

# Kent Academic Repository

## Full text document (pdf)

### Citation for published version

Pashley, CL and Morgan, GJ and Kalverda, AP and Thompson, GS and Kleanthous, C and Radford, SE (2012) Conformational Properties of the Unfolded State of Im7 in Nondenaturing Conditions. *Journal of Molecular Biology*, 416 . pp. 300-318. ISSN 0022-2836.

### DOI

<https://doi.org/10.1016/j.jmb.2011.12.041>

### Link to record in KAR

<https://kar.kent.ac.uk/71808/>

### Document Version

Publisher pdf

#### Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

#### Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

#### Enquiries

For any further enquiries regarding the licence status of this document, please contact:

[researchsupport@kent.ac.uk](mailto:researchsupport@kent.ac.uk)

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>



# Conformational Properties of the Unfolded State of Im7 in Nondenaturing Conditions

Clare L. Pashley<sup>1,2</sup>, Gareth J. Morgan<sup>1,2</sup>, Arnout P. Kalverda<sup>1,2</sup>, Gary S. Thompson<sup>1,2</sup>, Colin Kleanthous<sup>3</sup> and Sheena E. Radford<sup>1,2\*</sup>

<sup>1</sup>Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

<sup>2</sup>Institute of Molecular and Cellular Biology, University of Leeds, Leeds LS2 9JT, UK

<sup>3</sup>Department of Biology, University of York, York YO10 5YW, UK

Received 1 November 2011;  
received in revised form

19 December 2011;

accepted 20 December 2011

Available online

28 December 2011

Edited by C. R. Matthews

## Keywords:

protein folding;

NMR;

unfolded ensemble;

denatured state;

immunity protein

The unfolded ensemble in aqueous solution represents the starting point of protein folding. Characterisation of this species is often difficult since the native state is usually predominantly populated at equilibrium. Previous work has shown that the four-helix protein, Im7 (immunity protein 7), folds via an on-pathway intermediate. While the transition states and folding intermediate have been characterised in atomistic detail, knowledge of the unfolded ensemble under the same ambient conditions remained sparse. Here, we introduce destabilising amino acid substitutions into the sequence of Im7, such that the unfolded state becomes predominantly populated at equilibrium in the absence of denaturant. Using far- and near-UV CD, fluorescence, urea titration and heteronuclear NMR experiments, we show that three amino acid substitutions (L18A–L19A–L37A) are sufficient to prevent Im7 folding, such that the unfolded state is predominantly populated at equilibrium. Using measurement of chemical shifts, <sup>15</sup>N transverse relaxation rates and sedimentation coefficients, we show that the unfolded species of L18A–L19A–L37A deviates significantly from random-coil behaviour. Specifically, we demonstrate that this unfolded species is compact ( $R_h = 25 \text{ \AA}$ ) relative to the urea-denatured state ( $R_h \geq 30 \text{ \AA}$ ) and contains local clusters of hydrophobic residues in regions that correspond to the four helices in the native state. Despite these interactions, there is no evidence for long-range stabilising tertiary interactions or persistent helical structure. The results reveal an unfolded ensemble that is conformationally restricted in regions of the polypeptide chain that ultimately form helices I, II and IV in the native state.

© 2012 Elsevier Ltd. Open access under [CC BY license](http://creativecommons.org/licenses/by/3.0/).

\*Corresponding author. Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK.  
E-mail address: [s.e.radford@leeds.ac.uk](mailto:s.e.radford@leeds.ac.uk).

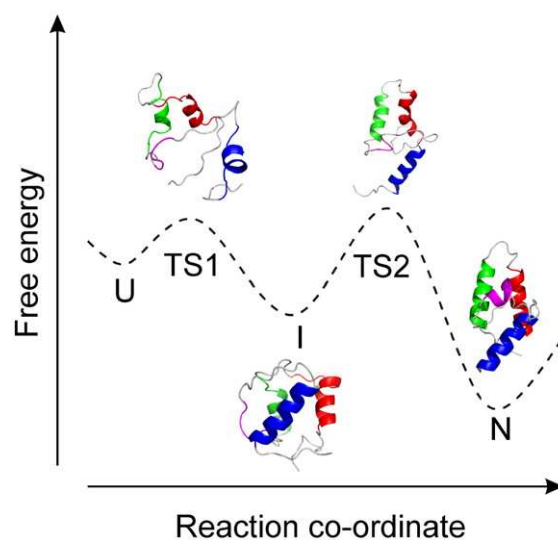
Present address: G. J. Morgan, Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA.

Abbreviations used: TS1, transition state 1; TS2, transition state 2; Cole7, colicin E7; SSP, secondary structure propensity; smFRET, single-molecule Förster resonance energy transfer; Im7, immunity protein 7; EDTA, ethylenediaminetetraacetic acid; HSQC, heteronuclear single quantum coherence; AUC, analytical ultracentrifugation; ITC, isothermal titration calorimetry; BMRB, Biological Magnetic Resonance Data Bank; NOE, nuclear Overhauser enhancement; AABUF, average area buried upon folding; PDB, Protein Data Bank.

## Introduction

Characterisation of all species formed during the folding process is required to understand how a protein folds in atomistic detail.<sup>1</sup> Conformations that are transiently sampled within the denatured ensemble represent an interesting, but experimentally challenging, starting point of folding. The unfolded state may dictate the conformations open to a protein during its lifetime by restricting the conformational space available to the polypeptide chain. Accordingly, the presence of residual structure in the unfolded state has been suggested as one property that could help resolve Levinthal's long-standing paradox.<sup>1,2</sup> Assuming folding takes place on a funnelled landscape,<sup>3</sup> residual structure could reduce the conformational space available to the unfolded protein chain at the commencement of folding, potentially reducing the time spent searching for the native conformation. The general consensus from many studies is that residual structure in the unfolded state could play an important role in protein folding.<sup>4–7</sup> Residual structure may provide initiation points for folding,<sup>5,8</sup> affect the stability of the native protein<sup>9</sup> and/or modulate folding mechanisms and rates.<sup>4</sup> These findings highlight the importance of delineating the structural, thermodynamic and kinetic properties of the unfolded state.

Assessing the nature of residual structure in the unfolded state of a natively folded protein is experimentally challenging since this species is rarely populated at equilibrium under conditions favouring folding. There are proteins that naturally lend themselves to the study of the unfolded state (i.e., using naturally unstable proteins or proteins requiring a cofactor to fold).<sup>10,11</sup> However, steps are usually required to increase the population of unfolded conformers under ambient conditions before biophysical studies can be undertaken. A number of different approaches have been utilised, including unfolding the protein at low pH, denaturation at low or high temperature or the addition of chaotrope.<sup>12–17</sup> Alternatively, mutagenesis, sometimes in combination with methionine oxidation, has been used to increase the population of the unfolded state at equilibrium.<sup>18–21</sup> Extrapolation to low denaturant concentrations has also been carried out using single-molecule Förster resonance energy transfer (smFRET) experiments.<sup>22–24</sup> The results of these studies revealed that nonrandom behaviour can persist in denatured states of proteins, even under harsh denaturing conditions. These results provide compelling evidence that unfolded species under ambient conditions also possess significant residual native or nonnative structure, albeit that these structures may be rare or transient. Since this is the starting point for protein folding *in vitro*,



**Fig. 1.** Schematic representation of the folding energy landscape of Im7. Species populated during folding are indicated as follows: U, unfolded state; I, intermediate state; N, native state; TS1, the first transition state; TS2, the rate-determining transition state. The structure of Im7 (PDB code: 1AYI<sup>33</sup>) is shown for the native state, and representative structures from the ensemble generated by restrained molecular dynamics simulations are shown for TS1, I and TS2.<sup>27,34</sup> Regions that form the four helices in the native state are coloured red (helix 1), green (helix 2), pink (helix 3) and blue (helix 4).

understanding the structural properties of the unfolded ensemble under non-denaturing conditions is fundamental to our understanding of the folding process.

The folding mechanism of the colicin immunity protein 7 (Im7) has been extensively characterised as a model for helical protein folding.<sup>24–30</sup> Despite its small size, this 87-residue protein has been shown to fold to its native state *via* an on-pathway intermediate (I), stabilised by both native and nonnative interactions.<sup>31,32</sup> Using a combination of protein engineering ( $\Phi$ -value analysis), hydrogen exchange protection experiments and restrained molecular dynamics simulations, it has been possible to obtain all-atom models of the ensembles for the early and late transition states (TS1 and TS2) and the populated intermediate (I), allowing an atomic model of the Im7 folding mechanism (Fig. 1).<sup>27</sup> These experiments revealed that TS1 is expanded ( $\beta_T^{\text{TS1}} = 0.24$ ), where the  $\beta$ -Tanford value ( $\beta_T$ ) is an indirect measure of the average solvent-exposed surface area in a given state relative to that of the denatured and native states.  $\beta_T^x$  therefore indirectly reflects the overall compactness of the species,  $x$ .<sup>35</sup> If  $\beta_T^x = 0$ , the species,  $x$ , has the same average solvent-exposed surface area as the denatured state,

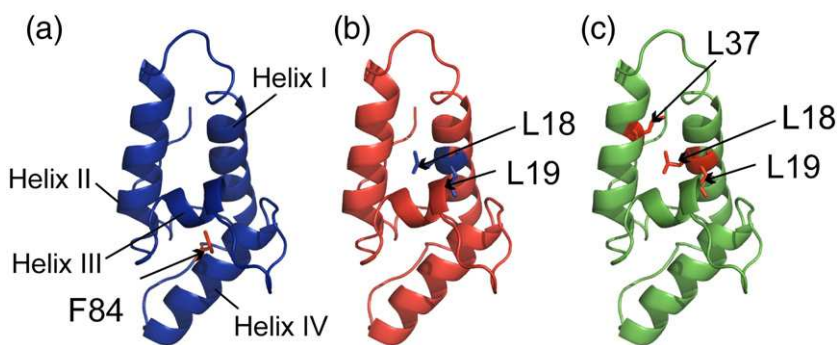


Fig. 2. Ribbon diagrams of Im7 (PDB code: 1AYI<sup>33</sup>) showing its four native helices. The location of residues mutated in the variants (a) F84A, (b) L18A–L19A and (c) L18A–L19A–L37A are indicated. The figure was created using PyMOL.

whereas if  $\beta_T^x = 1$ , the species has an average solvent-exposed surface area similar to that of the native state.<sup>35</sup> Furthermore,  $\Phi$ -value analysis indicated that TS1 lacks ordered secondary structure, although there is evidence for persistent long-range interactions between helices I and II in this species.<sup>27</sup> The transition from TS1 to the more ordered, on-pathway intermediate (I) is a key step in the folding of Im7 and involves substantial compaction ( $\beta_T^I = 0.74$ ); the formation of stable native-like secondary structure in helices I, II and IV; desolvation of the core; helix–helix docking; and the formation of both native-like and nonnative stabilising interactions.<sup>27</sup> The rate-limiting step occurs as the protein traverses TS2 from I to form the native state.<sup>27</sup> Subtle rearrangements of the helices and core along with further compaction ( $\beta_T^{TS2} = 0.89$ ) take place in TS2, enabling formation of the binding site for helix III. Helix III then forms and docks in the last step of folding (Fig. 1).<sup>27</sup>

Despite the extensive knowledge of the folding mechanism of Im7, our understanding of the starting point of folding (i.e., the unfolded state under non-denaturing conditions) remained limited. The urea-denatured state of Im7 has been characterised in 6 M urea and was shown to be highly expanded and to lack residual secondary structure or stable long-range interactions.<sup>29</sup> However, recent smFRET studies of Im7 suggested that the unfolded state in the absence of denaturant is significantly more compact than the urea-denatured state in 6 M urea in that the FRET efficiency between the donor and acceptor dyes (i.e., the proximity ratio) of the former approaches that of the native state in 0 M urea.<sup>24</sup> Here, protein engineering has been used to trap Im7 in an unfolded state in the absence of denaturant, enabling determination of its structural properties using fluorescence, CD spectroscopy and heteronuclear NMR methods under the solution conditions used to monitor folding (pH 7, 10 °C).<sup>24,26–28</sup> The results reveal a collapsed ensemble that lacks persistent helical structure. Folding of Im7 may then progress by the consolidation of the collapsed

regions to form cooperative stabilising interactions that promote secondary- and tertiary-structure formation once folding commences.

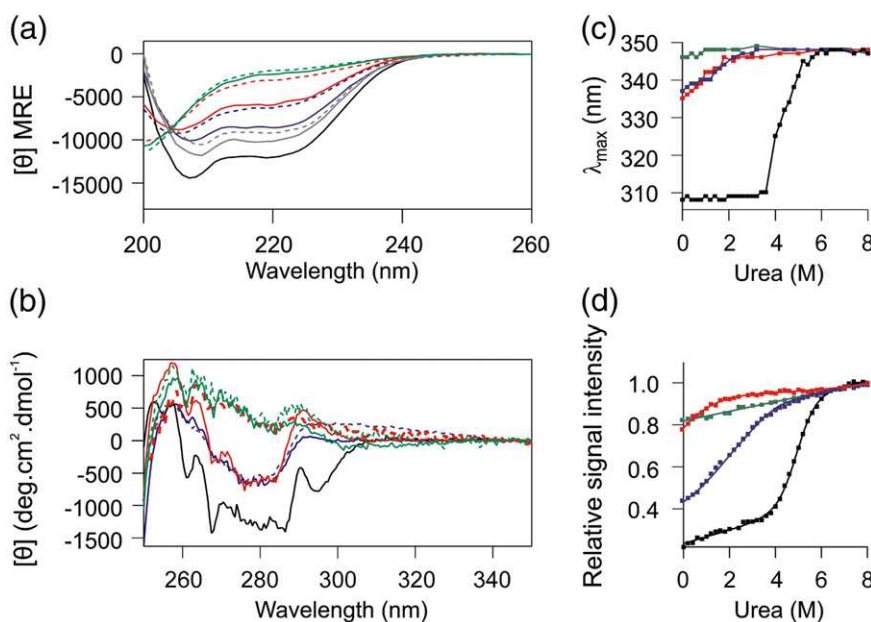
## Results

### Design of an unfolded Im7 variant

Wild-type Im7 has a folding free energy ( $\Delta G_{UN}^\circ$ ) of  $-24.9 \text{ kJ mol}^{-1}$  at 10 °C, pH 7.0, in 0.4 M  $\text{Na}_2\text{SO}_4$ , the conditions that have been used to study its folding mechanism.<sup>24,26–28</sup> Under these conditions, the partially folded intermediate is also stable ( $\Delta G_{UI}^\circ = -11.9 \text{ kJ mol}^{-1}$ ).<sup>30</sup> Significant destabilisation of both the native and intermediate states is required to populate the unfolded state at equilibrium in the absence of denaturant. Amino acid substitutions to the sequence of Im7 made previously during  $\Phi$ -value analysis were considered to achieve this.<sup>27,30</sup> The aim was to combine one or more of these substitutions to destabilise the native and intermediate states of the protein, such that these species are no longer significantly populated at pH 7.0 and 10 °C. Previous  $\Phi$ -value analysis demonstrated that the amino acid substitution F84A is highly destabilising, such that the folding kinetics of this variant could not be measured.<sup>30</sup> Substitution of this amino acid was considered desirable since this residue is buried in both the intermediate and native states but is not involved in native helical structure (Fig. 2a).<sup>27,33</sup>

Since substitution of a large aromatic group could potentially affect the residual structure in the unfolded state by reducing hydrophobic collapse, the more subtle substitutions L18A, L19A and L37A in helices I and II were also considered. These sequence alterations have each been shown to destabilise both N and I significantly ( $\Delta G_{UN}^\circ$  of  $-12.6$ ,  $-14.2$  and  $-11.9 \text{ kJ mol}^{-1}$  and  $\Delta G_{UI}^\circ = -3.7$ ,  $-5.8$  and  $-4.3 \text{ kJ mol}^{-1}$ , respectively).<sup>30</sup> Previous  $\Phi$ -value analysis reveals  $\Phi_{TS1}$  values of  $0.31 \pm 0.19$ ,  $0.38 \pm 0.19$  and  $0.28 \pm 0.15$ ;  $\Phi_I$  values of  $0.45 \pm 0.09$ ,





**Fig. 3.** Summary of CD and fluorescence spectroscopy of the Im7 variants in buffer A at 10 °C. (a) Far-UV CD spectra of the different variants of Im7 at 10 °C. Dotted lines indicate spectra in 0 M Na<sub>2</sub>SO<sub>4</sub> and a continuous line indicates spectra in the presence of 0.4 M Na<sub>2</sub>SO<sub>4</sub>. Spectra of wild-type Im7, which is independent of Na<sub>2</sub>SO<sub>4</sub> (black), F84A (blue), L18A–L19A (red) and L18A–L19A–L37A (green), and the trapped intermediate state I53A L54A (grey) are shown.<sup>32</sup> These colours are used to indicate the different variants in all panels. The units of mean residue ellipticity (MRE) are deg cm<sup>2</sup> dmol<sup>-1</sup> peptide bond<sup>-1</sup>. (b) Near-UV CD spectra of the different variants in the presence of 0 M Na<sub>2</sub>SO<sub>4</sub> (dotted line) or 0.4 M Na<sub>2</sub>SO<sub>4</sub>. The spectrum of wild-type Im7 in 0.4 M Na<sub>2</sub>SO<sub>4</sub> is shown in black. Urea-induced denaturation of wild-type Im7 and the variants measured (c) by the change in fluorescence emission  $\lambda_{\max}$  and (d) by the change in the CD signal at 222 nm in the presence of 0.4 M Na<sub>2</sub>SO<sub>4</sub>. CD data are normalised to the signal in 8 M urea before conversion to signal intensity relative to this value. Data for the wild-type protein (black) are fitted to a model describing a two-state transition [Eq. (2)], as described in [Materials and Methods](#). All other data were fitted using a straight line, or polynomial function, to guide the eye.

$0.56 \pm 0.09$  and  $0.28 \pm 0.11$ ; and  $\Phi_{\text{TS2}}$  values of  $0.67 \pm 0.42$ ,  $0.70 \pm 0.41$  and  $0.53 \pm 0.18$ , for the individual L18A, L19A and L37A amino acid substitutions, respectively.<sup>30</sup> A double-substitution variant L18A–L19A, along with a triple-substitution variant L18A–L19A–L37A, was created in order to try to populate the unfolded state predominantly at equilibrium in the absence of denaturant (Fig. 2b and c).<sup>30</sup> Residues involved in the binding interface between Im7 and its cognate binding partner colicin E7 (ColE7) were not altered in any design.<sup>36</sup> The Im7–E7 complex has a  $\Delta G^{\circ}_{\text{binding}}$  of  $-83.2$  kJ mol<sup>-1</sup>.<sup>36</sup> As a consequence, the ability of the variants to undergo binding-induced folding could be determined as a test of the ability of the variants to reach a functional state despite the amino acid substitutions introduced.<sup>37</sup> The Im7 variants F84A, L18A–L19A and L18A–L19A–L37A were expressed in *Escherichia coli* and purified to homogeneity as described previously ([Materials and Methods](#)).<sup>38</sup> Their structural characteristics, unfolding properties and stability were then deduced using a range of biophysical methods.

### Structural properties of the Im7 variants

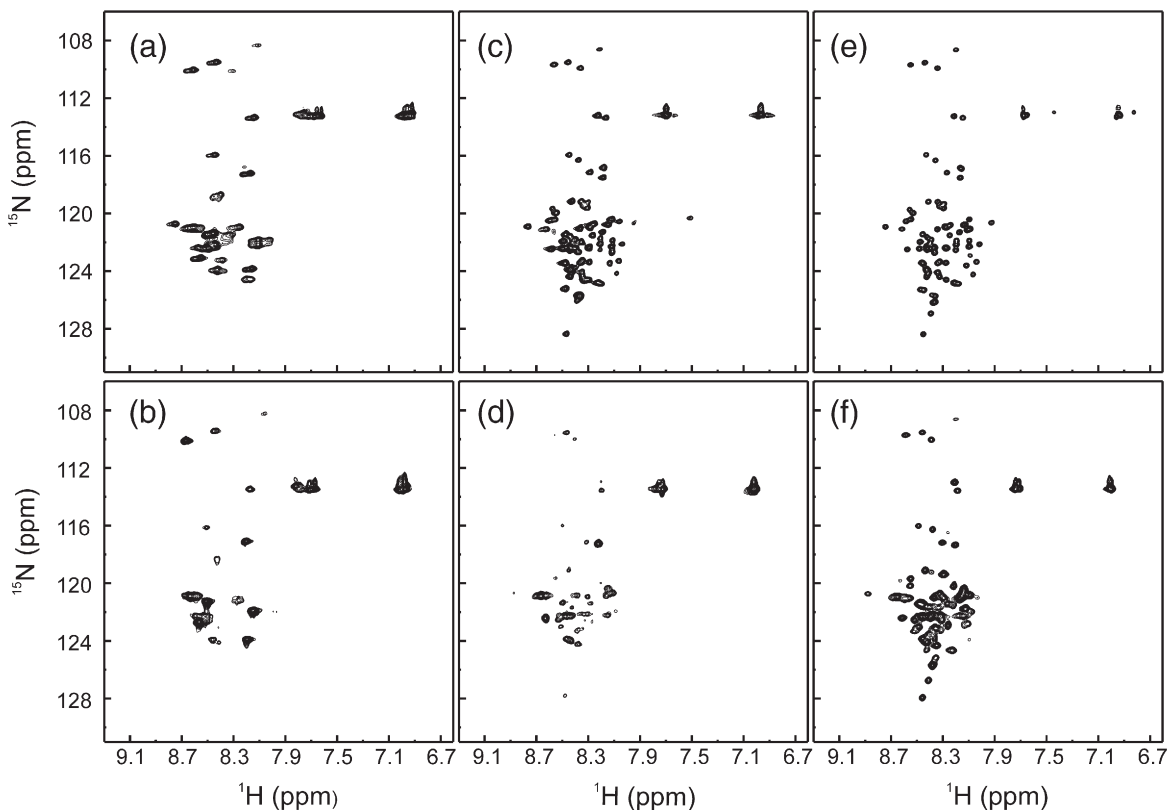
A series of experiments were performed in order to determine which, if any, of the variants created are unfolded under ambient conditions as a prelude to their more detailed structural characterisation using NMR. First, CD spectroscopy was used to assess the secondary and tertiary structure of each variant. Previous protein folding studies of Im7 were performed in 50 mM sodium phosphate (pH 7.0) containing 1 mM ethylenediaminetetraacetic acid (EDTA) (buffer A) and 0.4 M Na<sub>2</sub>SO<sub>4</sub> at 10 °C. Since Na<sub>2</sub>SO<sub>4</sub> is known to stabilise the intermediate and native states of Im7, but not to alter their structural properties,<sup>39</sup> CD spectra were recorded in the presence and absence of this kosmotrope (Fig. 3). Wild-type Im7 has a helical content of 50%, which does not change upon the addition of Na<sub>2</sub>SO<sub>4</sub>.<sup>33,39</sup> The far-UV CD spectrum of this protein was used to calculate the helical content of each of the variant proteins relative to that of the native state ([Materials and Methods](#)).<sup>32,33</sup> The far-UV CD spectrum of F84A shows that this variant contains significant helicity

(~25%) (Fig. 3a). Near-UV CD showed that this protein also retains fixed tertiary structure, although this differs from that of the native protein (Fig. 3b). The helical content of F84A increases slightly upon the addition of 0.4 M Na<sub>2</sub>SO<sub>4</sub> (~36% helicity) (Fig. 3a), while there is little effect of Na<sub>2</sub>SO<sub>4</sub> on the tertiary structure as indicated by near-UV CD (Fig. 3b). The L18A–L19A variant behaves differently. In the absence of Na<sub>2</sub>SO<sub>4</sub>, this variant has little helical content (~12%) and no detectable fixed tertiary structure (Fig. 3a and b). However, in the presence of 0.4 M Na<sub>2</sub>SO<sub>4</sub>, the helical content of L18A–L19A increases to ~25% and the near-UV CD spectrum indicates that this variant adopts fixed, nonnative tertiary interactions (Fig. 3b). By contrast, L18A–L19A–L37A has little helical structure (~9%) and lacks fixed tertiary interactions involving aromatic side chains, independent of the concentration of Na<sub>2</sub>SO<sub>4</sub> (Fig. 3a and b).

Superposition of the far-UV CD spectra of the variants (Fig. 3a) reveals an isodichroic point that is not shared by the wild-type protein. This suggests that the variants are in equilibrium between two or more conformations with distinct secondary structure, irrespective of the amino acid substitutions introduced. Significantly, the double mutant L53A

I54A, which has previously been shown to populate the I-state to greater than 90% in 0.4 M Na<sub>2</sub>SO<sub>4</sub>,<sup>32</sup> shares the isodichroic point of the variants (Fig. 3a). This suggests that the variants created are in equilibrium between the I and U states, with the population of each state being dependent on the concentration of Na<sub>2</sub>SO<sub>4</sub> and the amino acid substitutions introduced.

A cooperative unfolding transition is a key characteristic of natively folded proteins.<sup>40</sup> Accordingly, the absence of a cooperative unfolding transition upon the addition of urea was taken here as an additional criterion for population of an unfolded species. Urea titrations of each variant were performed using far-UV CD ( $\theta_{222\text{ nm}}$ ) and fluorescence emission ( $\lambda_{\text{max}}$ ) of the single tryptophan (W75) as structural probes (Fig. 3c and d). The data show that in the presence of 0.4 M Na<sub>2</sub>SO<sub>4</sub>, F84A and L18A–L19A exhibit evidence of a non-linear dependence of the CD and fluorescence signal on the denaturant concentration, indicative of an unfolding transition. However, L18A–L19A–L37A lacks evidence of significant nonlinearity in the urea titrations, demonstrating the absence of a detectable cooperative unfolding event (Fig. 3c and d). Consistent with the far- and near-UV CD spectra, in the



**Fig. 4.** <sup>1</sup>H–<sup>15</sup>N HSQC spectra (500 MHz) of Im7 variants. Spectra of F84A (a and b), L18A–L19A (c and d) and L18A–L19A–L37A (e and f) were acquired in buffer A (90% <sup>1</sup>H<sub>2</sub>O/10% <sup>2</sup>H<sub>2</sub>O) with 0 M Na<sub>2</sub>SO<sub>4</sub> (a, c and e) or 0.4 M Na<sub>2</sub>SO<sub>4</sub> (b, d and f) at 10 °C.

**Table 1.** Hydrodynamic properties of wild-type Im7 and the L18A–L19A–L37A variants at 10 °C

	$S_{20,w}^a$	$f/f_0$	$R_h$ (Å)	$C^b$	$MM_{\text{apparent}}$ (kDa)
Wild type <sup>c</sup>	1.45 ± 0.03	1.21 ± 0.02	17.8 ± 0.3	0.99	10.6 ± 0.3
L18A–L19A–L37A <sup>d</sup>	0.99 ± 0.18	1.79 ± 0.04	26.1 ± 0.6	0.28	13.3 ± 2.4
L18A–L19A–L37A <sup>d,e</sup> (0.4 M Na <sub>2</sub> SO <sub>4</sub> )	0.90 ± 0.03	1.70 ± 0.05	24.6 ± 0.8	0.41	10.4 ± 0.2
L18A–L19A–L37A <sup>d,f</sup> (6 M urea)	0.75 ± 0.04	2.33 ± 1.73	34.0 ± 1.7	<0	—

<sup>a</sup> Given as 10<sup>-13</sup> s.  
<sup>b</sup> Compaction factors calculated using the equation  $C = (R_D^h - R_h) / (R_D^h - R_N^h)$  as described by Wilkins *et al.*<sup>47</sup>  $R_N^h$  and  $R_D^h$  are the predicted values of hydrodynamic radii for the native and fully denatured states, respectively, and  $R_h$  is the experimentally determined hydrodynamic radius.  
<sup>c</sup> Error is standard deviation from three samples at 50 μM, 60 μM and 100 μM protein concentration.  
<sup>d</sup> Error is standard deviation from three samples at 50 μM, 60 μM and 400 μM protein concentration.  
<sup>e</sup> Sample in buffer A with 0.4 M Na<sub>2</sub>SO<sub>4</sub>.  
<sup>f</sup> Sample in buffer A with 6 M urea.

absence of Na<sub>2</sub>SO<sub>4</sub>, F84A and L18A–L19A appeared less stable than their Na<sub>2</sub>SO<sub>4</sub>-stabilised counterparts (data not shown). Interestingly, even though L18A–L19A–L37A lacks evidence of cooperative unfolding, there is a 2-nm blue shift in the  $\lambda_{\text{max}}$  of this variant in 0 M urea compared with 8 M urea in the presence of 0.4 M Na<sub>2</sub>SO<sub>4</sub> (Fig. 3c). This shift is not observed in the absence of Na<sub>2</sub>SO<sub>4</sub> (data not shown), consistent with Na<sub>2</sub>SO<sub>4</sub> inducing collapse of this highly destabilised variant.

### Sodium sulfate perturbs the unfolded ensemble

The data presented above suggest that the variants studied here either populate the unfolded state predominantly at equilibrium (L18A–L19A–L37A) or exhibit conformational exchange between the intermediate and the unfolded state (L18A–L19A and F84A). To assess the feasibility of characterising the unfolded ensemble of these proteins in more detail using NMR, we acquired the <sup>1</sup>H–<sup>15</sup>N heteronuclear single quantum coherence (HSQC) spectrum of each variant in 0 M urea and 0 M or 0.4 M Na<sub>2</sub>SO<sub>4</sub> at 10 °C (Fig. 4). The spectra of F84A show significant line broadening both in the presence and in the absence of Na<sub>2</sub>SO<sub>4</sub>, consistent with conformational exchange between different species occurring on the microsecond-to-millisecond timescale (Fig. 4a and b). The <sup>1</sup>H–<sup>15</sup>N HSQC spectrum of L18A–L19A in the absence of Na<sub>2</sub>SO<sub>4</sub> (Fig. 4c and d) shows sharp resonances and limited dispersion in the <sup>1</sup>H dimension, consistent with this variant predominantly populating the unfolded state in the absence of Na<sub>2</sub>SO<sub>4</sub> (Fig. 4c). However, significant line-broadening effects are observed for L18A–L19A in 0.4 M Na<sub>2</sub>SO<sub>4</sub>, in accord with the results presented in Fig. 3. Finally, the <sup>1</sup>H–<sup>15</sup>N HSQC spectrum of L18A–L19A–L37A in the absence of Na<sub>2</sub>SO<sub>4</sub> shows limited dispersion in the <sup>1</sup>H dimension and sharp resonances, characteristic of an unfolded protein (Fig. 4e). Upon the addition of 0.4 M Na<sub>2</sub>SO<sub>4</sub>, the <sup>1</sup>H–<sup>15</sup>N HSQC spectrum of L18A–L19A–L37A exhibits line broadening com-

pared with the spectrum in 0 M Na<sub>2</sub>SO<sub>4</sub> (Fig. 4f). This is likely caused by an additional contribution to the transverse relaxation rate caused by conformational exchange (i.e., U-to-I equilibrium) that is slow–intermediate exchange on the chemical shift timescale. The presence of Na<sub>2</sub>SO<sub>4</sub> may increase the population of the minor species in equilibrium with the unfolded state and/or alter the rate of exchange between species causing the line broadening observed in Fig. 4f. Overall, the L18A–L19A–L37A variant populates the unfolded state to the greatest extent. This variant was taken forward for more detailed structural analysis.

### The unfolded state is more compact than the urea-denatured state

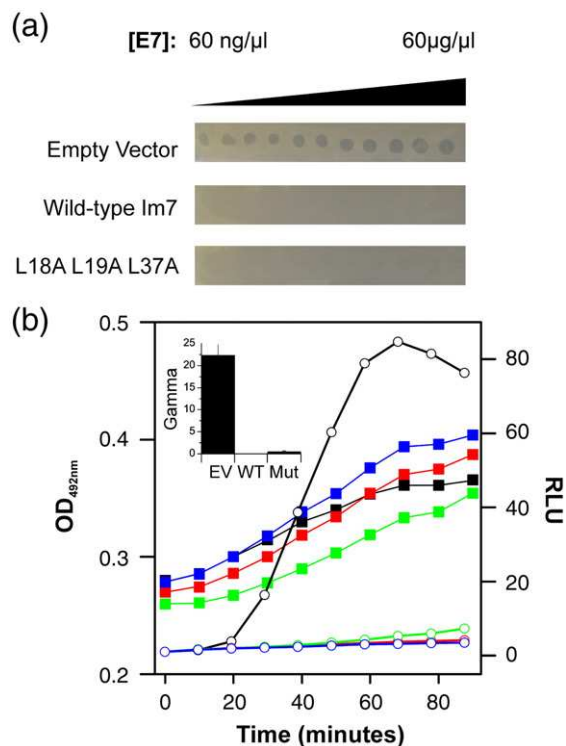
Compaction of the unfolded state of Im7 in the absence of denaturant has been observed previously using single-molecule experiments.<sup>24</sup> Compaction has also been observed in single-molecule experiments for many, although not all proteins.<sup>41</sup> Other proteins for which smFRET studies have reported collapse of the unfolded state as denaturant is diluted include RNaseH,<sup>42</sup> immunity protein 9,<sup>43</sup> cold shock protein CspTm,<sup>44</sup> chymotrypsin inhibitor,<sup>45</sup> acyl-co-enzyme binding protein<sup>45</sup> and barstar.<sup>46</sup> To obtain a quantitative measure of the extent of compaction of L18A–L19A–L37A, we performed sedimentation velocity analytical ultracentrifugation (AUC) experiments, from which an estimate of the radius of hydration ( $R_h$ ) in buffer A at 10 °C was made (Materials and Methods). The measured  $R_h$  value of the native wild-type protein is 17.8 ± 0.3 Å (Table 1), consistent with the value expected for a protein the size of Im7 using the empirical method of Wilkins *et al.* (17.7 Å).<sup>47</sup> The  $R_h$  measured for the L18A–L19A–L37A variant in buffer A with 6 M urea is 34.0 ± 1.73 Å, which is larger than 29.4 Å expected for an unfolded chain of this length (Table 1).<sup>47</sup> Nonetheless, these measurements are consistent with the urea-denatured state of Im7 being highly expanded, with dimensions that

indicate that the protein is as extended as expected for a random coil.<sup>47</sup> The  $R_h$  for L18A–L19A–L37A in the absence of urea in buffer A with 0 M or 0.4 M  $\text{Na}_2\text{SO}_4$ , determined using sedimentation velocity AUC, was  $26.1 \pm 0.6 \text{ \AA}$  and  $24.6 \pm 0.8 \text{ \AA}$ , respectively, consistent with these unfolded species being more compact than their urea-denatured counterparts. Minor differences in the  $R_h$  values of L18A–L19A–L37A in 0 M and 0.4 M  $\text{Na}_2\text{SO}_4$  are within experimental error. In comparison to the folded protein, the  $R_h$  of L18A–L19A–L37A is approximately 1.4 times the  $R_h$  measured for wild-type Im7. The frictional ratio for each species also supports the finding that L18A–L19A–L37A is more compact in the absence of urea (Table 1). Overall, these data suggest that L18A–L19A–L37A is approximately 30% more compact than its urea-denatured counterpart.

### Binding of ColE7 induces the folding of unfolded Im7

While the experiments described above suggest that L18A–L19A–L37A is unfolded, it is important to verify that the engineered unfolded sequence is capable of folding to the native state provided that suitable stabilising conditions can be found. Binding studies with ColE7 were performed using *in vivo* protection assays against ColE7 toxicity to investigate the potential of this variant to fold *in vivo*.<sup>48</sup> Since the amino acid substitutions introduced into the variants do not involve binding site residues,<sup>49</sup> the free energy of binding for Im7–E7 ( $\Delta G^\circ_{\text{binding}} = -83.2 \text{ kJ mol}^{-1}$  at 25 °C)<sup>36</sup> should provide sufficient energy to fold the variant, provided that the sequence is compatible with a native structure. A lawn of *E. coli* overexpressing the Im7 variant of interest was grown in the presence of differing concentrations of ColE7 to determine if this is the case for L18A–L19A–L37A.<sup>48</sup> When expression of wild-type Im7, or an Im7 variant, is induced by IPTG, cells survive only if Im7 is able to bind and inactivate the ColE7. Hence, zones of clearance are observed in this assay for cells carrying Im7 sequences incapable of folding and binding to ColE7. These experiments showed that expression of L18A–L19A–L37A in *E. coli* JM83 is able to provide protection against ColE7, consistent with binding-induced folding of L18A–L19A–L37A (Fig. 5a).

To further confirm the ability of L18A–L19A–L37A to provide protection against ColE7, we used a more sensitive, stress-induced reporter system (Materials and Methods) (Fig. 5b). This system utilises *E. coli* DPD1718, which has been engineered to contain a heterologous *luxCDABE* gene complex under the control of the cellular stress-induced *E. coli* *recA* promoter.<sup>50</sup> The *luxA*, *luxB*, *luxC*, *luxD* and *luxE* genes express the structural proteins responsible for bacterial bioluminescence. Hence, luminescence can be detected when the *luxCDABE* genes are



**Fig. 5.** *In vivo* cell protection experiments to assay colicin activity. (a) ColE7 was spotted onto lawns of *E. coli* carrying the empty pTrc99A vector or the same vector containing the gene for wild-type Im7 or L18A–L19A–L37A. Protein production was induced by the addition of 1 mM IPTG prior to the addition of the ColE7. The dark spots indicate regions where colicin has killed the lawn of *E. coli*, indicating the lack of inhibition of ColE7. (b) Cell growth monitored at OD<sub>492</sub> (filled squares) and relative luminescence units (RLU) (open circles) are shown for cultures of *E. coli* DPD1718 carrying the empty vector pTrc99A (black) or the same vector containing the genes for wild-type Im7 (red) or Im7 L18A–L19A–L37A (green) as a function of time after the addition of 4 nM ColE7. As a control, cells expressing wild-type Im7 were monitored without the addition of ColE7 (shown in blue). In all cases, 1 mM IPTG was used to induce expression of the Im7 gene 60 min prior to the addition of ColE7. Gamma values are shown in the inset and reflect the relative luminescence level induced by the addition of 4 nM ColE7 to cultures of *E. coli* DPD1718 carrying the empty vector pTrc99A (EV) or the same vector containing the gene for wild-type Im7 (WT) or the Im7 L18A–L19A–L37A variant (Mut). Gamma values were calculated as described in Materials and Methods. The error bars shown are the standard deviation from six repeat experiments.

expressed under the control of a stress-induced promoter activated by colicin-induced DNA damage.<sup>50</sup> To exploit this assay, we transformed *E. coli* DPD1718 cells with the pTrc99a vector containing the gene for L18A–L19A–L37A. Cells were grown to an OD<sub>600</sub> (optical density at 600 nm) of



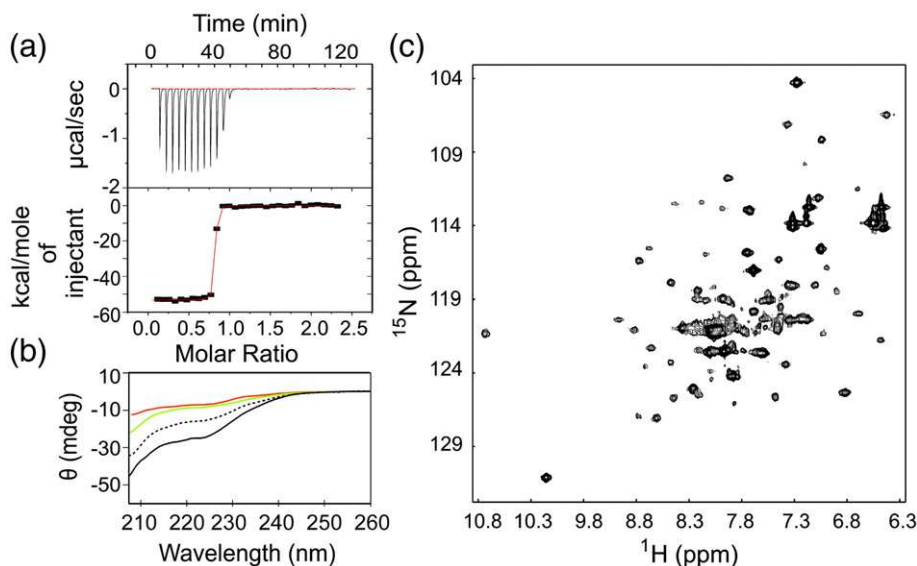
0.3, and protein expression was then induced by the addition of 1 mM IPTG for 1 h. The cells were then challenged by the addition of 4 nM ColE7, and luminescence, indicative of the induction of the stress response, was monitored (Fig. 5b). A “gamma value” was calculated to quantify colicin activity (Materials and Methods), with higher gamma values corresponding to a greater colicin activity within the cell.<sup>50</sup> These experiments also indicated that L18A–L19A–L37A provides protection against ColE7, to a level similar to that provided by wild-type Im7 (Fig. 5b). These results are in accord with the results of the colicin protection assays (Fig. 5a) and confirm that L18A–L19A–L37A is able to fold to a functional entity capable of inactivating ColE7 upon their binding.

To confirm binding-induced folding of L18A–L19A–L37A, we also performed *in vitro* binding studies of this variant to the DNase domain of ColE7 using far-UV CD, <sup>1</sup>H–<sup>15</sup>N NMR and isothermal titration calorimetry (ITC) (Fig. 6). Since Im7 binds only to the DNase domain of ColE7, the isolated ColE7 DNase domain was used for the *in vitro* binding experiments. ITC experiments (Fig. 6a) revealed that L18A–L19A–L37A binds to the DNase domain of ColE7 with a  $K_d$  in the subnanomolar range.<sup>51</sup> High-affinity binding is a common feature of cognate colicin DNase–Im protein complexes ( $K_d \leq 10^{-14}$  M), and the very tight binding

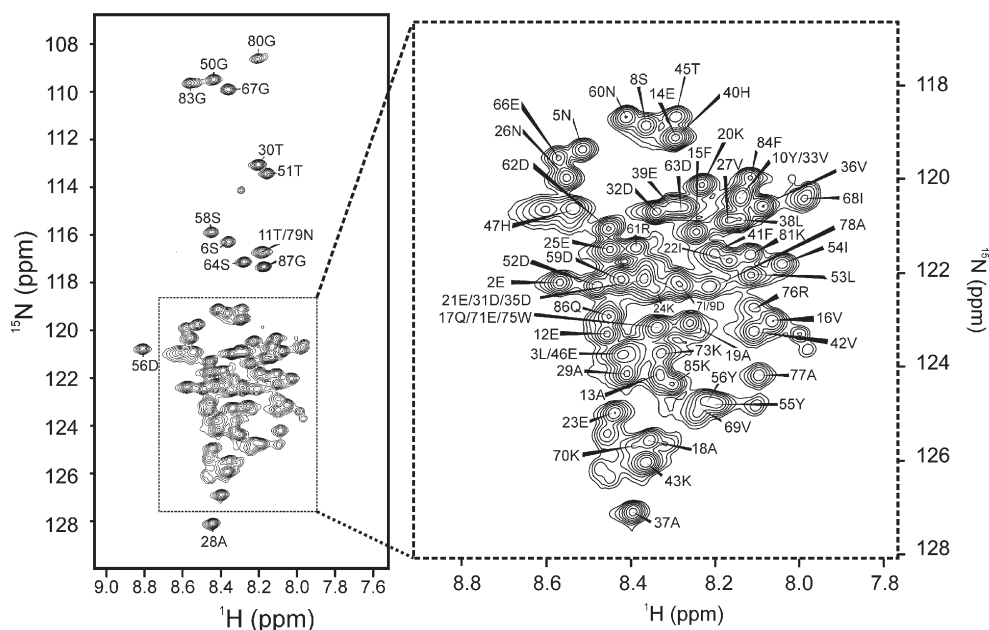
between L18A–L19A–L37A and ColE7 (which rules out accurate determination of the  $K_d$  using direct titration experiments) is consistent with these observations.<sup>52</sup> Such high-affinity binding is also consistent with the full protection against ColE7 observed *in vivo*.<sup>52</sup> Indicative of binding-induced folding, far-UV CD experiments revealed an increase in the helical content of L18A–L19A–L37A upon addition of the ColE7 DNase domain (Fig. 6b). The most compelling evidence for binding-induced folding of L18A–L19A–L37A, however, was obtained using <sup>1</sup>H–<sup>15</sup>N HSQC NMR, which showed a spectrum of the complex of unlabelled ColE7 DNase domain with <sup>15</sup>N-labelled L18A–L19A–L37A (Fig. 6c), which is consistent with the spectrum of a folded colicin-bound immunity protein observed by Whittaker *et al.*<sup>25</sup>

### Residue-specific characterisation of the structural properties of unfolded Im7

To determine the nature of residual structure in L18A–L19A–L37A and to characterise its dynamic properties, we obtained <sup>1</sup>H–<sup>15</sup>N HSQC spectra in buffer A with 0.2 M Na<sub>2</sub>SO<sub>4</sub>, pH 7.0, at 10 °C (Fig. 7). Under these conditions, line broadening of the resonances is reduced compared with that observed in 0.4 M Na<sub>2</sub>SO<sub>4</sub> (Fig. 4f). Although the folding pathway of Im7 has been shown to be independent



**Fig. 6.** *In vitro* assays of Im7 binding to the ColE7 DNase domain. (a) ITC binding isotherm obtained by the titration of Im7 L18A–L19A–L37A into the DNase domain of ColE7 (upper panel); the red line in the lower panel shows a fit of the data to a 1:1 binding model as a guide to the eye only (see Materials and Methods for details). (b) Far-UV CD spectra taken in buffer A at 10 °C of 40 µM ColE7 E7 DNase domain alone (green), 10 µM L18A–L19A–L37A alone (red) and the spectrum of both proteins mixed together (black). The black dotted line represents the summed spectra of ColE7 and L18A–L19A–L37A, indicating the expected result if the proteins do not interact. (c) <sup>1</sup>H–<sup>15</sup>N HSQC spectrum of a 500-µM solution of <sup>15</sup>N-labelled L18A–L19A–L37A mixed 1:1 with <sup>14</sup>N-labelled DNase domain of ColE7 at 10 °C in buffer A (90% <sup>1</sup>H<sub>2</sub>O/10% <sup>2</sup>H<sub>2</sub>O) with 0.2 M Na<sub>2</sub>SO<sub>4</sub>.



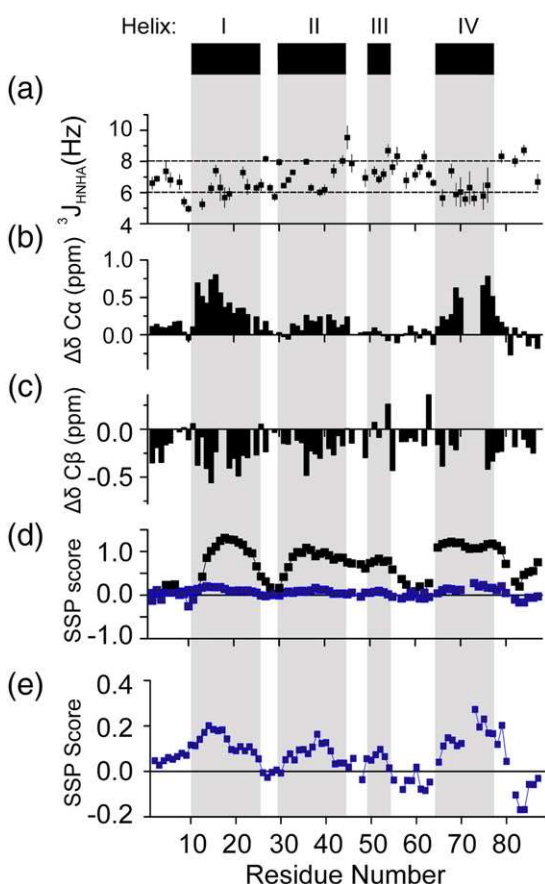
**Fig. 7.** The 500-MHz  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of L18A-L19A-L37A. The spectrum was acquired in buffer A (90%  $^1\text{H}_2\text{O}/10\%$   $^2\text{H}_2\text{O}$ ) with 0.2 M  $\text{Na}_2\text{SO}_4$  at 10 °C. Sequential NOEs are observed for NH groups of  $i$  and  $i+1$  residues for stretches of the protein involving residues 8-17, 18-37, 41-45, 51-52, 54-55, 60-63 and 67-69; however, no longer-range NOEs were observed in the three-dimensional HSQC-NOESY-HSQC experiment. The expanded region shows the assignments in crowded regions more clearly.

of the concentration of  $\text{Na}_2\text{SO}_4$ ,<sup>39</sup> these conditions were chosen to permit assignment in conditions that most closely emulate previous folding studies.<sup>24,26-28</sup> Using  $^{13}\text{C}/^{15}\text{N}$ -labelled protein, we assigned 82 of 83 non-proline residues of L18A-L19A-L37A in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of the protein (Fig. 7). Assignments for 81 CO, 84  $\text{C}^\alpha$ , 67  $\text{C}^\beta$  and 66  $\text{H}^\alpha$  resonances of the variant were also obtained [Biological Magnetic Resonance Data Bank (BMRB) entry 17513]. The chemical shifts were analysed by comparison with reference random-coil chemical shift values to ascertain whether helical structure persists in the unfolded variant.<sup>53,54</sup> This analysis revealed that no regions with regular secondary structure are present in the unfolded state (Fig. S1), consistent with analysis of this variant using far-UV CD (Fig. 3a). However, some small, systematic  $\text{C}^\alpha$  chemical shift deviations are observed in regions of the protein corresponding to the native helices I and IV, suggesting that the unfolded ensemble could sample  $\alpha$ -helical conformations transiently in these regions.

To obtain further evidence of whether secondary structure persists within the unfolded ensemble, we measured  $^3\text{J}_{\text{HNHA}}$  coupling constants for L18A-L19A-L37A in 0.2 M  $\text{Na}_2\text{SO}_4$ , pH 7.0, at 10 °C. A total of 56  $^3\text{J}_{\text{HNHA}}$  coupling constants could be measured, with values that range from 4.96 to 9.54 Hz (Fig. 8a). Of the 56  $^3\text{J}_{\text{HNHA}}$  coupling constants, 39 are less than 8 Hz (the minimum

value expected for ordered  $\beta$ -sheet structure) and 10 are less than 6 Hz (the maximum value expected for  $\alpha$ -helical structure).<sup>55</sup> The  $^3\text{J}_{\text{HNHA}}$  coupling constants also indicate a protein that lacks persistent secondary structure. Interestingly, the 10 residues with  $^3\text{J}_{\text{HNHA}}$  values that do fall below 6 Hz are found in regions of the sequence corresponding to the native helices I and IV, suggestive of an overall preference towards helical conformations in these regions (Fig. 8a).

The chemical shifts of L18A-L19A-L37A in 6 M urea were also obtained and used as a reference to compare with chemical shifts measured in 0 M urea. Using  $^{13}\text{C}/^{15}\text{N}$ -labelled protein, we assigned 87 residues in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of L18A-L19A-L37A in 0.2 M  $\text{Na}_2\text{SO}_4$  and 6 M urea at 10 °C (Fig. S2). Assignments for 82 CO, 82  $\text{C}^\alpha$  and 76  $\text{C}^\beta$  resonances were also obtained (BMRB entry 17513). The  $\text{C}^\alpha$  and  $\text{C}^\beta$  chemical shifts of L18A-L19A-L37A in buffer A with 0 M urea and 0.2 M  $\text{Na}_2\text{SO}_4$  were then compared with the chemical shifts obtained for the same protein in buffer A with 6 M urea and 0.2 M  $\text{Na}_2\text{SO}_4$  (Fig. 8b and c). The  $\text{C}^\alpha$  secondary shifts are extremely sensitive to secondary structure and exhibit downfield-shifted (positive) values in  $\alpha$ -helical regions.<sup>57</sup> Using this analysis, we observed positive  $\text{C}^\alpha$  secondary shifts for L18A-L19A-L37A in regions corresponding to the native helices I, II and IV (Fig. 8b). The  $\text{C}^\beta$  secondary shifts are also sensitive to secondary structure.<sup>57</sup> Interestingly, the



**Fig. 8.** NMR analysis of the residual secondary structure in L18A–L19A–L37A. (a)  $^3J_{\text{HNHA}}$  coupling constants measured in buffer A (90%  $^1\text{H}_2\text{O}/10\%$   $^2\text{H}_2\text{O}$ ) with 0.2 M  $\text{Na}_2\text{SO}_4$  at 10 °C. Error bars reflect the errors in determining the peak heights estimated from the spectral noise levels. Horizontal lines mark the upper and lower boundaries for residues lying in extended ( $\geq 8$  Hz) or helical regions ( $\leq 6$  Hz).<sup>55,56</sup> The positions of the helices in the native protein are indicated by the black bars above the plots with secondary structure taken from the X-ray structure of wild-type Im7 (PDB code: 1AY1<sup>33</sup>). Chemical shift differences between L18A–L19A–L37A in 0 M and 6 M urea for (b)  $\text{C}^\alpha$  and (c)  $\text{C}^\beta$  resonances measured in buffer A (90%  $^1\text{H}_2\text{O}/10\%$   $^2\text{H}_2\text{O}$ ) with 0.2 M  $\text{Na}_2\text{SO}_4$  at 10 °C. The values collected in 6 M urea have been subtracted from those collected in the absence of urea. (d) SSP analysis of L18A–L19A–L37A in buffer A with 0.2 M  $\text{Na}_2\text{SO}_4$  at 10 °C (blue) and wild-type Im7 (black) in buffer A. SSP analysis was performed using the published chemical shifts (BMRB entry: 7188) for wild-type Im7.<sup>29</sup> (e) The SSP analysis for L18A–L19A–L37A in buffer A with 0.2 M  $\text{Na}_2\text{SO}_4$  at 10 °C is shown expanded to highlight regions with significant, nonzero SSP scores.

$\text{C}^\beta$  secondary shifts of L18A–L19A–L37A are generally negative throughout the protein sequence (Fig. 8c). Although  $\text{C}^\beta$  chemical shifts are less sensitive to  $\alpha$ -helix formation,<sup>58</sup> the generally upfield-shifted  $\text{C}^\beta$  values are consistent with an

overall increase in  $\alpha$ -helical propensity upon dilution of the protein from denaturant (Fig. 8c).<sup>57</sup> Although there are clear deviations in the chemical shifts of L18A–L19A–L37A in 0 M and 6 M urea, the  $\text{C}^\alpha$  values all lie within 1 ppm of each other. By contrast, helix formation in native proteins generally results in  $\text{C}^\alpha$  chemical shifts that change by  $\sim 2.6$  ppm.<sup>57</sup> The small chemical shift differences of L18A–L19A–L37A in 0 M and 6 M urea indicate a significant, but weak, tendency towards helical structure in regions corresponding to the native helices I, II and IV in the absence of denaturant.

Next, the secondary structure propensity (SSP) scores were determined for L18A–L19A–L37A in 0 M urea using  $\text{C}^\alpha$  and  $\text{C}^\beta$  chemical shifts. These values were then compared with those determined using previously published  $\text{C}^\alpha$  and  $\text{C}^\beta$  chemical shifts for native wild-type Im7 in 0 M urea (BMRB entry 7188)<sup>29</sup> (Fig. 8d and e).<sup>59</sup> The SSP algorithm is designed to highlight conformational tendencies within an ensemble of unfolded or partially structured states.<sup>59</sup> An SSP score of 1 indicates 100%  $\alpha$ -helix propensity, whereas a score of  $-1$  indicates that the residue of interest has a 100% propensity to be in a  $\beta$ -strand conformation.<sup>59</sup> The SSP scores for native, wild-type Im7 accurately predict the regions that are known to be  $\alpha$ -helical in the native structure (Fig. 8d).<sup>33</sup> By contrast, the SSP scores for L18A–L19A–L37A are much smaller, indicating an overall lack of persistent secondary structure. Nonetheless, there are inclinations towards helical conformations. The average SSP scores for regions that correspond to the native helices in L18A–L19A–L37A are shown in Table 2. These values support the previous results, demonstrating that helices I and IV have the highest helical propensity in the unfolded state. Notably, the SSP score indicates some helical tendency in helix III, which is not formed until very late in folding. However, the 4.3% helical propensity predicted by SSP is very small in comparison to the values obtained for helices I and IV, and this propensity is not reflected in the  $^3J_{\text{HNHA}}$

**Table 2.** Helical propensity of the Im7 sequence

Helix	Residues <sup>a</sup>	Wild-type % helix <sup>b</sup>	Measured for isolated peptide <sup>c</sup>	SSP (L18A–L19A–L37A) (% helix) <sup>d</sup>
I	12–25	10.0	14.1	13.1
II	31–45	1.6	9.8	8.1
III	51–56	0.2	0	4.3
IV	65–79	5.4	11.2	16.3

<sup>a</sup> Native  $\alpha$ -helical regions in wild-type Im7 (taken from 1AY1<sup>37</sup>).

<sup>b</sup> AGADIR<sup>55</sup> prediction of the average helical content of wild-type Im7 for the region indicated at 10 °C, pH 7.0.

<sup>c</sup> Values quoted are from Knowling *et al.*<sup>28</sup>

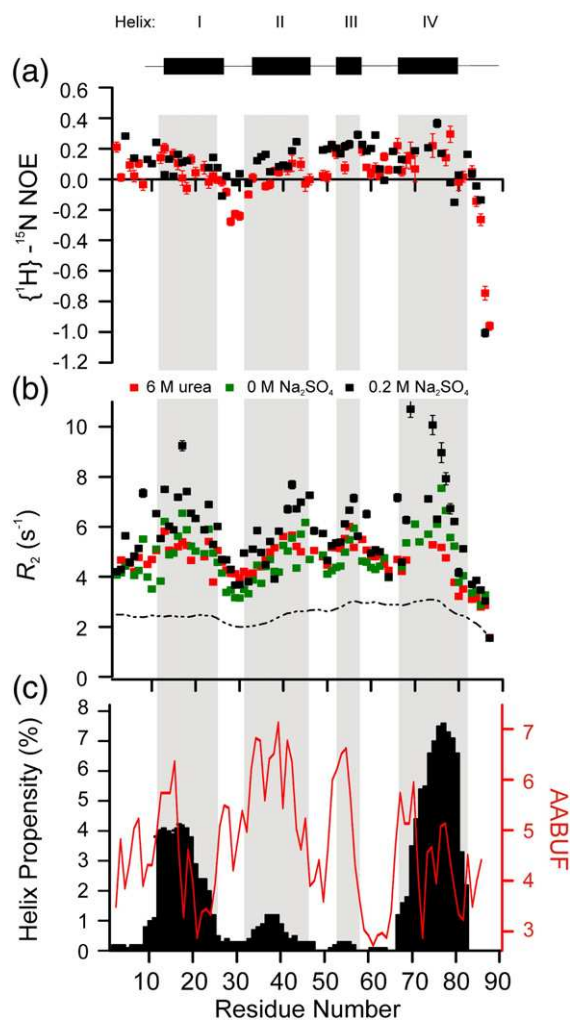
<sup>d</sup> Average SSP<sup>59</sup> score for this region of L18A–L19A–L37A, determined using  $\text{C}^\alpha$  and  $\text{C}^\beta$  chemical shifts.



coupling constants for residues for which their values could be determined (Fig. 8a). Estimates of helical propensity using the prediction algorithm AGADIR<sup>60</sup> and studies of synthetic peptides with sequences equivalent to the four helical regions of native Im7 also predict a weak helical propensity for these regions of the protein sequence that correlate with the SSP scores measured for L18A–L19A–L37A in 0 M urea (Table 2).<sup>28</sup> These results suggest that the helical structure present in the unfolded state of Im7 reflects the natural helical propensity of the polypeptide sequence, rather than additional helical structure that may be induced by the collapse of the polypeptide chain in the absence of chaotrope.

### Polypeptide chain dynamics of unfolded Im7

Backbone dynamics of L18A–L19A–L37A were investigated by measuring backbone transverse relaxation (<sup>15</sup>N  $R_2$ ) rates and <sup>1</sup>H–<sup>15</sup>N heteronuclear steady-state nuclear Overhauser enhancements (NOEs) (Fig. 9). <sup>1</sup>H–<sup>15</sup>N NOE data, determined for 62 residues, have values ranging from –0.96 to 0.29 (Fig. 9a). The values indicate flexibility throughout the protein, as all <sup>1</sup>H–<sup>15</sup>N NOEs are below the average value (+0.78) expected for the amide groups of a rigid globular protein that is tumbling isotropically.<sup>62</sup> The <sup>1</sup>H–<sup>15</sup>N NOEs measured in 6 M urea are very similar to those in 0 M urea (Fig. 9a), indicating that urea is not significantly modulating backbone dynamics that can be detected by this experiment (i.e., motion occurring on the picosecond-to-nanosecond timescale).<sup>63</sup> Na<sub>2</sub>SO<sub>4</sub> modulates line broadening in the <sup>1</sup>H–<sup>15</sup>N HSQC spectrum of L18A–L19A–L37A (Fig. 4e and f). Therefore, the <sup>15</sup>N  $R_2$  relaxation rates were measured in both 0 M and 0.2 M Na<sub>2</sub>SO<sub>4</sub>. <sup>15</sup>N  $R_2$  relaxation rates were measured for 75 out of 87 residues for this variant in 0 M Na<sub>2</sub>SO<sub>4</sub>/0 M urea, 76 out of the 87 residues of L18A–L19A–L37A in 0.2 M Na<sub>2</sub>SO<sub>4</sub>/0 M urea and 62 out of the 87 residues in 0.2 M Na<sub>2</sub>SO<sub>4</sub>/6 M urea (Fig. 9b). Analysis of the  $R_2$  rates reveals the presence of significant nonrandom structure in L18A–L19A–L37A indicated by increased  $R_2$  rates, relative to the intrinsic  $R_2$  values [estimated using Eq. (8)]. Interestingly, there is a small increase in the  $R_2$  rates of L18A–L19A–L37A in 0 M Na<sub>2</sub>SO<sub>4</sub>/0 M urea compared with the values obtained in 6 M urea, especially in residues located in helices I and IV (Fig. 9b). This could reflect decreased conformational flexibility in the unfolded state or an increase in the exchange between different species in the absence of the chaotrope. The  $R_2$  rates increase further in the presence of 0.2 M Na<sub>2</sub>SO<sub>4</sub>, consistent with further, now significant, conformational exchange between different species. Nonetheless, the clusters of residues with increased  $R_2$  rates in 0 M Na<sub>2</sub>SO<sub>4</sub>/0 M urea lie in regions that are helical in the native protein and correspond well



**Fig. 9.** Backbone <sup>15</sup>N relaxation rates and <sup>1</sup>H–<sup>15</sup>N NOEs of L18A–L19A–L37A measured in 0 M and 6 M urea. (a) <sup>1</sup>H–<sup>15</sup>N NOEs of the L18A–L19A–L37A in buffer A (90% <sup>1</sup>H<sub>2</sub>O/10% <sup>2</sup>H<sub>2</sub>O) with 0.2 M Na<sub>2</sub>SO<sub>4</sub> and 0 M urea (black), or 6 M urea (red). (b) <sup>15</sup>N transverse relaxation rates (50.66 MHz) in 0 M Na<sub>2</sub>SO<sub>4</sub> (green), 0.2 M Na<sub>2</sub>SO<sub>4</sub> (black) or 6 M urea with 0.2 M Na<sub>2</sub>SO<sub>4</sub> (red). All data were acquired in buffer A (90% <sup>1</sup>H<sub>2</sub>O/10% <sup>2</sup>H<sub>2</sub>O) at 10 °C. Data are plotted against residue number, with native helices indicated by the grey-shaded regions. The broken line in (b) represents the intrinsic <sup>15</sup>N  $R_2$  relaxation of a polypeptide chain with the sequence of L18A–L19A–L37A Im7 [determined using Eq. (8)]. Errors are shown but are smaller than the symbols for many of the data points. (c) Plots of AABUF (red) and the helix propensities (black) for L18A–L19A–L37A. AABUF values<sup>61</sup> were calculated with the ExpASY tool ProtScale (<http://www.us.expasy.org/tools/protscale.html>) and normalised from 0 to 10. Helix propensities were calculated with AGADIR (pH 7.0, 10 °C).<sup>60</sup>

with the average area buried upon folding (AABUF) (Fig. 9c). The AABUF is proportional to the hydrophobic contribution of a residue to the conformational free energy of a protein.<sup>61</sup> Previous



results have shown that the extent of residual secondary structure and AABUF correlate with sequence-dependent dynamic variations due to cluster formation in other denatured and unfolded states, such as acid-unfolded apomyoglobin,<sup>64</sup> the  $\Delta 131\Delta$  fragment of staphylococcal nuclease<sup>65</sup> and the lambda repressor denatured state.<sup>20</sup> This correlation for L18A–L19A–L37A suggests that hydrophobic interactions may be responsible for the deviations in  $R_2$  rates from those estimated from a random-coil model (Fig. 9b). The pattern of  $R_2$  rates of L18A–L19A–L37A in 0 M  $\text{Na}_2\text{SO}_4$ /0 M urea closely resembles the previously reported  $R_2$  rates of wild-type Im7 in 6 M urea, as well as those of L18A–L19A–L37A in 6 M urea (Fig. 9b).<sup>29</sup> Four clusters (centred at L18, V42, Y56 and K73, named clusters I, II, III and IV) with increased  $R_2$  rates were previously interpreted to indicate hydrophobic clustering within aliphatic/aromatic-rich regions of the Im7 sequence that encompass the helix-forming regions of the protein (Fig. 9c).<sup>29</sup> These results suggest that clustering of hydrophobic regions in the unfolded state of L18A–L19A–L37A is preserved even when the protein is dissolved in 6 M urea.

To determine whether the four hydrophobic clusters observed in L18A–L19A–L37A form long-range interactions, we created the amino acid substitutions F15A and F41A in L18A–L19A–L37A and we examined the effect of these substitutions on the dynamics of the entire polypeptide chain using  $^{15}\text{N}$   $R_2$  relaxation measurements. F15A was previously chosen as probe for long-range interactions within the urea-unfolded state of Im7, since this is the largest hydrophobic residue in cluster I. Should the clusters interact, a significant change in the dynamic properties of clusters II, III and IV would be expected to occur in this variant.<sup>29</sup> Similarly, F41A was chosen to probe for long-range interactions made by cluster II. Resonances of these variants were assigned by comparison with the spectrum of  $^1\text{H}$ – $^{15}\text{N}$  HSQC of L18A–L19A–L37A, and backbone  $^{15}\text{N}$   $R_2$  relaxation rates were determined for 69 residues of the variant F15A–L18A–L19A–L37A and 58 residues of the variant L18A–L19A–L37A–F41A (Fig. S3a and b). Chemical shift changes for resonances surrounding the hydrophobic amino acid substitution sites were observed, resulting in residues 13–16 in the F15A–L18A–L19A–L37A variant and residues 39–45 in the L18A–L19A–L37A–F41A variant not being assigned (Fig. S3a and b). Nonetheless, chemical shift differences for the residues that could be assigned in each of the two variants were small compared with those of L18A–L19A–L37A, with F15A–L18A–L19A–L37A exhibiting average chemical shift changes of  $\Delta\delta_{\text{N}}=0.16\pm 0.14$  ppm and  $\Delta\delta_{\text{H}}=0.03\pm 0.03$  ppm, while L18A–L19A–L37A–F41A exhibits average chemical shift differences of  $\Delta\delta_{\text{N}}=0.06\pm 0.06$  ppm and  $\Delta\delta_{\text{H}}=0.01\pm 0.02$  ppm. Similarly, no significant change in  $R_2$  relaxation rates was observed in regions

distant from the amino acid substitution site, while small deviations in  $R_2$  values close to the site of substitution were observed (Fig. S3). Together, the data suggest that the conformational properties of these variants are very similar to those of L18A–L19A–L37A and indicate that stabilising, intercluster hydrophobic interactions do not persist in the unfolded state of Im7 in the absence of urea. Instead, local clustering and hydrophobic residues are consistent with the chemical shift changes and deviations in  $R_2$  values close to the substitution site. Local hydrophobic clustering rather than long-range hydrophobic interactions were also observed for the wild-type Im7 in 6 M urea.<sup>29</sup> Taken together, the data suggest that L18A–L19A–L37A is a dynamic ensemble of species that contains four clusters of locally interacting residues that lack persistent long-range hydrophobic interactions between the clusters but retain a small, overall preference for  $\alpha$ -helical secondary structure.

## Discussion

### The unfolded state of Im7 is collapsed but lacks persistent secondary structure

The results above demonstrate that the sequence substitutions L18A, L19A and L37A destabilise Im7 significantly such that the unfolded state is predominantly populated under ambient conditions. The isodichroic point observed in the far-UV CD spectrum of all three variants in Fig. 3a indicates that all three destabilised variants are in conformational exchange between the unfolded state of Im7 and a more structured species. In addition, the similarity between the  $^1\text{H}$ – $^{15}\text{N}$  HSQC spectrum of the L18A–L19A and the L18A–L19A–L37A variant in the absence of  $\text{Na}_2\text{SO}_4$  provides compelling evidence, that despite the differences in amino acid substitutions, both the proteins populate comparable unfolded states. Furthermore, the ability of L18A–L19A–L37A to populate a native-like conformation has been demonstrated, indicating that the amino acid substitutions introduced into the protein have not disrupted the sequence information required to adopt the native fold. Together, these results suggest that the L18A–L19A–L37A variant is an excellent mimic of the unfolded state of Im7 in ambient conditions.

Measurements of the properties of this species using far- and near-UV CD, fluorescence and NMR indicate that this protein lacks significant secondary structure and persistent tertiary interactions. Nonetheless, the ensemble possesses an overall residual helicity of approximately 6–9% as measured by far-UV CD spectroscopy and NMR (overall SSP score) that results from the inherent properties of the Im7 sequence (local helical propensity and hydrophobicity). The unfolded state created here is approximately

30% more compact than its counterpart in 6 M urea, qualitatively in agreement with observations made by smFRET studies that revealed compaction of the unfolded state in the absence of urea.<sup>24</sup> The overall low helicity of the unfolded ensemble of L18A–L19A–L37A is consistent with measurements of the helical propensities of synthetic peptides with sequences corresponding to the individual helices of Im7 and using the secondary-structure prediction algorithm AGADIR.<sup>28,60</sup> Considering all of the experimental data obtained, the unfolded state of Im7 is envisaged to be collapsed, to lack persistent long-range interactions and to possess only inklings of helical structure.

### Implications for the folding of Im7

Recent extensive simulations of rapidly folding proteins suggest that there is a strong correlation between helical propensity within the unfolded state and the order of formation of local native-like structure along the folding transition path.<sup>66</sup> Additionally, the formation of only a subset of key long-range native contacts in the unfolded state may be sufficient to establish a native-like topology and stabilize the transiently formed secondary structure.<sup>66</sup> Despite lacking persistent interactions, residual structure in the unfolded state of Im7 could contribute to the determination of the rate and/or mechanism of folding to the native state. In this context, it is interesting to note that all four regions corresponding to helices in native Im7 populate helical conformations (Table 2). It is somewhat surprising that helix III shows a tendency towards  $\alpha$ -helical conformations given this is the last helix to form during folding and has little, if any, predicted helical propensity (Table 2). Studies on the folding mechanism of Im7 have shown that helices I and II form crucial interactions within the first transition state (TS1) for folding (Fig. S1), specifically involving residues L18, L19 and L37.<sup>27</sup> Despite substituting these three leucine residue with alanine in L18A–L19A–L37A, the sequence is still able to fold to a native conformation provided that suitable stabilising conditions are formed (herein, provided by binding to Cole7). Conformational exchange observed in the presence of Na<sub>2</sub>SO<sub>4</sub> suggests that L18A–L19A–L37A is able to fold into the intermediate state, albeit transiently. Previous  $\Phi$ -value analysis has shown that as Im7 folds from TS1 to the I-state, helix IV forms nonnative interactions with helices I and II, precluding the docking of helix III.<sup>27</sup> Helix IV is the most populated helix within L18A–L19A–L37A and its pre-organisation may contribute to the folding mechanism. Together, these data confirm the validity of using L18A–L19A–L37A as a mimic of the unfolded state of the wild-type protein, since the sequence is able to fold to both the on-pathway intermediate and native states.

Comparison of the results presented here, along with previous studies of urea-denatured Im7,<sup>29</sup> reveals that the unfolded states of Im7 in 0 M and 6 M urea have remarkably similar structural properties, since neither ensemble has significant secondary structure, both contain four elements of local hydrophobic clusters and neither contains persistent tertiary interactions. Nonetheless, significant changes in the unfolded state of Im7 occur upon dilution from denaturant. Firstly, the unfolded state becomes more compact in the absence of urea. Additionally, an overall increase in helicity, particularly in helices I and IV, occurs when the chaotrope is removed, which could drive, or be driven by, the concomitant collapse.<sup>67</sup> In the case of the wild-type unfolded sequence, the formation of stabilising long-range interactions, particularly involving side chains in helices I and II, would trigger folding.<sup>26,27</sup> Transient long-range interactions within the unfolded state of Im7 have not yet been probed and further investigation will be required to reveal whether such interactions prime the ensemble to fold by transiently sampling the persistent interactions that are formed in TS1. Persistent interactions define TS1 and, thereby, mark the first stable interactions that commit the sequence of Im7 to fold to a compact globular structure.

### Comparison with residual structure in other model proteins

Compaction of the polypeptide chain under non-denaturing conditions has been observed for many protein sequences placed under conditions that do not permit folding.<sup>10,19,68,69</sup> For example, a radius of gyration or hydration that is approximately 1.3–1.5 times that of the folded state has been reported for the *Drosophila* drk N-terminal (drkN) domain,<sup>10</sup> the N-terminal (NTL9) and the C-terminal domain (CTL9) of ribosomal protein L9<sup>19,68</sup> and the  $\Delta$ 131 $\Delta$  fragment of staphylococcal nuclease.<sup>65,69</sup> These measurements of compaction are comparable to the observations made here for L18A–L19A–L37A, for which the  $R_h$  is 1.4 times larger than that of native wild-type Im7. Interestingly, Anil *et al.* demonstrated that the same unfolded sequence could have markedly different levels of secondary structure yet share the same  $R_h$  values, suggesting that compaction should not be considered diagnostic of residual structure.<sup>19</sup> Furthermore, Kohn *et al.* have shown that even a random unfolded polypeptide chain exhibits compaction compared with its corresponding urea-denatured state, indicating that compaction of the unfolded state need not be mediated by sequence-dependent interactions.<sup>70</sup> Collapse is likely mediated by hydrophobic and electrostatic interactions that could be driven by a combination of nonspecific structural properties of the protein and also specific interactions that could be important for folding.<sup>71</sup> Further

experiments, for example, using paramagnetic relaxation enhancements and relaxation dispersion NMR methods combined with detailed simulation methods, such as those studies conducted in the Forman-Kay,<sup>10,72</sup> Vendruscolo,<sup>21</sup> Poulsen<sup>73</sup> and Blackledge<sup>74</sup> and co-workers, will be required to define the origins of the residual structure observed in the unfolded state of L18A–L19A–L37A in more detail.

Although it is difficult to classify and compare the extent of residual structure within unfolded states of different protein, it is becoming clear that these ensembles form a broad range of structures.<sup>76</sup> The extent of observed residual structure ranges from highly disordered states (e.g., chymotrypsin inhibitor 2<sup>76</sup>) to denatured states that possess significant residual structure and persistent tertiary contacts in non-denaturing conditions, such that they can be classed as folding intermediates (e.g., engrailed homeodomain<sup>18</sup>). The results presented here suggest that the unfolded state of Im7 lacks persistent structure and is nearer the “disordered” end of this spectrum of unfolded protein conformers. The lack of persistent helical structure in the unfolded state of Im7 is consistent with a nucleation–condensation mechanism for the intermediate formed from the unfolded state that defines the first stage of Im7 folding. In members of the homeodomain family, the extent of residual helical structure in the denatured state corresponds to a spectrum of folding mechanisms from diffusion–collision for highly structured denatured proteins to nucleation–condensation in proteins that lack significant helical propensity in their denatured state.<sup>77</sup> Hydrophobic interactions have also been implicated as important sites for initiation of protein folding.<sup>5</sup> The results presented here add to the emerging understanding of residual structure in the unfolded ensembles of  $\alpha$ -helical proteins. A combination of secondary structure propensity and hydrophobic interactions determines the nature of subsequent folding steps taken and ultimately leads to the malleability observed in the transition states populated during folding.<sup>77</sup> In the case of Im7, both hydrophobic clustering in specific regions of the polypeptide chain and local helical propensity define the starting point of folding, which ultimately leads to an efficient search to folded species that are required for colicin binding and cell survival.<sup>36,37</sup>

## Materials and Methods

### Expression, purification and mutagenesis

All variants were created from wild-type Im7 containing an N-terminal hexahistidine tag.<sup>30</sup> Mutants were created using the Stratagene QuikChange Site-Directed Mutagenesis Kit following the manufacturer's instructions, except that the incubation time with DpnI was increased from 1 h to overnight. Genes were sequenced to confirm that

mutagenesis had been successful. Proteins were over-expressed in *E. coli* cells and purified as described previously,<sup>38</sup> with the exception that anion-exchange chromatography was carried out using Source 15Q resin (GE Healthcare). Protein was loaded onto the column in 50 mM sodium phosphate (pH 6.0), the column was washed with 2 column volumes of this buffer and protein was eluted with a gradient of 0–0.65 M NaCl. <sup>15</sup>N/<sup>13</sup>C-labelled proteins were expressed in BL21(DE3) cells and grown in <sup>15</sup>N/<sup>13</sup>C-enriched medium containing 1 g/L of <sup>15</sup>NH<sub>4</sub>Cl and 2 g/L of uniformly labelled <sup>13</sup>C-glucose. Protein was shown to be > 95% pure using SDS-PAGE and of the expected mass (determined using electrospray ionisation mass spectrometry). Unless indicated otherwise, all proteins were dissolved in buffer A containing 50 mM sodium phosphate and 1 mM EDTA (pH 7.0).

### Circular dichroism

Data were collected on a Jasco J715 CD spectropolarimeter (Great Dunmow, Essex, UK). The CD signal for measurements in the far-UV region (200–260 nm) was recorded in a 1-mm-path-length cell with a protein concentration of 0.2 mg/ml. Far-UV CD spectra were recorded using a 1-nm bandwidth, a 1-nm resolution, a 20-nm/min scan speed and a response time of 8 s, averaging 5 scans. Measurements in the near-UV CD spectrum (250–350 nm) were recorded at a protein concentration of 1.5 mg/ml using a 1-cm-path-length cell, with a 1-nm bandwidth, a 0.5-nm resolution, a 10-nm/min scan speed and a response time of 4 s, averaging over 12 scans. The measured mean residue ellipticity was used to estimate helical content, using Eq. (1). The scaling factor 50 is included as 50% of the residues in wild-type Im7 are in a helical conformation, as defined in the Protein Data Bank (PDB) entry 1AYI.<sup>33</sup>

$$\% \text{ Helical Content} = \left( \frac{\text{MRE}_{222 \text{ nm}}^{\text{variant}}}{\text{MRE}_{222 \text{ nm}}^{\text{Im7}}} \right) \times 50 \quad (1)$$

### Equilibrium denaturation curves

Data were collected on a Photon Technology International fluorimeter (Ford, West Sussex, UK), with a 1-cm-path-length cuvette or a Jasco J715 CD spectropolarimeter with a 1-mm-path-length cuvette. Solutions of protein containing 0–8 M urea in 0.2-M increments were used, with a final protein concentration of 0.1 mg/ml for fluorescence samples and 0.2 mg/ml for far-UV CD samples. All samples were incubated overnight at 10 °C before measurements were taken. The CD signal at 222 nm was measured using a bandwidth of 1 nm, a response time of 1 s and the average signal taken over 1 min. Data for the wild-type protein were fitted to a model describing a two-state transition:

$$\text{Signal} = \frac{\left( (a[\text{urea}] + b)\exp(\Delta G_{\text{UF}}^{\circ} - M_{\text{UF}}[\text{urea}]) / RT + (c[\text{urea}] + d) \right)}{\left( 1 + \exp(\Delta G_{\text{UF}}^{\circ} - M_{\text{UF}}[\text{urea}]) / RT \right)} \quad (2)$$

where  $\Delta G_{\text{UF}}^{\circ}$  (kJ mol<sup>-1</sup>) is the equilibrium stability,  $M_{\text{UF}}$  is the equilibrium  $m$ -value,  $a$  and  $c$  represent the denaturant



dependence of the folded and unfolded signal intensities, respectively, and  $b$  and  $d$  are the signal intensities of the folded and unfolded states, respectively, in the absence of denaturant. The fluorescence intensity was an insensitive probe of protein folding for the variants and therefore fluorescence emission spectra were monitored, using an excitation wavelength of 280 nm. Fluorescence emission was recorded from 300 to 400 nm in 1-nm increments, integrating for 1 s at each wavelength.

### Analytical ultracentrifugation

Sedimentation velocity experiments were carried out with a Beckman Optima XL-I analytical ultracentrifuge (Beckman, Palo Alto, CA) using an An-50 Ti or An-60 Ti rotor at 10 °C, 50,000 rpm, and using conventional aluminum double-sector centerpieces with sapphire windows. Samples were prepared by dissolving lyophilised protein in buffer A followed by overnight dialysis against this buffer. The dialysate was used as a buffer blank solution. Radial absorbance scans ( $\lambda = 280$  or 296 nm) were collected at 300-s intervals. The sedimentation coefficients ( $S$ ) were analysed in terms of a size distribution function  $c(S)$  using the program SEDFIT.<sup>78</sup> Experimentally determined sedimentation coefficients were corrected to standard conditions (20 °C) in water to give corrected sedimentation coefficient ( $S_{20,w}$ ) values using the solvent density ( $\rho$ ) and protein partial specific volumes ( $\bar{v}$ ), which were calculated using SEDNTERP software.<sup>79</sup>  $S_{20,w}$  values were used according to the Svedberg equation to determine the translational friction coefficient ( $f$ ) of the protein using Eq. (3), where  $M$  is the molecular weight of the protein,  $\bar{v}$  and  $\rho_{20,w}$  are the partial specific volume of the protein and solvent density, respectively, calculated under standard conditions and  $N_A$  is Avogadro's number. Sedimentation coefficients are reported in Svedberg units ( $S$ ), which correspond to  $10^{-13}$  s.<sup>80</sup>

$$S_{20,w} = \frac{M(1 - \bar{v}\rho_{20,w})}{N_A f} \quad (3)$$

Experimentally determined frictional coefficients were compared with the frictional coefficient that would be expected for a rigid, hydrated spherical molecule of equal volume ( $f_0$ ) determined using Eq. (4), based on the Stokes equation,<sup>80</sup> where  $\eta$  is the viscosity of the solution and  $R_0$  is the radius of the sphere.

$$f_0 = 6\pi\eta R_0 \quad (4)$$

Frictional ratios ( $f/f_0$ ) provide a measure of the compaction of the protein compared with a perfect sphere.<sup>80</sup> Since the frictional coefficient of the protein is proportional to the radius of hydration ( $R_h$ ), it has been used to determine an estimate of the  $R_h$  for the proteins studied, by comparison with the Stokes radius ( $R_0$ ) using Eq. (5).

$$\frac{f}{f_0} \cdot R_0 = R_h \quad (5)$$

### In vivo colicin assays

JM83 *E. coli* cells were transformed with a pTrc99a plasmid encoding the Im7 gene of interest and inoculated into 0.7% LB agar containing 50  $\mu$ g/ml carbenicillin and 1 mM IPTG. This was poured onto a precast 1.5% LB agar plate, and once set, a serial dilution of E7 colicin was spotted onto the plate. This assay is described by Wallis *et al.*<sup>48</sup>

### In vivo LuxA assays

This assay makes use of an SOS-inducible chromosomal *lux* operon to detect DNA damage induced by ColE7 in reporter cells.<sup>50</sup> *E. coli* DPD1718 contain a fusion of the *E. coli* *recA* promoter region with the *Photobacterium luminescens* *luxCDABE* reporter integrated into the *lacZ* locus of *E. coli* DPD1692.<sup>50</sup> *E. coli* DPD1718 were transformed with a pTrc99a plasmid encoding the Im7 gene of interest, and overnight cultures were grown in LB at 37 °C in 10-ml culture volumes. Cultures were then diluted 1:50 into a final volume of 100 ml of LB media and grown for 2 h at 37 °C in the presence of 30  $\mu$ g/ml chloramphenicol and 50  $\mu$ g/ml carbenicillin. We then added 1 mM IPTG to the cultures, and they were allowed to grow for another hour until an OD<sub>600</sub> of 0.4–0.5 was reached within the culture flasks. The cells were then diluted 1:2 and 100  $\mu$ l was added to a 96-well optical bottom microtitre plate (Corning®) prewarmed to 37 °C. ColE7 (4  $\mu$ l; final concentration, 4 nM) was then added. All assays were performed in a PerkinElmer EnVision microtitre plate reader. Induction of luminescence was followed over a period of 90 min, with readings taken every 600 s. Cell densities were monitored by measuring the OD at 492 nm. The gamma values were determined at the 50 -min time point according to Eq. (6), where  $L_{\text{sample}}$  is the luminescence value of the cell strain of interest corrected for cell density and  $L_{\text{control}}$  is the cell density corrected luminescence value of control cells that were not challenged with ColE7.

$$\gamma = \frac{L_{\text{sample}} - L_{\text{control}}}{L_{\text{control}}} \quad (6)$$

### Isothermal titration calorimetry

Experiments were carried out using a VP-ITC (Microcal) calorimeter following the manufacturer's instructions. Briefly, lyophilised protein samples were extensively dialysed overnight at 4 °C in the assay buffer (50 mM MOPS at pH 7.0 containing 1 mM EDTA). The assay buffer and proteins were filtered (0.22  $\mu$ M pore size) and diluted to the working concentration (10  $\mu$ M ColE7 DNase domain and 120  $\mu$ M L18A–L19A–L37A) and degassed immediately prior to use. Experiments were performed at 298 K with an initial injection of 5  $\mu$ l followed by a series of injections of 7.5  $\mu$ l. The resulting binding isotherm was fitted to a 1:1 binding model using the software provided by the manufacturer, although the high affinity of the complex prohibited accurate measurement of the  $K_d$  from this experiment.

### NMR spectroscopy

All NMR experiments were carried out at 10 °C using Varian Unity Inova spectrometers operating at <sup>1</sup>H

† Available at <http://www.rasmb.bbri.org>



frequencies of 500, 600 and 750 MHz. Protein samples in the range of 0.2–0.5 mM were prepared in buffer with 10% D<sub>2</sub>O and in the presence of 0 M, 0.2 M or 0.4 M Na<sub>2</sub>SO<sub>4</sub>, or 0.2 M Na<sub>2</sub>SO<sub>4</sub> and 6 M urea, as indicated. Gradient-enhanced <sup>1</sup>H–<sup>15</sup>N HSQC spectra were acquired using 128 complex points and 16 scans per increment with spectral widths of 8511 Hz and 1800 Hz in the <sup>1</sup>H and <sup>15</sup>N dimensions, respectively. Watergate solvent suppression was used, and all NMR data were processed using NMRPipe and analysed in NMRView.<sup>81,82</sup>

Although the Im7 construct used has an additional histidine tag at the N-terminus, assignments have been labelled according to the corresponding residue number in the wild-type untagged sequence. Peak assignments of the L18A–L19A–L37A Im7 variant were determined in the presence of 0.2 M Na<sub>2</sub>SO<sub>4</sub> and in the same buffer with the addition of 6 M urea. <sup>1</sup>H chemical shifts were referenced directly to external DSS, and the <sup>13</sup>C and <sup>15</sup>N chemical shifts were referenced indirectly to DSS. <sup>15</sup>N, <sup>13</sup>C-labelled proteins were used to obtain sequential connectivity of C<sup>α</sup> and C<sup>β</sup> resonances by performing HNCACB, CBCA(CO)NH, HN(CO)CA and HNCA experiments. The assignments of the L18A–L19A–L37A Im7 variant in the absence of urea were further confirmed by performing a 120-ms three-dimensional <sup>1</sup>H–<sup>15</sup>N HSQC-NOESY (nuclear Overhauser enhancement spectroscopy)-HSQC experiment. An HNHA experiment was also performed to obtain H<sup>α</sup> resonance assignments for L18A–L19A–L37A in the absence of urea. Chemical shift differences were calculated using the temperature, pH and sequence-corrected chemical shifts of Kjaergaard *et al.*<sup>53,54</sup> SSP analysis was performed using C<sup>α</sup> and C<sup>β</sup> chemical shifts using the algorithm developed by Marsh *et al.* using the software provided.<sup>59</sup>

Homonuclear three-bond J<sub>(HNHA)</sub> coupling constants were determined by an HNHA experiment on <sup>15</sup>N-labelled L18A–L19A–L37A in buffer A (90% <sup>1</sup>H<sub>2</sub>O/10% <sup>2</sup>H<sub>2</sub>O) in the presence of 0.2 M Na<sub>2</sub>SO<sub>4</sub> at 10 °C. A mixing time (2Δ) of 25.08 ms was used in the pulse sequence described by the method developed by Vuister and Bax.<sup>83</sup> Coupling (J<sub>NH</sub>) constants can be determined by the diagonal (*I*<sub>diag</sub>) and cross-peak (*I*<sub>cross</sub>) intensities using Eq. (7), where Δ is the experimental delay in the pulse sequence.<sup>83</sup> Errors were determined from the error in measuring the peak heights determined from the noise level of the spectra.

$$\frac{I_{\text{cross}}}{I_{\text{diag}}} = -\tan^2(2\pi J_{\text{NH}}\Delta) \quad (7)$$

Backbone <sup>15</sup>N transverse relaxation measurements were carried out as previously described,<sup>62</sup> for samples in buffer A. An estimate of the error was determined using duplicate points and spectral noise levels. Relaxation measurements were performed at 500 MHz using a series of 11 experiments with mixing times ranging from 1.60 to 191.6 ms. Backbone <sup>15</sup>N R<sub>2</sub> relaxation rates were modelled and estimated using the approach previously described in Le Duff *et al.*<sup>29</sup> The per-residue sequence-dependent intrinsic R<sub>2</sub> values due to polymer hydrodynamics were calculated from the radius of gyration of each residue and its neighbours modulated by an exponential decrease in

persistence along the chain. Sequence-dependent relaxation rates in the absence of clustering were given by Eq. (8):

$$R_{2i} = k_{R_{2i}} \cdot \sum_{j=1}^N \tau_j \cdot e^{-\frac{|i-j|}{\lambda_j}} \quad (8)$$

where *k*<sub>R<sub>2i</sub></sub> is an empirical scaling coefficient, R<sub>2i</sub> is the intrinsic transverse relaxation rate of each residue, *N* is the number of residues and τ<sub>*j*</sub> and λ<sub>*j*</sub> are the intrinsic correlation time of each residue and the persistence length for segmental motion of the polypeptide, respectively. It was assumed that λ<sub>*j*</sub> was 2 for motion deriving from alanine and glycine and 7 for all other residue types. The intrinsic correlation time for each residue, τ<sub>*j*</sub>, was calculated from Stokes law using Eq. (9):

$$\tau_j = k_{R_g} \cdot R_g^3 \quad (9)$$

where *k*<sub>R<sub>g</sub></sub> is an arbitrary scaling factor that was amalgamated with *k*<sub>R<sub>2i</sub></sub> and R<sub>*g*</sub> is the radius of gyration for each residue. Values of R<sub>*g*</sub> were obtained from Levitt,<sup>84</sup> via the database<sup>§</sup> with the exception that the R<sub>*g*</sub> for proline was increased from 1.25 Å to 2 Å to allow for the reduction of conformational space due to ring closure in the peptide, previously described in the literature.<sup>85</sup>

Heteronuclear NOEs were measured with the procedure and pulse sequences described by Farrow *et al.*<sup>62</sup> Proton saturation was achieved with a 120° pulse applied every 5 ms over 3 s during the 5-s relaxation delay. NOE values were determined as the ratio of the average peak heights with and without proton saturation, with the uncertainties of the NOE values taken to be error in measuring the peak heights determined from the noise level of the spectra.

#### Data bank accession number

Sequence-specific assignments for L18A–L19A–L37A in buffer A with 0.2 M Na<sub>2</sub>SO<sub>4</sub> and 0 M or 6 M urea have been deposited in the BMRB as entry 17513.

#### Acknowledgements

We thank David Brockwell, Alice Bartlett and all members of the Radford and Brockwell groups for many helpful discussions. Many thanks to Richard James for providing the bacterial strain for the LuxA assay. We also acknowledge Amy Barker and Iain Manfield for their assistance with AUC and ITC experiments, respectively. We are grateful to Stuart Warriner and James Titchmarsh who provided the plate reader for luminescence readings. Joerg Gspöner kindly provided images for Fig. 1. This work was supported by Wellcome Trust (grants 062164/z/00/z and 075675) and the Biotechnology

‡ <http://pound.med.utoronto.ca/software.html>

§ <http://www.genome.jp/aaindex>

and Biological Sciences Research Council (grants BB/G019452/1 and BB/526502/1).

## Supplementary Data

Supplementary data to this article can be found online at [doi:10.1016/j.jmb.2011.12.041](https://doi.org/10.1016/j.jmb.2011.12.041)

## References

- Bartlett, A. I. & Radford, S. E. (2009). An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat. Struct. Mol. Biol.* **16**, 582–588.
- Levinthal, C. (1968). Are there pathways for protein folding. *J. Med. Phys.* **65**, 44–45.
- Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19.
- Morrone, A., McCully, M. E., Bryan, P. N., Brunori, M., Daggett, V., Gianni, S. & Travaglini-Allocatelli, C. (2011). The denatured state dictates the topology of two proteins with almost identical sequence but different native structure and function. *J. Biol. Chem.* **286**, 3863–3872.
- 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. USA*, **105**, 6278–6283.
- Modig, K., Jurgensen, V. W., Lindorff-Larsen, K., Fieber, W., Bohr, H. G. & Poulsen, F. M. (2007). Detection of initiation sites in protein folding of the four helix bundle ACBP by chemical shift analysis. *FEBS Lett.* **581**, 4965–4971.
- Bowler, B. E. (2011). Residual structure in unfolded proteins. *Curr. Opin. Struct. Biol.* [doi:10.1016/j.sbi.2011.09.002](https://doi.org/10.1016/j.sbi.2011.09.002).
- Bruun, S. W., Ieřmantavičius, V., Danielsson, J. & Poulsen, F. M. (2010). Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein. *Proc. Natl Acad. Sci. USA*, **107**, 13306–13311.
- Wallgren, M., Aden, J., Pylypenko, O., Mikaelsson, T., Johansson, L. B., Rak, A. & Wolf-Watz, M. (2008). Extreme temperature tolerance of a hyperthermophilic protein coupled to residual structure in the unfolded state. *J. Mol. Biol.* **379**, 845–858.
- Marsh, J. A. & Forman-Kay, J. D. (2009). Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *J. Mol. Biol.* **391**, 359–374.
- Bai, Y. W., Chung, J., Dyson, H. J. & Wright, P. E. (2001). Structural and dynamic characterization of an unfolded state of poplar apo-plastocyanin formed under nondenaturing conditions. *Protein Sci.* **10**, 1056–1066.
- Shan, B., McClendon, S., Rospigliosi, C., Eliezer, D. & Raleigh, D. P. (2010). The cold denatured state of the C-terminal domain of protein L9 is compact and contains both native and non-native structure. *J. Am. Chem. Soc.* **132**, 4669–4677.
- Rosner, H. I. & Poulsen, F. M. (2010). Residue-specific description of non-native transient structures in the ensemble of acid-denatured structures of the all-beta protein c-src SH3. *Biochemistry*, **49**, 3246–3253.
- Huang, J. R. & Grzesiek, S. (2010). Ensemble calculations of unstructured proteins constrained by RDC and PRE data: a case study of urea-denatured ubiquitin. *J. Am. Chem. Soc.* **132**, 694–705.
- Cliff, M. J., Craven, C. J., Marston, J. P., Hounslow, A. M., Clarke, A. R. & Waltho, J. P. (2009). The denatured state of N-PGK is compact and predominantly disordered. *J. Mol. Biol.* **385**, 266–277.
- Reed, M. A. C., Jelinska, C., Syson, K., Cliff, M. J., Splevins, A., Alizadeh, T. *et al.* (2006). The denatured state under native conditions: a non-native-like collapsed state of N-PGK. *J. Mol. Biol.* **357**, 365–372.
- Schwarzinger, S., Wright, P. E. & Dyson, H. J. (2002). Molecular hinges in protein folding: the urea-denatured state of apomyoglobin. *Biochemistry*, **41**, 12681–12686.
- Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. & Fersht, A. R. (2005). Solution structure of a protein denatured state and folding intermediate. *Nature*, **437**, 1053–1056.
- Anil, B., Li, Y., Cho, J. H. & Raleigh, D. P. (2006). The unfolded state of NTL9 is compact in the absence of denaturant. *Biochemistry*, **45**, 10110–10116.
- Chugha, P. & Oas, T. G. (2007). Backbone dynamics of the monomeric lambda repressor denatured state ensemble under nondenaturing conditions. *Biochemistry*, **46**, 1141–1151.
- Francis, C. J., Lindorff-Larsen, K., Best, R. B. & Vendruscolo, M. (2006). Characterization of the residual structure in the unfolded state of the  $\Delta 131\Delta$  fragment of staphylococcal nuclease. *Proteins*, **65**, 145–152.
- Merchant, K. A., Best, R. B., Louis, J. M., Gopich, I. V. & Eaton, W. A. (2007). Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl Acad. Sci. USA*, **104**, 1528–1533.
- Rosenkranz, T., Schlesinger, R., Gabba, M. & Fitter, J. (2011). Native and unfolded states of phosphoglycerate kinase studied by single-molecule FRET. *ChemPhysChem*, **12**, 704–710.
- Pugh, S. D., Gell, C., Smith, D. A., Radford, S. E. & Brockwell, D. J. (2010). Single-molecule studies of the Im7 folding landscape. *J. Mol. Biol.* **398**, 132–145.
- Whittaker, S. B., Clayden, N. J. & Moore, G. R. (2011). NMR characterisation of the relationship between frustration and the excited state of Im7. *J. Mol. Biol.* **414**, 511–529.
- Bartlett, A. I. & Radford, S. E. (2010). Desolvation and development of specific hydrophobic core packing during Im7 folding. *J. Mol. Biol.* **396**, 1329–1345.
- 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.
- Knowing, S. E., Figueiredo, A. M., Whittaker, S. B., 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.

29. Le Duff, C. S., Whittaker, S. B., 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.
30. 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.
31. Whittaker, S. B. M., Spence, G. R., Grossmann, J. G., 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.
32. Spence, G. R., Capaldi, A. P. & Radford, S. E. (2004). Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J. Mol. Biol.* **341**, 215–226.
33. 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.
34. Gsponer, J., Hopearuoho, H., Whittaker, S. B. M., Spence, G. R., Moore, G. R., Paci, E. *et al.* (2006). Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc. Natl Acad. Sci. USA*, **103**, 99–104.
35. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant  $m$  values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148.
36. Keeble, A. H., Kirkpatrick, N., Shimizu, S. & Kleanthous, C. (2006). Calorimetric dissection of colicin DNase-immunity protein complex specificity. *Biochemistry*, **45**, 3243–3254.
37. Foit, L., Morgan, G. J., Kern, M. J., Steimer, L. R., von Hacht, A. A., Titchmarsh, J. *et al.* (2009). Optimizing protein stability in vivo. *Mol. Cell*, **36**, 861–871.
38. 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.
39. Cobos, E. S. & Radford, S. E. (2006). Sulfate-induced effects in the on-pathway intermediate of the bacterial immunity protein Im7. *Biochemistry*, **45**, 2274–2282.
40. Huang, F., Sato, S., Sharpe, T. D., Ying, L. & Fersht, A. R. (2007). Distinguishing between cooperative and unimodal downhill protein folding. *Proc. Natl Acad. Sci. USA*, **104**, 123–127.
41. Sosnick, T. R. & Barrick, D. (2011). The folding of single domain proteins—have we reached a consensus? *Curr. Opin. Struct. Biol.* **21**, 12–24.
42. Kuzmenkina, E. V., Heyes, C. D. & Nienhaus, G. U. (2005). Single-molecule Forster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc Natl Acad. Sci. USA*, **102**, 15471–15476.
43. Tezuka-Kawakami, T., Gell, C., Brockwell, D. J., Radford, S. E. & Smith, D. A. (2006). Urea-induced unfolding of the immunity protein Im9 monitored by spFRET. *Biophys. J.* **91**, 42–44.
44. Muller-Spath, S., Soranno, A., Hirschfeld, V., Hofmann, H., Ruegger, S., Reymond, L. *et al.* (2010). Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl Acad. Sci. USA*, **107**, 14609–14614.
45. Laurence, T. A., Kong, X., Jager, M. & Weiss, S. (2005). Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc. Natl Acad. Sci. USA*, **102**, 17348–17353.
46. Hofmann, H., Golbik, R. P., Ott, M., Hubner, C. G. & Ulbrich-Hofmann, R. (2008). Coulomb forces control the density of the collapsed unfolded state of barstar. *J. Mol. Biol.* **376**, 597–605.
47. 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.
48. Wallis, R., Leung, K. Y., Pommer, A. J., Videler, H., Moore, G. R., James, R. & Kleanthous, C. (1995). Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 2. Cognate and noncognate interactions that span the millimolar to femtomolar affinity range. *Biochemistry*, **34**, 13751–13759.
49. Ko, T. P., Liao, C. C., Ku, W. Y., Chak, K. F. & Yuan, H. S. (1999). The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein. *Structure*, **7**, 91–102.
50. Vankemmelbeke, M., Healy, B., Moore, G. R., Kleanthous, C., Penfold, C. N. & James, R. (2005). Rapid detection of colicin E9-induced DNA damage using *Escherichia coli* cells carrying SOS promoter-lux fusions. *J. Bacteriol.* **187**, 6601.
51. Liang, Y. (2008). Applications of isothermal titration calorimetry in protein science. *Biochim. Biophys. Acta*, **40**, 565–576.
52. Li, W., Keeble, A. H., Giffard, C., James, R., Moore, G. R. & Kleanthous, C. (2004). Highly discriminating protein-protein interaction specificities in the context of a conserved binding energy hotspot. *J. Mol. Biol.* **337**, 743–759.
53. 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**, 139–149.
54. 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**, 157–165.
55. Pardi, A., Billeter, M. & Wuthrich, K. (1984). Calibration of the angular dependence of the amide proton-C $\alpha$  proton coupling constants,  $^3\text{JHN}\alpha$ , in a globular protein. *J. Mol. Biol.* **180**, 741–751.
56. Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M. & Dobson, C. M. (1996). Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol.* **255**, 494–506.
57. Spera, S. & Bax, A. (1991). Empirical correlation between protein backbone conformation and C $\alpha$  and C $\Delta$ .  $^{13}\text{C}$  nuclear magnetic resonance chemical shifts. *J. Am. Chem. Soc.* **113**, 5490–5492.
58. 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. USA*, **101**, 17394–17397.



59. Marsh, J. A., Singh, V. K., Jia, Z. C. & Forman-Kay, J. D. (2006). Sensitivity of secondary structure propensities to sequence differences between  $\alpha$ - and  $\gamma$ -synuclein: implications for fibrillation. *Protein Sci.* **15**, 2795–2804.
60. 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**, 173–191.
61. 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.
62. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1994). A heteronuclear correlation experiment for simultaneous determination of  $^{15}\text{N}$  longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR*, **4**, 727–734.
63. Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone dynamics of proteins as studied by  $^{15}\text{N}$  inverse detected heteronuclear NMR spectroscopy—application to staphylococcal nuclease. *Biochemistry*, **28**, 8972–8979.
64. 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.
65. Alexandrescu, A. T. & Shortle, D. (1994). Backbone dynamics of a highly disordered 131 residue fragment of staphylococcal nuclease. *J. Mol. Biol.* **242**, 527–546.
66. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. (2011). How fast-folding proteins fold. *Science*, **334**, 517–520.
67. Ziv, G., Thirumalai, D. & Haran, G. (2009). Collapse transition in proteins. *Phys. Chem. Chem. Phys.* **11**, 83–93.
68. Li, Y., Picart, F. & Raleigh, D. P. (2005). Direct characterization of the folded, unfolded and urea-denatured states of the C-terminal domain of the ribosomal protein L9. *J. Mol. Biol.* **349**, 839–846.
69. Shortle, D. & Ackerman, M. S. (2001). Persistence of native-like topology in a denatured protein in 8 M urea. *Science*, **293**, 487–489.
70. Kohn, J. E., Gillespie, B. & Plaxco, K. W. (2009). Non-sequence-specific interactions can account for the compaction of proteins unfolded under “native” conditions. *J. Mol. Biol.* **394**, 343–350.
71. Ziv, G. & Haran, G. (2009). Protein folding, protein collapse, and tanford’s transfer model: lessons from single-molecule FRET. *J. Am. Chem. Soc.* **131**, 2942–2947.
72. Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D. & Kay, L. E. (2001). Slow dynamics in folded and unfolded states of an SH3 domain. *J. Am. Chem. Soc.* **123**, 11341–11352.
73. Kristjansdottir, S., Lindorff-Larsen, K., Fieber, W., Dobson, C. M., Vendruscolo, M. & Poulsen, F. M. (2005). Formation of native and non-native interactions in ensembles of denatured ACBP molecules from paramagnetic relaxation enhancement studies. *J. Mol. Biol.* **347**, 1053–1062.
74. Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M. R., Zweckstetter, M. & Blackledge, M. (2010). NMR characterization of long-range order in intrinsically disordered proteins. *J. Am. Chem. Soc.* **132**, 8407–8418.
75. Daggett, V. & Fersht, A. (2003). The present view of the mechanism of protein folding. *Nat. Rev., Mol. Cell Biol.* **4**, 497–502.
76. Kazmirski, S. L., Wong, K. B., Freund, S. M. V., Tan, Y. J., Fersht, A. R. & Daggett, V. (2001). Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl Acad. Sci. USA*, **98**, 4349–4354.
77. Banachewicz, W., Religa, T. L., Schaeffer, R. D., Daggett, V. & Fersht, A. R. (2011). Malleability of folding intermediates in the homeodomain superfamily. *Proc. Natl Acad. Sci. USA*, **108**, 5596–5601.
78. Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619.
79. Schuck, P. (2003). On the analysis of protein self-association by sedimentation velocity analytical ultracentrifugation. *Anal. Biochem.* **320**, 104–124.
80. Lebowitz, J., Lewis, M. S. & Schuck, P. (2002). Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci.* **11**, 2067–2079.
81. 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.
82. Johnson, B. A. & Blevins, R. A. (1994). NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
83. Vuister, G. W. & Bax, A. (1993). Quantitative J correlation: a new approach for measuring homonuclear three-bond  $J(\text{HNH}\alpha)$  coupling constants in  $^{15}\text{N}$ -enriched proteins. *J. Am. Chem. Soc.* **115**, 7772–7777.
84. Levitt, M. (1976). A simplified representation of protein conformations for rapid simulation of protein folding. *J. Mol. Biol.* **104**, 59–107.
85. Schwarzwinger, S., Kroon, G. J. A., 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.