Collapse Precedes Folding in Denaturant-Dependent Assembly of Ubiquitin

Govardhan Reddy*,† and D. Thirumalai‡

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

‡*Department of Chemistry, University of Texas at Austin, Austin, TX 78712*

E-mail: greddy@sscu.iisc.ernet.in

Abstract

Despite the small size the folding of Ubiquitin (Ub), which plays an indispensable role in targeting proteins for degradation and DNA damage response, is complex. A number of experiments on Ub folding have reached differing conclusions regarding the relation between collapse and folding, and whether intermediates are populated. In order to resolve these vexing issues, we elucidate the denaturant-dependent thermodynamics and kinetics of Ub folding in low and neutral pH as a function of Guanidinium chloride and Urea using coarse-grained molecular simulations. The changes in the fraction of the folded Ub, and the radius of gyration (R_g) as a function of the denaturant concentration, $[C]$, are in quantitative agreement with experiments. Under conditions used in experiments, R_g of the unfolded state at neutral pH changes only by $\approx 17\%$ as the $[GdmCl]$ decreases from 6 M to 0 M. We predict that the extent of compaction of the unfolded state increases as temperature decreases. A two-dimensional folding landscape as a function of R_g and a measure of similarity to the folded state reveals unambiguously that the native state assembly is preceded by collapse, as discovered in fast mixing experiments on several proteins. Analyses of the folding trajectories,

under mildly denaturing conditions ($[GdmCl]$=1.0M or $[Urea]$=1.0M), shows that Ub folds by collision between preformed secondary structural elements involving kinetic intermediates that are primarily stabilized by long-range contacts. Our work explains the results of Small Angle X-Ray Scattering (SAXS) experiments on Ub quantitatively, and establishes that evolved globular proteins are poised to collapse. In the process, we explain the discrepancy between SAXS and single molecule fluorescent resonant energy transfer (smFRET) experiments, which have arrived at a contradicting conclusion concerning the collapse of polypeptide chains.

Introduction

Mono and poly-ubiquitin play an important role in cell signaling pathways. Ubiquitination of specific lysine residues in the target proteins[1] is a signal for triggering cellular processes such as protein degradation by the proteosomes,[2] and DNA damage response needed for genome stability.[3] Thus, understanding how Ubiquitin (Ub) folds is important in describing its cellular functions. The crystal structure[4] (PDB ID: 1UBQ) of the 76 residue monomeric Ub shows that the folded state has five β-sheets and two helices (Figure 1A). The C_α contact-map based on α-carbon atom shows that there are short-range contacts between the amino acid residues in the helices (α_1 and α_2), and the hairpin $\beta_1\beta_2$. Long range contacts connect the strands β_1-β_5, β_3-β_5, and the loops L_1-L_2 (Figure S1). The predicted complex Ub folding kinetics[5] as a function of temperature is attributed to the multiple long-range contacts in the folded state of Ub setting the stage for an in depth investigation of how chemical denaturants modulate the folding landscape of Ub.

A combination of experiments[6–9] and theory[10–15] has yielded insights into the generic mechanisms by which proteins fold. However, it has been difficult to provide a molecular description of folding mechanisms, which are experimentally probed using denaturants whereas most of the folding simulations are done by varying temperature. Because of the large differences in the impact of temperature and denaturants on the various states of proteins direct comparisons between experiments and computations is largely elusive. This is a major bottleneck in assessing the efficacy of simulations. A major advance in overcoming this bottleneck was made with the introduction of the phenomenological Molecular Transfer Model (MTM)[16,17] that combines the classical transfer model and coarse-grained representation of the polypeptide chain. In several previous studies[16–21] MTM has been applied to *quantitatively* describe denaturant-dependent folding of a number of proteins containing between 50-250 amino acid residues.

Here, we investigate the thermodynamics and kinetics of Ub folding, which has been investigated by a variety of experiments[22–39] and computations.[5,40–54] However, the effects of

denaturants such as Guanidinium chloride and Urea, have not been considered in simulations. This is important because there are few central experimental controversies, such as the extent of collapse in the denatured ensemble of Ub as the denaturant concentration is lowered and if intermediates are present, that remain unresolved. We provide quantitative insights into these problems by dissecting Ub folding using simulations accounting for both denaturant and pH effects.

Ub has a significant number of charged residues, making the folding properties dependent on pH.[5,38,39,55] By incorporating electrostatic interactions within the framework of the SOP-SC model[19,56] we performed exhaustive simulations over a wide range of temperature and denaturant concentrations. Various thermodynamic properties of Ub computed from simulations, such as the fraction of the protein in the unfolded basin of attraction f_{UBA}, the free energy difference between the folded and unfolded state ΔG_{NU}, and the radius of gyration R_g, as a function of the denaturant concentrations are in excellent agreement with experiments.[36,39,57] The pair distance distribution functions $P(r)$, and R_g computed for the burst phase of Ub folding show that the size of Ub decreases only modestly in the early stages of folding, in agreement with the Small Angle X-ray scattering experiments (SAXS).[57] The equilibrium values of the radius of gyration (R_g^{UBA}) of the unfolded ensemble decreases continuously as the denaturant concentration is lowered. Folding landscape as a function of R_g and the structural overlap function (χ) shows that compaction precedes folding. Ub in highly stabilizing conditions folds through the diffusion-collision model by populating distinct kinetic intermediates with partially folded structures. Our study tidily resolves the apparent controversies between conclusions reached using SAXS and smFRET experiments, and suggests that the tendency to collapse is universal of all evolved globular proteins.

Methods

Self Organized Polymer-Side Chain (SOP-SC) model: As described in detail else-

where,[5] we modeled Ub using a coarse-grained SOP-SC model.[19,56] Each residue is represented using two interaction centers, one for the backbone atoms and the other for the side chain (SC). The interaction centers are at the C_α atom position of the residue, and the center of mass of the side chain. The SCs interact via a residue-dependent statistical potential.[58] Acidic residues (Figure 1A) are protonated at low pH, minimizing the effect of electrostatic interactions. To mimic Ub folding in neutral pH, we added charges by placing them on the side chains of the charged residues. The SOP-SC models for Ub are constructed using the crystal structure.[4]

The force field in the SOP-SC model is a sum of bonded and non-bonded interactions. The bonded interactions (E_B), between a pair of connected beads (two successive C_α atoms or a SC connected to a C_α atom), account for chain connectivity. The non-bonded interactions are a sum of native (E_{NB}^N) and non-native (E_{NB}^{NN}) interactions. If two beads are separated by at least three bonds, and if the distance between them in the coarse-grained crystal structure is less than a cutoff distance R_c (Table S1) then their interactions are considered native. The rest of the pairs of beads, not covalently linked or native, are classified as non-native interactions. Electrostatic effects at neutral pH are modeled using the screened Coulomb potential (E^{el}).[5]

The coarse-grained force-field of a protein conformation in the SOP-SC model represented by the coordinates $\{\mathbf{r}\}$ at $[C] = 0$ is given by

$$E_{CG}(\{\mathbf{r}\}, 0) = E_B + E_{NB}^N + E_{NB}^{NN} + \lambda E^{el}. \quad (1)$$

Description of the various energy terms in Equation 1 can be found in the SI. The parameter λ can take the values 0 or 1 to either switch off or switch on the electrostatic effects. The parameters used in the energy function are in Table S1 in the SI.

Molecular Transfer Model: We used the Molecular Transfer Model (MTM) model[16,17] to compute the Ub folding thermodynamics and kinetics in the presence of denaturants, Urea

and Guanidinium Hydrochloride. In a solution at denaturant concentration $[C]$, the effective coarse-grained force field for the protein using MTM framework is,

$$E_{CG}(\{\mathbf{r}\},[C]) = E_{CG}(\{\mathbf{r}\},0) + \Delta G_{tr}(\{\mathbf{r}\},[C]), \tag{2}$$

where $E_{CG}(\{\mathbf{r}\},0)$ is given by Equation 1. The protein-denaturant interaction free energy ($\Delta G_{tr}(\{\mathbf{r}\},[C])$) in an aqueous solution at a denaturant concentration [C] is given by

$$\Delta G_{tr}(\{\mathbf{r}\},[C]) = \sum_{k=1}^{N} \delta g_{tr,k}([C])\alpha_k(\{\mathbf{r}\})/\alpha_{Gly-k-Gly}, \tag{3}$$

where $N(=N_{res}\times 2 = 152)$ is the number of residues in Ub, $\delta g_{tr,k}([C])$ is the transfer free energy of bead k, $\alpha_k(\{\mathbf{r}\})$ is the solvent accessible surface area (SASA) of the site k in a protein conformation described by positions $\{\mathbf{r}\}$, $\alpha_{Gly-k-Gly}$ is the SASA of the bead k in the tripeptide $Gly-k-Gly$. The radii for side chains of amino acids needed to compute $\alpha_k(\{\mathbf{r}\})$ are given in Table S2 in Ref.[19] The experimental[16,18,59] transfer free energies $\delta g_{tr,i}([C])$, which depend on the chemical nature of the denaturant, for backbone and side chains are listed in Table S3 in Ref.[19] The values for $\alpha_{Gly-k-Gly}$ are listed in Table S4 in Ref.[19]

Simulations: We used low friction Langevin dynamics simulations[60] to obtain the thermodynamic properties. The average value of a physical quantity, A, at any temperature, T and [C] is calculated[16,17] using the Weighted Histogram Method (WHAM),[61]

$$\langle A([C],T)\rangle = Z([C],T)^{-1}\sum_{k=1}^{R}\sum_{t=1}^{n_k}\frac{A_{k,t}e^{-(E_{k,t}(\{\mathbf{r}_{k,t}\},[0])+\Delta G_{tr}(\{\mathbf{r}_{k,t}\},[C]))/k_BT}}{\sum_{m=1}^{R} n_m e^{f_m - E_{k,t}(\{\mathbf{r}_{k,t}\},[0])/k_BT_m}}. \tag{4}$$

In Equation 4, R is the number of simulation trajectories, n_k is the number of protein conformations from the k^{th} simulation, $A_{k,t}$ is the value of the property of the t^{th} conformation from the k^{th} simulation, T_m and f_m are the temperature and free energy respectively from the m^{th} simulation, $E_{k,t}(\{\mathbf{r}_{k,t}\},[0])$ and $\Delta G_{tr}(\{\mathbf{r}_{k,t}\},[C]))$ are the internal energy at $[C] = 0$ M

and MTM energy respectively of the t^{th} conformation from the k^{th} simulation. The partition function $Z([C],T)$ is,

$$Z([C],T) = \sum_{k=1}^{R}\sum_{t=1}^{n_k} \frac{e^{-(E_{k,t}(\{\mathbf{r}_{k,t}\},[0])+\Delta G_{tr}(\{\mathbf{r}_{k,t}\},[C]))/k_BT}}{\sum_{m=1}^{R} n_m e^{f_m - E_{k,t}(\{\mathbf{r}_{k,t}\},[0])/k_BT_m}}. \quad (5)$$

We performed Brownian dynamics simulations[62] to simulate the protein folding kinetics. To obtain the folding trajectories at the concentration $[C]$, the full Hamiltonian (Equation 2) with a friction coefficient corresponding to water (see SI for details) is used.

Data Analysis: The structural overlap function,[63] $\chi = 1 - \frac{1}{N_{tot}}\sum_{i=1}^{N_{tot}} \Theta\left(\delta - |r_i - r_i^0|\right)$, distinguishes the naive basin of attraction (NBA) and unfolded basin of attraction (UBA); $N_{tot}(= 11026)$ is the number of pairs of interaction centers in the SOP-SC model of Ub separated by at least 2 bonds, r_i is the distance between the i^{th} pair of beads, and r_i^0 is the distance in the folded state, Θ is the Heaviside step function, and $\delta = 2$ Å. A plot of χ as a function of time, and the probability distribution of χ at the melting temperature T_m, in both neutral and acidic pH are shown in Figure S2 in the SI. The fraction of molecules in the unfolded basin of attraction (UBA) as a function of the denaturant concentration, $[C]$, is calculated using χ as an order parameter (see SI for details). The radius of gyration is calculated using $R_g = (1/2N^2)(\sum_{i,j}\vec{r}_{ij}^{\,2})^{1/2}$, where $\vec{r}_{ij}$ is the vector connecting interaction centers i and j.

In order to determine the order of formation of contacts that are separated along the sequence, we used the local contact order,[64,65] S_i. We define S_i associated with residue i as,

$$S_i = \frac{1}{M}\sum_{j=1}^{N_{res}} \frac{|i-j|}{N_{res}} \delta^0_{|r_i - r_j|}, \quad (6)$$

where $|i-j|$ is the length of the protein sequence between the residues i and j, M is the number of native contacts that the i^{th} residue forms with the other residues in the folded state, $\delta^0_{|r_i-r_j|}$ is the Kronecker delta function used to count the native contacts formed by

the residue i, and superscript 0 indicates that only native contacts are considered.

Results

Thermodynamics of Ubiquitin folding: We computed the fraction of Ub in the unfolded basin of attraction (f_{UBA}), radius of gyration (R_g), and the free energy difference between the native and unfolded state ($\Delta G_{NU} = G_{NBA} - G_{UBA}$) as a function of the denaturant concentration ($[C]$) in both low pH and neutral pH (Figure 1).

Linear extrapolation method is valid in Urea: In order to compute the thermodynamic properties as a function of $[Urea]$, we choose $T = 343.25\ K$, since $\Delta G_{NU}(T = 343.25\ K,\ [C] = 0\ M) \approx 3.23\ kcal/mol$ is approximately equal to the experimental value $\Delta G_{NU}(T = 298\ K,\ [Urea] = 0\ M)$ at pH = 2.0.[39] We choose $T = 333\ K$ to calculate the $[GdmCl]$-dependent properties, since $\Delta G_{NU}(T = 333\ K,\ [C] = 0\ M) \approx 6.6\ kcal/mol$ is approximately equal to the $\Delta G_{NU}(T = 298\ K,\ [GdmCl] = 0\ M)$ found in experiments.[39] The simulation temperatures in the two denaturants are different because experiments[39] at pH = 2.0 show that the $\Delta G_{NU}([C] = 0\ M)$ values extracted using Urea and GdmCl experiments do not coincide. The value $\Delta G_{NU}([Urea] = 0\ M) = 3.23\ kcal/mol$ obtained from Urea denaturation experiments agrees with the differential scanning calorimetry experiments[66] in water indicating that the linear extrapolation method (LEM) used to extrapolate $\Delta G_{NU}([Urea])$ as function of $[Urea]$ data to obtain $\Delta G_{NU}([Urea] = 0\ M)$, is reliable. The LEM method fails for $[GdmCl]$ experiments, as the plot of $\Delta G_{NU}([GdmCl])$ as a function of $[GdmCl]$ for concentrations $[GdmCl] < 1.0\ M$ is not linear. When this plot is linearly extrapolated to $[GdmCl] = 0\ M$, it yields $\Delta G_{NU}([GdmCl] = 0\ M) = 6.6\ kcal/mol$, approximately twice the value obtained in differential scanning experiments.[66]

To compute thermodynamic properties as a function of denaturant concentration in neutral pH we choose $T = 333.25\ K$ ensuring that $\Delta G_{NU}(T = 333.25\ K,\ [C] = 0\ M) \approx 7.5\ kcal/mol$ is approximately equal to $\Delta G_{NU}(T = 298\ K,\ [GdmCl] = 0\ M)$ from exper-

iments[36] at pD = 5.0. Experiments for Ub denaturation in neutral pH by Urea are not available. We used this procedure of choosing simulation temperatures to compute properties in different denaturing conditions because we cannot get absolute free energies using CG models. It is worth emphasizing that this is the only adjustable parameter in the MTM model.

Population of the unfolded state as a function of $[C]$**:** The calculated f_{UBA} in low and neutral pH as a function of $[GdmCl]$ is in excellent agreement with experiments[36,39] (Figure 1B and 1C) as are the predictions for Urea denaturation in low pH[39] (Figure 1B). It is interesting that in low pH, Urea acts as a better denaturing agent than GdmCl. The denaturing ability of the Guanidinium ions depends on the anion[67] as well as pH, which modulates the electrostatic interactions. In neutral pH the large concentrations for both Urea and GdmCl required to denature Ub reflects the extraordinary stability of Ub. The free energy difference between the folded and unfolded states, ΔG_{NU}, as a function of $[C]$ in both low and neutral pH is linear in both GdmCl and Urea (Figure 1D). The slopes of the lines, representing the m values, in low pH for GdmCl and Urea are $\approx 1.8\ kcal\ mol^{-1}\ M^{-1}$ and $1.3\ kcal\ mol^{-1}\ M^{-1}$, respectively. The m-values from experiments[39] in low pH for GdmCl and Urea are 1.8 and 1.0 $kcal\ mol^{-1}\ M^{-1}$ respectively. The computed m-values for GdmCl and Urea in neutral pH are $\approx 1.7\ kcal\ mol^{-1}\ M^{-1}$ and $1.3\ kcal\ mol^{-1}\ M^{-1}$, respectively. The m-value for GdmCl denaturation computed from simulations compares well with the experimental[36] value of $1.9\ kcal\ mol^{-1}\ M^{-1}$.

Radius of gyration as a function of $[C]$**:** The equilibrium radius of gyration, $\langle R_g \rangle$, as a function of $[GdmCl]$ from simulations is in excellent agreement with SAXS measurements[57] at neutral pH (Figure 2A). Our simulations also reproduce the midpoint value of the denaturant concentration ($[GdmCl] \approx 3.8$ M) where the protein folding-unfolding transition occurs in neutral pH (Figure 2A). Experimental data for R_g as a function of $[Urea]$ or $[GdmCl]$ in low pH is not available for comparison with the simulation data (Figure 2B). The $\langle R_g \rangle$ of the unfolded ensemble depends on pH. Experiments[57,68,69] estimate that R_g of the unfolded

state at pH 2.5 is ≈ 32 Å whereas it is more compact at pH 7.0 with $R_g \approx 26$ Å. In neutral pH, electrostatic interactions play a dominant role and contribute to the compaction of the protein.[5] In low pH, the acidic residues are protonated minimizing the role of electrostatic interactions and the protein samples conformations with higher R_g values.[5]

Unfolded states are compact under native conditions: Despite considerable evidence to the contrary there is doubt, based largely on SAXS experiments on protein L and Ub, that the radius of gyration of the UBA, R_g^{UBA} remains a constant even at low $[C]$. The excellent agreement between simulations and experiments in Figure 2A allows us to shed light on the puzzling conclusions reached elsewhere.[57] To this end we calculated R_g^{UBA} using only the ensemble of unfolded conformations. The results in Figure 3 unambiguously show that R_g^{UBA}, decreases in both low and neutral pH as $[GdmCl]$ decreases from 6 M to 0 M indicating that the protein on an average decreases in size. At pH = 7.0, R_g^{UBA} decreases from ≈ 26 Å to ≈ 22 Å, where as in low pH, R_g decreases from ≈ 30 Å to ≈ 23 Å (Figure 2). In low pH, on dilution of $[Urea]$ from 6 M to 0 M, the change in R_g is only ≈ 3 Å, which is less than the ≈ 7 Å decrease upon $[GdmCl]$ dilution. A lower temperature ($T = 333\ K$) is used $[GdmCl]$ simulations ($T = 333\ K$) compared to $[Urea]$ ($T = 343.25\ K$) accounting (in part) for the larger compaction as $[GdmCl]$ is decreased.

To explore temperature effects.[70] we calculated R_g^{UBA} as a function of $[GdmCl]$ at different values of T. The results in Figure 3 show that there is considerable compaction with $\frac{R_g^{UBA}([C]=6M)-R_g^{UBA}([C]=1M)}{R_g^{UBA}([C]=6M)}$ reaching nearly 30% below $T = 333\ K$ (Figure 3A). However, under experimental conditions the change is small, which would be difficult to determine accurately in typical scattering experiments. Our predictions, amenable to experimental tests, show that collapsibility can only be demonstrated by varying external conditions (for example temperature, denaturants, and force).

Ub compaction in the the burst phase: Just as in experiments we triggered folding by decreasing $[C]$ to a value below $[C_m]$ from a high value at which Ub is unfolded. From these dilution simulations we calculated the changes in $R_g(t_B|[C])$ during the burst-phase,

t_B, (t_B = 100 μs after initiating folding) to ascertain the extent of collapse. Each data point in Figure 4A is computed from at least 40,000 conformations from different folding trajectories. The initial unfolded protein conformations are taken from the simulations performed at $[GdmCl] = 6$ M.

The data shows that during the early phase of folding, R_g decreases by ≈2-3 Å in neutral pH and T = 335 K as $[GdmCl]$ is diluted from 6 M to 1 M (Figure 4A). The standard deviation of R_g is ≈5 Å, exceeding the observed changes in R_g. This is because on the t_B time scale Ub samples conformations with widely varying R_g (Figure 5A). Similar behavior is observed when $[Urea]$ is used as a denaturant in low pH (see Figure S3 in the SI). The experimental[57] burst phase R_g data of Ub in neutral pH is in good agreement with the simulation data within the error bars. The decrease in R_g at low denaturant conditions is due to the partial or full folding of 2-3 trajectories within t_B after folding is initiated (for example trajectory in Figure 5B). During the burst phase, the structural over lap function, χ (defined in SI) shows that there is a small decrease in χ for [Urea] and [GdmCl] between 1M and 3M (Figure 4B and S3B). The standard deviation of the data between 1M and 3M is larger as only 2-3 trajectories partially or fully fold to form native contacts.

Distance distribution functions: The pair distance distribution functions, $P(r)$s, the inverse Fourier transform of the scattering intensity, computed using conformations sampled at t_B as a function of $[GdmCl]$ are in good agreement with the SAXS experiments[57] (Figure 4C). We find that $P(r)$ for the folded Ub spans between 0 Å $< r <$ 40 Å with a peak at ≈ 16.5 Å. In contrast, $P(r)$ in the burst phase at both low and neutral pH spans 0 Å $< r <$ 100 Å with a peak at ≈ 25.5 Å showing that the protein samples expanded conformations with a large variation in R_g. In agreement with the $P(r)$, the probability distribution of χ, $P(\chi)$, also shows that native-like contacts are absent on the t_B time scale (Figure 4D). However, detectable compaction of Ub can be achieved by varying solvent conditions. For example, SAXS experiments[33] performed at -20 ^{o}C and 4 ^{o}C in the presence of 45% ethylene Glycol show that R_g of the protein decreased from ≈ 26 Å to ≈ 15 Å during the burst phase of

folding. Our results in Figure 3 support the conclusion reached elsewhere that environmental factors are critical in the compaction of proteins.[33]

Compaction precedes folding: The relationship between native state formation and collapse, displayed as a two dimensional plot in R_g and χ (Figure 6), shows vividly that only upon significant decrease in R_g does the search for the native state begins. This finding is not only consistent with a number of rapid mixing experiments[71–73] but also is in accord with theoretical predictions that search for the folded state occurs by sampling minimum energy compact structures.[74] It should also be noted that even $\chi \approx 0.8$, indicating that Ub is unfolded, R_g has decreased to about 23 Å. This shows a decrease of $\approx$ 8 Å (3 Å) from an average equilibrium value of 32 Å (26 Å) in low (neutral) pH (Figure 6). Thus, kinetic folding trajectories also show that compaction occurs continuously as folding conditions are changed.

Extent of compaction in the burst phase and at equilibrium is similar: The equilibrium R_g^{UBA} decreases continuously as the denaturation concentration decreases although the extent of compaction depends on T (Figure 3). Despite large errors there is detectable change in R_g^{UBA} in the individual kinetic folding trajectories during the initial folding stages. Greater compaction of R_g^{UBA} under folding conditions occurs at times that exceed the burst times, arbitrarily defined here as 100 μs. To illustrate the connection between compaction and formation of a minimum number of contacts we monitored the dynamics of non-local contact formation in denaturant dilution simulations.

Formation of a few such contacts is needed for compaction. On the time scale of about 100 μs these contacts are formed only in some folding trajectories (Figure 5). To illustrate the importance of these contacts in inducing compaction we computed the time-dependent changes in the local contact order,[64,65] S_i (see methods), of the residues ILE3 (β_1), LEU43 (β_3) and LEU67 (β_5). These residues also participate in stabilizing the hydrophobic core of the folded protein. The S_i for the residues ILE3, LEU43 and LEU67 shows that the long range contacts are not formed even after 800 μs in the folding trajectory in Figure 5A and

5C. The long range contacts, that also drive compaction, are formed only in fast folding trajectories shown in Figure 5B and 5D. More generally, although proteins that have a large fraction of non-local contacts are more collapsible (equilibrium R_g^{UBA} decreases more as the denaturant concentration decreases compared to proteins that are largely helical)[75] the time scale for compaction is likely to be longer. Taken together our results firmly establish that R_g^{UBA} is smaller under native conditions than at high denaturant concentrations. Because the changes in R_g^{UBA} are small due to the finite size of proteins, accurate experimental measurements are needed to reach firm conclusions about the collapsibility of Ub and proteins in general.

R_g^{UBA} inferred from smFRET and direct simulations is qualitatively consistent: We computed the equilibrium FRET efficiency, $\langle E \rangle$, as a function of $[C]$ in low pH by assuming that the donor and acceptor dyes are attached near the N and C termini of Ub (Figure 7A). The FRET efficiency is related to the end-to-end distance (R_{ee}) probability distribution $P(R_{ee})$ by,

$$\langle E \rangle = \int_0^L \frac{P(R_{ee})}{1 + (\frac{R_{ee}}{R_0})^6} dR_{ee}, \tag{7}$$

where $L(= 285$ Å) is the contour length of Ub, and $R_0(= 54$ Å) is the Forster radius for the donor-acceptor dyes assuming the donor and acceptor dyes are AlexaFluor 488 and AlexaFluor 594, respectively.[76] These are reasonable dyes to perform FRET experiments on Ub as R_{ee} for Ub varies between 20 Å and 130 Å, and this is approximately in the range $0.5R_o$ to $2R_o$.[77] The FRET efficiency in the UBA, $\langle E^{UBA} \rangle$, increases as the denaturant concentration, $[C]$, is diluted implying that Ub becomes compact (Figure 7A and 7B).

We followed the standard practice used in smFRET experiments[76,78] to calculate $\langle R_{g,FRET}^{UBA} \rangle$ from $\langle E \rangle$. We assume that the unfolded state ensemble can be modeled as a Gaussian polymer with $P(R_{ee})$ given by,

$$P(R_{ee}) = 4\pi R_{ee}^2 \left(\frac{3}{2\pi \langle R_{ee}^2 \rangle} \right)^{3/2} \exp\left(-\frac{3R_{ee}^2}{2\langle R_{ee}^2 \rangle} \right). \tag{8}$$

The $P(R_{ee})$ in Equation 8 when used in Equation 7 yields the average end-to-end distance square, $\langle R_{ee}^2 \rangle$. The FRET estimate for $\langle R_{g,FRET}^{UBA} \rangle$ is calculated using the relation,[79] $\langle R_{g,FRET}^{UBA} \rangle = \sqrt{\langle R_{ee}^2 \rangle / 6}$.

The predicted results in low and high pH as a function of $[GdmCl]$ concentration are shown in Figure 7C and 7D. Two important points are worth making: (1) In both low and neutral pH the values of $\langle R_{g,FRET}^{UBA} \rangle$ estimated from $\langle E^{UBA} \rangle$ decrease as $[GdmCl]$ is reduced. Thus, compaction of the unfolded proteins under folding conditions can be inferred from smFRET experiments, as we recently showed for the PDZ domain.[80] (2) Interestingly, there is no one to one correspondence between direct calculation of $\langle R_g^{UBA} \rangle$ and $\langle R_{g,FRET}^{UBA} \rangle$. At low pH, $\langle R_{g,FRET}^{UBA} \rangle$ is greater than $\langle R_g^{UBA} \rangle$ at high $[GdmCl]$ whereas at a lower $[GdmCl]$ the results are just the opposite. A similar finding was reported for R17 spectrin.[81] However, at neutral pH, $\langle R_g^{UBA} \rangle$ is greater than $\langle R_{g,FRET}^{UBA} \rangle$ at all values of $[GdmCl]$. (3) In both low and neutral pH the extent of compaction inferred from $\langle E \rangle$ is greater than $\langle R_g^{UBA} \rangle$, which explains the differences in the estimates of the dimensions of the unfolded state ensemble from different experimental probes.

Although the Gaussian chain end-to-end distribution gives a reasonable estimate of the R_g extracted from the FRET efficiency data (Figure 7C and 7D), caution should be used to make quantitative comparison with the R_g data estimated from SAXS. Recent FRET experiments[82] on Ub further indicated that, although the protein shows a 2-state transition, the fluorophores lead to significant decrease in the stability of the protein pointing to the possibility that they could interact with the protein, thus effecting its stability. Furthermore for the fluorophores used in the experiment,[82] the FRET efficiency for the protein in folded state should be ≈ 1. However, the measured value is 0.77 indicating that quenching by the residues in the vicinity of the acceptor or the orientation of the fluorophores is restricted when the protein is in the folded state. In addition, there are problems associated with the use of Gaussian chain $P(R_{ee})$ to obtain absolute values of $\langle R_g^{UBA} \rangle$ from smFRET efficiencies.[21,83–87] Nevertheless, the qualitative conclusions obtained from smFRET experiments, which have

the advantage of separating the folded and unfolded states, are valid.

Intermediates are populated during Ub folding: In low and neutral pH and under mildly denaturing conditions (T = 300 K and $[GdmCl]$=1.0 M or $[Urea]$=1.0 M), Ub folding is well described by the diffusion-collision mechanism.[88] In the early stages of folding, kinetic intermediates with folded micro domains in different parts of the protein are populated stabilized by the formation of some of the long range contacts ($\beta_1\beta_5$, $\beta_3\beta_5$ and L_1L_2) leading to structures I1-I4 (Figure 8) These micro domains diffuse for about 10-100 μs and, finally collide and coalesce to form the fully folded structure.

Neutral pH: In highly stabilizing conditions (low T = 300 K and $[Urea]$ = 1.0 M]) Ub folds through 4 different pathways (Figure 8A) suggesting that assembly occurs by the kinetic partitioning mechanism.[89] In zero denaturant conditions at neutral pH, the first step in three of the folding pathways (KIN1-3) is the establishment of contacts between the loops L_1L_2 (Figure S4) resulting in intermediate I1[5] (Figure 8). The formation of L_1L_2 contacts is driven primarily by the charged residues at the interface of the loops L_1L_2 (Figure 1A). Subsequent to the formation of L_1L_2 contacts, either of the long range contacts $\beta_1\beta_5$ or $\beta_3\beta_5$ form giving rise to the intermediates I2 and I3 respectively (Figure 8). In the pathway KIN4, L_1L_2 contacts are established within $\approx$ 10 μs after the formation of $\beta_1\beta_5$ contacts leading to the intermediate I4. In the presence of $[GdmCl]$ = 1.0 M at T = 300 K similar structured intermediates I1-3 are observed in other folding trajectories KIN1-3 (Figure 8B and S5). The intermediates I1-3 are also observed[5] at T = 300 K in zero denaturant conditions. At higher temperatures T = 335 K and $[Urea]$ = 1.0 M, the L_1L_2 contacts by themselves are not very stable[5] (Figure S6). They are formed simultaneously along with either $\beta_1\beta_5$ or $\beta_3\beta_5$ again leading to the intermediates I2 and I3 (Figure 8).

Low pH: In low pH and highly stabilizing conditions (T = 300 K and $[Urea]$=1.0 M or $[GdmCl]$=1.0 M), the loop contacts L_1L_2 form first followed by $\beta_1\beta_5$ contacts, and subsequently $\beta_3\beta_5$ contacts (Figure S7). This folding pathway gives rise to I1 and I3 (Figure 8, S8 and S9) . However, at higher temperatures (T = 335 K and $[Urea]$ or $[GdmCl]$=1.0 M) loop

closure in the protein enabled by the formation of the $\beta_1\beta_5$ contacts are formed first followed by L_1L_2 and $\beta_3\beta_5$. This pathway populates I4 and is the dominant Ub folding pathway at the melting temperature in the absence of denaturants.[5] This is also in agreement with the minor folding pathway inferred from the Ψ-analysis experiments[90,91] performed at pH 7.5.

Multiple experiments pointed out the presence of kinetic intermediates in the Ub folding pathways.[24,27,33,34,92–94] However, it is difficult to obtain a fully resolved three-dimensional structures of the intermediates from the experiments alone. There is evidence[24,27,34,92,93] that $\beta_1\beta_2$ hairpin and the α_1 helix is stable in the intermediates. In accord with this observation, our simulations show that in all the intermediates (Figure 8 there is substantial presence of the $\beta_1\beta_2$ hairpin and the α_1 helix. The variation in the structures is due to the unstable β_3, β_4 and β_5 strands, which require non-local contacts for stability. There is also experimental evidence[94] that the small β_4 strand is unstructured in the intermediates, which is also in accord with our findings (Figure 8, I1-I4). The simulations not only support the inferences made from experiments about the secondary structures in Ub kinetic intermediates but also provide additional information about the tertiary contacts in these intermediates.

Discussion:

Unfolded Ub is compact under native conditions: Our work shows that the unfolded R_g^{UBA} decreases continuously as $[C]$ changes from a high to a low value (Figure 3). However, the extent of compaction depends on temperature. For example, R_g^{UBA} changes only by ≈ 3 Å at $T = 335$ K, a condition chosen to obtain the experimentally measured stability. The change in R_g^{UBA} is only $\approx 10\%$ reduction relative to $[GdmCl] = 7$M. By lowering T to 325 K, a 30% reduction in R_g^{UBA} is predicted when $[GdmCl]$ decreases from 7M to 1M. These findings show that R_g^{UBA} always decreases as the denaturation concentration is lowered but the extent of compaction depends on the temperature.

An argument used in the previous studies claiming the absence of collapse in Ub[57] is that the equilibrium R_g and the burst phase R_g coincide. This is indeed the case in our simulations as well. The magnitude of change in R_g during the burst phase in our simulations is just

as small as under equilibrium but is computable. Given that that there is only a modest decreases in R_g^{UBA} at $T = 335$ K there ought to have been a more precise analysis of the SAXS data with error estimates to rule out the claim that Ub does not collapse.[57] Absent such an analysis we believe that the conclusions reached elsewhere are at best tenuous.[57] Our findings suggest that additional experiments are needed to assess the propensity of Ub to collapse.

Finite size effects and Experimental issues: Using theoretical arguments we showed[95,96] that finite size of proteins plays a major role in restricting the magnitude of changes in the dimensions of unfolded proteins under folding conditions. In particular, theory and simulations have shown that $\Delta = \frac{R_g^{UBA}([C_h]) - R_g^{UBA}([C_l])}{R_g^{UBA}([C_h])}$ is typically small under conditions explored in experiments. For Ub, we find that $\Delta \approx 0.13$ (see Table 1). Recent experiments[81] on R17 spectrum show that $\Delta \approx 0.17$ is also small. Indeed, several proteins for which data or reliable simulation results are available the values of Δ are not large (Table 1). Thus, experiments that can distinguish the structural characteristics of the unfolded states from the folded conformations at $[C] < [C_m]$ are needed to establish the extent of collapse. Clearly this is more easily realized in smFRET experiments. However, the current method of extracting $R_g^{UBA}([C])$ using polymer model for $P(R_{ee})$ could overestimate (underestimate) the size of the unfolded states of proteins at high (low) values of $[C]$, thus exaggerating the extent of compaction. In addition, the attachment of dyes could also compromise the stability of the protein as suggested in a recent study on Ub.[82] Nevertheless, the conclusions inferred from smFRET experiments appear to be qualitatively robust, and are in line with SAXS measurements.[81]

Table 1 shows that there is little correlation between Δ and the length of the proteins, which have been studied experimentally to quantitatively estimate the extent of compaction. Not only is the variation in length small so are the changes in $R_g^{UBA}([C])$ as $[C]$ is decreased from a high to a low value less than C_m. The extent of relative compaction (Δ) varies from about 8% to 33%. The maximum change occurs in cold shock protein with a β-

sheet architecture in the folded state. This observation accords well with recent theoretical predictions[75] establishing that collapsibility in proteins with β-sheet structure is greater than in α-helical proteins.

Fate of Ub in the early stages of folding: The folding trajectory in Figure 5 shows that on the time scale of about 100 μs there is a reduction in $\langle R_g^{UBA} \rangle$ without accumulation of native-like structure as measured by χ. The extent of reduction coincides with the estimates based on results obtained at equilibrium. To set this observation in a broader context, it is useful to consider the relevant time scales when refolding is initiated by diluting the denaturant concentration. Broadly, we need to consider the interplay of three time scales. These are the folding time (τ_F), the collapse time (τ_c), and the time for forming the first tertiary contacts (τ_{tc}) required to decrease $\langle R_g^{UBA} \rangle$. The folding trajectory (Figure 8 and S4) shows that when $\tau_{tc} \approx \tau_c \approx 50$ μs the radius of gyration decreases while the energy per residue is at least four times greater than the value in the folded state. Thus, compaction occurs without any sign of folding. In addition, the $[R_g, \chi]$ plot shows that on the time scale τ_c the value of χ is far greater than that in the folded state. Taken together, these results imply that Ub is compact on the burst time scale without being in the NBA.

Our predictions for Ub are consistent with several previous experiments on other proteins. (1) In a recent study on Monellin, Udgaonkar has shown[97] that on the $\tau_c \approx 37$ μs (the folding time is considerably longer) the polypeptide chain contracted without much structure. These compact structures have been suggested to be minimum energy compact structures (MECS) predicted theoretically.[74] (2) Using site specific hydrogen-deuterium (H/D) in horse Cytochrome-c (Cyt-c) Roder and coworkers[98] showed that on a time scale of $\tau_c \approx 140$ μs ($< \tau_F$, which exceeds several ms) the protein collapses with establishment of native-like local contacts. This work also showed that Cyt-c compaction is unlikely to be due to the presence of the heme group. Rather, it is an intrinsic property of the protein. Interestingly, our simulations on Ub show that specific contacts are necessary for a reduction in $\langle R_g^{UBA} \rangle$. Thus, it is likely that a universal mechanism of compaction of polypeptide chain involves

formation of few native contacts. (3) In order to rationalize the interpretation based on SAXS experiments that collapse is absent in globular proteins it has been asserted that the exceptions hold only for proteins with disulfide bonds in the native state or proteins with prosthetic groups undergo collapse. This argument contradicts the recent discovery, using single molecule pulling experiments,[99] showing that disulfide bond forms only after compaction of the polypeptide chain. The experiments are supported by theoretical predictions of folding of BPTI,[100] which have been further substantiated using simulations.[101] Thus, both experiments and computations established that prior to the formation of the first disulfide bond there is reduction in the dimensions of the protein and not the other way around. The overwhelming evidence suggests that unfolded states of proteins are compact under native conditions with the extent of compaction being modest (Table 1).

Kinetic intermediates in refolding of Ub: It has remained controversial if Ub folds by populating discernible intermediates, whose formation of intermediates does depend on a number of factors. For example, Khorasanizadeh[35] showed that under stabilizing conditions the refolding kinetic data on Ub can only be explained by a three state model. Our simulations suggest that the kinetics could be even more complex. Folding occurs by parallel routes described by the KPM in which a fraction of molecules folds in a two-state manner whereas others reach the native state by populating distinct intermediates. These results suggest that only by varying the stability conditions and using high temporal resolution experiments can the complexity of Ub folding be fully elucidated.

Conclusions:

We performed molecular dynamics simulations using the coarse-grained SOP-SC model and molecular transfer model of Ubiquitin folding in Guanidinium Chloride and Urea solutions to address two questions of fundamental importance in protein folding. One of them is concerned with the size of the unfolded states under native conditions. We showed that under all conditions Ub does become compact as the denaturant concentration is decreased with the extent of compaction being dependent on stability of the folded state. There is

complete consistency between equilibrium and burst phase values of the radius of gyration of the unfolded states. Interestingly, the mechanism of Ub collapse is due to the formation of specific contacts in the unfolded state. In other words, the structure of the unfolded state under folding conditions is more compact than at high denaturant concentration. These and related experimental and computational studies show that the propensity for globular proteins to collapse is universal.

Our simulations, which are in quantitative agreement with experiments,[36,39] also addresses the second question, namely, are there intermediates in Ub folding? The answer to this question is in the affirmative. The kinetic intermediates observed during the folding pathways are identical to the ones observed in the temperature induced folding pathways in the absence of denaturants. The structures of the kinetic intermediates are determined by a combination of the long range contacts between the secondary structural elements L_1L_2, $\beta_1\beta_5$ and $\beta_3\beta_5$. Thus, folding of Ub is more complicated than previously thought but is well described by the kinetic partitioning mechanism.[89] These predictions can be validated using experiments with high temporal resolution. Finally, the combination of coarse-grained simulations and the use of MTM to account for denaturants is a major advance in examining folding of a variety of proteins in quantitative detail.

Acknowledgement

GR acknowledges startup grant from Indian Institute of Science-Bangalore, and funding from Nano mission, Department of Science and Technology, India. DT acknowledges a grant from the National Science Foundation through grant CHE 16-36424 and the Welch Chair at the University of Texas at Austin. Some of the computations are performed by C-DAC's PARAM Yuva-II. A portion of this research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Supporting Information Available

Description of the simulation methods; Table S1; Figures S1-S9. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Mattiroli, F.; Sixma, T. K. Lysine-targeting specificity in ubiquitin and ubiquitin-like modification pathways. *Nat. Struct. Mol. Biol.* **2014**, *21*, 308–316.

(2) Lecker, S.; Goldberg, A.; Mitch, W. Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. *J. Am. Soc. Nephrol.* **2006**, *17*, 1807–1819.

(3) Jackson, S. P.; Durocher, D. Regulation of DNA Damage Responses by Ubiquitin and SUMO. *Mol. Cell* **2013**, *49*, 795–807.

(4) Vjaykumar, S.; Bugg, C.; Cook, W. Structure of ubiquitin refined at 1.8 A resolution. *J. Mol. Biol.* **1987**, *194*, 531–544.

(5) Reddy, G.; Thirumalai, D. Dissecting Ubiquitin folding using the Self-Organized Polymer model. *J. Phys. Chem. B* **2015**, *119*, 11358–11370.

(6) Schuler, B.; Eaton, W. A. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **2008**, *18*, 16–26.

(7) Sosnick, T. R.; Barrick, D. The folding of single domain proteins - have we reached a consensus? *Curr. Opin. Struct. Biol.* **2011**, *21*, 12–24.

(8) Žoldák, G.; Rief, M. Force as a single molecule probe of multidimensional protein energy landscapes. *Curr. Opin. Struct. Biol.* **2013**, *23*, 48–57.

(9) Udgaonkar, J. B. Polypeptide chain collapse and protein folding. *Arch. Biochem. Biophys.* **2013**, *531*, 24–33.

(10) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. Navigating the Folding Routes. *Science* **1995**, *267*, 1619–1620.

(11) Dill, K.; Chan, H. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **1997**, *4*, 10–19.

(12) Thirumalai, D.; Klimov, D. Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr. Opin. Struct. Biol.* **1999**, *9*, 197–207.

(13) Shakhnovich, E. Protein folding thermodynamics and dynamics: Where physics, chemistry, and biology meet. *Chem. Rev.* **2006**, *106*, 1559–1588.

(14) Thirumalai, D.; O'Brien, E. P.; Morrison, G.; Hyeon, C. Theoretical Perspectives on Protein Folding. *Ann. Rev. Biophys.* **2010**, *39*, 159–183.

(15) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, *338*, 1042–1046.

(16) O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13403–13408.

(17) Liu, Z.; Reddy, G.; Thirumalai, D. Theory of the Molecular Transfer Model for Proteins with Applications to the Folding of the src-SH3 Domain. *J. Phys. Chem. B* **2012**, *116*, 6707–6716.

(18) O'Brien, E. P.; Brooks, B. R.; Thirumalai, D. Molecular Origin of Constant m-Values, Denatured State Collapse, and Residue-Dependent Transition Midpoints in Globular Proteins. *Biochemistry* **2009**, *48*, 3743–3754.

(19) 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. USA* **2011**, *108*, 7787–7792.

(20) Reddy, G.; Liu, Z.; Thirumalai, D. Denaturant-dependent folding of GFP. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 17832–17838.

(21) Maity, H.; Reddy, G. Folding of Protein L with Implications for Collapse in the Denatured State Ensemble. *J. Am. Chem. Soc.* **2016**, *138*, 2609–2616.

(22) Candotti, M.; Esteban-Martin, S.; Salvatella, X.; Orozco, M. Toward an atomistic description of the urea-denatured state of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 5933–5938.

(23) Vallee-Belisle, A.; Michnick, S. W. Visualizing transient protein-folding intermediates by tryptophan-scanning mutagenesis. *Nat Struct Mol Biol* **2012**, *19*, 731–736.

(24) Chung, H. S.; Shandiz, A.; Sosnick, T. R.; Tokmakoff, A. Probing the Folding Transition State of Ubiquitin Mutants by Temperature-Jump-Induced Downhill Unfolding. *Biochemistry* **2008**, *47*, 13870–13877.

(25) Rea, A. M.; Simpson, E. R.; Meldrum, J. K.; Williams, H. E. L.; Searle, M. S. Aromatic Residues Engineered into the beta-Turn Nucleation Site of Ubiquitin Lead to a Complex Folding Landscape, Non-Native Side-Chain Interactions, and Kinetic Traps. *Biochemistry* **2008**, *47*, 12910–12922.

(26) Rea, A. M.; Simpson, E. R.; Crespo, M. D.; Searle, M. S. Helix mutations stabilize a late productive intermediate on the folding pathway of ubiquitin. *Biochemistry* **2008**, *47*, 8225–8236.

(27) Schanda, P.; Forge, V.; Brutscher, B. Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11257–11262.

(28) Vallee-Belisle, A.; Michnick, S. W. Multiple tryptophan probes reveal that ubiquitin folds via a late misfolded intermediate. *J. Mol. Biol.* **2007**, *374*, 791–805.

(29) Crespo, M. D.; Simpson, E. R.; Searle, M. S. Population of on-pathway intermediates in the folding of ubiquitin. *J. Mol. Biol.* **2006**, *360*, 1053–1066.

(30) Larios, E.; Li, J.; Schulten, K.; Kihara, H.; Gruebele, M. Multiple probes reveal a

native-like intermediate during low-temperature refolding of ubiquitin. *J. Mol. Biol.* **2004**, *340*, 115–125.

(31) Went, H.; Benitez-Cardoza, C.; Jackson, S. Is an intermediate state populated on the folding pathway of ubiquitin? *FEBS Lett.* **2004**, *567*, 333–338.

(32) Kitahara, R.; Akasaka, K. Close identity of a pressure-stabilized intermediate with a kinetic intermediate in protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3167–3172.

(33) Qin, Z.; Ervin, J.; Larios, E.; Gruebele, M.; Kihara, H. Formation of a compact structured ensemble without fluorescence signature early during ubiquitin folding. *J. Phys. Chem. B* **2002**, *106*, 13040–13046.

(34) Cordier, F.; Grzesiek, S. Temperature-dependence properties as studied by of protein hydrogen bond high-resolution NMR. *J. Mol. Biol.* **2002**, *317*, 739–752.

(35) Khorasanizadeh, S.; Peters, I.; Roder, H. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat Struct Biol* **1996**, *3*, 193–205.

(36) Khorasanizadeh, S.; Peters, I.; Butt, T.; Roder, H. Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* **1993**, *32*, 7054–7063.

(37) Briggs, M.; Roder, H. Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 2017–2021.

(38) Wintrode, P.; Makhatadze, G.; Privalov, P. Thermodynamics of ubiquitin unfolding. *Proteins* **1994**, *18*, 246–253.

(39) Makhatadze, G. Thermodynamics of protein interactions with urea and guanidinium hydrochloride. *J. Phys. Chem. B* **1999**, *103*, 4781–4785 .

(40) Sirovetz, B. J.; Schafer, N. P.; Wolynes, P. G. Water Mediated Interactions and the Protein Folding Phase Diagram in the Temperature-Pressure Plane. *J. Phys. Chem. B* **2015**, *119*, 11416–11427.

(41) Mandal, M.; Mukhopadhyay, C. Microsecond molecular dynamics simulation of guanidinium chloride induced unfolding of ubiquitin. *Phys. Chem. Chem. Phys.* **2014**, *16*, 21706–21716.

(42) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Atomic-Level Description of Ubiquitin Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 5915–5920.

(43) Craig, P. O.; Laetzer, J.; Weinkam, P.; Hoffman, R. M. B.; Ferreiro, D. U.; Komives, E. A.; Wolynes, P. G. Prediction of Native-State Hydrogen Exchange from Perfectly Funneled Energy Landscapes. *J. Am. Chem. Soc.* **2011**, *133*, 17463–17472.

(44) Chung, H. S.; Tokmakoff, A. Temperature-de pendent downhill unfolding of ubiquitin. II. Modeling the free energy surface. *Proteins* **2008**, *72*, 488–497.

(45) Kony, D. B.; Hunenberger, P. H.; van Gunsteren, W. F. Molecular dynamics simulations of the native and partially folded states of ubiquitin: Influence of methanol cosolvent, pH, and temperature on the protein structure and dynamics. *Protein Sci* **2007**, *16*, 1101–1118.

(46) Zhang, J.; Qin, M.; Wang, W. Multiple folding mechanisms of protein ubiquitin. *Proteins* **2005**, *59*, 565–579.

(47) Irback, A.; Mitternacht, S.; Mohanty, S. Dissecting the mechanical unfolding of ubiquitin. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13427–13432.

(48) Dastidar, S.; Mukhopadhyay, C. Unfolding dynamics of the protein ubiquitin: Insight from simulation. *Phys. Rev. E* **2005**, *72*, 051928.

(49) Fernandez, A.; Colubri, A.; Berry, R. Three-body correlations in protein folding: the origin of cooperativity. *Physica A* **2002**, *307*, 235–259.

(50) Marianayagam, N.; Jackson, S. The folding pathway of ubiquitin from all-atom molecular dynamics simulations. *Biophys Chem* **2004**, *111*, 159–171.

(51) Alonso, D.; Daggett, V. Molecular dynamics simulations of protein unfolding and limited refolding - characterization of partially unfolded states of ubiquitin in 60-percent methnol and in water. *J. Mol. Biol.* **1995**, *247*, 501–520.

(52) Alonso, D.; Daggett, V. Molecular dynamics simulations of hydrophobic collapse of ubiquitin. *Protein Sci* **1998**, *7*, 860–874.

(53) Irback, A.; Mitternacht, S. Thermal versus mechanical unfolding of ubiquitin. *Proteins* **2006**, *65*, 759–766.

(54) Sorensen, J.; Head-Gordon, T. Toward minimalist models of larger proteins: A ubiquitin-like protein. *Proteins* **2002**, *46*, 368–379.

(55) Tripathi, S.; Garcia, A. E.; Makhatadze, G. I. Alterations of Nonconserved Residues Affect Protein Stability and Folding Dynamics through Charge-Charge Interactions. *J. Phys. Chem. B* **2015**, *119*, 13103–13112.

(56) 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.

(57) Jacob, J.; Krantz, B.; Dothager, R.; Thiyagarajan, P.; Sosnick, T. Early collapse is not an obligate step in protein folding. *J. Mol. Biol.* **2004**, *338*, 369–382.

(58) Betancourt, M.; Thirumalai, D. Pair potentials for protein folding: Choice of reference states and sensitivity of predicted native states to variations in the interaction schemes. *Prot. Sci.* **1999**, *8*, 361–369.

(59) Auton, M.; Bolen, D. Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale. *Biochemistry* **2004**, *43*, 1329–1342.

(60) Veitshans, T.; Klimov, D.; Thirumalai, D. Protein folding kinetics: Timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold Des* **1997**, *2*, 1–22.

(61) Kumar, S.; J.M., R.; Bouzida, D.; Swendsen, R.; Kollman, P. The weighted histogam analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(62) Ermak, D. L.; Mccammon, J. A. Brownian Dynamics with Hydrodynamic Interactions. *J Chem Phys* **1978**, *69*, 1352–1360.

(63) Guo, Z.; Thirumalai, D. Kinetics and Thermodynamics of Folding of a de novo Designed four Helix Bundle. *J. Mol. Biol.* **1996**, *263*, 323–343.

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

(65) Klimov, D.; Thirumalai, D. Lattice models for proteins reveal multiple folding nuclei for nucleation-collapse mechanism. *J. Mol. Biol.* **1998**, *282*, 471–492.

(66) Mayr, L.; Schmid, F. Stabilization of a Protein by Guanidinium Chloride. *Biochemistry* **1993**, *32*, 7994–7998.

(67) Makhatadze, G.; Lopez, M.; Richardson, J.; Thomas, S. Anion binding to the ubiquitin molecule. *Protein Sci.* **1998**, *7*, 689–697.

(68) Huang, J.-r.; Gabel, F.; Jensen, M. R.; Grzesiek, S.; Blackledge, M. Sequence-Specific Mapping of the Interaction between Urea and Unfolded Ubiquitin from Ensemble

Analysis of NMR and Small Angle Scattering Data. *J. Am. Chem. Soc.* **2012**, *134*, 4429–4436.

(69) Gabel, F.; Jensen, M. R.; Zaccai, G.; Blackledge, M. Quantitative Modelfree Analysis of Urea Binding to Unfolded Ubiquitin Using a Combination of Small Angle X-ray and Neutron Scattering. *J. Am. Chem. Soc.* **2009**, *131*, 8769–8771.

(70) Nettels, D.; Muller-Spath, S.; Kuster, F.; Hofmann, H.; Haenni, D.; Ruegger, S.; Reymond, L.; Hoffman, A.; Kubelka, J.; Heinz, B.; Gast, K.; Best, R.; Schuler, B. Single-molecule spectroscopy of the temperature-induced collapse of unfolded proteins. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 20740–20745.

(71) Matsumoto, S.; Yane, A.; Nakashima, S.; Hashida, M.; Fujita, M.; Goto, Y.; Takahashi, S. A rapid flow mixer with 11-mu s mixing time microfabricated by a pulsed-laser ablation technique: Observation of a barrier-limited collapse in cytochrome c folding. *J. Am. Chem. Soc.* **2007**, *129*, 3840–3841.

(72) Lapidus, L. J.; Yao, S.; McGarrity, K. S.; Hertzog, D. E.; Tubman, E.; Bakajin, O. Protein hydrophobic collapse and early folding steps observed in a microfluidic mixer. *Biophys. J.* **2007**, *93*, 218–224.

(73) Kathuria, S. V.; Kayatekin, C.; Barrea, R.; Kondrashkina, E.; Graceffa, R.; Guo, L.; Nobrega, R. P.; Chakravarthy, S.; Matthews, C. R.; Irving, T. C.; Bilsel, O. Microsecond Barrier-Limited Chain Collapse Observed by Time-Resolved FRET and SAXS. *J. Mol. Biol.* **2014**, *426*, 1980–1994.

(74) Camacho, C. J.; Thirumalai, D. Minimum energy compact structures of random sequences of heteropolymers. *Phys. Rev. Lett.* **1993**, *71*, 2505–2508.

(75) Samanta, H.; Zhuravlev, P.; Hinczewski, M.; Hori, N.; Chakrabarti, S.; Thirumalai, D. Protein collapse in encoded in the folded state architecture. *Phys. Rev. X* **2016**, , .

(76) Sherman, E.; Haran, G. Coil-globule transition in the denatured state of a small protein. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11539–11543.

(77) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; Springer Science & Business Media, 2013.

(78) Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1528–1533.

(79) Rubinstein, M.; Colby, R. H. *Polymer physics*; OUP Oxford, 2003.

(80) Liu, Z.; Reddy, G.; Thirumalai, D. Folding PDZ2 Domain Using the Molecular Transfer Model. *J. Phys. Chem. B* **2016**, *120*, 8090–8101.

(81) Borgia, A.; Zheng, W.; Buholzer, K.; Borgia, M. B.; Schler, A.; Hofmann, H.; Soranno, A.; Nettels, D.; Gast, K.; Grishaev, A.; Best, R. B.; Schuler, B. Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods. *J. Am. Chem. Soc.* **2016**, *138*, 11714–11726.

(82) Saito, M.; Kamonprasertsuk, S.; Suzuki, S.; Nanatani, K.; Oikawa, H.; Kushiro, K.; Takai, M.; ting Chen, P.; Chen, E. H.-L.; Chen, R. P.-Y.; Takahashi, S. Significant Heterogeneity and Slow Dynamics of the Unfolded Ubiquitin Detected by the Line Confocal Method of Single-Molecule Fluorescence Spectroscopy. *J. Phys. Chem. B* **2016**, *120*, 8818–8829.

(83) O'Brien, E. P.; Morrison, G.; Brooks, B. R.; Thirumalai, D. How accurate are polymer models in the analysis of Forster resonance energy transfer experiments on proteins? *J. Chem. Phys.* **2009**, *130*, 124903.

(84) Sinha, K. K.; Udgaonkar, J. B. Dissecting the non-specific and specific components of

the initial folding reaction of barstar by multi-site FRET measurements. *J. Mol. Biol.* **2007**, *370*, 385–405.

(85) Laurence, T.; Kong, X.; Jager, M.; Weiss, S. Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 17348–17353.

(86) Goldenberg, D. P. Computational simulation of the statistical properties of unfolded proteins. *J. Mol. Biol.* **2003**, *326*, 1615–1633.

(87) Zhou, H.-X. Dimensions of denatured protein chains from hydrodynamic data. *J. Phys. Chem. B* **2002**, *106*, 5769–5775.

(88) Karplus, M.; Weaver, D. Diffusion-collision model for protein folding. *Biopolymers* **1979**, *18*, 1421–1437.

(89) Guo, Z. Y.; Thirumalai, D. Kinetics of Protein-Folding - Nucleation Mechanism, Time Scales, and Pathways. *Biopolymers* **1995**, *36*, 83–102.

(90) Krantz, B.; Dothager, R.; Sosnick, T. Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J. Mol. Biol.* **2004**, *337*, 463–475.

(91) Krantz, B.; Dothager, R.; Sosnick, T. Discerning the structure and energy of multiple transition states in protein folding using psi-analysis (vol 337, pg 463, 2004). *J. Mol. Biol.* **2005**, *347*, 1103.

(92) Chung, H.; Khalil, M.; Smith, A.; Ganim, Z.; Tokmakoff, A. Conformational changes during the nanosecond-to-millisecond unfolding of ubiquitin. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 612–617.

(93) Chung, H. S.; Ganim, Z.; Jones, K. C.; Tokmakoff, A. Transient 2D IR spectroscopy of ubiquitin unfolding dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14237–14242.

(94) Zheng, Z.; Sosnick, T. R. Protein Vivisection Reveals Elusive Intermediates in Folding. *J. Mol. Biol.* **2010**, *397*, 777–788.

(95) Camacho, C. J.; Thirumalai, D. Kinetics and Thermodynamics of Folding in Model Proteins. *Proc. Natl Acad Sci USA* **1993**, *90*, 6369–6372.

(96) Li, M.; Klimov, D.; Thirumalai, D. Finite size effects on thermal denaturation of globular proteins. *Phys. Rev. Lett.* **2004**, *93*, 268107.

(97) Goluguri, R. R.; Udgaonkar, J. B. Microsecond Rearrangements of Hydrophobic Clusters in an Initially Collapsed Globule Prime Structure Formation during the Folding of a Small Protein. *J. Mol. Biol.* **2016**, *428*, 3102–3117.

(98) Fazelinia, H.; Xu, M.; Cheng, H.; Roder, H. Ultrafast Hydrogen Exchange Reveals Specific Structural Events during the Initial Stages of Folding of Cytochrome c. *J. Am. Chem. Soc.* **2014**, *136*, 733–740.

(99) 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.

(100) Camacho, C.; Thirumalai, D. Theoretical predictions of folding pathways by using the proximity rule, with applications to Bovine Pancreatic Trypsin-Inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 1277–1281.

(101) Qin, M.; Wang, W.; Thirumalai, D. Protein folding guides disulfide bond formation. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 11241–11246.

Table 1: Relative changes in the radius of gyration of various proteins in the UBA ensemble (R_g^{UBA}) between high and low denaturant concentrations

Protein	$R_g^{UBA}([C_h])$[d] ([GdmCl] = 7 M)	$R_g^{UBA}([C_l])$[e] ([GdmCl] = 1M)	$\Delta = \frac{R_g^{UBA}([C_h]) - R_g^{UBA}([C_l])}{R_g^{UBA}([C_h])}$
Protein L[21] (N_{res}=64)	26.3 Å	22.5 Å ([C]=0.25 M)	14.4%
Monellin (N_{res}=96; Neutral pH)[a]	27.8 Å	25.5 Å ([C]=0.25 M)	8.3%
PDZ2 Domain[80] (N_{res}=94)[b]	32.2Å	29.8Å ([C]=0.25 M)	7.5%
Ubiquitin (N_{res}=76; Neutral pH)	25.9 Å	22.5 Å ([C]=0.25 M)	13.1%
Ubiquitin (N_{res}=76; Low pH)	30.4 Å	23.3 Å ([C]=0.25 M)	23.4%
SH3[19] (N_{res}=56)	23.7 Å	20.3 Å ([C]=0.2 M)	14.3%
Cold Shock[16] (N_{res}=70)	26.4 Å	17.8 Å ([C]=1 M)	32.6%
ACTR[81] (N_{res}=73)[c]	29.7 Å	24.7 Å ([C]= 0.3 M)	16.8%
R17d[81] (N_{res}=116)[c]	40.3 Å	33.2 Å ([C]= 0.6 M)	17.6%

[a] Unpublished data

[b] Denaturant used is Urea

[c] Data from experiments

[d] $R_g^{UBA}([C_h])$ denotes R_g of the protein in the UBA ensemble in high denaturant concentrations

[e] $R_g^{UBA}([C_l])$ denotes R_g of the protein in the UBA ensemble in low denaturant concentrations

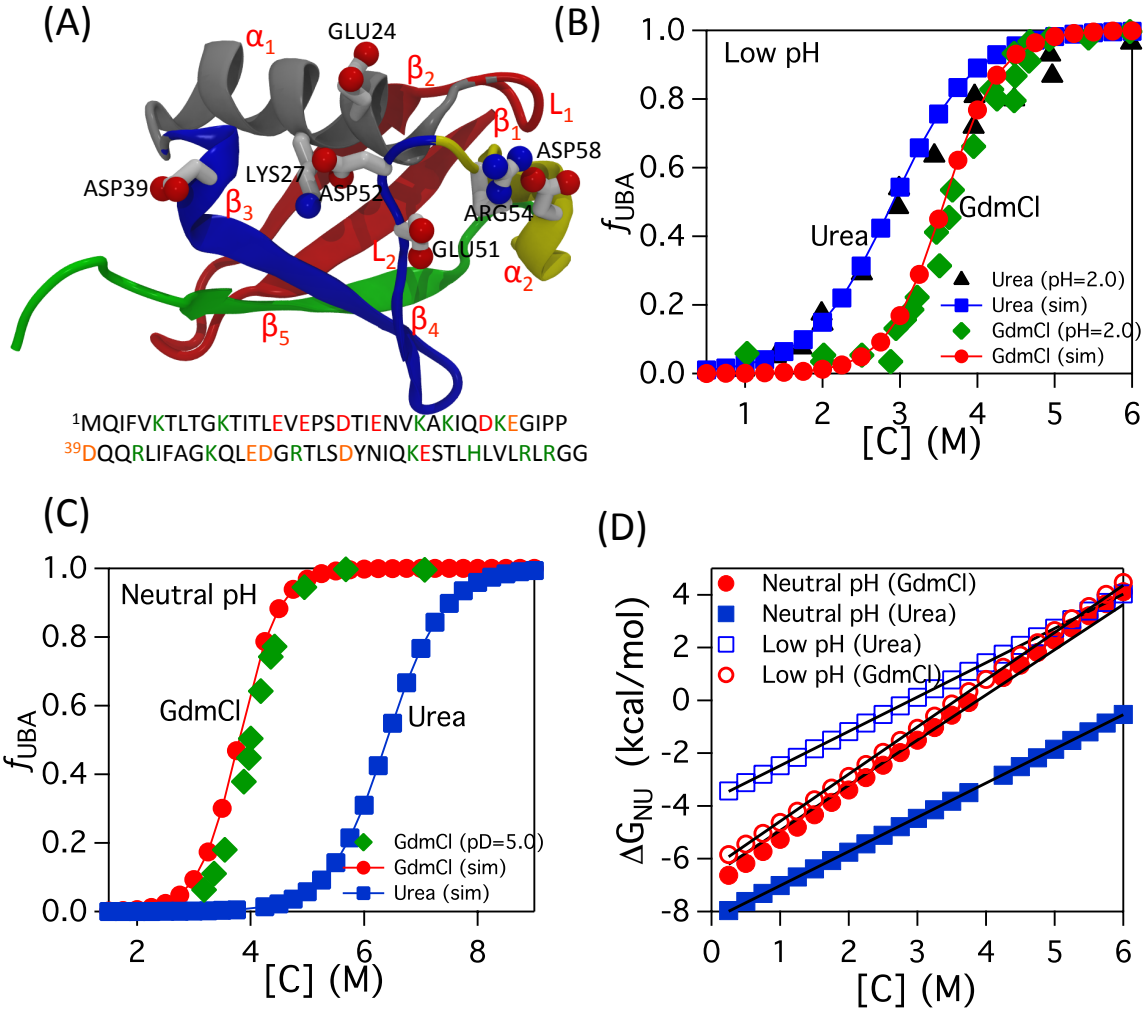

Figure 1: Denaturant-dependent folding of Ub. (A) Crystal structure of Ub (PDB ID: 1UBQ). The folded structure consists of 5 β-sheets labeled β_1 (red), β_2 (red), β_3 (blue), β_4 (blue) and β_5 (green). Two helices are shown in α_1 (silver) and α_2 (yellow). Charged residues at the interface of the helix α_1 and loops L_1 and L_2 are shown. At the bottom the single letter codes for the amino acid residues present in Ub is shown. The letters in green and red are positively and negatively charged amino acids, respectively. (B) The fraction of the protein in the UBA, f_{UBA} as a function of the denaturant concentration $[C]$ in low pH. The data in solid green diamonds and solid black triangles is f_{UBA} as function of $[GdmCl]$ and $[Urea]$ respectively from experiments[39] performed at pH = 2.0. The data in solid red circles and blue squares is $[GdmCl]$ and $[Urea]$ data respectively from simulations using the low pH Ub model. (C) The data in solid green diamonds is f_{UBA} as a function of $[GdmCl]$ from experiments[36] performed at pD = 5.0. Data in solid red circles and blue circles is $[GdmCl]$ and $[Urea]$ data respectively from simulations using the neutral pH Ub model. (D) Free energy difference between the folded and unfolded states, ΔG_{NU} as a function of $[C]$.

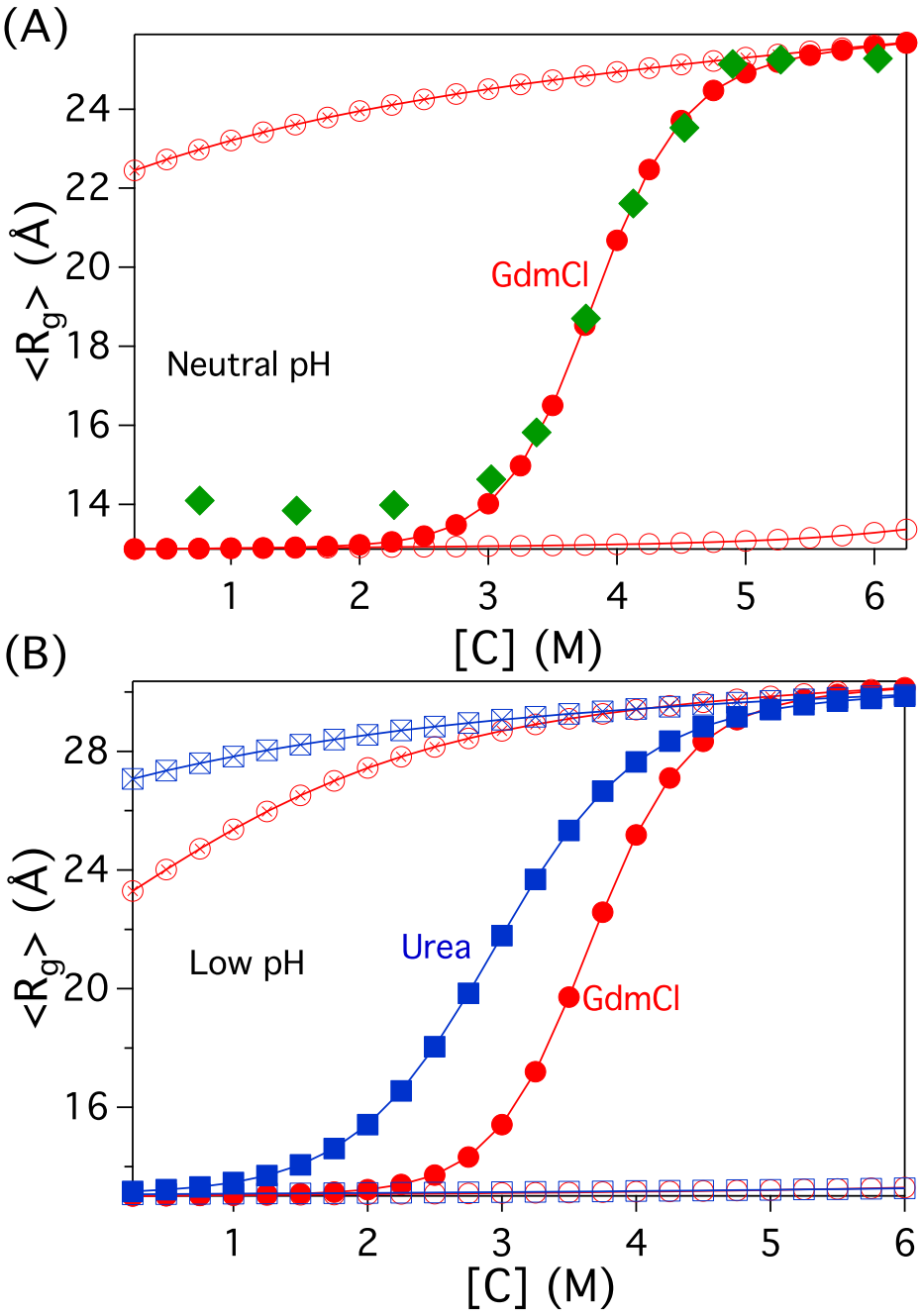

Figure 2: The radius of gyration, R_g, as a function of denaturant concentration, $[C]$. (A) Neutral pH. Data in red solid circles and green diamonds are from simulations and experiments,[57] respectively for $[GdmCl]$. $\langle R_g \rangle$ of UBA and NBA basins computed from simulations are shown in red circles with cross and red empty circles, respectively. (B) Low pH. Symbols in red circles and blue squares are for $[GdmCl]$ and $[Urea]$, respectively. Empty symbols and symbols with cross represent NBA and UBA basins, respectively.

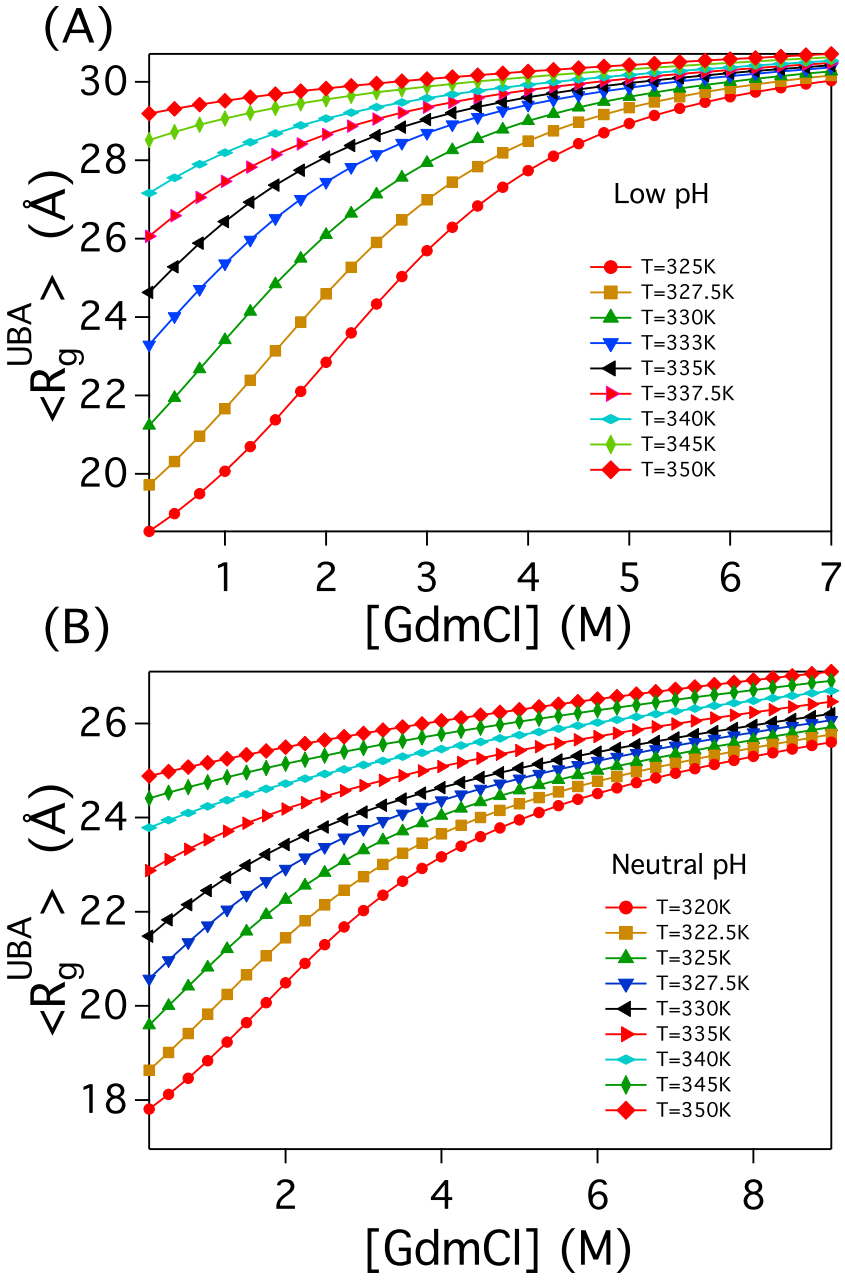

Figure 3: $\langle R_g^{UBA} \rangle$ as function of $[GdmCl]$ in (A) low pH and (B) neutral pH at different temperatures. The average size of the protein in the UBA basin decreases with the temperature in low $[GdmCl]$.

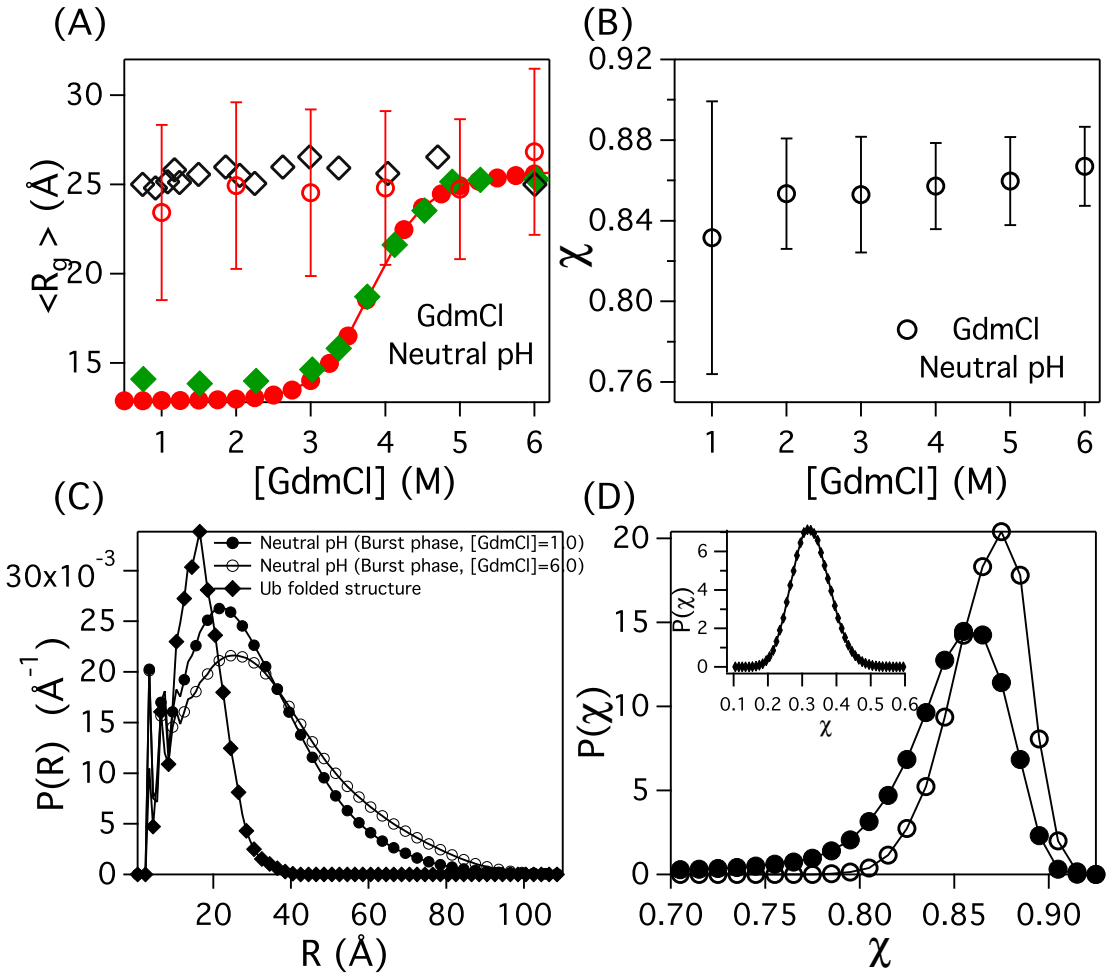

Figure 4: (A) Radius of gyration, R_g, as a function of denaturant concentration. Equilibrium $\langle R_g \rangle$ from coarse-grained simulations in neutral pH as a function of $[GdmCl]$ is in red solid circles. Data in green diamonds and black empty diamonds is equilibrium and burst-phase $\langle R_g \rangle$ as function of $[GdmCl]$ from experiments[57] at pH = 7.0. Data in empty red circles (neutral pH, T = 335 K, $[GdmCl]$) is $\langle R_g \rangle$ during the burst phase of Ub folding. (B) Structural overlap function, χ, plotted as a function of $[C]$. The empty circle symbols in the plot represent the same conditions described in panel-(A). (C) Pair distance distribution function, $P(R)$, plotted as a function of distance, R, for the Ub native structure, and during the burst phase of Ub folding. (D) Probability distribution of χ, $P(\chi)$. The symbols represent the same conditions described in panel-(C). Inset shows $P(\chi)$ for the folded structure in neutral pH at $T = 335$ K and $[C] = 0$ M conditions.

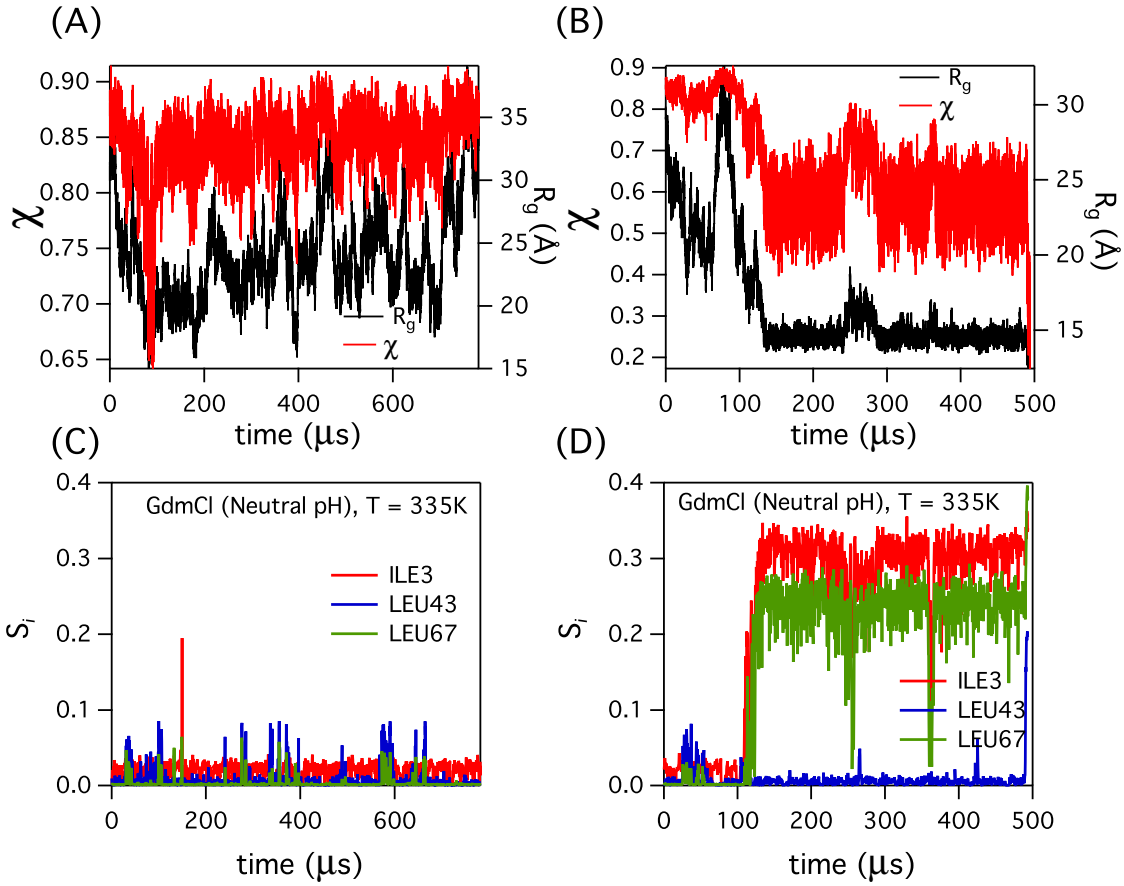

Figure 5: χ and R_g of a Ub folding trajectory in neutral pH is plotted as a function of time for conditions $[GdmCl]$=1.0 M and $T = 335\ K$. (A) Trajectory where Ub does not fold in 800 μs and (B) trajectory where Ub folds in 500 μs. The contact order, S_i, for the residues ILE3, LEU43 and LEU67 plotted as a function of time for the (C) unfolded trajectory shown in panel-(A), and (D) folded trajectory in panel-(B).

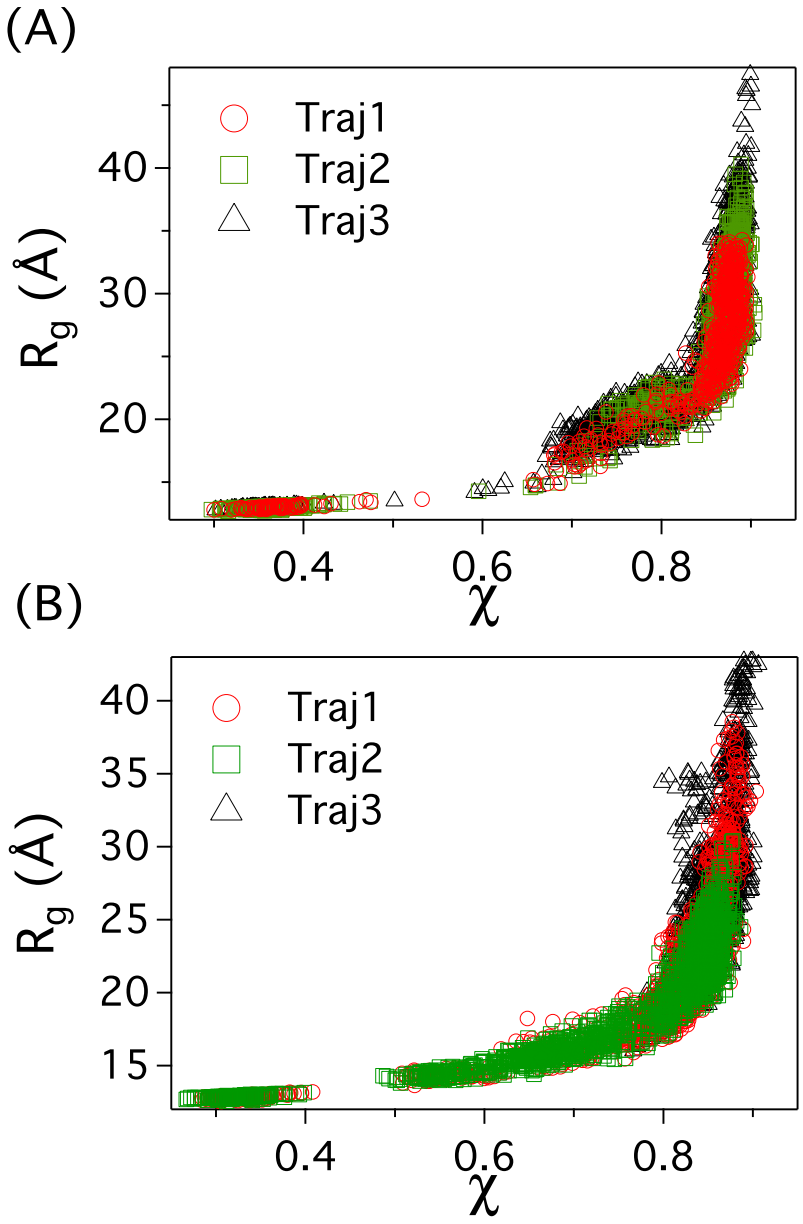

Figure 6: R_g as a function of χ. Three different folding trajectories are shown in red circles, green squares, and black triangles, respectively. Each data point is obtained by averaging the folding trajectory data for $\approx 0.25\mu s$. (A) Low pH, T=332K, [C]=0M; (B) Neutral pH, T=335K, [C]=0M.

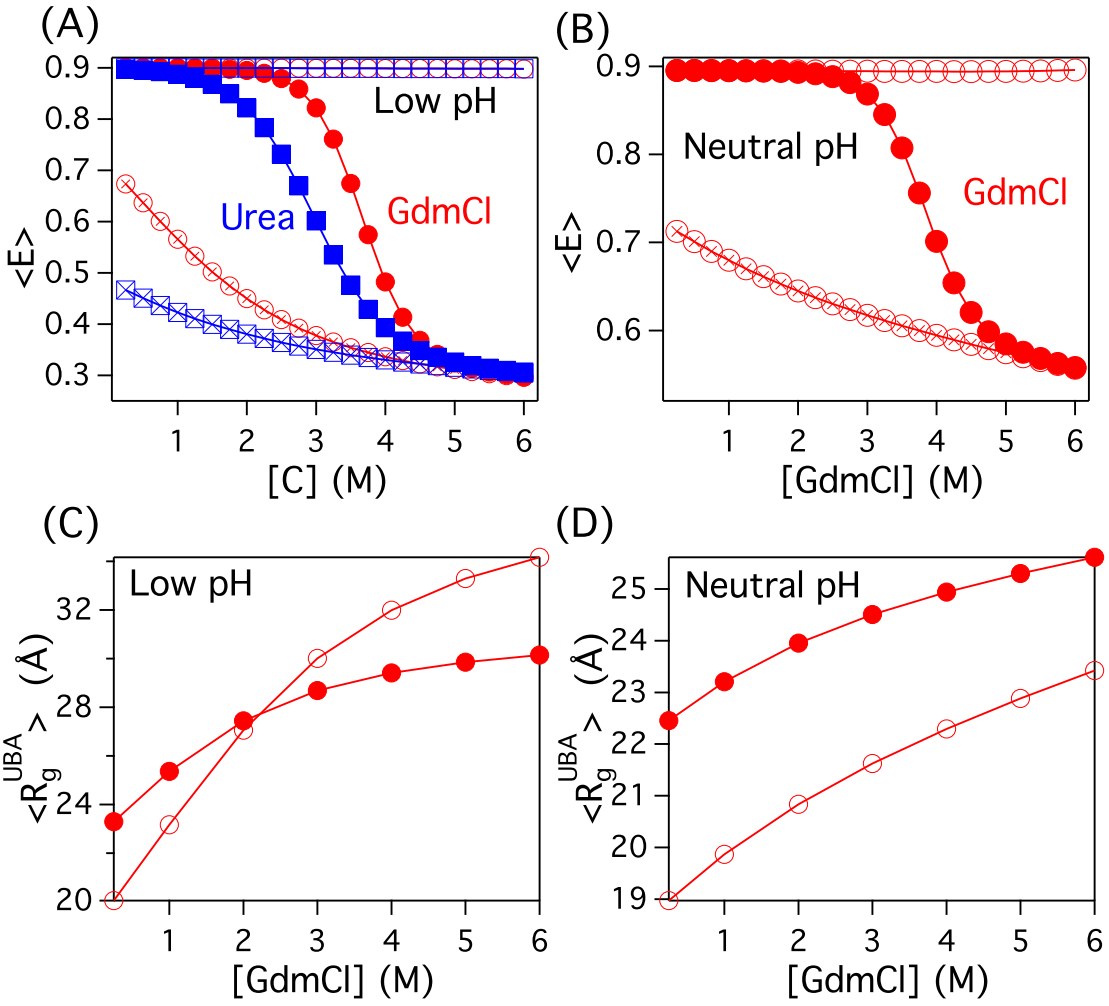

Figure 7: (A) Equilibrium FRET efficiency, $\langle E \rangle$, as a function of $[GdmCl]$ and $[Urea]$ in low pH is in red circles and blue squares, respectively. $\langle E \rangle$ for the protein conformations in the UBA and NBA basins as a function of $[GdmCl]$ ($[Urea]$) is shown in cross circles (cross squares) and empty circles (empty squares), respectively. (B) $\langle E \rangle$, as a function of $[GdmCl]$ in neutral pH. The symbols represent the same as in A. (C) Data in solid red circles is $\left\langle R_g^{UBA} \right\rangle$ as a function of $[GdmCl]$ at $T = 333\ K$ in low pH. The data is empty red circles is $\left\langle R_{g,FRET}^{UBA} \right\rangle$ estimated from the FRET efficiency data in the UBA ensemble, $\langle E^{UBA} \rangle$. (D) $\left\langle R_{g,FRET}^{UBA} \right\rangle$ as a function of $[GdmCl]$ in neutral pH at $T = 333.25\ K$. The symbols represent the same conditions as in panel-C.

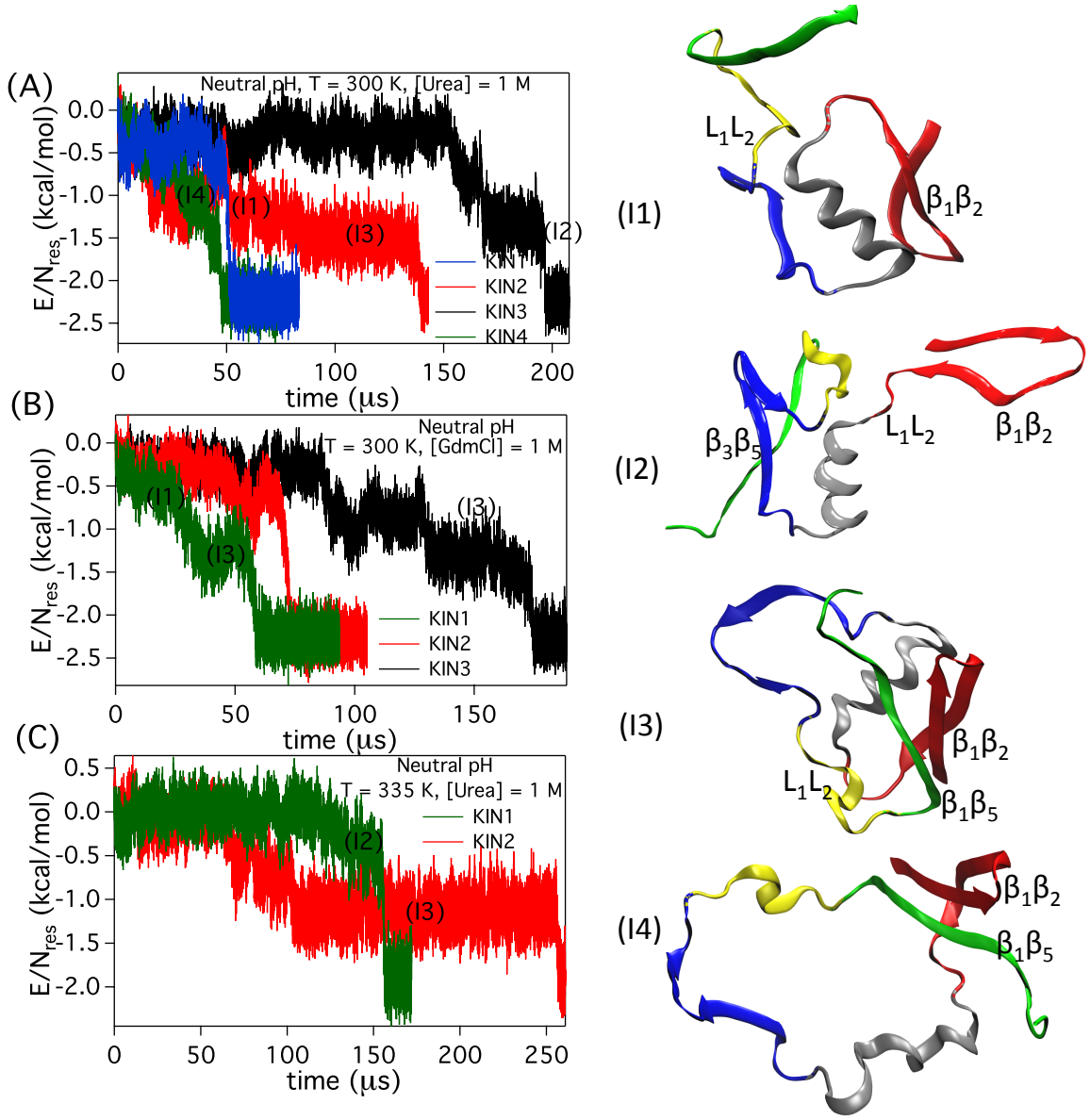

Figure 8: Ub folding kinetics in neutral pH. The folding pathways are inferred from change in energy per residue, E/N_{res}, as a function of time at conditions (A) $T = 300\ K$, $[Urea]$=1.0 M (B) $T = 300\ K$, $[GdmCl]$=1.0 M, and (C) $T = 335\ K$, $[Urea]$=1.0 M. Four kinetic intermediates labeled I1, I2, I3 and I4 are identified in the folding pathways. Representative structures of the kinetic intermediates are on the right.

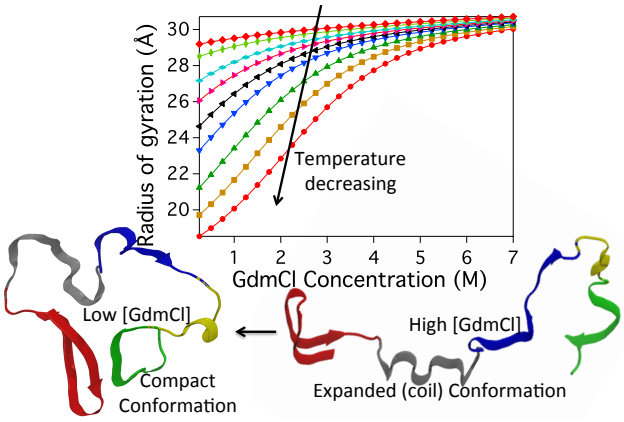

Table of Contents (TOC) figure