

# SCIENTIFIC REPORTS



OPEN

## Equilibrium folding dynamics of *meACP* in water, heavy water, and low concentration of urea

Yang Zhou &amp; Daiwen Yang

Many proteins fold in apparent two-state behavior, as partially folded intermediates only transiently accumulate and easily escape detection. Besides a native form and a mainly unfolded form, we captured a partially unfolded form of an acyl carrier protein from *Micromonospora echinospora* (*meACP*) in the folding/unfolding equilibrium using chemical exchange saturation transfer NMR experiments. The C-terminal region of the partially unfolded form is mainly folded and the N-terminal is unfolded. Furthermore, to understand how the folding process of *meACP* is influenced by solvent environments, we compared the folding dynamics of *meACP* in D<sub>2</sub>O, H<sub>2</sub>O and low concentration of urea. As the environment becomes more denaturing from D<sub>2</sub>O to H<sub>2</sub>O and then to urea, the unfolded state becomes increasingly populated, and the folding rate decreases. Adding a small amount of urea, which does not change solvent viscosity, has little effects on the unfolding rates, while changing H<sub>2</sub>O to D<sub>2</sub>O reduces the unfolding rates possibly due to the increase of solvent viscosity. The quantified solvent effects on the protein folding Gibbs energy and activation energy suggest that the transition state of folding may have a similar structure to the native state of the protein.

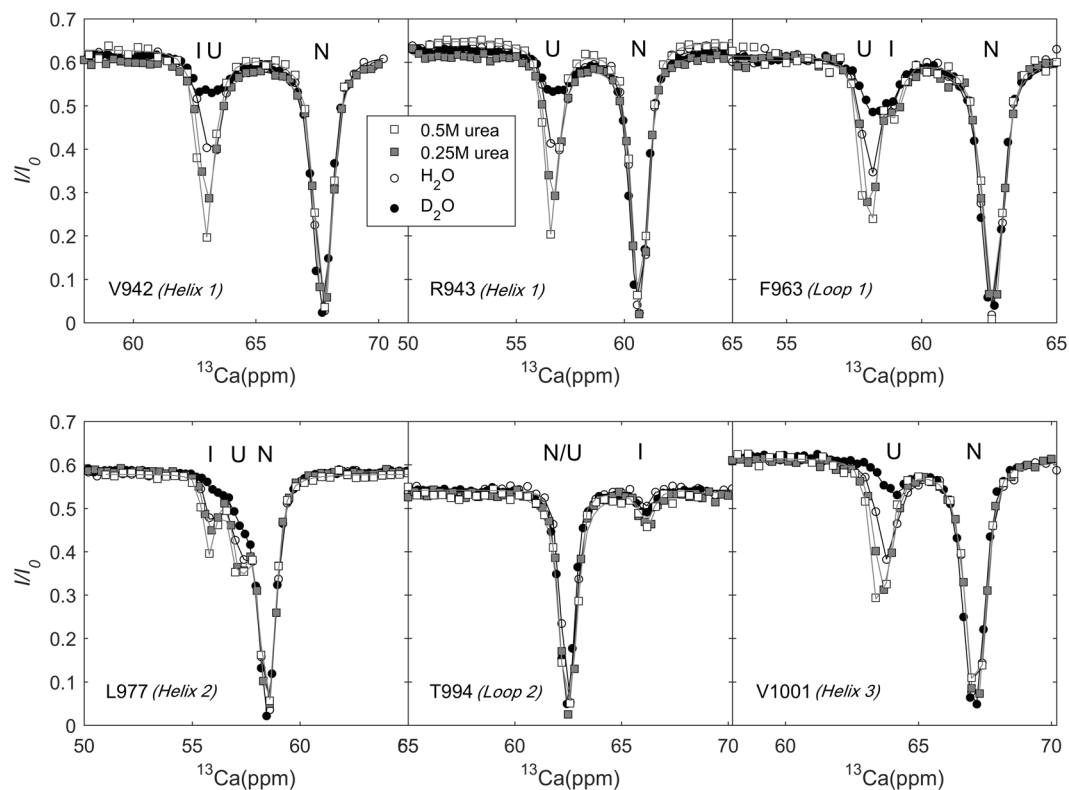
Protein molecules exist in multiple conformational states under native conditions. The dominant folded state with the lowest free energy is in dynamic equilibrium with minor states such as unfolded state and partially unfolded intermediate states<sup>1–6</sup>. Studying the conformational exchanges among different states of a protein can reveal structure-function relationship<sup>6–8</sup> or/and help to understand protein folding<sup>3,4,9</sup>.

Protein folding to biologically functional conformations is important to all living organisms. Misfolding may lead to fatal diseases. Our understanding of protein folding has advanced in the past half-century, yet a general mechanism that applies to most proteins remains missing<sup>10</sup>. The earliest classic view that proteins fold non-randomly through intermediate states on “predetermined” folding pathways, came from the famous Levinthal paradox that was proposed in 1969<sup>11</sup>. Later studies have well demonstrated the existence of folding intermediates, on-pathway and off-pathway. Different models such as the framework model<sup>12</sup>, the diffusion-collision<sup>13</sup>, and the jigsaw puzzle model<sup>14</sup> were proposed based on experimental and theoretical studies. However, a consensus has not been reached yet; some models were even contradictory to each other. The “folding funnel energy landscape” view, proposed in the 1990’s, emphasizes the ensemble nature of protein conformations, describing macroscopic “states” as a distribution of various conformations during parallel folding events<sup>15</sup>. In this “funnel” theory, the so-called off-pathway states are energy traps that are direct and “on-route” for the population involved in these pathways. In more recent years, different hierarchical hypotheses have been revived, such as the “zipping and assembly”<sup>16</sup> and “foldon” views<sup>3,17–20</sup>. The concept of “foldon”, small cooperative folding unit, emerged since the folding of many proteins was detected to be sub-globally cooperative. The “foldon” view embraces the idea of predetermined macroscopic folding pathways and suggests the sequential nature of “foldon” forming processes.

In most of the views, folding intermediates are the keys to understand the folding processes. Although they have been detected in many protein systems, obtaining their specific structural information has been challenging, as high energy intermediate states are often transiently formed and sparsely populated in the native unfolding-folding equilibrium<sup>21</sup>. The situation is changing now. The advancement of powerful analytical techniques in recent years such as nuclear magnetic resonance (NMR)<sup>3,22–24</sup> and mass spectrometry<sup>9</sup> has made it possible to characterize the intermediate states at high temporal and spatial resolution.

Shifting protein folding equilibrium by co-solvents such as heavy water (D<sub>2</sub>O) and urea has been well characterized in the literature. Most of the protein molecules are more stable (Table S1) and more rigid<sup>25</sup> and fold more

Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore, 117543, Singapore. Correspondence and requests for materials should be addressed to D.Y. (email: [dbsydw@nus.edu.sg](mailto:dbsydw@nus.edu.sg))



**Figure 1.** Representative  $^{13}\text{Ca}$  CEST profiles obtained with a RF weak field of 15 Hz in  $\text{H}_2\text{O}$ ,  $\text{D}_2\text{O}$ , 0.25 M urea, and 0.5 M urea. The data obtained with a field of 30 Hz are not included for visual clarity. Solid lines are the best fits.

rapidly<sup>26</sup> in  $\text{D}_2\text{O}$  than in  $\text{H}_2\text{O}$ , because of stronger D-bond than H-bond in neutral water<sup>27</sup>. Urea is a widely used denaturant although its mechanism of denaturing remains not clearly known. Studies suggest that urea denatures proteins by a “direct interaction” mechanism, as urea has preferential binding to all regions of proteins and can form H-bond with protein backbones<sup>28,29</sup>.

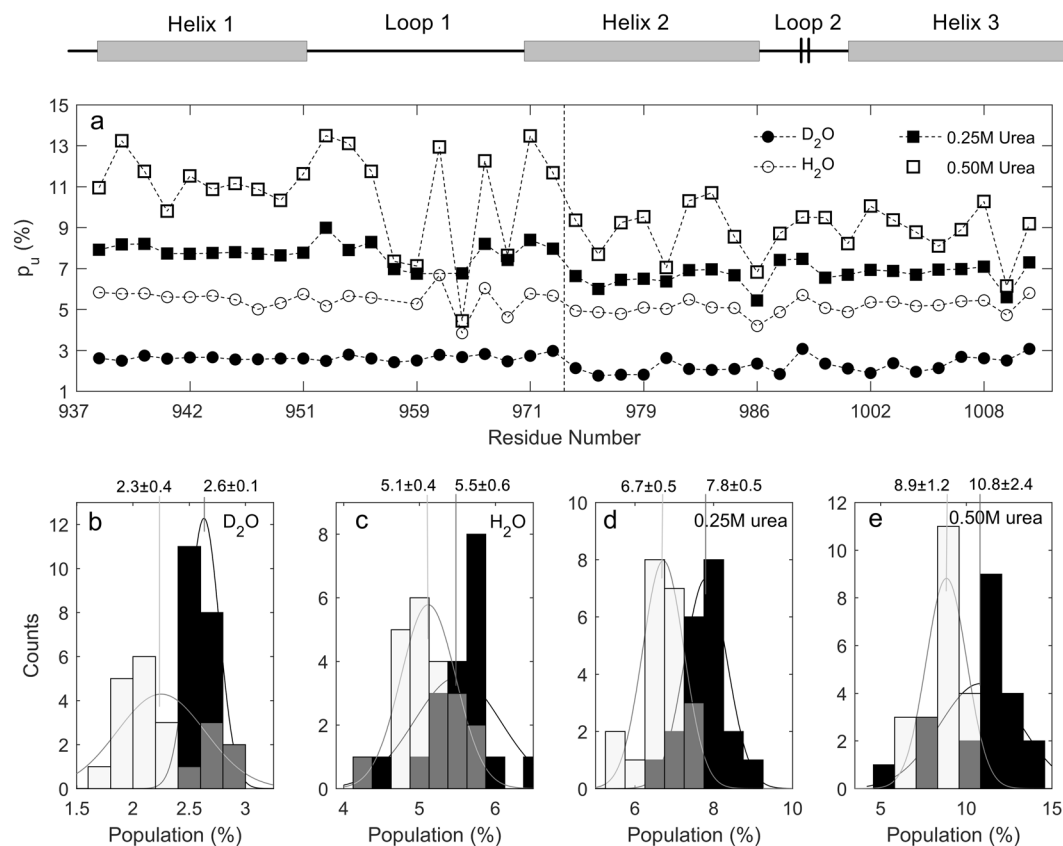
Recently we have found that the acyl carrier protein from *Micromonospora echinospora* (*meACP*, PDB code: 2I9f) exists in one major native folded state (N), one minor largely unfolded state (U), and one minor intermediate state (I) under native conditions<sup>3,30</sup>. The three states are in relatively slow conformational exchange on the sub-second timescale, following the pathway  $\text{N} \leftrightarrow \text{U} \leftrightarrow \text{I}$  or  $\text{U} \leftrightarrow \text{N} \leftrightarrow \text{I}$ . Beyond this, the overall structural features of this state I were mostly uncertain. In this study, we deepened our study of *meACP* folding to explore structural features of state I using chemical exchange saturation transfer (CEST) NMR. We extracted the backbone  $^{13}\text{Ca}$  chemical shifts of *meACP* in the state I, and demonstrated that this state is a partially unfolded form. In an effort to understand the folding dynamics thoroughly, the folding equilibria were compared under varied conditions:  $\text{D}_2\text{O}$ ,  $\text{H}_2\text{O}$ , and low concentration of urea (0.25 M and 0.5 M). The comparison provides insights into the folding of *meACP*.

## Results and Discussions

$^{13}\text{Ca}$  and  $^{15}\text{N}$  CEST experiments on *meACP* in  $\text{H}_2\text{O}$ ,  $\text{D}_2\text{O}$ ,  $\text{H}_2\text{O} + 0.25\text{ M urea}$ , and  $\text{H}_2\text{O} + 0.5\text{ M urea}$  were performed. 43 residues displayed two dips in their  $^{13}\text{Ca}$  CEST profiles, showing those residues existed in at least two forms. As demonstrated previously on the basis of chemical shifts, the minor form was mainly unfolded (U) with  $\alpha$ -helical propensity<sup>2,31</sup>. Moreover, a third state (I) was clearly seen in their CEST profiles for several residues (Figs. 1 and S1).

For the residues displaying two-dip CEST profiles, we first performed individual fitting using the two-state model. The population values of state U ( $p_u$ ) extracted from the  $^{13}\text{Ca}$  CEST are shown in Fig. 2. The results from the  $^{15}\text{N}$  CEST are given in Fig. S4. Interestingly, the average  $p_u$  values obtained by individual fitting with the two-state model for the residues in the C-terminal half region (C-terminal region of helix 2 and helix 3: from V974 to A1010) were smaller than those for the residues in the N-terminal half region (helix 1, loop 1 and N-terminal region of helix 2: from A938 to I972). The t-tests (p-value < 1.4% for all data sets) confirmed that the differences were significant. The population differences between the two regions showed that the folding of *meACP* was not globally all or none. Is there a partially folded state where the C-terminal region remains folded while the N-terminal region is unfolded?

When the dip for state I overlaps with that for state U (i.e.,  $\Omega_U \approx \Omega_I$ ) or overlaps with that for state N ( $\Omega_N \approx \Omega_I$ ), the CEST profiles of a three-state system look like two-state profiles, but cannot be fitted well to the global two-state model. The bad global fitting of our data by the two-state model indicates the global presence of



**Figure 2.** Population of state U  $p_u$  (a),  $p_u$  distribution of the two half regions (N-terminal: black bar; C-terminal: white bar; overlapped portion of black and white bars: grey bar) of *meACP* in  $D_2O$  (b),  $H_2O$  (c), 0.25 M urea (d) and 0.5 M urea (e). The averages and standard deviations of  $p_u$  for the two halves in different solvents are shown on the top of each panel.

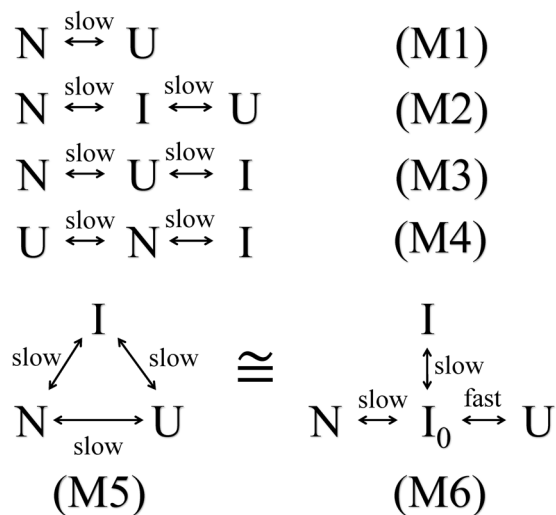
state I (Fig. S2). In addition, 18 (out of 93) residues displayed three-dip  $^{15}N$  CEST profiles with clear evidence of a third state  $I^2$ . Moreover, the dip depth for state I in CEST profiles increased with urea concentration (Figs. 1 and S1), and the difference of  $p_u$  values between the N- and C-terminal regions increased as well (Fig. 2). Taken together, we suggest that the difference of  $p_u$  between the two regions shown in Fig. 2 comes from the global presence of state I.

For a system with three conformational states, there are four possible three-state exchange models (M2~M5 in Fig. 3). As discussed in our previous paper<sup>2</sup>, the exchange rates between N and U and between N and I were around  $250\text{ s}^{-1}$  and  $100\text{ s}^{-1}$ , respectively, and model M2 was excluded based on the  $^{15}N$  data. In the on-pathway model (M2 in Fig. 3), the exchange between N and I must be faster than the exchange between N and U, which is contradictory to the result derived from the  $^{15}N$  data. It is noteworthy that models M3 and M4 are two special cases of M5, where the exchange rate between I and N or the rate between I and U is zero.

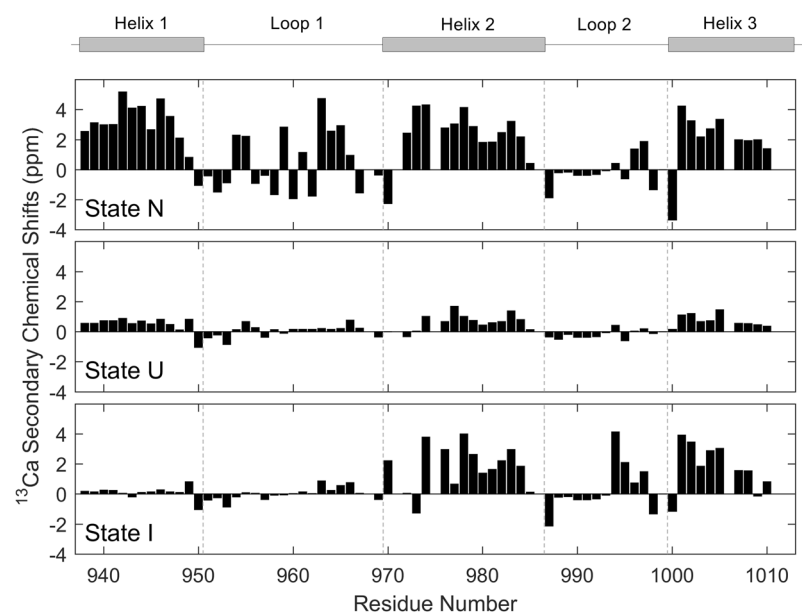
For each data set, global fittings based on the two-state model and three-state models were conducted to compare the fitting goodness, and to extract the global exchange rates and populations of different states (shared by all residues) as well as the per-residue chemical shifts of minor states ( $\Omega_U$  only in two-state model;  $\Omega_U$  and  $\Omega_I$  in three-state models), transverse relaxation rates ( $R_{2N}$ ) and longitudinal relaxation rates ( $R_{1N}$ ) of the major state. The reduced  $\chi^2$  values for global fittings by three possible three-state models M3~M5 were significantly smaller than those for the two-state model M1 and the on-pathway three-state model M2 (see Table S2). The F-tests showed that the global three-state models M3~M5 described the CEST data significantly better than models M1 and M2 ( $p$ -value  $< 10^{-4}$  for all data sets). Indeed, the two-state model failed obviously for a few residues that displayed three dips. Considering the possibility that global minima might not be reached in the global fittings by the three-state models (due to large search space) and the fact that the fitting residuals by M3~M5 were similar, we think that the models M3~M5 matched equally well with the data (Table S2). So we could not determine which model is true for *meACP*. Fortunately, the  $^{13}Ca$  chemical shifts of state I ( $\Omega_I$ ) obtained with these three three-state models were almost identical.

As the  $\Omega_I$  values were extracted by minimizing  $\chi^2$ , it is likely that  $\Omega_I$  could be positioned incorrectly at  $\Omega_U$  in a few cases when  $\Omega_I$  and  $\Omega_U$  were very similar and  $\chi^2$  values were not sensitive to  $\Omega_I$  positions. Nevertheless,  $\Omega_I$  could be determined at high confidence level from the  $\chi^2$  distribution for most residues.

**Structural features of the two minor states.** The secondary chemical shifts (CSs) of each residue in three states were calculated by subtracting the random coil chemical shift (CS) values, which were experimentally



**Figure 3.** Possible exchange models examined for *meACP*.



**Figure 4.**  $^{13}\text{Ca}$  secondary chemical shifts of state N, U and I in  $\text{D}_2\text{O}$ .  $^{13}\text{Ca}$  chemical shifts experimentally measured in 4 M urea (fully denatured) were used as the reference.

measured in the fully denatured sample (*meACP* in 4 M urea), from the CS values of the corresponding residues which were obtained in the CEST data. The analysis of  $^{13}\text{Ca}$  secondary CSs is a simple method to identify the secondary structures of a protein in different states<sup>32</sup> (Fig. 4).

For state U, the  $^{13}\text{Ca}$  secondary CSs in three helical regions had an average value of around +1 ppm, as compared with state N whose  $^{13}\text{Ca}$  secondary CSs in the helical region were in the range of +2~+3 ppm (Fig. 4), indicating the presence of residual helical structures. The residual helical pattern in state U was the same as that in the native state, implying that the three native helices were not fully unfolded in state U. The  $^{13}\text{Ca}$  secondary CS amplitudes for the helical regions in state U were about one-third of those in state N, suggesting that the three helices in state U are populated for about 30% of the time.

For state I, the  $^{13}\text{Ca}$  secondary CSs of its N-terminal region (helix 1, loop 1, and N-terminal region of helix 2) were close to zero, indicating this region exists in a nearly fully unfolded form. On the other hand, the  $^{13}\text{Ca}$  secondary CSs of the C-terminal region (helix 3 and C-terminal region of helix 2) in state I were mostly similar to those in state N, showing that the C-terminal region of state I is native-like (Fig. 4) (or the helices 2 and 3 remain mainly folded in state I).

Obviously, the state I is overall a partially unfolded form (PUF). This is consistent with the results shown in Fig. 2. Based on the two-state model, a residue can stay as either a folded or an unfolded form. As the N-terminal

| Solvents                       | Expt.            | $p_u$ (%)  | $p_I$ (%) | $\Delta G_{NU}$ (kcal/mol) | $\Delta G_{NI}$ (kcal/mol) | $k_{NU}$ (s <sup>-1</sup> ) | $k_{UN}$ (s <sup>-1</sup> ) | $k_{UI} + k_{IU}$ (s <sup>-1</sup> ) |
|--------------------------------|------------------|------------|-----------|----------------------------|----------------------------|-----------------------------|-----------------------------|--------------------------------------|
| D <sub>2</sub> O               | <sup>13</sup> Ca | 2.0 ± 0.01 | ~2.0*     | 2.3                        | ~2.3*                      | 10.4 ± 2.3                  | 495.2 ± 7.9                 | ~73*                                 |
| H <sub>2</sub> O               | <sup>13</sup> Ca | 4.6 ± 0.02 | ~1.9      | 1.8                        | ~2.3                       | 16.9 ± 2.6                  | 356.6 ± 10.4                | ~141                                 |
| H <sub>2</sub> O               | <sup>15</sup> N  | 4.8 ± 0.01 | ~0.8      | 1.8                        | ~2.8                       | 15.0 ± 0.3                  | 297.6 ± 4.9                 | ~85                                  |
| 0.25 M Urea + H <sub>2</sub> O | <sup>13</sup> Ca | 6.5 ± 0.04 | ~2.1      | 1.6                        | ~2.2                       | 14.9 ± 0.8                  | 220.0 ± 5.5                 | ~93                                  |
| 0.25 M Urea + H <sub>2</sub> O | <sup>15</sup> N  | 6.9 ± 0.05 | ~6.9      | 1.5                        | ~1.5                       | 15.7 ± 0.3                  | 199.6 ± 2.7                 | ~34                                  |
| 0.50 M Urea + H <sub>2</sub> O | <sup>13</sup> Ca | 7.9 ± 0.04 | ~8.5      | 1.4                        | ~1.4                       | 15.6 ± 0.6                  | 176.5 ± 4.1                 | ~31                                  |

**Table 1.** Folding and unfolding based on three-state model N-U-I. \*, “~” is used to indicate that the results were estimated as  $p_I$  and  $(k_{UI} + k_{IU})$  are correlated in the fitting.

region has a higher unfolded population than the C-terminal region (Fig. 2), some protein molecules must exist as a PUF. Based on <sup>15</sup>N CEST data alone, we previously suggested that state I adopted a locally altered conformation in which the N-terminal region of helix 2 differs from both helical and random coil structures while other helices are native-like<sup>2</sup>. Based on <sup>13</sup>Ca and <sup>15</sup>N data, it is now clear that the state I resembles state N in the C-terminal half, while it resembles state U in the N-terminal half (Fig. 4). It is noteworthy that state I still differs significantly from state N in the N-terminal region of helix 2.

**The partially unfolded form (PUF or state I).** In one scenario, the folding follows the triangle model (M5 in Fig. 3). It might suggest parallel folding pathways. The terms  $p_I * k_{IN}$  and  $p_u * k_{UN}$  can represent the contributions to folding from apparent folding pathways U-I-N and U-N, respectively.  $\frac{p_I * k_{IN}}{p_u * k_{UN}}$  was estimated to be smaller than ~0.06 for all data sets (Table S3), showing that the pathway U-I-N contributes significantly less than the U-N pathway in the protein folding. Although the triangle model fits the experimental data, it is not necessary for the folding to be parallel, because a four-state model with state I as an off-pathway product (M6 in Fig. 3) still fits the data. Therefore, the state I might be an “on-pathway” folding intermediate state and it is also possible to be an “off-pathway” intermediate state.

The triangle model can be approximated as the single-pathway model N-U-I (M3) when  $k_{IN} + k_{NI}$  is much smaller than  $k_{IU} + k_{UI}$ . It can also be approximated as the model U-N-I (M4) when  $k_{IN} + k_{NI}$  is much larger than  $k_{IU} + k_{UI}$ . For some of our data sets  $k_{IN} + k_{NI} \gg k_{IU} + k_{UI}$ , and for other data sets  $k_{IN} + k_{NI} \ll k_{IU} + k_{UI}$  (Table S3). This is caused by the fact that each set of data could be fitted equally well by M3 and M4 (Table S2). Because the folding can follow the two possible single-pathway models, state I might be an “off-pathway” state of the folding from state U to N. We previously suggested that there might be a dynamic equilibrium between monomeric and oligomeric forms, and the oligomers might result from this state I<sup>2</sup>. The N-terminal region of helix 2 (around S970-L977) in state I is very non-native (Fig. 4), which might be prone to aggregate. The misfolding of this region might cause the failure of the further folding of state I.

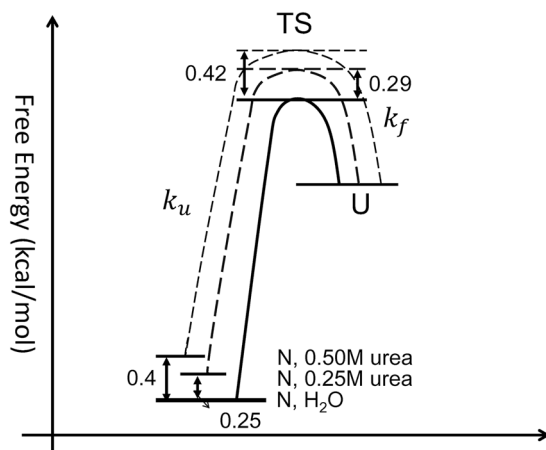
If state I<sub>0</sub> is on the folding path from U to I<sub>0</sub> to N, but local misfolding happens for a fraction of this state and blocks its folding to state N due to “the occurrence of an on-pathway optional error”<sup>33</sup>, the misfolded fraction (state I) appears as an off-pathway product (“dead-end”) (M6 in Fig. 3). Therefore, the “off-pathway” state I may represent the misfolded population, which is only a fraction of the constructive state I<sub>0</sub> with optional local misfolding errors. This putative “on-pathway” intermediate state I<sub>0</sub> would be similar to the observed state I in overall structure, but it would interconvert with state U so fast that it could not be detected by CEST and relaxation dispersion.

**Comparison of folding dynamics in D<sub>2</sub>O, H<sub>2</sub>O, and urea.** The effects of co-solvent D<sub>2</sub>O and urea on shifting the folding equilibria were examined. The <sup>13</sup>Ca CEST profiles recorded in H<sub>2</sub>O, D<sub>2</sub>O and urea varied mainly in the depths of the minor dips (Fig. 1). The dips for both states U and I were larger in urea and smaller in D<sub>2</sub>O, compared with those in H<sub>2</sub>O (see Fig. 1 and Fig. S1).

Table 1 shows the extracted folding parameters in D<sub>2</sub>O, H<sub>2</sub>O, 0.25 M urea and 0.50 M urea by following M3 (N-U-I). Models M4 and M5 gave similar  $p_u$ ,  $k_{UN}$ ,  $k_{NU}$ ,  $\Omega_U$ , and  $\Omega_I$  values as M3 did (see Tables S3 and S4). As buffer condition became more denaturing (from D<sub>2</sub>O to H<sub>2</sub>O, to higher concentration of urea), the population of unfolded state ( $p_u$ ) increased, reflecting the decrease of Gibbs energy of unfolding ( $\Delta G_{N-U}$ ). We do expect unfolded states (U) to be more populated in urea, less in D<sub>2</sub>O, as D<sub>2</sub>O and urea have been widely known for their stabilizing and destabilizing effects on proteins, respectively. *meACP* was ~0.5 kcal/mol more stable in D<sub>2</sub>O (Table 1). This is consistent with literature studies (summarized in Table S1). Previous data suggest that the larger the proteins are, the more prominent the stabilizing effect is (Fig. S3). The stability enhanced by D<sub>2</sub>O agrees well with the size of *meACP*.

The exchange between states N and U can be viewed as the major folding/unfolding process, therefore the conversion rates from U to N ( $k_{UN}$ ) and from N to U ( $k_{NU}$ ) were denoted as the folding ( $k_f$ ) and unfolding ( $k_u$ ) rates, respectively. According to the results in Table 1, Tables S3 and S4,  $k_f$  ( $k_{UN}$ ) decreased as solvent became more denaturing;  $k_u$  ( $k_{NU}$ ) increased when solvent was changed from D<sub>2</sub>O to H<sub>2</sub>O, and there were no significant effects on  $k_u$  by a low concentration of urea.

There were large uncertainties in the population ( $p_I$ ) and exchange rates of state I ( $k_{X-I}$  and  $k_{I-X}$ , which are the exchange rates between I and N and between I and U) because the fitting  $\chi^2$  values changed little when  $p_I$ ,  $k_{X-I}$ , and  $k_{I-X}$  were within certain ranges. Therefore, we only estimated  $p_I$  and the sum of  $k_{X-I}$  and  $k_{I-X}$ . Despite the large uncertainty in the estimated  $p_I$ , we observed in the raw profiles that the dips for state I were larger in urea, and



**Figure 5.** Schematic plot of relative free energy profiles of *meACP* folding in varied solvents. The free energies in states TS and N are relative to that in state U. (Energy values were calculated based on Table 1).

smaller in  $D_2O$  than in  $H_2O$  (Figs. 1 and S1). The variations of state I dips in the raw profiles indicate that the population of state I increases with the destabilization strength of the solvent (in current range of low concentrations of urea).

**Effects of urea on *meACP* folding and unfolding.** High concentration of urea is denaturing, while the effects of low concentration of urea on protein stability seem to vary for different proteins and solvent environments. It was shown that low concentration urea stabilizes protein ferrocycytochrome *c*<sup>34</sup>. Here we found that *meACP* was destabilized, even at the concentration of urea as low as 0.25 M. The populations of both state U and state I increased with urea concentration.

According to Eq. 4 (see Materials and Methods section), the effect of urea on  $k_f$  mainly originates from the change of energy barrier (activation energy  $G^\ddagger$ ), because the changes of viscosity and internal friction at low concentration of urea buffer (0.25 M) are negligible ( $\approx 1\%$ )<sup>35,36</sup>. Therefore, the change of activation energy is given by,

$$\Delta\Delta G_{urea-H_2O}^{\ddagger,f} = \Delta G_{urea}^{\ddagger,f} - \Delta G_{H_2O}^{\ddagger,f} \approx -RT \ln \left( \frac{k_f^{urea}}{k_f^{H_2O}} \right) \quad (1)$$

Using the results in Table 1,  $\Delta\Delta G_{urea-H_2O}^{\ddagger,f}$  was estimated as  $\sim 0.29$  kcal/mol and  $\sim 0.42$  kcal/mol for *meACP* at 0.25 M and 0.5 M urea, respectively (Fig. 5). Similarly,  $\Delta\Delta G_{urea-H_2O}^{UN}$  (the change of the Gibbs energy of folding,  $\Delta G_{urea}^{UN} - \Delta G_{H_2O}^{UN}$ ) was calculated from extracted  $p_u$  values with Eq. 3.  $\Delta\Delta G_{urea-H_2O}^{UN}$  was  $\sim 0.25$  kcal/mol and  $\sim 0.40$  kcal/mol for *meACP* in 0.25 M and 0.5 M urea (relative to  $H_2O$ ), respectively (Fig. 5). Accordingly,

$$\phi_f = \frac{\Delta\Delta G_{urea-H_2O}^{\ddagger,f}}{\Delta\Delta G_{urea-H_2O}^{UN}} \approx 1 \quad (2)$$

The  $\phi_f$  factor, or so-called Tanford's  $\beta$  value, is viewed as an indicator of the average degree of changes in accessible surface areas ( $\Delta ASA$ ) of the transition state (TS) relative to that of the unfolded state during the protein folding<sup>37</sup>.  $\phi_f$  generally ranging from 0 to 1, represents the ratio of TS stability change to native state (N) stability change. A  $\phi_f$  value of 1, meaning that the relative stability change of TS equals that of N, suggests that the overall  $\Delta ASA$  of TS and that of N are similar during unfolding. Thus,  $\phi_f \approx 1$  indicates that the TS may have a native-like structure. According to the results obtained in  $D_2O$ , we found that  $\phi_f > 1/2$ , further supporting that the TS is native-like (see discussion in the supplementary information).

Overall, the results here suggest that the TS could be placed close to state N on the path  $N \leftrightarrow U$  in the possible models (M3~M6) shown in Fig. 3. In M3, the pathway would be  $N \leftrightarrow TS \leftrightarrow U \leftrightarrow I$  when considering the TS; in M4 it would be  $U \leftrightarrow TS \leftrightarrow N \leftrightarrow I$ ; in M5 it would be  $N (\leftrightarrow I) \leftrightarrow TS \leftrightarrow U (\leftrightarrow I)$ ; in M6 it would be  $U \leftrightarrow I_0 (\leftrightarrow I) \leftrightarrow TS \leftrightarrow N$ , according to their structural features (or overall folded percentages). The TS is suggested to be a folded transition state that is close to the native state, sharing a similar structure with the native state.

## Conclusion

We found that there exists a partially unfolded form of *meACP*, where the N-terminal half is unfolded while the C-end half remains mainly folded. The existence of PUF was confirmed by  $^{15}N$  and  $^{13}Ca$  CEST results in different solvents. The detected PUF could be either an “on-pathway” intermediate state in the triangle model or an “off-pathway” state in the other three-state models and four-state model. This PUF might represent a fraction of the constructive folding intermediate, and occurred to be misfolded. The structure features of this PUF might suggest the sequential folding order in which the C-terminal region folds before N-terminal region does. Furthermore, the H/D isotope effect and urea effect on *meACP* folding dynamics were investigated by comparing CEST results in different solvent conditions. Consistent with literature studies on H/D isotope effects, the native



| No. | Sample labeling | Solvent Condition  | Experiment            |
|-----|-----------------|--|-----------------------|
| 1   | $^{13}\text{C}$ | 95% $\text{H}_2\text{O}$ , 5% $\text{D}_2\text{O}$               | $^{13}\text{Ca}$ CEST |
| 2   | $^{13}\text{C}$ | 100% $\text{D}_2\text{O}$  | $^{13}\text{Ca}$ CEST |
| 3   | $^{13}\text{C}$ | 95% $\text{H}_2\text{O}$ , 5% $\text{D}_2\text{O}$ + 0.25 M urea | $^{13}\text{Ca}$ CEST |
| 4   | $^{13}\text{C}$ | 95% $\text{H}_2\text{O}$ , 5% $\text{D}_2\text{O}$ + 0.50 M urea | $^{13}\text{Ca}$ CEST |
| 5   | $^{15}\text{N}$ | 95% $\text{H}_2\text{O}$ , 5% $\text{D}_2\text{O}$               | $^{15}\text{N}$ CEST  |
| 6   | $^{15}\text{N}$ | 95% $\text{H}_2\text{O}$ , 5% $\text{D}_2\text{O}$ + 0.25 M urea | $^{15}\text{N}$ CEST  |

**Table 2.** CEST experiments.

state of *meACP* was more stable in  $\text{D}_2\text{O}$ , folding was accelerated, and unfolding was decelerated compared to that in  $\text{H}_2\text{O}$ . Low concentration of urea destabilized the native state of the protein and seemed to stabilize state PUF. It also slowed down the folding rates while had no effects on the unfolding rates, suggesting the structural similarity between the transition state and native state of *meACP*.

## Materials and Methods

**Sample preparation and NMR spectroscopy.** Different acyl carrier protein samples (Table 2) were prepared by following previous protocols<sup>38</sup>. All NMR samples contained 0.6–1 mM protein, 50 mM NaCl, 5 mM EDTA, and 50 mM phosphate at pH 6.9.  $^{13}\text{Ca}$  CEST and  $^{15}\text{N}$  CEST experiments were performed at 25 °C on a Bruker 800 MHz NMR spectrometer equipped with a cryoprobe. The experimental parameters used were the same as those described previously<sup>2,31</sup>.

**CEST data analysis.** *Individual and global fitting using the two-state model.* The residues displaying two dips separated by more than 1 ppm were chosen to extract kinetics parameters. Their CEST profiles were first fitted individually using the two-state model (M1 in Fig. 3) to obtain parameters for each residue, i.e., folding and unfolding rates ( $k_f$ ,  $k_u$ ), population of the minor state ( $p_u$ ), chemical shift in the minor state ( $\Omega_U$ ), longitudinal relaxation rates of major ( $R_{1N}$ ) and minor states ( $R_{1U}$ ), transverse relaxation rates of the major ( $R_{2N}$ ) and minor states ( $R_{2U}$ ). In the fitting, the J coupling effect was taken into account as described previously<sup>39</sup>. We assumed that  $^1J_{\text{CO}\alpha} = 55$  Hz and  $^1J_{\text{CO}\beta} = 35$  Hz for all residues in  $^{13}\text{C}$  labelled protein samples; and set  $R_{1U} = R_{1N}$  for each residue. When  $R_{2U}$  was used as an independent fitting parameter, the extracted  $R_{2U}$  for most residues deviated significantly from the expected values for our system due to the presence of a third state that overlaps with state N or U as well as deviations of  $^1J_{\text{CO}\alpha}$  and  $^1J_{\text{CO}\beta}$  from the assumed values. To simplify the fitting, we assumed all the residues have the same  $R_{2U}$  values in the unfolded state and set  $R_{2U} = 6.5$  and  $8.1 \text{ s}^{-1}$  for  $^{13}\text{Ca}$  in  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  respectively and  $3.0 \text{ s}^{-1}$  for  $^{15}\text{N}$  in  $\text{H}_2\text{O}$  at 25 °C. These values were estimated based on the  $^{15}\text{N}$   $R_2$  values of intrinsically disordered protein  $\alpha$ -synuclein ( $\sim 3.5 \text{ s}^{-1}$ ), which were measured at 15 °C on an 800 MHz spectrometer<sup>40</sup>. The same assumption was made for three-state fittings. In global fitting, all residues shared a common exchange rate ( $k_{ex}$ ) and a common population of the minor state ( $p_u$ ), but they each had unique  $R_{1N}$ ,  $R_{2N}$ , and  $\Omega_U$  values.

*Individual fitting and global fitting using three-state models.* The procedures for fitting the CEST data of individual residues to three-state models (M2 ~ M5) were the same as that described above. The residues displaying two and three dips were also fitted globally to the three-state models. In the fitting, we set  $R_{2I} = R_{2N}$  and  $R_{1I} = R_{1U} = R_{1N}$  for each residue, and assumed  $^1J_{\text{CO}\alpha} = 55$  Hz and  $^1J_{\text{CO}\beta} = 35$  Hz for all residues. For residues displaying three well separated dips,  $\Omega_U$  and  $\Omega_I$  values were already certainly known. For residues displaying two-dip profiles,  $\Omega_U$  values were quite certain and  $\Omega_I$  should be close to either  $\Omega_N$  or  $\Omega_U$ . Optimization was done by extensive grid-search of  $\Omega_I$ . To extract the  $\Omega_I$  value as accurate as possible, we used the following procedure:

- The  $\Omega_I$  of each residue was first estimated by individual fitting via grid search of  $\Omega_I$ ;
- $\Omega_I$  values obtained in step *a* were used as the input values in the global fitting to obtain global exchange rates between states N and U ( $k_1$ ) and between I and U or I and N ( $k_2$ ), populations of states N ( $p_N$ ), and U ( $p_u$ );
- $\Omega_I$  values were re-calculated by individual fitting with fixed global  $k_1$ ,  $k_2$ ,  $p_N$ , and  $p_u$  values (obtained in step *b*);
- Repeat steps *b* and *c*, until  $\chi^2$  of the global fitting decreased to a stable value.

**Error estimation.** For each CEST profile, the uncertainty ( $\delta$ ) in intensity ratio ( $I/I_0$ ) was estimated by calculating the standard deviation of data points in the ‘non-saturation area’. The following Monte Carlo simulations were used to estimate fitting errors of extracted CEST parameters for each residue:

- Generate 120 sets of profiles using extracted parameters, add random noise with a standard deviation of  $\delta$  and mean of 0;
- The 120 set of profiles were fitted to extract 120 sets of fitting parameters. The standard deviation of each parameter was considered as the fitting error.

To determine global fitting errors, 80% of the residues used in the global fitting were randomly taken to extract global parameters and repeat 120 times to obtain standard deviations.

**Energy calculations.** According to the population of each state, the Gibbs energy change ( $\Delta G$ ) of process A to B was calculated by

$$\Delta G_{AB} = -RT \ln(p_B/p_A) \quad (3)$$

where  $p_A$  and  $p_B$  are the populations of states A and B, respectively. The subscript AB is short for  $A \rightarrow B$ .

The folding rate ( $k_f$ ) is related to the activation energy ( $G^\ddagger$ ) by<sup>41,42</sup>

$$k_f = C(\sigma + \eta)^{-1} e^{-\Delta G^\ddagger/RT} \quad (4)$$

where C is the frequency factor for the folding process,  $\sigma$  refers to the “internal friction”, and  $\eta$  is the viscosity of solvent.  $\sigma$  reflects the contribution of the energy landscape ruggedness to the reaction rate and is dominated by the structure of a protein when denaturant concentrations are low<sup>36</sup>.

**Data Availability.** The datasets used in the current study are available in the supplementary information file, and the Matlab scripts for data fitting are available from the corresponding author upon request.

## References

- Vallurupalli, P., Bouvignies, G. & Kay, L. E. Studying “invisible” excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* **134**, 8148–8161, <https://doi.org/10.1021/ja3001419> (2012).
- Lim, J., Xiao, T., Fan, J. & Yang, D. An Off-Pathway Folding Intermediate of an Acyl Carrier Protein Domain Coexists with the Folded and Unfolded States under Native Conditions. *Angew. Chem.* **53**, 2358–2361 (2014).
- Maity, H., Maity, M., Krishna, M. M., Mayne, L. & Englander, S. W. Protein folding: the stepwise assembly of foldon units. *Proc. Natl. Acad. Sci. USA* **102**, 4741–4746 (2005).
- Bollen, Y. J., Kamphuis, M. B. & van Mierlo, C. P. The folding energy landscape of apoflavodoxin is rugged: hydrogen exchange reveals nonproductive misfolded intermediates. *Proc. Natl. Acad. Sci. USA* **103**, 4095–4100 (2006).
- Yu, B. & Yang, D. Coexistence of multiple minor states of fatty acid binding protein and their functional relevance. *Sci Rep* **6** (2016).
- Xiao, T., Fan, J. S., Zhou, H., Lin, Q. & Yang, D. Local Unfolding of Fatty Acid Binding Protein to Allow Ligand Entry for Binding. *Angew Chem Int Ed Engl.* **55**, 6869–6872 (2016).
- Manglik, A. & Kobilka, B. The role of protein dynamics in GPCR function: Insights from the  $\beta$ 2 AR and rhodopsin. *Curr. Opin. Cell Biol.* **27**, 136–143 (2014).
- Ye, L., Van Eps, N., Zimmer, M., Ernst, O. P. & Prosser, R. S. Activation of the A2A adenosine G-protein-coupled receptor by conformational selection. *Nature* (2016).
- Hu, W. *et al.* Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. USA* **110**, 7684–7689 (2013).
- Dill, K. A. & MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **338**, 1042–1046 (2012).
- Levinthal, C. Are there pathways for protein folding. *J. Chim. Phys.* **65**, 44–45 (1968).
- Udgaonkar, J. B. & Baldwin, R. L. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* **335**, 694–699 (1988).
- Karplus, M. & Weaver, D. L. Protein-folding dynamics. *Nature* **260**, 404–406 (1976).
- Harrison, S. C. & Durrin, R. Is there a single pathway for the folding of a polypeptide chain? *Proc. Natl. Acad. Sci. USA* **82**, 4028–4030 (1985).
- Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).
- Ozkan, S. B., Wu, G. A., Chodera, J. D. & Dill, K. A. Protein folding by zipping and assembly. *Proc. Natl. Acad. Sci. USA* **104**, 11987–11992 (2007).
- Chamberlain, A. K., Handel, T. M. & Marqusee, S. Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat. Struct. Mol. Biol.* **3**, 782–787 (1996).
- Vu, N.-D., Feng, H. & Bai, Y. The folding pathway of barnase: the rate-limiting transition state and a hidden intermediate under native conditions. *Biochemistry* **43**, 3346–3356 (2004).
- Cellitti, J., Bernstein, R. & Marqusee, S. Exploring subdomain cooperativity in T4 lysozyme II: Uncovering the C-terminal subdomain as a hidden intermediate in the kinetic folding pathway. *Protein Sci.* **16**, 852–862 (2007).
- Fitzkee, N. C., Torchia, D. A. & Bax, A. Measuring rapid hydrogen exchange in the homodimeric 36 kDa HIV-1 integrase catalytic core domain. *Protein Sci.* **20**, 500–512 (2011).
- Sánchez, I. E. & Kiefhaber, T. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325**, 367–376 (2003).
- Korzhnev, D. M. & Kay, L. E. Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. *Acc. Chem. Res.* **41**, 442–451 (2008).
- Bouvignies, G. & Kay, L. E. Measurement of proton chemical shifts in invisible states of slowly exchanging protein systems by chemical exchange saturation transfer. *J. Phys. Chem. B* **116**, 14311–14317 (2012).
- Hansen, A. L., Bouvignies, G. & Kay, L. E. Probing slowly exchanging protein systems via  $^{13}\text{C}\alpha$ -CEST: monitoring folding of the Im7 protein. *J. Biomol. NMR* **55**, 279–289 (2013).
- Cioni, P. & Strambini, G. B. Effect of Heavy Water on Protein Flexibility. *Biophys. J.* **82**, 3246–3253 (2002).
- Parker, M. J. & Clarke, A. R. Amide backbone and water-related H/D isotope effects on the dynamics of a protein folding reaction. *Biochemistry* **36**, 5786–5794 (1997).
- Scheiner, S. & Cuma, M. Relative stability of hydrogen and deuterium bonds. *J. Am. Chem. Soc.* **118**, 1511–1521 (1996).
- Lim, W. K., Rösger, J. & Englander, S. W. Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. *Proc. Natl. Acad. Sci. USA* **106**, 2595–2600 (2009).
- Hua, L., Zhou, R., Thirumalai, D. & Berne, B. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc. Natl. Acad. Sci. USA* **105**, 16928–16933 (2008).
- Lim, J. *et al.* Rigidifying acyl carrier protein domain in iterative type I PKS CalE8 does not affect its function. *Biophys. J.* **103**, 1037–1044 (2012).
- Zhou, Y. & Yang, D.  $^{13}\text{C}\alpha$  CEST experiment on uniformly  $^{13}\text{C}$ -labeled proteins. *J. Biomol. NMR* **61**, 89–94 (2015).
- Wishart, D. S. & Sykes, B. D. The  $^{13}\text{C}$  chemical-shift index: a simple method for the identification of protein secondary structure using  $^{13}\text{C}$  chemical-shift data. *J. Biomol. NMR* **4**, 171–180 (1994).



33. Englander, S. W. & Mayne, L. The nature of protein folding pathways. *Proc. Natl. Acad. Sci. USA* **111**, 15873–15880 (2014).
34. Bhuyan, A. K. Protein stabilization by urea and guanidine hydrochloride. *Biochemistry* **41**, 13386–13394 (2002).
35. Kawahara, K. & Tanford, C. Viscosity and density of aqueous solutions of urea and guanidine hydrochloride. *J. Biol. Chem.* **241**, 3228–3232 (1966).
36. Volk, M., Milanesi, L., Waltho, J. P., Hunter, C. A. & Beddard, G. S. The roughness of the protein energy landscape results in anomalous diffusion of the polypeptide backbone. *PCCP* **17**, 762–782 (2015).
37. Ramos, C. H., Weisbuch, S. & Jamin, M. Diffusive motions control the folding and unfolding kinetics of the apomyoglobin pH 4 molten globule intermediate. *Biochemistry* **46**, 4379–4389 (2007).
38. Lim, J. *et al.* Solution Structures of the Acyl Carrier Protein Domain from the Highly Reducing Type I Iterative Polyketide Synthase CalE8. *Plos One* **6**, e20549 (2011).
39. Zhou, Y. & Yang, D. Effects of J couplings and unobservable minor states on kinetics parameters extracted from CEST data. *J Magn Reson* **249C**, 118–125 (2014).
40. Kim, S., Wu, K. P. & Baum, J. Fast hydrogen exchange affects  $^1\text{uN}$  relaxation measurements in intrinsically disordered proteins. *J. Biomol. NMR* **55**, 249–256 (2013).
41. Pabit, S. A., Roder, H. & Hagen, S. J. Internal friction controls the speed of protein folding from a compact configuration. *Biochemistry* **43**, 12532–12538 (2004).
42. Ansari, A., Jones, C. M., Henry, E. R., Hofrichter, J. & Eaton, W. A. The role of solvent viscosity in the dynamics of protein conformational changes. *Science* **256**, 1796–1798 (1992).

## Acknowledgements

This work was supported by a Grant from Singapore Ministry of Education Academic Research Fund Tier 3 (MOE2012-T3-1-008).

## Author Contributions

D.Y. designed research; Y.Z. performed research; Y.Z. and D.Y. wrote the paper.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-16449-4>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017