This is a repository copy of *Conformational dynamics is more important than helical propensity for the folding of the all alpha-helical protein Im7*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/83270/

Version: Accepted Version

Article:

Frustration and dynamics are more important than helical propensity for the folding of the all α-helical protein Im7

Angelo Miguel Figueiredo[1,2*], Sara B.-M. Whittaker[3], Stuart E. Knowling[4,5], Sheena E. Radford[4], Geoffrey R. Moore[1*]

[1] Centre for Structural and Molecular Biochemistry, School of Chemistry, University of East Anglia, Norwich, NR4 7TJ, UK

[2] Present address: REQUIMTE, CQFB, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516 Monte de Caparica, Portugal

[3] The Henry Wellcome Building for Biomolecular NMR Spectroscopy, School of Cancer Sciences, University of Birmingham, Vincent Drive, Birmingham B15 2TT, UK

[4] Astbury Centre for Structural Molecular Biology, School of Molecular and Cellular Biology, University of Leeds, Leeds, LS2 9JT, UK

[5] Present address: Immunology and Microbial Science, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California, 92037

* To whom correspondence may be addressed. E-mail: am.figueiredo@fct.unl.pt or g.moore@uea.ac.uk

Abbreviations: AABUF, average area buried upon folding; fid, free induction decay; HSQC, heteronuclear single quantum coherence; Im7, the immunity protein for colicin E7; MD, molecular dynamics; NOE, nuclear Overhauser enhancement; ppm, parts per million; DSS, 2,2-(dimethylsilyl)propanesulfonic acid; Im7*, His-tagged Im7; Im7H3M3, Im7 variant containing an engineered helix III;

Abstract

The 87-residue four α-helical protein Im7 folds *via* a three-state mechanism from its urea-denatured state, through an on-pathway intermediate state which contains three of the four native α-helices oriented in a non-native manner. The helix of the native protein not formed in the intermediate, helix III, is the shortest, consisting of only six residues, and its failure to be formed until late in the folding pathway is thought to be related to frustration in the structure. Im7H3M3 is a 94-residue variant of Im7 in which residue substitutions and a polyalanine helix extension were added in order to make helix III the longest of the four native helices. Surprisingly, this protein also folds in a three-state mechanism from its urea-denatured state *via* the same on-pathway intermediate. In order to investigate the structural basis for these findings we calculated the frustration in the structure of Im7H3M3 and used NMR spectroscopy to characterise its conformational properties. Equilibrium peptide N^1H/N^2H exchange showed that the native state of Im7H3M3 is in equilibrium with an intermediate state that lacks helix III, similar to Im7. Model-free analysis of the backbone ^{15}N relaxation parameters of Im7H3M3 identified residues experiencing chemical exchange contributions to their relaxation that aligned with the residues predicted to have highly frustrated interactions, also like Im7. Finally, we determined conformational properties of urea-denatured Im7H3M3 and identified four independent clusters of interacting residues that corresponded to the four α-helices of the native protein. In Im7 the cluster sizes were related to the lengths of the corresponding α-helices with cluster III being the smallest but in Im7H3M3, by contrast, cluster III was also the smallest, despite this region forming the longest helix in the native state. These results suggest that the common conformational properties of the urea-denatured states promote the rapid formation of inter-cluster interactions that are responsible for

formation of a three-helix intermediate of Im7 and Im7H3M3 in which the residues that form helix III remain non-helical. Thus it appears that for Im7 and Im7H3M3 features of the native structure are formed early in folding linked to hydrophobic collapse of the unfolded state.

Introduction

The influence of α-helical propensity on how proteins fold has been well explored [1–5]. In some cases it appears that the unfolded state of the protein contains nascent helical structure that favours folding by the diffusion-collision mechanism [6], in which marginally stable elements of secondary structure dock together aiding their stabilization and promoting formation of the native state. Other helical proteins fold via a hydrophobic collapse mechanism [7] in which collapse of the chain preceded helix formation [8], thus again linking the propensity of an amino acid sequence to form secondary structure with its mechanism of folding. Daggett and Fersht [9] view the hydrophobic collapse and diffusion-collision models as extremes of the nucleation-condensation mechanism in which secondary and tertiary structure form concomitantly and note that where a protein falls within the continuum is determined by the conformational preferences of the residues in the amino acid sequence.

The groundwork that led to current understanding of protein folding mechanisms has involved kinetic, thermodynamic and computational studies of many small proteins. Amongst these the colicin immunity proteins[10,11], which are inhibitors of DNase bacteriocins and provide immunity to the producing cells [12], have played an important part. The family of immunity proteins are highly homologous, with Im7 and Im9 having 57% sequence identity and sharing a common distorted four α-helical structure [13,14]. Despite their high structural similarity, Im9 folds *via* a two-state mechanism from its urea-denatured state [10], while Im7 folds in a three-state manner (Fig. 1A) *via* an on-pathway kinetic intermediate [15]. Φ-analysis [16], NMR spectroscopy [17–19] and MD simulations [11,20] have revealed this intermediate to be a compact structure that contains helices I, II and IV of the native state, arranged in a manner

that allows both native and non-native inter-helical contacts. Building on these observations, Sutto *et al* [21] showed that Im7 has a native structure that is not minimally frustrated and hence has an energy landscape that is rough [22,23] and gives rise to a low-lying excited state close to the native state which is populated during folding and is manifested as the intermediate. Even in the absence of chaotropes this excited state is populated at equilibrium which has allowed it to be probed by equilibrium NH exchange [17] and relaxation-dispersion NMR [19].

A significant feature of the Im7 folding pathway is that the final step in folding is the formation of helix III of the native state (Fig. 1A). This is the smallest of the four helices of Im7 and comprises only six residues in the native structure compared to the 13 or 14 residues of the other helices I, II and IV. Helix III exhibits the lowest helical propensity of all the helices and this led to considerations of whether it is the last helical element to form because it has the lowest helix propensity, or because there are specific features of the amino acid sequence that promote formation of the three helix intermediate. To explore this question Knowling *et al.* [24] engineered Im7 to create a variant in which helix III was lengthened *via* insertion of a polyalanine helix, designed to extend into the original residues of helix III and be stabilised *via* an internal solvent exposed salt bridge. Upon redesign helix III had the highest predicted helical propensity. Knowling *et al.* [24] showed that the resulting variant, Im7H3M3 (Fig. 1B and C), had a three-dimensional structure little altered from that of native Im7 despite the increased length of helix III, and also that this variant folds *via* the same on-pathway intermediate as the wild-type protein. Thus, it appears that folding of Im7 *via* a three-helical intermediate is independent of the helical propensity of helix III.

NMR studies of both urea-denatured Im7 [25] and a triple-variant of Im7 that is unfolded under non-denaturing conditions [26] have revealed that the unfolded states of these proteins

contain four non-interacting hydrophobic clusters that align with the helices of the native state. On dilution of the urea to initiate folding of urea-denatured native Im7 the clusters that are associated with helices I, II and IV are thought to interact in a form of hydrophobic collapse with the consequent rapid formation of the three helix intermediate [11,25–27]. To explore this possibility and to determine whether there is frustration in the structure of Im7H3M3 similar to that of native Im7, we here report NMR studies of the folded and urea-denatured states of Im7H3M3. Our results confirm that the presence of the on-pathway folding intermediate is connected with the presence of frustration in the structure of Im7H3M3, and shows that the conformational properties of the denatured protein play a key role in determining the details of the folding landscape and the topology of the native state.

Results and Discussion

Frustration in the structure of Im7H3M3

The structure of Im7H3M3 was determined by NMR as described in Knowling *et al*. [24]. The core region of Im7H3M3 that has the same amino acid sequence as Im7 (Fig. 1B), residues 2-55 and 72 -93, has an identical fold to Im7 (Fig. 1C) as revealed by backbone RMSDs of 0.8 Å, and, importantly, the long helix III of Im7H3M3 overlays well with the shorter helix III of native Im7, as shown by the backbone RMSDs for the common residues, 50-55, of 0.4 Å. Having demonstrated that the structures of Im7 and Im7H3M3 are strikingly similar we carried out an analysis of frustration in the structure of Im7H3M3 using the approach of Sutto *et al.* [21] with the protein Frustratometer Server (http://www.frustratometer.tk/) [29]. Energy landscape theory states that sites of minimal frustration are associated with stable folding cores of proteins and that these minimal frustrations result when inter-residue interactions in a polypeptide chain are not in conflict with each other and cooperatively lead to a low-energy conformation [23]. In such cases the protein's statistical energy landscapes may have a roughness reflecting the occurrence of favourable non-native interactions during the folding process but the consequences of this are not likely to be significant to the folding pathway. However, where the landscape is more rugged due to considerable roughness a relatively long-lived non-native state may arise that acts as a kinetic folding intermediate, as with Im7. Since analysis of frustration in protein structures highlights apparently unfavourable inter-atomic contacts, which might also come about through errors in the structure coordinates, in what

follows it is pertinent to note that the structures of Im7 and Im7H3M3 were determined entirely independently, the former by X-ray crystallography [14] and the latter by NMR spectroscopy [24].

As Sutto *et al* reported for Im7 [21] we found that frustration is considerable and not randomly scattered across the Im7H3M3 primary sequence (Fig. 2A). However, residues with minimal frustration are found within all four of the α-helices, generally located in inter-helix contacts so that the core of Im7H3M3 itself is largely minimally frustrated (Fig. 2B). Importantly, residues 51-55 of Im7H3M3, which are equivalent to the same residues of Im7 and contribute to helix III in both proteins, are highly frustrated in both structures. Despite the similarity between them, the engineering of Im7 to create Im7H3M3 increased the number of frustrated positions (Fig. 2A).

NMR relaxation studies of Im7 and Im7H3M3

Backbone NH groups of ^{15}N-labelled proteins act as isolated IS spin systems with the relaxation of the ^{15}N nuclei dominated by dipole-dipole interaction with their attached ^{1}H and the chemical shift anisotropy, both of which are modulated by changes in orientation of the NH bond vector with time, and thus they are good probes of protein dynamics [30,31]. Detailed relaxation analyses have been previously reported for Im7 and its' His-tagged variant, Im7*, leading to the identification of residues undergoing chemical exchange on a time scale that impacts the measured R_2 rates [18]. Whittaker *et al* [19] reported a direct correlation between residues involved in conformational exchange and those that Sutto *et al* [21] reported as experiencing frustration, and showed that the correlation resulted from exchange between the

9

native state of Im7 and a low-lying excited state that is populated as a consequence of frustration. To explore whether a similar excited state is present for Im7H3M3 we undertook ^{15}N relaxation analyses (Table 1). Data for 86 of the 94 residues of Im7H3M3 expected to have a detectable ^{1}H-^{15}N HSQC resonance (i.e. excluding the N-terminal and Pro residues) were obtained (Fig. 3), with residues for which relaxation data are not reported excluded because their resonances were too overlapped with others for accurate determination of signal intensities. Surprisingly, the average relaxation parameters for Im7H3M3 suggest the protein is behaving as if it were smaller than Im7* (Table 1) despite its additional 7 residues. Consistent with this, the hydrodynamic radii of Im7* [18] and Im7H3M3, determined as described in Materials and Methods, 19.3 ± 0.4 Å and 17.8 ± 0.3 Å, respectively, indicates that Im7H3M3 is more compact than Im7*. This does not appear to be simply a consequence of a more restricted His-tag for Im7H3M3 than Im7* as the hydrodynamic radii calculated with HYDROPRO [32] from the structures without His-tags (Fig. 1C) indicates: 18.5 Å and 18.0 Å, for Im7 and Im7H3M3, respectively.

The sequence variations of ^{15}N R_1, ^{15}N R_2 and {^{1}H}-^{15}N NOE values (Fig. 3) are consistent with Im7H3M3 being a well-structured globular protein. With the exception of the residues close to the termini and the inter-helix loop regions, particularly the Gly-rich linker between helices III and IV the profiles are largely featureless. The sequence variation of the R_2/R_1 ratios for the backbone ^{15}N resonances (Fig. 3D) highlights those residues, which have relaxation properties significantly different from the majority of residues. The same plot for Im7 (see Fig. 4 of reference [19]) also identifies the corresponding residues to have heightened R_2/R_1 ratios, suggestive of similar motions in both proteins.

In order to proceed with a model-free analysis [33–35] of the Im7H3M3 relaxation data

the diffusion tensor was determined. This was performed by calculating the relative lengths of the principal axes of the inertia tensor using the program pdbinertia [36]. These were 1.00:0.82:0.70, which indicates that the rotational diffusion tensor is either axially symmetric, as it is for Im7, or fully anisotropic. When the ^{15}N R_2/R_1 ratio is independent of the rapid internal motions and magnitude of the chemical shift anisotropy, it can be used to derive the rotational diffusion tensor. Excluding relaxation parameters for residues Val 27, Asp 35, His 40, Phe 41, Ile 44, Thr 45, Glu 46 and Ile 54, which have ^{15}N R_2/R_1 ratios indicating significant contributions from internal motions, the data gave an estimate of the correlation time (τ_c) of 5.98 ns calculated with ModelFree4 [30,37]. This value is consistent with the 6.06 ns calculated from the structure coordinates with HYDRONMR [38]. Since the fully anisotropic model does not provide an improvement, relative to the axially symmetric model (see Table S1 in Supporting Information), the axially symmetric diffusion tensor model was chosen to best represent the motion of Im7H3M3 in solution, which we assign to be rotation as a prolate ellipsoid. The description of the rotational diffusion tensor of Im7H3M3 as a prolate ellipsoid is consistent with the distribution of R_2/R_1 ratios according to the method developed by Clore [39] (see Supporting Information, Fig. S1).

The backbone model-free parameters, S^2 and R_{ex} (Fig. 4), were determined from the relaxation data (Fig. 3) using the axially symmetric diffusion parameters as described in Materials and Methods. High values are seen for the average S^2 giving a picture of a largely rigid protein, except for the termini and Gly-rich linker between helices III and IV. Of most significance in terms of the concept of frustration, the analysis reveals a considerable number of residues that exhibit sizable R_{ex} terms. When these are mapped onto the structure of Im7H3M3 (Fig. 2C) it is obvious that they align with the residues predicted to have highly

frustrated interactions (Fig. 2B).

Equilibrium peptide N^{1}H/N^{2}H exchange rates of Im7 and Im7H3M3

Peptide hydrogen exchange experiments are routinely used to investigate conformational dynamics of proteins [40], including determining structural features of transiently populated intermediates such as that formed in Im7 [17]. Provided that hydrogen exchange occurs by an EX2 mechanism, free energies of exchange can be extracted from the observed rates of exchange and it is these free energies, which provide the key structural insights. Experimental exchange rates, k_{ex}, and free energies of exchange for Im7M3H3, ΔG_{HX}, are summarised in Table 2 with the sequence dependence of ΔG_{HX} given in Fig. 5 and compared with similar data obtained previously for Im7 [17].

The ΔG_{HX} values obtained for Im7H3M3 are remarkably similar to those of Im7. Thus, residues in unstructured regions have NH exchange rates that are too fast for their ΔG_{HX} values to be measured. Residues in helices I and IV have ΔG_{HX} values similar to the corresponding $\Delta G°_{UN}$, indicating that their exchange requires global unfolding, while residues in helix II have ΔG_{HX} values that are similar to ΔG_{UI} determined previously [24] using Φ-value analysis. For helix III $\Delta G_{HX} < \Delta G°_{UI}$ for both Im7 and Im7M3H3 indicating that helix III is not formed in the intermediate ensemble.

NMR characterization of urea-unfolded Im7H3M3

Having characterized the conformational dynamics of the native state of Im7H3M3 and probed the conformation of the intermediate state that is in equilibrium with the native state,

we next turned our attention to urea-denatured Im7H3M3. The limited chemical shift dispersion of the ^{1}H-^{15}N HSQC spectrum of Im7H3M3 denatured in 6 M urea (Fig. 6), particularly in the ^{1}H dimension, shows the protein to be unfolded, as expected from the fluorescence and CD studies of Knowling *et al* [24]. Despite the poor dispersion almost complete assignments (88 peptide NH groups; 94% completeness excluding Met1 and the three proline residues) were obtained from standard triple resonance experiments (CBCANH, CBCA(CO)NH, HNCO) supplemented by a 3D HNN spectrum.

Since the unfolded states of proteins are highly dynamic, the observed NMR parameters are a population-weighted average over all structures in the conformational ensemble. Nevertheless, deviations of chemical shifts from their expected random coil values, secondary chemical shifts ($\Delta\delta = \delta_{obs} - \delta_{rc}$, where the chemical shift δ is referenced to a random coil shift δ_{rc}), are a useful measure of transient secondary structure [41]. However, determination of secondary chemical shifts is dependent on an appropriate choice of random coil chemical shifts. We have used the latest random coil values reported in the literature [42,43], which take into account a set of sequence corrections to the random coil values for all nuclei (for pH, temperature and neighbouring residues), following the approach of Schwarzinger *et al*. [44]. The widely used method for identification of protein secondary structure elements uses ^{13}C chemical shift data [41], which reflect the relative population of backbone dihedral angles in the α and β regions of conformational space [45]. The secondary chemical shifts incorporating sequence corrections (Fig. 7) suggest that though the protein is largely unfolded in 6 M urea there are regions that may be involved in transient secondary structure. This is suggested by the predominantly positive secondary shifts for ^{13}Cα and ^{13}CO, particularly for regions of the

13

protein corresponding to native helices, since positive values are indicative of α-helices [46]. Though Cβ chemical shifts are less sensitive to the presence of α-helices [47], $\Delta\delta C\alpha$ - $\Delta\delta C\beta$ values are a useful tool to reveal secondary structure propensities. For Im7H3M3 in 6M urea, $\Delta\delta C\alpha$ - $\Delta\delta C\beta$ values (Fig. 7D) suggest that residues forming helices I, III and IV of native Im7H3M3 have a preference for ϕ/ψ angles close to those required for α-helices, because they are positive. In contrast, the negative $\Delta\delta C\alpha$ - $\Delta\delta C\beta$ values (Fig. 7D) for residues that form helix II in native Im7H3M3 indicate that these residues have a preference for ϕ/ψ angles in the β region. As Pashley *et al*. [26] note in their NMR study of a mutant of Im7 that is unfolded in the absence of chaotrophs, though positive Cα chemical shift changes indicate formation of helices the magnitude of the shifts are less than the ~2.6 ppm change observed in folded proteins [46] which means that the tendency for some residues in unfolded proteins, and here we include Im7H3M3, to adopt helical structure is weak. Pashley *et al*. [26] found that residues in all four helices of their unfolded mutant Im7 had helical character with those from the native helices I and IV having most. Given that their sample was unfolded in the absence of urea and had 0.2 M Na_2SO_4 present our findings for Im7H3M3 in the presence of urea and absence of Na_2SO_4 are in reasonable agreement.

Even with disordered proteins NOEs can provide valuable structural information though they are not readily interpreted quantitatively because of the conformational averaging. However, as Yao *et al*., [48] showed, $d_{NN(i,i+1)}$ NOEs measured at long mixing times are good indicators of helical content in disordered proteins. For Im7H3M3 in 6 M urea $d_{NN(i,i+1)}$ NOEs were observed (Fig. 7E) for regions that are helices in the folded structure, correlating well with the regions suggested by the chemical shift analyses to have transient helical content.

Long range NOEs indicative of the presence of preferred topologies, were not observed for Im7H3M3 in 6 M urea.

Polypeptide chain dynamics of urea-unfolded Im7H3M3

Backbone dynamics of urea-unfolded Im7H3M3 were investigated with ^{15}N R_1, ^{15}N R_2 and $\{^{1}H\}$-^{15}N heteronuclear NOE data recorded at ^{1}H frequencies of 600 and 800 MHz at 10 °C. Relaxation parameters were determined for 80 of the 94 backbone amides (Fig. 8) as described in Materials and Methods. For residues 2 to 94, the average R_2 values at 600 MHz and 800 MHz are 4.464 ($\pm$ 0.079) s^{-1} and 5.025 ($\pm$ 0.126) s^{-1}, respectively, while the average R_1 values at 600 and 800 MHz are 1.757 ($\pm$ 0.037) s^{-1} and 1.538 ($\pm$ 0.043) s^{-1}, respectively. The $\{^{1}H\}$-^{15}N NOE values alone indicate considerable flexibility throughout the urea-unfolded Im7H3M3 sequence, below the average value of +0.78 expected for backbone amides of a rigid globular protein tumbling isotropically [49].

The slight increase in the R_2 values for Im7H3M3 compared to those of urea-unfolded Im7 suggests that a change in the conformational ensemble has occurred. To explore this we determined the hydrodynamic radius (R_h) of Im7H3M3 from NMR diffusion experiments (see Materials and Methods). At 25 °C and 10 °C, respectively, it was 29.7 $\pm$ 0.4 Å and 25.5 $\pm$ 0.6 Å compared to the theoretical maximum value of 30.7 Å, calculated as described by Wilkins *et al.* [50], and the R_h of urea-unfolded Im7* at 10 °C, 29.8 $\pm$ 1.6 Å [25]. Thus, as with the native folded proteins, Im7H3M3 appears to be more compact than Im7* in their urea-unfolded states. Im7H3M3 in 6 M urea shows a similar degree of compaction as the Im7 mutant L18A-

15

L19A-L37A unfolded in the absence of urea since the hydrodynamic radius of the latter at 10 °C is 26.1 ± 0.6 Å.

Dynamics of a polypeptide chain can be deduced from the backbone NH relaxation parameters R_1, R_2, and $\{^1H\}$-^{15}N NOE through the use of the reduced spectral density functions $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$ [51–53]. The magnitudes of the spectral density functions are sensitive to motions at the corresponding frequencies, zero, ω_N and $0.87\omega_H$. Thus, $J(0)$ reflects slow internal motions on the millisecond to microsecond time scale as well as slow global rotational diffusion while $J(0.87\omega_H)$ reports on the presence of fast internal motions, on the picosecond timescale [51–53]. In the case of urea-unfolded Im7H3M3, $J(0)$ is most informative as it shows that many residues involved in secondary structure elements in the native state have restricted mobility, with $J(0)$ values above the average (Fig. 9A). However, those residues comprising helix III of the native state of Im7H3M3 fall into two distinct groups: the N-terminal segment from residues Gly 50 to Glu 56, whose $J(0)$ values (Fig. 9A) indicate restricted motion, and the C-terminal segment from residues Ala 57 to Asn 64 which appears to have largely random fast motions on the picosecond timescale unrestricted by whatever perturbs the motions of the N-terminal segment. Comparing the $J(0)$ values between the urea-unfolded states of Im7H3M3 and wild-type Im7 (compare Fig. 9A with Fig. 6 of [25]), both measured at 600 MHz, gives further insight into the motional variations due to the elongation of helix III. While residues forming helices I and II of native Im7 have similar $J(0)$ values for urea-unfolded Im7H3M3 and wild-type Im7, smaller $J(0)$ values were observed for the residues forming the C-terminal region of helix III in native Im7H3M3 and the adjacent loop, indicating that these residues are more flexible than the corresponding regions of urea-unfolded Im7.

16

As reported for urea-unfolded Im7 [25], and shown here to facilitate comparison (Fig. 10A and C), the maxima in the sequence profile of $J(0)$ (Fig. 9A), indicating motional restrictions on the backbone NH groups, can be accounted for by clusters of side chains coming together to restrict the motions of the polypeptide backbone (Fig. 10). The correlation between the clusters and the average area buried upon folding (AABUF) [54], which is proportional to the hydrophobic contribution of a residue to the conformational free energy of a protein, and not the helix propensity as determined by AGADIR [55] (Fig. 10B and D), confirms that it is the hydrophobicity of the amino acid sequence and not the helix propensity that is the driving force for cluster formation. Nevertheless, as was observed previously with Im7 the clusters are associated with residues forming α-helices in the native structure, which is a consequence of many of the residues that promote cluster formation also promoting helix formation. This is also shown by the correspondence between the location of the α-helices of the native state and hydrophobic clusters identified by HCADraw [56].

Characteristics of the clusters can be obtained from fitting the observed R_2 rates to models for polypeptide motion. We have used the segmental motion model [57,58] and the volume dependent model [59], as described in Materials and Methods, because there is not clear agreement in the literature on which is most applicable. However, the key features extracted about the clusters (Table 3) were the same for both models: clusters I, II and IV of urea-unfolded Im7H3M3 are the same size as the corresponding clusters of wild-type Im7, which is not surprising since the engineered insert into Im7H3M3 is not in these sequence regions and the clusters are largely non-interacting. Cluster III, the smallest in Im7 is still one of the smallest in Im7H3M3. We return to this observation below. Despite, Im7H3M3 cluster III

17

being the smallest, the sequence elongation did increase the sequence separation between helices and therefore the sequence separation between side chains contacts, the folding core still resembles as the one observed in wild-type Im7 (Figure 10), guaranteeing the same fidelity during the folding landscape. These results suggest that transient hydrophobic clustering caused by local side-chain interactions may drive the early folding events, as observed on denatured apomyoglobin [59]. Definitely, regions of great flexibility as it is the C-terminus (Ala-rich) and the contiguous glycine linker between helix III and IV in Im7H3M3, act to the so-called "molecular hinges" [59] process for the folding mechanism of the protein.

Implication for the folding mechanism of Im7

In the previous study of Knowling *et al.* [24] the notion that helix III in Im7 is the last to fold because it has the lowest helical propensity was dispelled by engineering the helix to contain an extended poly-Ala sequence in Im7M3H3. The data reported here add to these previous studies by demonstrating that it is the properties of the unfolded ensemble that favour the folding of Im7 *via* a three helical intermediate. The observation that the clusters in the urea-denatured state of Im7H3M3 mirror those of urea-denatured Im7 (Table 3), as well as Im7 denatured in the absence of chaotrope [26] are the critical findings that underpin this conclusion. However, cluster III of Im7H3M3 is the smallest of the four clusters, as it is in Im7 despite this region having the highest helix propensity (Fig. 10). The reason is clear; the high helix propensity has been achieved largely by inserting a polyalanine helix, and because Ala is small it has a low AABUF [54] (Fig. 10) and thus does not give rise to a large cluster. Thus,

hydrophobic collapse involving the interaction of the largest clusters early in folding creates the three helical intermediate that is common to the folding pathways of both Im7 and Im7H3M3 in which their largest clusters, I, II and IV, interact. Since such an interaction promotes these clusters adopting their preferred helical conformations the similarity of the collapsed states leads to similar three-helical intermediate states.

General implication for protein folding

Numerous studies of many small proteins have contributed to the current view that the rates of folding for proteins that do not involve kinetic intermediates are determined by the topology of the native state. Following the initial analyses of Baker and his colleagues [60–62], who showed there was a direct correlation between the rate at which such a protein folds and the average sequence separation between contacting residues expressed as an absolute value or relative to the sequence length, which they called the contact order, there have been other analyses confirming that the long-range order of the native state is an important determinant of folding rate [5,63–66]. Grantcharova et al., [62] discussed some of the implications of the correlation of the folding rate with contact order and pointed out that this correlation implies that the contact order of the native state is correlated with the contact order of the transition state ensemble. The work presented here adds to this view, suggesting that the conformational properties of the folding intermediate of Im7 is determined by the nature of hydrophobic clusters in the denatured state. Consistent with this view, Nishimura et al. [67], and Felitsky et al. [68], used NMR measurements to show that transient long-range contacts in unfolded apomyoglobin, some of which are non-native but some native-like, are important for folding,

suggesting that the contact order of the native state does indeed start to appear early on the folding pathway. Calculations also support the idea that native-like contacts are formed early in protein folding linked to hydrophobic collapse [69]. Overall, therefore, the detailed analyses of Im7H3M3 presented here, combined with previous NMR analyses of wild-type Im7 [18,19], urea denatured Im7 [25], Im7 denatured in the absence of chaotrope [26] and Im9 [19] all point to the collapsed status of the denatured protein playing a role in determining the details of the folding landscape and the topology of the native state. Since sequence determines both collapse in the denatured state and its inherent secondary structure propensity, the correlation of structure in the denatured state, the ruggedness of the folding energy landscape and the rate of folding to the native state is perhaps not surprsing.

Materials and Methods

Sample preparation

[15]N labelled and ^{13}C/^{15}N double labelled samples of Im7H3M3 were produced and purified as described previously [70]. For NMR experiments lyophilised samples were resuspended in 50 mM phosphate buffer, pH 7, 10% ^2H_2O/90% H_2O at a concentration of ~0.5 - 1 mM. For urea-unfolded studies the lyophilised protein was dissolved in 50 mM phosphate buffer, 10%

^{2}H$_2$O/90% H$_2$O containing 6 M urea, pH 7.0. The urea concentration was determined using a refractometer, as described by Pace [71].

Frustration analysis of Im7 and Im7H3M3

The Im7 crystal structure (1AYI.pdb) and the Im7H3M3 NMR solution structure (2K0D.pdb) were used in the calculation of the residue-based configurational frustration using the web server at http://www.frustratometer.tk/. The algorithm quantifies the degree of frustration manifested in spatially local interatomic interactions [29].

NMR spectroscopy

All NMR experiments were performed at 25 ºC (unless otherwise specified) and acquired with Avance III 800 MHz, Bruker Avance II+ 600, Bruker Avance II+ 400 and spectrometers or with Varian Unity Inova spectrometers operating at 500 and 600 MHz proton Larmor frequencies. The Avance II+ 600 MHz spectrometer at Lisbon was additionally equipped with a cryogenic probe. Proton chemical shifts were referenced against external DSS while nitrogen and carbon chemical shifts were referenced indirectly to DSS using absolute frequency ratios. All NMR data were processed using NMRPipe [72] or Bruker TopSpin 2.1 software and analysed with CCPNMR [73] or NMRView [74].

^{15}N relaxation measurements of native Im7H3M3

^{15}N R_1 and R_2 relaxation rates and {^{1}H} -^{15}N heteronuclear NOE values for native Im7H3M3

were measured at a ^{1}H frequency of 600 MHz and 400 MHz at 25 °C by standard procedures [75,76]. The R_1 measurements included a recycle delay between scans of 4.0 s and an array of ten different relaxation delays: 0.01 (in duplicate), 0.05, 0.08, 0.2 (in duplicate), 0.5, 0.75, 1, 2 s. The R_2 relaxation delays were: 0.01 (in duplicate), 0.03, 0.05 (in duplicate), 0.07, 0.11, 0.15 (in duplicate), 0.25 s. A 3 s saturation delay was applied during d1 in all {^{1}H}-^{15}N steady-state NOE experiments with a total recycle delay of 5 s to allow the longitudinal magnetization to relax back to equilibrium. Data were recorded with 128 complex t_1 (^{15}N) increments and 1024 complex t_2 (^{1}H) points, with spectral windows of 8000 Hz for ^{1}H and 1680 Hz for ^{15}N at 600 MHz and, with spectral windows of 5331 Hz for ^{1}H and 1120 Hz for ^{15}N at 400 MHz, accumulating 64 scans per fid. All data were processed using NMRPipe [72] and analysed with CCPNMR analysis software [73]. For both R_1 and R_2 data, mono-exponential 2-parameter decay functions were fit to peak intensity versus measured relaxation delay profiles using the CURVEFIT program freely available from Arthur G. Palmer, III [77]. Uncertainties in the derived R_1 and R_2 values were estimated using Monte-Carlo simulations with 1000 random Gaussian noise iterations, taking into account the root mean square noise in the spectra [30]. Heteronuclear NOE values were calculated as the ratio of peak volumes in spectra recorded with and without saturation. In the experiment without saturation, a total recycle delay, d1, of 5 sec was used in place of the saturation delay to ensure the same recycle delay between scans for both experiments. Errors in the NOE values were calculated from the uncertainties in the peak volume measurements estimated by the root mean square noise in each of the two spectra.

For model-free analysis, an initial estimate of the rotational diffusion tensor was obtained from the R_2/R_1 ratios of the individual residues and the PDB coordinates of the

solution structure of Im7H3M3 (2K0D.pdb) using the programs *pdbinertia*, *r2r1_diffusion* and *quadric_diffusion* distributed by Arthur G. Palmer, III [36]. To exclude highly mobile residues or residues with chemical exchange contributions the criteria proposed by Tjandra [78] were applied: residues with NOE < 0.65 were excluded as well as those for which

$$\left(\frac{\langle T_2 \rangle - T_{2,n}}{\langle T_2 \rangle} - \frac{\langle T_1 \rangle - T_{1,n}}{\langle T_1 \rangle} \right) > 1.5 \, SD$$

where $<T_2>$ and $<T_1>$ are averages over the residues that have NOE > 0.65, $T_{2,n}$ is the value of T_2 of residue n and SD is the standard deviation of the distribution of the values in brackets over all residues. The reduced set of rigid residues was used to fit all dynamic models described by the model-free approach using the program ModelFree 4.20 [30]. The latter automatically performs the rigorous statistical testing protocol for the assignment of model functions for each individual residue. For comparison, the rotational diffusion tensor was also predicted using HYDRONMR [38] using an atomic element radius of 3.3 Å.

N^1H/N^2H exchange

[15]N- labelled samples were used to analyse the decay of amide proton signal intensities due to hydrogen exchange with ^2H_2O. A ^{15}N sample lyophilised from water was dissolved into 100% ^2H_2O buffer, 50 mM phosphate buffer, with 0.4 M Na_2SO_4 and containing 0.01% sodium azide, pH* 6.96, (* indicating direct meter reading uncorrected for any isotope effects). Spectra were acquired at a sample temperature of 10 °C. To reduce the time required for the sample to reach temperature the buffer solution was pre-equilibrated at 10 °C for 40 minutes before dissolution of the lyophilised protein. The dissolved protein was then immediately

placed in the NMR tube and inserted into the NMR spectrometer, previously tuned and shimmed using a sample with the same buffer characteristics. The dead time elapsed between dissolving the sample in ^2H_2O and recording the first spectrum was approx. 2 min. Consecutive ^1H-^{15}N HSQC spectra were recorded on a Varian Unity Inova spectrometer operating at 600 MHz using the Biopack pulse sequence gNfhsqc, 512 x 48 complex points, and spectral widths of 7700 x 1650 Hz, for the ^1H and ^{15}N dimensions, respectively. The first two spectra were recorded with 2 scans, the next six with 4 scans, the next seventeen with 8 scans and the last five spectra with 16 scans. In this way, acceptable signal/noise ratios were maintained as peak intensities decreased with amide hydrogen exchange for deuterium. After approximately 6 hours the majority of amide protons had exchanged completely. Cross-peak volumes were obtained using NMRPipe [72] and normalised over the number of scans of each ^1H-^{15}N HSQC spectrum. To calculate the exchange rates, the normalised peak volumes corresponding to each amide peak acquired as a function of the exchange time (defined as the period from the suspension of the lyophilised sample in ^2H_2O) were fitted to a three-parameter single-exponential decay function with equation [5] using Origin (OriginLab, Northampton, MA).

$$I(t) = I_0 e^{-k_{ex} t} + C \qquad [5]$$

where: C is the baseline noise offset, I_0 is the amplitude of the exchange curve at zero time, t is the time in minutes and k_{ex} is the exchange rate. The exchange reaction is described by [79]:

$$(NH)_{closed} \underset{kop}{\overset{kcl}{\rightleftarrows}} (NH)_{open\ states} \overset{kint}{\longrightarrow} N-D \qquad [6]$$
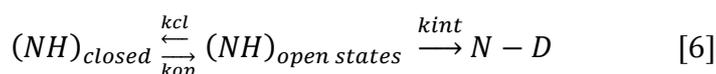
in which an amide group undergoes structural opening (k_{op}) and closing (k_{cl}) to an exchange form, $(N-H)_{open}$. From this state, exchange occurs with solvent with an intrinsic rate constant

for exchange (k_{int}). Under steady-state conditions, the exchange rate (k_{ex}) of equation [6] is given by:

$$k_{ex} = \frac{k_{op}k_{int}}{k_{op} + k_{cl} + k_{int}}$$
[7]

where k_{int} is the known intrinsic chemical HX rate calibrated for the equivalent residue amide at the experimental condition. Herein, intrinsic exchange rates, k_{int}, were obtained by using the web program SPHERE [80] with default activation energies: $E_{acid} = 15$ kcal/mol, $E_{base} = 2.6$ kcal/mol. In the transiently open condition, a kinetic competition between exchange and reclosing ensues. If reclosing is faster $k_{cl} \gg k_{int}$, the structural opening reaction appears as a preequilibrium step prior to the rate-limiting chemical exchange, and the observed rate constant (k_{ex}) is $k_{ex} = K_{op} k_{int}$ where K_{op} is the equilibrium constant for structural opening ($K_{op} = k_{op} / k_{cl}$) – EX2 mechanism. From the Boltzmann relationship ($\Delta G_{HX} = -RT \ln K_{op}$) one can then calculate the free energy change for structural opening reaction that exposes the hydrogen to exchange. To guarantee the same exchange mechanism (EX2) as previously reported for wild-type Im7 [17] the same conditions were used (ionic strength, pH and temperature) to monitored the amide hydrogen exchange for Im7H3M3.

Urea-unfolded Im7H3M3

[13]C/^{15}N labelled and ^{15}N labelled samples of Im7H3M3 in 6 M urea were used for backbone resonance assignment and for relaxation studies, respectively. All NMR measurements were done with freshly prepared samples that were allowed to reach equilibrium before NMR acquisition over a period of 5 hours. Standard triple resonance experiments for backbone

assignment (CBCANH, CBCA(CO)NH, HNCO and HNN) were measured at 25 °C on Bruker Avance III 800 MHz and Varian INOVA 500 MHz spectrometers in the School of Chemistry, University of East Anglia, equipped with room temperature triple resonance probes. The spectral widths for the 3D NMR experiments recorded at 500 MHz (CBCA(CO)NH, HNCO and HNN) were 5629 Hz for ^{1}H, 1320 Hz for ^{15}N, 7535 Hz for ^{13}Cαβ and 1294 Hz for ^{13}CO; at 800 MHz spectral widths were 11161 Hz for ^{1}H, 1953 Hz for ^{15}N and 12500 Hz for ^{13}Cαβ (CBCANH). The ^{1}H and ^{15}N carrier frequencies were set at 4.71 ppm (water) and 120 ppm, respectively. The ^{13}C carrier frequency was set to 56 ppm for HNN, 45 ppm for CBCANH and CBCA(CO)NH, and 174 ppm for HNCO. The 3D HNN spectrum was recorded with 32 complex points along t_1 (^{15}N) and t_2 (^{15}N) and 1024 complex points along t_3 (H^N), 64 scans for each fid, and $T_N = T_C = 12.5$ ms, i.e., the time that ^{13}Cα magnetisation is transferred to the ^{15}N nuclei spins via scalar coupling. The CBCA(CO)NH spectrum was acquired with 1024 (H^N), 32 (^{15}N) and 44 (^{13}C) complex points with 32 scans; the CBCANH spectrum was acquired with 2048 (H^N), 32 (^{15}N) and 80 (^{13}C) complex points with 32 scans, and the HNCO spectrum was acquired with 1024 (H^N), 32 (^{15}N) and 64 (^{13}CO) complex points with 16 scans for each fid.

To probe the existence of inter- and/or intra-residue NOEs in urea-unfolded Im7H3M3, a 3D ^{1}H-^{1}H-^{15}N NOESY-HSQC experiment was recorded at 800 MHz with a mixing time of 200 msec, with spectral widths of 10,000 Hz for ^{1}H (t_3), 2200 Hz for ^{15}N (t_2), 10,000 Hz for ^{1}H (t_1), 2048 x 22 x 64 complex points, respectively, and with 16 scans for each fid. The buffer conditions were the same as used for the backbone assignment but the temperature was lowered to 10 °C. To monitor for temperature dependence of the chemical

shifts, 2D ^{1}H-^{15}N HSQC spectra were recorded from 10 to 25 °C, which allowed us to follow completely the full backbone assignment of the protein.

Residue-specific backbone amide ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates and steady-state heteronuclear {^{1}H}-^{15}N NOE were collected on uniformly ^{15}N-enriched Im7H3M3 in 6 M urea at two static magnetic fields strengths, 600 and 800 MHz, respectively, and at 10 °C using standard procedures described in the literature [75,76]. ^{1}H-^{15}N HSQC spectra were recorded with spectral widths of 5435 Hz for ^{1}H, 1583 Hz for ^{15}N, at 600 MHz and at 800 MHz, spectral widths of 7246 Hz for ^{1}H, and 2110 Hz for ^{15}N. All spectra were acquired with 1024 (t_2) x 128 (t_1) complex data points and four scans per fid. Duplicate time points were used to obtain an estimate of the error. ^{15}N R_1 data were acquired with the following relaxation delay times: 10 (duplicate, 2x), 50, 80, 200 (2x), 500 (2x), 750, 1000 and 2000 msec. Similarly, ^{15}N R_2 values were obtained from a series of 20 experiments with a CPMG delay cycle of 16 msec, recorded in multiple interleaved cycles of: 1 (2x), 2 (2x), 3, 4 (2x), 5, 6, 7, 8 (2x), 9, 10 (2x), 12, 13, 15, 20 and 25. The rates were fit with the program CURVEFIT [77], as for the relaxation studies in the native state. Steady-state {^{1}H}-^{15}N NOE values were obtained by recording spectra with and without ^{1}H saturation applied during the last 5 sec of a 7 sec delay between successive transients. Proton saturation was achieved with 120° ^{1}H pulses applied every 5 msec for 3 sec. {^{1}H}-^{15}N NOE values were calculated as the ratio of peak volumes from spectra recorded with and without saturation. The errors in the NOE values were calculated from the uncertainties in the peak volume measurements estimated by the root mean square noise in each spectrum.

Following the procedure in Le Duff *et al.* [25], R_2 relaxation rate profiles were fitted to a segmental motion model [57,58] and to a segmental motion model incorporating a residue volume dependence [59]. The first model predicts a bell-shaped profile distribution for the dynamics of a linear peptide, with increased flexibility at the termini, as shown by the first term of equation [8]

$$R_2^{exp}(i) = R_{int} \sum_{j=1}^{N} e^{\frac{|i-j|}{\lambda_0}} + \sum_{cluster} R_{cluster}\ e^{-\left(\frac{i-x_{cluster}}{2\lambda_{cluster}}\right)^2} \qquad [8]$$

It assumes that the influence of the neighbouring residues in a polypeptide chain is independent of side chain volume or hydrophobicity, and decays exponentially as the distance from a given residue increases; R_{int} is the intrinsic relaxation rate, which depends on temperature and viscosity, λ_0 is the persistence length of the polypeptide chain (in terms of number of residues) and N is the total chain length. The second term of the equation accounts for the residue volume dependence. This Gaussian term is characterised by the position of the cluster in the protein (residue number) $x_{cluster}$, the cluster width $\lambda_{cluster}$, and a distinct relaxation rate for each cluster, $R_{cluster}$. Overall, the first term of the equation characterises the baseline, whereas the second term fits clusters yielding the deviation from the baseline relaxation profile.

The spectral density at zero frequency, $J(0)$, was calculated as described by Lefevre [52] using the reduced spectral density. From the relaxation parameters ^{15}N R_1, R_2 and $\{^1H\}$-^{15}N NOE, reduced spectral densities were calculated using the jw_mapping.py python script incorporated in the *relax* program [81]. The spectral density functions were obtained assuming that at higher frequencies $J(\omega_H) \approx J(\omega_H + \omega_N) \approx J(\omega_H - \omega_N) \approx J(<\omega_H>)$ [52] and that $J(<\omega_H>)$ is

equivalent to $J(0.87\omega_H)$ or $J(\omega_H + \omega_N)$, where $\omega_H + \omega_N < \omega_H$ since the Larmor frequencies of proton and nitrogen have opposite sign. Thus, $J(0)$ is represented as follows:

$$J(0) = \frac{3}{2(3d^2 + c^2)}\left[\frac{1}{2}R_1 + R_2 - \frac{3}{5}R_{NOE}\right]$$

$$J(\omega_N) = \frac{1}{3d^2 + c^2}\left[R_1 - \frac{7}{5}\sigma\right]$$

$$J(\omega_H + \omega_N) = J(\omega_H) = \frac{1}{5d^2}\sigma$$

where,

$$\sigma = \left[(\{^1H\} - {}^{15}N)NOE - 1\right]R_1\frac{\gamma_N}{\gamma_H}$$

The constants c^2 and d^2 are approximately equal to 1.25×10^9 $(rad/s)^2$ and 1.35×10^9 $(rad/s)^2$, respectively, at 14.1 T ($\omega_H = 600$ MHz), and 2.25×10^9 $(rad/s)^2$ and 1.35×10^9 $(rad/s)^2$ at 18.8 T ($\omega_H = 800$ MHz). ^{15}N chemical shift anisotropy was considered to be -170 ppm and the NH bond length 1.02 Å. Uncertainties in the spectral density values were estimated from 500 Monte Carlo simulations using the *relax* program [81].

NMR diffusion experiments

Pulse-field gradient diffusion NMR experiments were carried out with lyophilised Im7H3M3 dissolved in 100% D_2O (580 µl) and with 20 µl 1,4-dioxane added as internal molecular radius standard. The PG-SLED pulse sequence [50] was used to collect pulse-field-gradient diffusion experiments on a Bruker Avance III 800MHz at 10 °C, respectively. 15 gradient experiments were acquired for each data set, with the gradient strengths augmented linearly

through the acquisition from 0 to 30 G/cm and all other delays and pulses held constant. Gradient pulses (δ) were applied for 6.3 msec with a recovery time of 0.7 msec, and diffusion delay (Δ) of 100 msec. This was found to be adequate to give a total decay of more than 90%. 32 transients were acquired per gradient experiment. Each experiment was acquired with a spectral width of 13,227.5 Hz, and 16-k complex points. The data were zero filled to 32-k complex points and apodized with a sine-squared function. Data were analysed using the variable gradient fitting routines in Bruker TopSpin 2.1 software and in all cases protein resonances were fit with a single exponential decay function using peak intensities. Theoretical hydrodynamic radii (R_h) values were calculated from the empirical equation for folded and denatured proteins [50]. Experimental R_h values for Im7H3M3 were determined as follows: ($D_{ref}/D_{protein}$) x $R_{h(ref)}$, where D_{ref} and $D_{protein}$ are the measured diffusion coefficients of dioxane and the protein, respectively, and $R_{h(ref)}$ is the effective hydrodynamic radius of dioxane, taken to be 2.12 Å [50].

Hydrophobic analysis

The per-residue average area buried upon folding (AABUF) was calculated using the method described by Rose *et al.* [54] using the ExPaSy tool ProtScale (http://us.expasy.org/tools/protscale.html), with a window size of 7 residues and normalised from 0 to 1. Hydrophobic cluster analysis was performed using the program HCADraw [56] on the ExPaSy tool web server (http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=HCA).

Acknowledgements

Figure captions

Fig. 1 (A)   Schematic comparison of the energetics of folding of Im7 and Im7H3M3 based on the thermodynamic and kinetic parameters reported by Knowling *et al*. [24] for Im7H3M3. Except for the cartoon of the structure of folded Im7H3M3 the structure cartoons represent Im7. The first is the urea-unfolded state of Im7 which is extended, contains four independent clusters of residues with each cluster containing interacting hydrophobic side chains and has urea molecules (blue circles) associated with it. In the absence of urea the unfolded state rapidly collapses to transition state 1 (TS1), which is almost devoid of secondary structure and lacks a stable hydrophobic core. TS1 leads to the intermediate state (I) which is followed by transition state 2 (TS2). The four helices of the native structures are coloured differently, with the important helix III coloured green. Even though helix III has been elongated in Im7H3M3 this protein folds via an on-pathway intermediate[24] similarly to wild-type Im7 demonstrating that the formation of an intermediate is an integral feature of the folding mechanism of Im7 that does not result from the short length and low helical propensity of the native helix III. Overall, the elongation of helix III of Im7 to create Im7H3M3 has marginally destabilised the intermediate state by less than 2.5 kJ mol^{-1} and stabilised the native state by the same amount[24].

(B)   The Im7 and Im7H3M3 amino acid sequences. The redesigned helix III in Im7H3M3 is longer than its counterpart in Im7. Residues in boxes correspond to α-helices in the native structures.

(C) Overlay of the most representative conformer (closest to the average model judged by global RMSDs) from the NMR solution structure ensemble of Im7H3M3 (green)[24] and the X-ray crystal structure of Im7 (blue)[14]. The NMR ensemble of 30 conformers of Im7H3M3 was superimposed onto the crystal structure of Im7 by the SuperPose web server [28]: global RMSDs over all residues 1.3Å, heavy atoms 1.1Å, and backbone atoms 0.8Å, respectively.

Fig. 2 (A) Frustration contact maps of Im7H3M3 (left) and wild-type Im7 (right), with highly frustrated residues coloured black and minimally frustrated residues coloured green. The frustration index (F_{ij}) [21] gives the energetic fitness for a given set of residues to interact. An interaction is minimally frustrated when $F_i > 0.78$ and highly frustrated when $F_{ij} < -1$.

(B) Configurational frustration represented on atomic structures of Im7H3M3 (PDB code: 2K0D) and wild-type Im7 (PDB code: 1AYI). A cluster of minimally frustrated contacts (green) defines the core of the protein, which involves all the α-helices, with highly frustrated contacts shown in red.

(C) Backbone fold of Im7H3M3 with the residues exhibiting R_{ex} terms derived from model-free analyses (see below) shown in blue. Residues without R_{ex} terms are coloured grey.

Fig. 3 ^{15}N relaxation data of Im7H3M3 at 25 ºC, in 50 mM phosphate buffer, pH 7, 10% ^{2}H$_2$O, at a ^{1}H frequency of 600 MHz (black) and 400 MHz (red). The sequence

33

dependence of (A) R_1 s^{-1}, (B) R_2 s^{-1}, (C) heteronuclear {^{1}H}-^{15}N NOE values, and (D) R_2/R_1 ratios). The secondary structure elements are represented by horizontal bars at the top of the figure.

Fig. 4   ModelFree relaxation analysis of Im7H3M3 at 25 °C using data measured at 400 and 600 MHz. The generalised order parameter S^2 (A) provides a measure of atomic flexibility of the ^{1}H-^{15}N bond vector on the ps-ns time scale. The chemical exchange terms R_{ex} at 600 MHz (B) describe residues affected by motions occurring on the μs-ms time scale. Residues with R_{ex} values >3 s^{-1} are identified. The horizontal bars at the top of the figure indicate secondary structure elements.

Fig. 5   Gibbs free energy of exchange (ΔG_{HX}) for residues of wild type Im7 (gray) and Im7H3M3 (blue) measured at 10 °C, in 50 mM phosphate buffer, pH 7, and 0.4 M Na$_2$SO$_4$. The dashed lines represent the ΔG^0_{UI} (14.8 kJ/mol for Im7H3M3 and 12.4 kJ/mol for Im7) and the dotted lines represent the ΔG^0_{UN} (23.3 kJ/mol for Im7H3M3 and 25.6 kJ/mol for Im7) [24]. Secondary structure elements are illustrated at the top of the figure, with the extended helix 3 on the mutant Im7H3M3 displayed in red. The ΔG_{HX} values for Im7 were taken from Gorski *et al.* [17].

Fig. 6   600 MHz ^{1}H-^{15}N HSQC spectrum of urea-unfolded (6 M urea) Im7H3M3 dissolved in 50 mM phosphate buffer at 10 °C, pH 7, 10% ^{2}H$_2$O. The backbone assignments have been deposited in the BMRB under accession code 15666.

Fig. 7 Sequence dependence of the secondary chemical shifts of (A) ^{13}CO, (B) ^{13}Cα and (C) ^{13}Cβ of Im7H3M3 in 50 mM phosphate buffer, 6 M urea, 10% D$_2$O at pH 7 and 25 °C. Intrinsic random coil chemical shifts were used according to reference [42,43]. Summary of the assigned cross-peaks in the ^{1}H-^{1}H-^{15}N NOESY-HSQC spectrum of Im7H3M3 (E) showing NOEs between the HN of residue i and the HN of residue i+1 (black squares). The horizontal bars on the top of the figure indicate the secondary structure elements present in the native state.

Fig. 8 ^{15}N relaxation data for urea-unfolded Im7H3M3 acquired at ^{1}H frequencies of 600 MHz (black) and 800 MHz (red). The data were acquired at pH 7 and 10 °C. (A) R_1, (B) R_2, (C) heteronuclear {^{1}H}-^{15}N NOE values and (D) R_2/R_1 ratios as a function of residue number. The black bars at the top of the figure indicate the location of the α-helices in the native protein.

Fig. 9 Reduced spectral density mapping of urea-unfolded Im7H3M3. Data recorded at 600 MHz (black) and 800 MHz (red), at pH 7 and 10 °C; (A) $J(0)$, (B) $J(\omega_N)$ and (C) $J(\omega_H)$ calculated as described in Materials and Methods. The black bars at the top of the figure indicate the location of the α-helices in the native protein.

Fig. 10 R_2 relaxation rates for (A) Im7H3M3 in 6 M urea and (C) Im7 in 6 M urea, pH 7 and at 10 °C, and the average area buried upon folding (AABUF) [54] and helix propensity

(AGADIR) [55] for (B) Im7H3M3 and (D) Im7, plotted as a function of residue number. Measured R_2 rates are shown as black squares with error bars; red lines in (A) and (C) shows the R_2 values calculated according to the volume-dependent model [59] incorporating only the side chain radius of gyration and with persistence length $\lambda_0 = 7$ except for Gly and Ala for which the value of λ_0 was set to 2, and a Gaussian term incorporating four clusters centered at residues Leu 18, Val 42, Gly 56, and Lys 80 (Im7H3M3). The blue lines in (A) and (C) show the R_2 values described by the segmental motion model [57,58]. Black bars at the bottom of graphs (A) and (C) represent d_{NN} (i, $i+1$) NOEs found in the urea-unfolded states; green bars, show hydrophobic clusters predicted by HCA_Draw [56]; and the positions of glycine residues (filled circles) and alanine residues (open circles) in the sequence of each protein is indicated. Open vertical bars in (B) and (D), AGADIR [55] prediction of helix propensity; the average area buried upon folding (red line) was normalised from 0 to 1 with a window size of 7 residues.

Table 1. Parameters describing global dynamics of Im7 and Im7H3M3 from ^{15}N relaxation analyses at 600 MHz

	Im7* (Whittaker *et al.* [18])	Im7H3M3 (this work)
Average ^{15}N R_1 (s^{-1})	1.71 ± 0.03	1.89± 0.04
Average ^{15}N R_2 (s^{-1})	10.18 ± 0.13	8.59 ± 0.08
Average {^{1}H}-^{15}N NOE	0.68 ± 0.09	0.71 ± 0.02
τ_m (ns)	5.27	5.98
$D_{\parallel}$ / $D\perp$	0.784	1.22
Average S^2	0.87 ± 0.09	0.88 ± 0.02

Table 2. Amide hydrogen exchange stabilities for Im7H3M3 measured at 10 ºC.

Residue	Position in the Native Structure	k_{ex} (s^{-1})	ΔG_{HX} (kJ mol^{-1})
Ser 6	N-terminus	2.42×10^{-3}	18.80
Asp 9	N-terminus	1.85×10^{-4}	21.12
Tyr 10	N-terminus	3.60×10^{-5}	22.54
Thr 11	N-terminus	3.97×10^{-5}	24.64
Phe 15	Helix I	1.33×10^{-5}	25.21
Val 16	Helix I	8.26×10^{-6}	24.97
Gln 17	Helix I	1.58×10^{-5}	26.48
Leu 18	Helix I	2.25×10^{-5}	24.01
Leu 19	Helix I	7.86×10^{-6}	24.27
Lys 20	Helix I	1.91×10^{-5}	25.11
Glu 21	Helix I	9.47×10^{-5}	20.59
Ile 22	Helix I	4.69×10^{-5}	19.58
Glu 23	Helix I	4.38×10^{-4}	15.08
Val 33	Helix II	3.97×10^{-4}	14.55
Leu 37	Helix II	4.03×10^{-4}	15.38
Leu 38	Helix II	1.22×10^{-4}	17.81
Phe 41	Helix II	2.20×10^{-3}	15.43
Val 42	Helix II	9.02×10^{-4}	13.92
Lys 43	Helix II	2.39×10^{-3}	14.11
Leu 53	Helix III	2.27×10^{-3}	12.07
Ile 54	Helix III	3.43×10^{-4}	14.57
Tyr 55	Helix III	4.84×10^{-4}	16.14
Glu 56	Helix III	4.52×10^{-3}	11.10
Ile 75	Helix IV	5.42×10^{-3}	10.13
Val 76	Helix IV	1.76×10^{-4}	16.19
Lys 77	Helix IV	2.82×10^{-4}	19.14
Glu 78	Helix IV	2.59×10^{-4}	18.21
Ile 79	Helix IV	3.79×10^{-5}	20.09
Lys 80	Helix IV	1.10×10^{-4}	20.87
Glu 81	Helix IV	1.30×10^{-4}	19.85
Trp 82	Helix IV	8.39×10^{-5}	19.94
Arg 83	Helix IV	1.12×10^{-4}	22.13
Ala 84	Helix IV	1.98×10^{-4}	22.14
Ala 85	Helix IV	4.78×10^{-4}	18.88
Lys 88	C-terminus	1.78×10^{-4}	21.91
Lys 92	C-terminus	5.43×10^{-4}	18.69

Table 3. Clusters in urea-unfolded Im7H3M3 and urea-unfolded wild-type Im7 determined from fits of ^{15}N R_2 relaxation data to models for polypeptide motion.

	Helical residues [a]	Cluster Centre	Cluster Width
Im7H3M3	12-27 (Helix I)	Leu 18	9
	32-45 (Helix II)	Val 42	9
	50-64 (Helix III)	Glu 56	3
	72-86 (Helix IV)	Lys 80	6
Im7 [b]	12-24 (Helix I)	Leu 18	9
	32-45 (Helix II)	Val 42	9
	51-56 (Helix III)	Tyr 56	3
	66-79 (Helix IV)	Lys 73	6

[a] from the corresponding native structure: Im7H3M3 (2K0D.pdb) and wild-type Im7 (1AYI.pdb), respectively

[b] from [25]

References:

1. Sosnick, T.R., Jackson, S., Wilk, R.R., Englander, S.W. & DeGrado, W.F. (1996). The role of helix formation in the folding of a fully alpha-helical coiled coil. *Proteins* **24**, 427–432

2. Lopez-Hernandez, E., Cronet, P., Serrano, L. & Munoz, V. (1997). Folding kinetics of Che Y mutants with enhanced native alpha-helix propensities. *J Mol Biol* **266**, 610–620

3. Islam, S.A., Karplus, M. & Weaver, D.L. (2002). Application of the diffusion-collision model to the folding of three-helix bundle proteins. *J Mol Biol* **318**, 199–215

4. Meisner, W.K. & Sosnick, T.R. (2004). Fast folding of a helical protein initiated by the collision of unstructured chains. *Proc Natl Acad Sci U S A* **101**, 13478–13482

5. Ivankov, D.N. & Finkelstein, A. V (2004). Prediction of protein folding rates from the amino acid sequence-predicted secondary structure. *Proc Natl Acad Sci U S A* **101**, 8942–8944

6. Karplus, M. & Weaver, D.L. (1976). Protein-folding dynamics. *Nature* **260**, 404–406

7. Baldwin, R.L. (1989). How does protein folding get started? *Trends Biochem Sci* **14**, 291–4

8. Fernandez, A., Kardos, J.J., Goto, Y. & Fernández, A. (2003). Protein folding: could hydrophobic collapse be coupled with hydrogen-bond formation? *FEBS Lett* **536**, 187–192

9. Daggett, V. & Fersht, A.R. (2003). Is there a unifying mechanism for protein folding? *Trends Biochem Sci* **28**, 18–25

10. Ferguson, N., Capaldi, a P., James, R., Kleanthous, C. & Radford, S.E. (1999). Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol* **286**, 1597–608

11. Friel, C.T., Smith, D.A., Vendruscolo, M., Gsponer, J. & Radford, S.E. (2009). The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints. *Nat Struct Mol Biol* **16**, 318–324

12. James, R., Penfold, C.N., Moore, G.R. & Kleanthous, C. (2002). Killing of E coli cells by E group nuclease colicins. *Biochimie* **84**, 381–9

13. Osborne, M.J., Breeze, A.L., Lian, L.Y., Reilly, A., James, R., Kleanthous, C. & Moore, G.R. (1996). Three-dimensional solution structure and 13C nuclear magnetic resonance assignments of the colicin E9 immunity protein Im9. *Biochemistry* **35**, 9505–9512

14. Dennis, C.A., Videler, H., Pauptit, R.A., Wallis, R., James, R., Moore, G.R. & Kleanthous, C. (1998). A structural comparison of the colicin immunity proteins Im7 and Im9 gives new insights into the molecular determinants of immunity-protein specificity. *Biochem J* **333**, 183–191

15. Capaldi, A.P., Shastry, M.C., Kleanthous, C., Roder, H. & Radford, S.E. (2001). Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat Struct Biol* **8**, 68–72

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

17. Gorski, S.A., Le Duff, C.S.C.S., Capaldi, A.P., Kalverda, A.P., Beddard, G.S., Moore, G.R. & Radford, S.E. (2004). Equilibrium hydrogen exchange reveals extensive hydrogen bonded secondary structure in the on-pathway intermediate of Im7. *J Mol Biol* **337**, 183–193

18. Whittaker, S.B.-M., Spence, G.R., Günter Grossmann, J., Radford, S.E. & Moore, G.R. (2007). NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. *J Mol Biol* **366**, 1001–1015

19. Whittaker, S.B.-M., Clayden, N.J. & Moore, G.R. (2011). NMR characterisation of the relationship between frustration and the excited state of Im7. *J Mol Biol* **414**, 511–29

20. Gsponer, J., Hopearuoho, H., Whittaker, S.B.-M., Spence, G.R., Moore, G.R., Paci, E., Radford, S.E. & Vendruscolo, M. (2006). Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc Natl Acad Sci U S A* **103**, 99–104

21. Sutto, L., Latzer, J., Hegler, J.A., Ferreiro, D.U. & Wolynes, P.G. (2007). Consequences of localized frustration for the folding mechanism of the IM7 protein. *Proc Natl Acad Sci U S A* **104**, 19825–19830

22. Ueda, Y., Taketomi, H. & Go, N. (1978). Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme. *Biopolymers* **17**, 1531–1548

23. Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. (1995). Funnels, pathways, and the energy landscape of protein-folding - a synthesis. *Proteins* **21**, 167–195

24. Knowling, S.E., Figueiredo, A.M., Whittaker, S.B.-M., Moore, G.R. & Radford, S.E. (2009). Amino acid insertion reveals a necessary three-helical intermediate in the folding pathway of the colicin E7 immunity protein Im7. *J Mol Biol* **392**, 1074–1086

25. Le Duff, C.S.C.S., Whittaker, S.B.-M., Radford, S.E. & Moore, G.R. (2006). Characterisation of the conformational properties of urea-unfolded Im7: implications for the early stages of protein folding. *J Mol Biol* **364**, 824–835

26. Pashley, C.L., Morgan, G.J., Kalverda, A.P., Thompson, G.S., Kleanthous, C. & Radford, S.E. (2012). Conformational properties of the unfolded state of Im7 in nondenaturing conditions. *J Mol Biol* **416**, 300–18

27. Figueiredo, A.M., Moore, G.R. & Whittaker, S.B.-M. (2012). Understanding how small helical proteins fold: conformational dynamics of Im proteins relevant to their folding landscapes. *Biochem Soc Trans* **40**, 424–8

28. Maiti, R., Van Domselaar, G.H., Zhang, H. & Wishart, D.S. (2004). SuperPose: a simple server for sophisticated structural superposition. *Nucleic Acids Res* **32**, W590–4

29. Jenik, M., Parra, R.G., Radusky, L.G., Turjanski, A., Wolynes, P.G. & Ferreiro, D.U. (2012). Protein frustratometer: a tool to localize energetic frustration in protein molecules. *Nucleic Acids Res* **40**, W348–51

30. Mandel, A.M., Akke, M. & Palmer, A.G. (1995). Backbone dynamics of Escherichia-coli ribonuclease HI - correlations with structure and function in an active enzyme. *J Mol Biol* **246**, 144–163

31. Palmer, A.G. 3rd (2004). NMR characterization of the dynamics of biomacromolecules. *Chem Rev* **104**, 3623–3640

32. García De La Torre, J., Huertas, M.L. & Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys J* **78**, 719–730

33. Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc* **104**, 4546–4559

34. Lipari, G. & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J Am Chem Soc* **104**, 4559–4570

35. Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C. & Gronenborn, A.M. (1990). Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J Am Chem Soc* **112**, 4989–4991

36. Diffusion. at <http://www.palmer.hs.columbia.edu/software/diffusion.html>

37. Palmer, A.G., Rance, M. & Wright, P.E. (1991). Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance C-12 heteronuclear NMR-spectroscopy. *J Am Chem Soc* **113**, 4371–4380

38. García de la Torre, J., Huertas, M.L., Carrasco, B. & De la Torre, J. (2000). HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson* **147**, 138–146

39. Clore, G.M., Gronenborn, A.M., Szabo, A. & Tjandra, N. (1998). Determining the magnitude of the fully asymmetric diffusion tensor from heteronuclear relaxation data in the absence of structural information. *J Am Chem Soc* **120**, 4889–4890

40. Englander, S.W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu Rev Biophys Biomol Struct* **29**, 213–238

41. Wishart, D.S. & Sykes, B.D. (1994). The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR* **4**, 171–180

42. Kjaergaard, M., Brander, S. & Poulsen, F.M. (2011). Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *J Biomol NMR* **49**, 139–49

43. Kjaergaard, M. & Poulsen, F.M. (2011). Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J Biomol NMR* **50**, 157–65

44. Schwarzinger, S., Kroon, G.J., Foss, T.R., Chung, J., Wright, P.E. & Dyson, H.J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc* **123**, 2970–2978

45. Dyson, H.J. & Wright, P.E. (1991). Defining solution conformations of small linear peptides. *Annu Rev Biophys Biophys Chem* **20**, 519–538

46. Spera, S. & Bax, A. (1991). Empirical correlation between protein backbone conformation and C.alpha. and C.beta. 13C nuclear magnetic resonance chemical shifts. *J Am Chem Soc* **113**, 5490–5492

47. Avbelj, F., Kocjan, D. & Baldwin, R.L. (2004). Protein chemical shifts arising from alpha-helices and beta-sheets depend on solvent exposure. *Proc Natl Acad Sci U S A* **101**, 17394–7

48. Yao, J., Chung, J., Eliezer, D., Wright, P.E. & Dyson, H.J. (2001). NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. *Biochemistry* **40**, 3561–3571

49. Farrow, N.A., Zhang, O., Forman-Kay, J.D. & Kay, L.E. (1994). A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J Biomol NMR* **4**, 727–34

50. Wilkins, D.K., Grimshaw, S.B., Receveur, V., Dobson, C.M., Jones, J.A. & Smith, L.J. (1999). Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques. *Biochemistry* **38**, 16424–16431

51. Peng, J.W. & Wagner, G. (1992). Mapping of the spectral densities of N-H bond motions in eglin c using heteronuclear relaxation experiments. *Biochemistry* **31**, 8571–8586

52. Lefevre, J.F., Dayie, K.T., Peng, J.W. & Wagner, G. (1996). Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions. *Biochemistry* **35**, 2674–2686

53. Farrow, N. a, Zhang, O., Szabo, A., Torchia, D. a & Kay, L.E. (1995). Spectral density function mapping using 15N relaxation data exclusively. *J Biomol NMR* **6**, 153–162

54. Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H. & Zehfus, M.H. (1985). Hydrophobicity of amino acid residues in globular proteins. *Science* **229**, 834–838

55. Lacroix, E., Viguera, A.R. & Serrano, L. (1998). Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J Mol Biol* **284**, 173–91

56. Gaboriaud, C., Bissery, V., Benchetrit, T. & Mornon, J.P. (1987). Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. *FEBS Lett* **224**, 149–155

57. Schwalbe, H., Fiebig, K.M., Buck, M., Jones, J.A., Grimshaw, S.B., Spencer, A., Glaser, S.J., Smith, L.J. & Dobson, C.M. (1997). Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* **36**, 8977–8991

58. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S.B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L.J., Dobson, C.M. & Schwalbe, H. (2002). Long-range interactions within a nonnative protein. *Science* **295**, 1719–1722

59. Schwarzinger, S., Wright, P.E. & Dyson, H.J. (2002). Molecular hinges in protein folding: the urea-denatured state of apomyoglobin. *Biochemistry* **41**, 12681–12686

60. Plaxco, K.W., Simons, K.T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277**, 985–994

61. Plaxco, K.W., Simons, K.T., Ruczinski, I. & Baker, D. Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry* **39**, 11177–11183

62. Grantcharova, V., Alm, E.J., Baker, D. & Horwich, A.L. (2001). Mechanisms of protein folding. *Current Opinion In Structural Biology* **11**, 70–82

63. Gromiha, M.M. & Selvaraj, S. (2001). Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. *J Mol Biol* **310**, 27–32

64. Zhou, H. & Zhou, Y. (2002). Folding rate prediction using total contact distance. *Biophys J* **82**, 458–63

65. Nölting, B., Schälike, W., Hampel, P., Grundig, F., Gantert, S., Sips, N., Bandlow, W. & Qi, P.X. (2003). Structural determinants of the rate of protein folding. *J Theor Biol* **223**, 299–307

66. Ouyang, Z. & Liang, J. (2008). Predicting protein folding rates from geometric contact and amino acid sequence. *Protein Sci* **17**, 1256–63

67. Nishimura, C., Lietzow, M.A., Dyson, H.J. & Wright, P.E. (2005). Sequence determinants of a protein folding pathway. *J Mol Biol* **351**, 383–392

68. Felitsky, D.J., Lietzow, M.A., Dyson, H.J. & Wright, P.E. (2008). Modeling transient collapsed states of an unfolded protein to provide insights into early folding events. *Proc Natl Acad Sci U S A* **105**, 6278–6283

69. Camilloni, C., Sutto, L., Provasi, D., Tiana, G. & Broglia, R.A. (2008). Early events in protein folding: Is there something more than hydrophobic burst? *Protein Sci* **17**, 1424–33

70. Gorski, S.A., Capaldi, A.P., Kleanthous, C. & Radford, S.E. (2001). Acidic conditions stabilise intermediates populated during the folding of Im7 and Im9. *J Mol Biol* **312**, 849–863

71. Pace, C.N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* **131**, 266–280

72. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPIPE - a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277–293

73. Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. & Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–96

74. Johnso, B.A. & Blevins, R.A. (1994). NMR VIEW - a computer-program for the visualization and analysis of NMR data. *J Biomol NMR* **4**, 603–614

75. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. & Kay, L.E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984–6003

76. Kay, L.E., Nicholson, L.K., Delaglio, F., Bax, A. & Torchia, D.A. (1992). Pulse sequences for removal of the effects of cross-correlation between dipolar and chemical-shift anisotropy relaxation mechanism on the measurement of heteronuclear T1 and T2 values in proteins. *J Magn Reson* **97**, 359–375

77. CurveFit. at <http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/palmer/software/curvefit.html>

78. Tjandra, N., Feller, S.E., Pastor, R.W. & Bax, A. (1995). Rotational diffusion anisotropy of human ubiquitin from N-15 NMR relaxation. *J Am Chem Soc* **117**, 12562–12566

79. Krishna, M.M.G., Hoang, L., Lin, Y. & Englander, S.W. (2004). Hydrogen exchange methods to study protein folding. *Methods* **34**, 51–64

80. Hydrogen Exchange Prediction. at <http://www.fccc.edu/research/labs/roder/sphere/sphere.html>

81. d'Auvergne, E.J. & Gooley, P.R. (2008). Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. *J Biomol NMR* **40**, 107–19

Supporting Information

Table S1 – Diffusion parameters for Im7H3M3.

Model	$\tau_{m,eff}$ [a]	$2D_z/(D_x+D_y)$ [b]	D_x/D_y	$\theta^{(c)}$ (°)	ϕ (°)	φ (°)	χ^2	F_x [d]	P_F [e]
Isotropic	6.10	1	1					616	
Axially symmetric	5.98	1.22	1	6.2	88.0		380	8.97	0.0001
Fully asymmetric	6.01	1.22	0.86	10.2	76.3	-79.0	355	1.44	0.2489
Hydrodynamic	6.06	1.12	0.90						

a) Effective correlation time, defined as $(6D)^{-1}$ in ns

b) Also defined as $D_\parallel/D\perp$

c) Euler angles describing the orientation of the diffusion tensor in the PDB coordinate frame

d) F_x ratios for assessing the validity of a reduction in $\chi2$ when additional parameters are added to the model

e) Probability that the reduction in $\chi2$ is achieved by chance

Note: for the axially symmetric model both the oblate ($D\parallel < D\perp$; $Dx \geq Dy > Dz$) and prolate models ($D\parallel > D\perp$; $Dz > Dy \geq Dx$) were considered, however it is clear that Im7H3M3 in solution is best represented by a prolate model, since $D\parallel / D\perp > 1$.

To corroborate the idea that Im7H3M3 in solution rotates as a prolate ellipsoid, the method developed by Clore and co-workers [39] was applied. This method uses a histogram as a way to obtain the magnitude of the diffusion tensor from the distribution of heteronuclear T_1/T_2 ratios (see Figure S1) without the need for any prior structural information. From the average of the highest, the lowest and the most frequently occurring T_1/T_2 ratios in the distribution, the tensor components were estimated. The final value obtained for the anisotropy, A was 1.35 and for

the rhombicity, η was 0.86. Since an average of the extreme values is used in the first instance, the anisotropy is rather under- than over-estimated, but it does give an idea of the anisotropy. Overall, this method confirms that Im7H3M3 in solution rotates as a prolate ellipsoid.

Figure S1 – Histogram of T_1/T_2 ratios for Im7H3M3