Role of Disulfide Bonds and Topological Frustration in the Kinetic Partitioning of Lysozyme Folding Pathways

Aswathy N. Muttathukattil,[†] Prashant Chandra Singh,[‡] and Govardhan Reddy^{*,†}

[†]*Solid State and Structural Chemistry Unit, Indian Institute of Science, Bengaluru, Karnataka, India 560012*

[‡]*School of Chemical Science, Indian Association for the Cultivation of Science, 2A & 2B Raja S.C. Mullick Road, Jadavpur, Kolkata 700032, India*

E-mail: greddy@iisc.ac.in

Phone: +91-80-22933533. Fax: +91-80-23601310

Abstract

Disulfide bonds in proteins can strongly influence the folding pathways by constraining the conformational space. Hen egg white lysozyme has four disulfide bonds and is widely studied for its antibacterial properties. Experiments on lysozyme infer that the protein folds through a fast and a slow pathway. However, the reasons for the kinetic partitioning in the folding pathways are not completely clear. Using a coarse-grained protein model and simulations, we show that two out of the four disulfide bonds, which are present in the α -domain of lysozyme, are responsible for the slow folding pathway. In this pathway, a kinetically trapped intermediate state, which is close to the native state is populated. In this state, the orientation of α -helices present in the α -domain are misaligned relative to each other. The protein in this state has to partially unfold by breaking down the inter-helical contacts between the misaligned helices to fold to

the native state. However, the topological constraints due to the two disulfide bonds present in the α -domain make the protein less flexible, and it is trapped in this conformation for hundreds of milliseconds. On disabling these disulfide bonds, we find that the kinetically trapped intermediate state and slow folding pathway disappear. Simulations mimicking the folding of protein without disulfide bonds in oxidative conditions show that the native disulfide bonds are formed as the protein folds indicating that folding guides the formation of disulfide bonds. The sequence of formation of the disulfide bonds is Cys64-Cys80 $\rightarrow$ Cys76-Cys94 $\rightarrow$ Cys30-Cys115 $\rightarrow$ Cys6-Cys127. Any disulfide bond, which forms before its precursor in the sequence has to break and follow the sequence for the protein to fold. These results show that lysozyme also serves as a very good model system to probe the role of disulfide bonds and topological frustration in protein folding. The predictions from the simulations can be verified by single molecule fluorescence resonance energy transfer (FRET) or single molecule pulling experiments, which can probe heterogeneity in the folding pathways.

Introduction

Proteins with disulfide bonds are generally found on the cell surface or extracellular space and they played an important role in elucidating the principles of protein folding.¹⁻³ The monumental demonstration that denatured protein chains fold into a unique native functional state under renaturing conditions is established using ribonuclease, which has four disulfide bonds.⁴ A reduced and denatured ribonuclease, which is enzymatically inactive is shown to always fold into a structure with activity equivalent to that of a native protein.⁵ The disulfide bonds in proteins enhance the thermodynamic stability of the folded state and also promote resistance to proteases, which degrades them. The breaking and forming of disulfide bonds are shown to induce conformational changes, which regulate protein activity.⁶

The presence of disulfide bonds can introduce frustration in the folding landscape of evolved proteins, which is funnel-like.⁷⁻⁹ Frustration in biomolecular folding can lead to the population of kinetic intermediates resulting in slow folding pathways.⁸ Proteins inherently have topological frustration, which is due to the requirement that contacts between specific residue pairs should form for the protein to fold while maintaining the chain connectivity.^{8,10} Frustration is also observed during the oxidative folding of proteins with disulfide bonds, when a significant part of the protein folds with the premature burial of the thiol group in cysteine before the disulfide bond formation. In these cases, local unfolding of the protein is required for the buried thiol group to get exposed and form the specific disulfide bond present in the protein native state.^{2,3}

In this paper, using coarse-grained protein models and molecular dynamics simulations, we show that the folding of lysozyme with four intact disulfide bonds exhibits kinetic partitioning in its folding pathways¹¹⁻¹³ due to the frustration introduced by the presence of disulfide bonds. The simulations show that lysozyme folds through a fast and a slow folding pathway in agreement with the experiments.¹⁴⁻¹⁷ The slow folding pathway is due to the population of a long-lived kinetic intermediate during the late stages of folding. This intermediate survives due to the frustration arising from the restriction in the protein confor-

mational space because of the disulfide bonds, which narrow the folding funnel decreasing the number of escape routes from the trapped state. The slow folding pathway disappears, when lysozyme folding is initiated from an unfolded ensemble without the disulfide bonds, and the formation of disulfide bonds is disallowed mimicking folding in the presence of reducing agents.

Figure 1: (A) C_{α} contact map of the native structure of lysozyme. Red, blue and cyan represents contacts between residues present in the α -domain, β -domain and inter-domain, respectively. The four disulfide bonds are marked as yellow circles. The important secondary and tertiary contacts are marked in the contact map. The secondary structure present along the contour of the protein chain is shown as a schematic along the right and upper axes: α -helices are red boxes, 3^{10} -helices are purple boxes and β -strands are blue arrows. (B) Cartoon representation of the crystal structure¹⁸ of lysozyme (protein data bank (PDB) ID: 1AKI). β -domain, α -domain and disulfide bonds are shown in blue, red and yellow, respectively.

Lysozyme due to its antibacterial properties is one of the most widely studied proteins.^{19–22} It is used as a model system to probe the effect of cosolvents^{23–30} and salts³¹ on protein folding and aggregation,^{32–34} study the population of transient intermediates,³⁵ properties of water in the protein hydration shell,^{36–38} and enzymatic activity.^{39,40} Hen egg white lysozyme is a two domain protein with 129 residues (Figure 1). One domain is rich in α -helices and is referred to as the α -domain (Lys1 to Asn39 and Ser86 to Leu129), and

the second domain contains β -sheets and is called the β -domain (Thr40 to Ser85). The α -domain contains three large α -helices (α_1 - α_3), each having more than ten residues, and one small helix with six residues (α_4). This domain further includes both the N and C-terminal segments of the protein. The β -domain contains triple-stranded anti-parallel β -sheets (β_1 - β_3) with a long loop. In addition to the α -helices and β -sheets, there is a short 3^{10} helix in both the domains. The native structure of lysozyme is further stabilized by four disulfide bonds: two in α domain (Cys6-Cys127 and Cys30-Cys115), one in β -domain (Cys64-Cys80) and another connecting the two domains (Cys76-Cys94).

A number of experiments have inferred that lysozyme with four intact disulfide bonds folds through a fast and a slow folding pathway.¹⁴⁻¹⁷ These experiments have also concluded that an intermediate is populated in the slow folding pathway.⁴¹ However, a complete understanding of the slow folding pathway is still lacking. The reasons for kinetic partitioning in the folding pathways are unknown and the structure of the intermediate populated in the slow folding pathway is not available. Questions such as how do the disulfide bonds influence lysozyme folding pathways, and whether protein folding guides disulfide formation or vice versa during oxidative folding still persist.⁴²⁻⁴⁴ To address these questions we performed Brownian dynamics simulations to study lysozyme folding kinetics using the coarse-grained self-organized polymer-side chain (SOP-SC) model.^{45,46} To comprehensively understand the role of disulfide bonds in lysozyme folding, we performed simulations mimicking three different conditions. In the first set of simulations, we studied lysozyme folding with all the four intact disulfide bonds, in the second set we disabled the disulfide bond formation mimicking lysozyme folding in reduced conditions, and in the third set the disulfide bonds form as the protein folds mimicking oxidative folding.

Figure 2: Representative simulation trajectories of fast (A) and slow (B) folding pathways of lysozyme. Time evolution of the radius of gyration (R_g) shows that in fast folding pathways, protein folds to the native state ($R_g \approx 15 \text{ \AA}$) within 15 ms. In the slow pathways, protein does not fold within 500 ms of the simulation time. In the slow pathways, protein is found to be kinetically trapped in a state with $R_g \approx 17 \text{ \AA}$. The initial stages of folding in slow trajectories are shown in the inset of panel-B. (C) R_g plotted as a function of time for the five Brownian dynamics folding simulations spawned using the same initial unfolded conformation. Trajectories KIN2 and KIN5 fold within 10 ms using the fast folding pathway. The remaining three trajectories, KIN1, KIN3, and KIN4, fold through the slow folding pathway and are trapped in the intermediate state. This shows that stochasticity in the folding dynamics is responsible for kinetic partitioning in lysozyme folding pathways. (D) Time evolution of the ensemble average of R_g , ($\langle R_g(t) \rangle$) of all the folding trajectories. $\langle R_g \rangle$ as a function of time for the subset, fast and slow trajectories, is also shown. The dotted line shows R_g of the protein in the native state. All the fast trajectories (49 out of 100) fold to the native state within ~ 15 ms, while slow trajectories (51 out of 100) are trapped in a conformation with $R_g \sim 17 \text{ \AA}$.

Results and Discussion

We spawned 100 Brownian dynamics simulation trajectories to probe the folding of lysozyme with four intact disulfide bonds at $T = 320$ K ($\approx 0.8 T_M$, where T_M is the melting temperature). A detailed description of the coarse-grained protein model and simulation methods are in the supporting information (SI). In these conditions, the radius of gyration (R_g) of the protein as a function of time for different trajectories shows that lysozyme folds through a fast and a slow folding pathways exhibiting kinetic partitioning^{12–17} (Figure 2). In 49 out of the 100 folding trajectories, lysozyme folds through a fast folding pathway (Figure 2A). In rest of the trajectories, lysozyme folds through a slow folding pathway, where a partially folded intermediate state is populated (Figure 2B). The average folding time of the fast folding trajectories is 9.7 ± 4.0 ms, and in contrast, the folding time for the slow trajectories is on the order of hundreds of milliseconds, which is in agreement with the experiments.^{15,47} In 49 out of the 51 slow folding pathways, the protein did not fold completely on a time scale of 500 ms, which is the maximum run time of the simulations, and is trapped in a conformation with $R_g \sim 17$ Å for hundreds of milliseconds (Figure 2B).

The initial unfolded conformation in all the folding trajectories are different, and these conformations are generated using a simulation performed at temperature, $T = 420$ K ($> T_M$) (Figure S1). To rule out the possibility that the kinetic partitioning is due to a heterogeneous mixture of unfolded molecules,^{15,48} *i.e.* one set of unfolded molecules fold through the fast pathway and a second set folds through the slow pathway, we also performed multiple folding simulations starting from the same unfolded conformation. Even in this case, the protein folds using both the slow and fast folding routes exhibiting kinetic partitioning (Figure 2C). Thus, it is the stochasticity in the folding dynamics, which is responsible for the protein choosing either the fast or slow folding route. The simulations broadly support the triangular folding mechanism proposed by Guo and Thirumalai,¹³ and Wildegger and Kiefhaber,¹⁶ with a trapped kinetic state as shown below

where N, I and U refer to lysozyme native, intermediate and unfolded states, respectively.

On initiating folding, lysozyme forms compact structures in both the fast and slow folding pathways within ~ 5 ms (Figure 2D). During this compaction, the average radius of gyration, $\langle R_g \rangle$, of lysozyme decreases by ≈ 8 Å. The compact state has an average R_g of ~ 20 Å, which is in agreement with the experimentally⁴⁹ measured value of 19.6 Å. The formation of β -domain (Figure 3A) and helices (Figure 3B,S2) contribute to the compaction in protein dimensions. The compacted state mainly lacks the tertiary contacts present in α -domain and the inter-domain contacts (Figure S3). After the compaction, bifurcation in the folding pathways to fast and slow folding routes is observed. In the fast folding pathway, the protein continues to fold to its native structure with $R_g \approx 15$ Å, on an average time scale of ~ 15 ms with the formation of inter-domain contacts.⁴⁹ However, in the slow folding pathways, the protein gets trapped in an intermediate state with lifetimes on the order of hundreds of milliseconds. The R_g of this intermediate state is ≈ 17 Å, and it is in good agreement with the experimentally⁴⁹ measured value of 16.7 Å. The R_g value of the intermediate state indicates that the protein is close to the folded state (Figure 2D).

To probe the folding of α and β -domains, we computed $\langle R_g \rangle$ of these individual domains as a function of time. In both the fast and slow folding pathways, the β -domain folds completely within 5 ms reaching the native state $\langle R_g \rangle$ value of ~ 10 Å (Figure 3A). The time evolution of $\langle R_g \rangle$ in the folding of α -domain shows that it is the rate determining step in the folding of lysozyme. In the fast folding trajectory, the α -domain folds on a time scale of ~ 15 ms reaching the native state $\langle R_g(t) \rangle$ value of ~ 12.5 Å (Figure 3B). Whereas in the slow folding trajectories, the α -domain is stuck in an intermediate state with a $\langle R_g \rangle$ value of ~ 17.5 Å (Figure 3B). This shows that α -domain folding in lysozyme is responsible for the

Figure 3: Time evolution of the ensemble average of radius of gyration ($\langle R_g(t) \rangle$) of β -domain (A) and α -domain (B) in fast and slow folding trajectories. β -domain folds fast in both the pathways, while α -domain shows differences in the folding kinetics between fast and slow pathways. The dotted line shows R_g of the respective domains in the native state.

population of slow folding pathways in lysozyme folding.

In the folding of α -domain, the three large α -helices (α_1 - α_3) influence the folding kinetics. In both the fast and slow folding pathways, the formation of individual helices precedes the formation of inter-helical contacts. In 70% of the fast folding pathways, we find that the inter helical contacts between α_2 and α_3 ($\alpha_2\alpha_3$) form first, and then the inter helical contacts $\alpha_1\alpha_2$ and $\alpha_1\alpha_3$ form almost simultaneously along with the inter-domain contacts between the α and β -domains leading to the complete folding of the protein (Figure 4A). In the remaining 30% of fast folding trajectories, $\alpha_1\alpha_2$ contacts form first, followed by the simultaneous formation of $\alpha_2\alpha_3$, $\alpha_1\alpha_3$ and inter domain contacts (Figure 4B). In the fast folding trajectories, either $\alpha_2\alpha_3$ or $\alpha_1\alpha_2$ contacts form first followed by the simultaneous formation of other two sets of inter helical contacts and inter domain contacts. This mechanism of formation of inter-helical contacts with the orientation of the helices similar to the orientation present in the native state makes lysozyme folding smooth without any kinetic traps.

In the slow folding pathways, a significant fraction of $\alpha_1\alpha_2$ contacts form first (Figure 4C). This structure also prevents the formation of $\alpha_2\alpha_3$ and $\alpha_1\alpha_3$ contacts. By visually observing this trapped conformation, we noticed that the orientation of helix α_1 is in the opposite direction relative to the helices α_2 and α_3 compared to the native folded state (see the schematics in the inset of Figures 4B,C). For the protein trapped in this state to fold correctly, α_1 has to reorient by breaking the $\alpha_1\alpha_2$ contacts, and this leads to a long-lived kinetic intermediate (Figure 4D, S2). To quantitatively show that α_1 is misaligned relative to α_2 , we computed the scalar triple product $\Theta_{12} = (\vec{\alpha}_1 \times \vec{\alpha}_2) \cdot \hat{z}$, where $\vec{\alpha}_1$ is the vector joining the backbone beads of residues Arg5 and Arg14 in α_1 , $\vec{\alpha}_2$ is the vector joining the backbone beads of the residues Leu25 and Ser36 in α_2 , and $\hat{z}$ represents the unit vector along z -axis. The protein conformations before computing Θ_{12} are aligned to the native state. The plot of Θ_{12} as a function of time for the fast and slow folding pathways clearly shows that the helices α_1 and α_2 have a different orientation. The contact map of kinetically trapped intermediate state shows that the β -domain is completely folded, and α -domain is partially folded with

Figure 4: Time evolution of the fraction of native inter-helical ($f_{\alpha\alpha'}$) and inter-domain ($f_{\alpha\beta}$) contacts in representative fast (A and B) and slow (C) folding trajectories. (A) In 70% of the fast folding trajectories, $\alpha_2\alpha_3$ contacts form first followed by the formation of $\alpha_1\alpha_2$, $\alpha_1\alpha_3$ and $\alpha\beta$. (B) In the remaining 30% of the fast folding trajectories, $\alpha_1\alpha_2$ contacts form first followed by the formation of $\alpha_1\alpha_3$, $\alpha_2\alpha_3$ and $\alpha\beta$. (C) In the slow folding pathways, the protein is trapped in a kinetic intermediate, and in this state a significant fraction of $\alpha_1\alpha_2$ and $\alpha\beta$ contacts are formed. (D) The relative orientation of α_1 and α_2 helices in fast (panel-B) and slow (panel-C) folding pathways is shown using the scalar triple product, $\Theta_{12} = (\vec{\alpha}_1 \times \vec{\alpha}_2) \cdot \hat{z}$, where $\vec{\alpha}_n$ is the vector along α_n and $\hat{z}$ is the unit vector along z -axis. In both, fast and slow folding pathways (panel-B and C), $\alpha_1\alpha_2$ contacts form early, but their relative orientations are different. In the fast pathways, the relative orientation of α_1 and α_2 is similar to that of native state.

$\alpha_2\alpha_3$, $\alpha_1\alpha_3$ and some of the inter-domain contacts missing (Figure 5). The population of a misfolded intermediate leading to a slow lysozyme folding pathway was also speculated using a model four helix bundle by Thirumalai and Guo.¹²

Figure 5: (A) C_α contact map of the kinetic intermediate populated in the slow folding pathways. Contacts between the residues present in the native and intermediate state are shown in red and gray. The inter helical contacts, $\alpha_1\alpha_3$ and $\alpha_2\alpha_3$, and some of the inter-domain contacts ($\alpha\beta$) are missing in the intermediate state compared to the native state. (B) Representative structure of the kinetically trapped intermediate in the slow folding trajectories. The α and β -domains are shown in red and blue, respectively. In the intermediate state, β -domain is folded and α -domain is partially folded.

At 320 K, we observed complete folding of lysozyme in 2 out of the 51 slow folding trajectories within the simulation time of 500 ms. The simulations show that the kinetically trapped misfolded intermediate has to partially unfold before folding back to the native state as inferred in some of the experiments^{50,51} (Figure 6, S4). In these trajectories, the $\alpha_1\alpha_2$ contacts present between the misaligned helices in the intermediate state break down, allowing the protein to escape from this trapped state (Figure 6A). The disruption in the $\alpha_1\alpha_2$ contacts, allows α_1 to reorient to the correct alignment relative to α_2 , and the protein folds back to the native state (Figure 6B). If disruption in the $\alpha_1\alpha_2$ contacts is a primary bottle

neck for the protein to escape from this trapped state, then increased thermal fluctuations at a higher temperature should facilitate the protein to escape from the trapped state. To verify this we performed additional 50 Brownian dynamics simulations of lysozyme folding at a higher temperature, $T=380$ K ($< T_M$). Out of the 50 trajectories, the kinetically trapped intermediate state is populated in 14 trajectories. In 10 out of the 14 trajectories, on a time scale of 200 ms, the protein transitioned from the intermediate state by disrupting the $\alpha_1\alpha_2$ contacts, as observed in the simulations performed at the lower temperature (Figure 6,S4). Even in experiments,⁵⁰ the folding of lysozyme at a higher temperature is found to be more cooperative because the thermal fluctuations at a higher temperature will facilitate the protein to easily unfold locally and escape from the kinetic intermediate state. The population of the intermediate state is not due to the formation of non-native interactions as hypothesized,^{51,52} as the coarse-grained native-centric protein model^{45,46,53} used in this study has no energetic frustration. In this model, non-native interactions cannot form, as all the pairs of residues, which are not native, interact only with a repulsive potential (see SI for more details). As a result, only topological frustration, and no energetic frustration, exists in these models. This shows that it should be the topological frustration and with a possible role of the disulfide bonds, which constrain the conformational space available for the protein should facilitate the population of the kinetically trapped intermediate state.

To test the possible role of disulfide bonds in the population of kinetic intermediate, we performed additional 100 folding simulations of lysozyme without the presence of disulfide bonds at the same temperature $T = 320$ K. This mimics the folding of lysozyme in the presence of reducing agents such as dithiothreitol, which prevents the formation of disulfide bonds. In these simulations we did not observe any kinetically trapped intermediate states and all the trajectories reached the folded state within 200 ms (Figure 7A). This shows that in lysozyme conformations, where α_1 is misaligned relative to α_2 (Figure 5), the topological constraints due to the disulfide bonds makes the protein less flexible for structural inter-conversions, and these conformations result in kinetic traps in the folding pathways. To

Figure 6: (A) A representative slow folding trajectory, where the protein escapes from the kinetically trapped state and folds to the native state. The $\alpha_1\alpha_2$ contacts between the helices in the incorrect orientation fall apart at ~ 30 ms, allowing the protein to fold back to the native state. (B) The relative orientation of helices quantified using the scalar triple product, $\Theta_{ij} = (\vec{\alpha}_i \times \vec{\alpha}_j) \cdot \hat{z}$, where $\vec{\alpha}_i$ is the vector along α_i and $\hat{z}$ is the unit vector along z -axis, shows that α_1 is incorrectly oriented relative to α_2 and α_3 in the kinetically trapped state. As the $\alpha_1\alpha_2$ contacts fall apart at ~ 30 ms, the orientation changes and the protein folds to the native state. (C) Schematic showing the orientation of the helices in the trapped and native state. α_1 orientation in trapped and native states is different relative to α_2 and α_3 . The residues forming the disulfide bonds are shown as yellow dots.

escape from this conformation, local unfolding of lysozyme, which is the unwinding of $\alpha_1\alpha_2$ inter-helical contacts is required. In renaturing conditions, the time scales to cross the barrier for local unfolding is on the order of hundreds of milliseconds, and this results in the slow folding pathway of lysozyme. The population of a misfolded kinetic intermediate due to topological frustration, which requires local unfolding or back tracking to escape from the kinetic trap is also observed in the folding of interleukin-1 β ,^{54,55} green fluorescent protein (GFP),^{56,57} cytochrome *c*⁵⁸ and a model four helix bundle.^{12,13}

To probe the role of individual disulfide bonds in the population of the intermediate state, we performed four sets of simulations and in each set only one particular disulfide bond is kept intact and the other three bonds are disabled. In each set, we spawned 50 folding trajectories at $T = 320$ K. The simulations clearly shows that the two disulfide bonds, Cys6-Cys127 and Cys30-Cys115, present in the α -helical domain of the protein are responsible for the kinetic partitioning in the folding pathways and facilitate the population of the kinetic intermediate in the slow folding pathway in agreement with the experiments⁵⁹⁻⁶¹ (Figure 7B). When only these two disulfide bonds are present, the system shows a similar percentage of the population of slow trajectories compared to that of a protein with all the four disulfide bonds present. This observation is completely in agreement with the experiments of Giez *et al.*,⁶¹ and since the population of the kinetic intermediate is due to the restriction in the rearrangement of helices α_1 and α_2 , and the two disulfide bonds, Cys6-Cys127 and Cys30-Cys115, are attached to these helices.

Studies^{42,43} show evidence that protein folding guides disulfide bond formation. We performed simulations mimicking oxidative folding of lysozyme to probe the sequence of disulfide bond formation, and whether protein folding guides disulfide bond formation or vice versa. We used the method developed by Qin *et al.*⁴³ to mimic the disulfide bond formation. In this method, we assume that a disulfide bond is formed when three criteria related to proximity, orientation and solvent exposure of cystine residues are satisfied (see SI for details). When this criteria is satisfied, a FENE potential and an angle potential corresponding to

Figure 7: (A) Percentage of simulation trajectories, where the protein has folded as a function of simulation time. All the simulation trajectories fold within 200 ms when the disulfide bonds are absent, and their formation is disabled. In the presence of all the four disulfide bonds, $\sim 50\%$ of the trajectories fold, and the rest are stuck in a kinetically trapped intermediate state. This shows that disulfide bonds are responsible for the slow folding pathway and population of the kinetic intermediate. (B) Percentage of slow folding trajectories where kinetically trapped intermediate state is populated when different combinations of disulfide bonds are allowed in the protein. S_0 and S_4 represent the protein with zero disulfide bonds, and all the four disulfide bonds present, respectively. S_1^1 , S_1^2 , S_1^3 , and S_1^4 represent the protein with only one disulfide bond Cys64-Cys80, Cys76-Cys94, Cys6-Cys127, and Cys30-Cys115, respectively. Slow folding pathways are observed in proteins with Cys6-Cys127 or Cys30-Cys115 disulfide bonds. The number of slow trajectories observed when these two disulfide bonds (S_2^{34}) are present is equal to the number of slow trajectories observed when all the four disulfide bonds are present. This implies the disulfide bonds Cys6-Cys127 and Cys30-Cys115 present in the α -domain are responsible for the population of the kinetically trapped intermediate resulting in slow folding pathway.

the disulfide bond is added to the side chains of the cystine residues mimicking the bond formation.

Figure 8: A representative folding trajectory of lysozyme in oxidative folding conditions. In the initial unfolded conformation, none of the disulfide bonds are present. The distance between various pairs of Cys residues, which form disulfide bonds in the native state are plotted as a function of time. The sequence of disulfide bond formation is as follows: Cys64-Cys80 $\rightarrow$ Cys76-Cys94 $\rightarrow$ Cys30-Cys115 $\rightarrow$ Cys6-Cys127. Representative snapshots of lysozyme conformations show the formation of disulfide bonds as the protein folds.

We have spawned 100 lysozyme folding trajectories at $T = 320$ K mimicking folding in oxidative conditions. In all the folding trajectories, we find that Cys64-Cys80 disulfide bond present in the β -domain forms within 10 ms after folding is initiated (Figure 8). The second disulfide bond formed is the Cys76-Cys94, and it is located at the interface of the α and β -domains. The disulfide bonds, Cys64-Cys80 and Cys76-Cys94, are short-ranged native contacts between the side chains of the Cys residues forming these bonds. The number of

residues separating the Cys residues forming the bonds along the contour of lysozyme chain is less than 18. As a result, during the initial protein collapse, these Cys residues come closer and become solvent inaccessible leading to the formation of the disulfide bond.

The other two disulfide bonds, Cys30-Cys115 and Cys6-Cys127, which form in the last stages of folding are in the α -domain, and these bonds are long-ranged native contacts between the side chains of the Cys residues forming these bonds, as they are well separated along the contour of the lysozyme chain. The Cys30-Cys115 and Cys6-Cys127 disulfide bonds are either formed simultaneously (only 6 % - Figure S4), or Cys30-Cys115 is formed before Cys6-Cys127. The order of disulfide bond formation in lysozyme during oxidative folding is Cys64-Cys80, Cys76-Cys94, Cys30-Cys115, and Cys6-Cys127.

If any one of the disulfide bonds form before its precursor in the order, then the protein does not fold to its native state (Figure S5). The disulfide bond, which has formed prematurely in the order gets solvent exposed, *i.e.* in this case, the bond is not buried in the protein core and the solvent accessible surface area (SASA) of the bond increases not satisfying the disulfide bond formation criteria, and it breaks. This allows the possibility that the bonds form in the order specified, allowing the protein fold to its native state (Figure S5D). In the folding pathways, we did not observe any non-native disulfide bond formation, and this shows that protein folding guides disulfide bond formation in agreement with the previous work.^{42,43}

Conclusions

In summary, using a coarse-grained protein model and molecular dynamics simulations we elucidated the role of disulfide bonds and topological frustration in the folding pathways of lysozyme (Figure 9). The kinetic partitioning¹³ in lysozyme folding pathways to fast and slow routes is due to the two disulfide bonds present in the α -domain of the protein. In the slow pathway, a kinetically trapped intermediate state is populated in the late stages

Figure 9: Schematic of dominant folding pathways of hen egg white lysozyme. The three α -helices, α_1 - α_3 , that influence the folding kinetics are shown in purple, red and green, respectively. The triple stranded β -sheets, which fold early are shown as one blue arrow. After folding is initiated from an unfolded state (U), the R_g of the protein decreases by $\sim 8 \text{ \AA}$ in size in the first 5 ms due to the folding of β -domain and α -helices. From this compacted state (C), the folding bifurcates to three different pathways: 1 slow and 2 fast pathways. In the slow pathway, a kinetically trapped state (I) is observed. In this state, the inter-helical contacts α_{12} are formed, but the orientation of α_1 relative to α_2 is in the opposite direction compared to the native state. For the protein to fold from this kinetically trapped state, the protein has to backtrack by breaking the α_{12} contacts, which will facilitate reorientation of α_1 allowing the protein fold to the native state (F). This rearrangement is restricted by the presence of two disulfide bonds, Cys6-Cys127 and Cys30-Cys115, present in the α -domain (Figure 7). In the fast folding pathways, transient intermediate I_1 or I_2 are populated. In I_1 , α_{23} contacts are formed initially, followed by α_{12} , α_{13} and inter-domain contacts. In I_2 , α_{12} contacts are formed initially and helices (α_1 and α_2) have the correct orientation as observed in the native state. In the later stages, α_{13} , α_{23} and inter-domain contacts form and the protein folds to the native state (F).

of folding due to misalignment in the packing of α -helices present in the α -domain. The protein has to partially unfold by breaking the inter-helical contacts between the misaligned helices to escape from this trapped state. However, due to the topological restraints imposed by the disulfide bonds in the α -domain, the protein is less flexible and it is trapped in the intermediate state for hundreds of milliseconds leading to the slow folding pathway. The schematic describing lysozyme folding pathways is shown in Figure 9, and the energy surface corresponding to this schematic inferred from various experiments was shown by Dobson *et al.* in Ref.⁶² On disabling the disulfide bonds, the kinetically trapped intermediate state is not populated and the slow folding pathway disappeared. On mimicking lysozyme folding in oxidative conditions by starting the simulations with unfolded conformations without disulfide bonds, we find that it is protein folding, which guides the formation of disulfide bonds.^{42,43}

The kinetically trapped intermediate state populated in the slow folding pathway, which has a lifetime of hundreds of milliseconds, can have implications in the aggregation of lysozyme. In this state, the interface of the α and β -domains are not completely formed. As the concentration of the kinetically trapped misfolded intermediate increases, it can lead to aggregation through the exposed β -sheets.^{32,33} Experiments³³ show that mutants, which enhance the population of nonnative states in lysozyme tend to increase the propensity to form fibrils. This shows that disulfide bonds can enhance frustration in the folding landscape of proteins through topological constraints, which can lead to the population of long lived intermediates that have a propensity to aggregate.

Materials and Methods

To study the folding kinetics of lysozyme we used the SOP-SC (self-organized polymer-side chain)^{45,46} coarse-grained model for proteins. SOP-SC is a native-centric model⁵³ in which each amino acid in protein is represented by two beads, one for the backbone and the other

for side chain. We used the crystal structure¹⁸ of lysozyme with PDB ID: 1AKI to construct the coarse-grained structures. A detailed description of the SOP-SC model and force field is described in the supplementary information (SI). Brownian dynamics simulations⁶³ are performed to probe lysozyme folding kinetics. To mimic the folding of protein in oxidative conditions, where the disulfide bonds form as the protein folds we used the model developed by Qin *et al.*⁴³ Details about the simulation methodology and disulfide bond model are described in the SI.

Acknowledgement

A part of this work is funded by the grants to G.R. and P.C.S. from Science and Engineering Research Board, Department of Science and Technology, India through the grants EMR/2016/001356 and EMR/2015/001605, respectively. The computations are performed using the HPC at TUE-CMS in IISc funded by Nano mission. A.N.M. acknowledges research fellowship from Council of Scientific and Industrial Research, India.

Supporting Information Available

Detailed description of the protein model, criteria for disulfide bond formation and simulation methods; Figures S1-S6

References

- (1) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the folding routes. *Science* **1995**, *267*, 1619–1620.
- (2) Narayan, M.; Welker, E.; Wedemeyer, W.; Scheraga, H. Oxidative folding of proteins. *Accounts Chem. Res.* **2000**, *33*, 805–812.

- (3) Chang, J.-Y. Diverse pathways of oxidative folding of disulfide proteins: Underlying causes and folding models. *Biochemistry* **2011**, *50*, 3414–3431.
- (4) Anfinsen, C. Principles that govern folding of protein chains. *Science* **1973**, *181*, 223–230.
- (5) Haber, E.; Anfinsen, C. Side-chain interactions governing pairing of half-cystine residues in ribonuclease. *J. Biol. Chem.* **1962**, *237*, 1839.
- (6) Fass, D. Disulfide bonding in protein biophysics. *Annu. Rev. Biophys.* **2012**, *41*, 63–79.
- (7) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* **1995**, *21*, 167–195.
- (8) Thirumalai, D.; Woodson, S. Kinetics of folding of proteins and RNA. *Accounts Chem. Res.* **1996**, *29*, 433–439.
- (9) Dill, K.; Chan, H. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **1997**, *4*, 10–19.
- (10) Ferreira, D. U.; Komives, E. A.; Wolynes, P. G. Frustration in biomolecules. *Q. Rev. Biophys.* **2014**, *47*, 285–363.
- (11) Thirumalai, D.; Klimov, D.; Woodson, S. Kinetic partitioning mechanism as a unifying theme in the folding of biomolecules. *Theor. Chem. Acc.* **1997**, *96*, 14–22.
- (12) Thirumalai, D.; Guo, Z. Nucleation mechanism for protein-folding and theoretical predictions for hydrogen-exchange labeling experiments. *Biopolymers* **1995**, *35*, 137–140.
- (13) Guo, Z. Y.; Thirumalai, D. Kinetics of protein-folding - Nucleation mechanism, time Scales, and pathways. *Biopolymers* **1995**, *36*, 83–102.
- (14) Radford, S.; Dobson, C.; Evans, P. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* **1992**, *358*, 302–307.

- (15) Kiefhaber, T. Kinetic traps in lysozyme folding. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 9029–9033.
- (16) Wildegger, G.; Kiefhaber, T. Three-state model for lysozyme folding: Triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.* **1997**, *270*, 294–304.
- (17) Matagne, A.; Radford, S.; Dobson, C. Fast and slow tracks in lysozyme folding: Insight into the role of domains in the folding process. *J. Mol. Biol.* **1997**, *267*, 1068–1074.
- (18) Artymiuk, P.; Blake, C.; Rice, D.; Wilson, K. The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme at 6Å resolution. *Acta Crystallogr. Sect. B-Struct. Commun.* **1982**, *38*, 778–783.
- (19) Dobson, C.; Evans, P.; Radford, S. Understanding how proteins fold - The lysozyme story so far. *Trends Biochem.Sci.* **1994**, *19*, 31–37.
- (20) Matagne, A.; Dobson, C. The folding process of hen lysozyme: A perspective from the ‘new view’. *Cell. Mol. Life Sci.* **1998**, *54*, 363–371.
- (21) Baase, W. A.; Liu, L.; Tronrud, D. E.; Matthews, B. W. Lessons from the lysozyme of phage T4. *Protein Sci.* **2010**, *19*, 631–641.
- (22) Stirnemann, G.; Kang, S.-g.; Zhou, R.; Berne, B. J. How force unfolding differs from chemical denaturation. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 3413–3418.
- (23) Makhatadze, G.; Privalov, P. Protein interactions with urea and guanidinium chloride - A calorimetric study. *J. Mol. Biol.* **1992**, *226*, 491–505.
- (24) Ibarra-Molero, B.; Sanchez-Ruiz, J. Are there equilibrium intermediate states in the urea-induced unfolding of hen egg-white lysozyme? *Biochemistry* **1997**, *36*, 9616–9624.
- (25) Sasahara, K.; Demura, M.; Nitta, K. Equilibrium and kinetic folding of hen egg-white lysozyme under acidic conditions. *Proteins* **2002**, *49*, 472–482.

- (26) Hua, L.; Zhou, R.; Thirumalai, D.; Berne, B. J. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 16928–16933.
- (27) Xia, Z.; Das, P.; Shakhnovich, E. I.; Zhou, R. Collapse of unfolded proteins in a mixture of denaturants. *J. Am. Chem. Soc.* **2012**, *134*, 18266–18274.
- (28) Roy, S.; Jana, B.; Bagchi, B. Dimethyl sulfoxide induced structural transformations and non-monotonic concentration dependence of conformational fluctuation around active site of lysozyme. *J. Chem. Phys.* **2012**, *136*, 115103.
- (29) Bisht, M.; Kumar, A.; Venkatesu, P. Refolding effects of partially immiscible ammonium-based ionic liquids on the urea-induced unfolded lysozyme structure. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12419–12422.
- (30) Biswas, B.; Muttathukattil, A. N.; Reddy, G.; Singh, P. C. Contrasting effects of guanidinium chloride and urea on the activity and unfolding of lysozyme. *ACS Omega* **2018**, *3*, 14119–14126.
- (31) Zhang, Y.; Cremer, P. S. The inverse and direct Hofmeister series for lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 15249–15253.
- (32) Booth, D.; Sunde, M.; Bellotti, V.; Robinson, C.; Hutchinson, W.; Fraser, P.; Hawkins, P.; Dobson, C.; Radford, S.; Blake, C.; Pepys, M. Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **1997**, *385*, 787–793.
- (33) Buell, A. K.; Dhulesia, A.; Mossuto, M. F.; Cremades, N.; Kumita, J. R.; Dumoulin, M.; Welland, M. E.; Knowles, T. P. J.; Salvatella, X.; Dobson, C. M. Population of non-native states of lysozyme variants drives amyloid fibril formation. *J. Am. Chem. Soc.* **2011**, *133*, 7737–7743.

- (34) Wu, S.; Ding, Y.; Zhang, G. Mechanic insight into aggregation of lysozyme by ultra-sensitive differential scanning calorimetry and sedimentation velocity. *J. Phys. Chem. B* **2015**, *119*, 15789–15795.
- (35) Bouvignies, G.; Vallurupalli, P.; Hansen, D. F.; Correia, B. E.; Lange, O.; Bah, A.; Vernon, R. M.; Dahlquist, F. W.; Baker, D.; Kay, L. E. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **2011**, *477*, 111–U134.
- (36) Yokomizo, T.; Higo, J.; Nakasako, M. Patterns and networks of hydrogen-bonds in the hydration structure of human lysozyme. *Chem. Phys. Lett.* **2005**, *410*, 31–35.
- (37) Jana, B.; Pal, S.; Bagchi, B. Hydrogen bond breaking mechanism and water reorientational dynamics in the hydration layer of lysozyme. *J. Phys. Chem. B* **2008**, *112*, 9112–9117.
- (38) Yamamoto, N.; Ohta, K.; Tamura, A.; Tominaga, K. Broadband dielectric spectroscopy on lysozyme in the sub-gigahertz to terahertz frequency regions: Effects of hydration and thermal excitation. *J. Phys. Chem. B* **2016**, *120*, 4743–4755.
- (39) Vocadlo, D.; Davies, G.; Laine, R.; Withers, S. Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **2001**, *412*, 835–838.
- (40) Ringe, D.; Petsko, G. A. How enzymes work. *Science* **2008**, *320*, 1428–1429.
- (41) Dhulesia, A.; Cremades, N.; Kumita, J. R.; Hsu, S.-T. D.; Mossuto, M. F.; Dumoulin, M.; Nietlispach, D.; Akke, M.; Salvatella, X.; Dobson, C. M. Local cooperativity in an amyloidogenic state of human lysozyme observed at atomic resolution. *J. Am. Chem. Soc.* **2010**, *132*, 15580–15588.
- (42) Kosuri, P.; Alegre-Cebollada, J.; Feng, J.; Kaplan, A.; Ingles-Prieto, A.; Badilla, C. L.; Stockwell, B. R.; Sanchez-Ruiz, J. M.; Holmgren, A.; Fernandez, J. M. Protein folding drives disulfide formation. *Cell* **2012**, *151*, 794–806.

- (43) Qin, M.; Wang, W.; Thirumalai, D. Protein folding guides disulfide bond formation. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 11241–11246.
- (44) Lv, J.-M.; Lu, S.-Q.; Liu, Z.-P.; Zhang, J.; Gao, B.-X.; Yao, Z.-Y.; Wu, Y.-X.; Potempa, L. A.; Ji, S.-R.; Long, M.; Wu, Y. Conformational folding and disulfide bonding drive distinct stages of protein structure formation. *Sci Rep* **2018**, *8*, 1494.
- (45) Hyeon, C.; Dima, R. I.; Thirumalai, D. Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. *Structure* **2006**, *14*, 1633–1645.
- (46) Liu, Z.; Reddy, G.; O'Brien, E. P.; Thirumalai, D. Collapse kinetics and chevron plots from simulations of denaturant-dependent folding of globular proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 7787–7792.
- (47) Radford, S.; Dobson, C. Insights into protein-folding using physical techniques - studies of lysozyme and alpha-lactalbumin. *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.* **1995**, *348*, 17–25.
- (48) Kotik, M.; Radford, S.; Dobson, C. Comparison of the refolding of hen lysozyme from dimethyl-sulfoxide and guanidinium chloride. *Biochemistry* **1995**, *34*, 1714–1724.
- (49) Segel, D.; Bachmann, A.; Hofrichter, J.; Hodgson, K.; Doniach, S.; Kiefhaber, T. Characterization of transient intermediates in lysozyme folding with time-resolved small-angle X-ray scattering. *J. Mol. Biol.* **1999**, *288*, 489–499.
- (50) Matagne, A.; Chung, E.; Ball, L.; Radford, S.; Robinson, C.; Dobson, C. The origin of the alpha-domain intermediate in the folding of hen lysozyme. *J. Mol. Biol.* **1998**, *277*, 997–1005.
- (51) Rothwarf, D.; Scheraga, H. Role of non-native aromatic and hydrophobic interactions in the folding of hen egg white lysozyme. *Biochemistry* **1996**, *35*, 13797–13807.

- (52) Kazmirski, S.; Daggett, V. Non-native interactions in protein folding intermediates: Molecular dynamics simulations of hen lysozyme. *J. Mol. Biol.* **1998**, *284*, 793–806.
- (53) GO, N. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183–210.
- (54) Gosavi, S.; Chavez, L.; Jennings, P.; Onuchic, J. Topological frustration and the folding of interleukin-1 beta. *J. Mol. Biol.* **2006**, *357*, 986–996.
- (55) Capraro, D. T.; Roy, M.; Onuchic, J. N.; Jennings, P. A. Backtracking on the folding landscape of the beta-trefoil protein interleukin-1 beta? *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 14844–14848.
- (56) Andrews, B. T.; Gosavi, S.; Finke, J. M.; Onuchic, J. N.; Jennings, P. A. The dual-basin landscape in GFP folding. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 12283–12288.
- (57) Reddy, G.; Liu, Z.; Thirumalai, D. Denaturant-dependent folding of GFP. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17832–17838.
- (58) Weinkam, P.; Zimmermann, J.; Romesberg, F. E.; Wolynes, P. G. The folding energy landscape and free energy excitations of cytochrome C. *Accounts Chem. Res.* **2010**, *43*, 652–660.
- (59) Denton, M.; Rothwarf, D.; Scheraga, H. Kinetics of folding of guanidine-denatured hen egg-white lysozyme and coarboxymethyl(Cys(6),Cys(L27))-lysozyme - A stopped-flow absorbency and fluorescence study. *Biochemistry* **1994**, *33*, 11225–11236.
- (60) Eyles, S.; Radford, S.; Robinson, C.; Dobson, C. Kinetic consequences of the removal of a disulfide bridge on the folding of hen lysozyme. *Biochemistry* **1994**, *33*, 13038–13048.
- (61) Guez, V.; Roux, P.; Navon, A.; Goldberg, M. Role of individual disulfide bonds in hen lysozyme early folding steps. *Protein Sci.* **2002**, *11*, 1136–1151.

- (62) Dobson, C.; Sali, A.; Karplus, M. Protein folding: A perspective from theory and experiment. *Angew. Chem.-Int. Edit.* **1998**, *37*, 868–893.
- (63) Ermak, D. L.; Mccammon, J. A. Brownian dynamics with hydrodynamic interactions. *J Chem Phys* **1978**, *69*, 1352–1360.