secondary structure forming simultaneously with hydrophobic collapse [35]. Thus native contacts are likely to form early on in the hydrophobic collapse of Im7. Further consolidation of the hydrophobic core, together with the development of stable hydrogen-bonded structure in helices I, II and IV then leads to the KIS.

**Implications for the folding of other proteins**

NMR studies of many proteins show that their unfolded states are not completely disordered ([26] and references therein). In some cases, the order appears to be native-like, but in others it is non-native. The question now is: will the order present in these unfolded states promote formation of native-like contacts early in folding in a similar way to that which we propose for Im7? The clearest examples where it seems likely are with other helical proteins, such as apomyoglobin, whose unfolded state at low pH contains hydrophobic clusters that mirror the location of helices in the folded apomyoglobin [30]. These clusters interact in the absence of chaotropes, leading to formation of helices in the kinetic folding intermediate which has a native-like topology [36,37]. The structure in the acid denatured state of the 4-helix bundle acyl-CoA-binding protein is even more striking, consisting of native-like interhelix contacts [38]. As with acid denatured apomyoglobin [30], the absence of a chaotrope

---

[Figure 3] $^{15}$N backbone $R_2/R_1$ ratios for Im7 (left-hand panel) and Im9 (right-hand panel) measured at 500 MHz [25°C, 50 mM potassium phosphate buffer, pH 7 (Im7) or pH 6.2 (Im9)]

Black bars at the top of each column represent the position of helices along the sequences of Im7 and Im9. Adapted from [18] with permission.
probably leads to clusters in the unfolded acyl-CoA-binding protein interacting, with a consequent promotion of helical hydrogen bond formation. Another example is the helical Rd-apocytochrome b₅₅₅, which populates partially unfolded intermediates during folding that contain both native-like secondary structure and non-native hydrophobic interactions [39]. Although an NMR study of its unfolded state has not been reported, the correlation between clusters detected by ¹⁵N relaxation and the sequence profile of the AABUF (average area buried upon folding) parameter [40] for other proteins [16,29,30] lead us to predict that unfolded Rd-apocytochrome b₅₅₅ contains clusters similar to those of Im7 and apomyoglobin (results not shown). A final example is the α/β protein barnase, whose unfolded state has been shown by NMR to contain residual non-random structure, some of which involves the regions of the protein that will form the three helices of the native state [41].

Acknowledgements

We thank Sheena Radford, Graham Spence and Stu Knowling (Leeds) and Lewis Kay and Alex Hansen (Toronto) for valuable discussions, and our UEA colleagues Colin Macdonald and Nick Cull for their technical assistance.

Funding

We gratefully acknowledge the Wellcome Trust, the Wolfson Foundation and the Biotechnology and Biological Sciences Research Council for its support of our Im protein work. A.M.F. thanks the Fundação para a Ciência e Tecnologia, Portugal for a Ph.D. fellowship (grant number SFRH/BD/19312/2004).

References

4 Fontanesi, F., Soto, I.C., Horn, D. and Barrientos, A. (2006) Assembly of Rd-apocytochrome b₅₅₅, which populates partially unfolded intermediates during folding that contain both native-like secondary structure and non-native hydrophobic interactions [39]. Although an NMR study of its unfolded state has not been reported, the correlation between clusters detected by ¹⁵N relaxation and the sequence profile of the AABUF (average area buried upon folding) parameter [40] for other proteins [16,29,30] lead us to predict that unfolded Rd-apocytochrome b₅₅₅ contains clusters similar to those of Im7 and apomyoglobin (results not shown). A final example is the α/β protein barnase, whose unfolded state has been shown by NMR to contain residual non-random structure, some of which involves the regions of the protein that will form the three helices of the native state [41].

Received 8 September 2011
doi:10.1042/BST20110739