THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Dissecting Ubiquitin Folding Using the Self-Organized Polymer Model

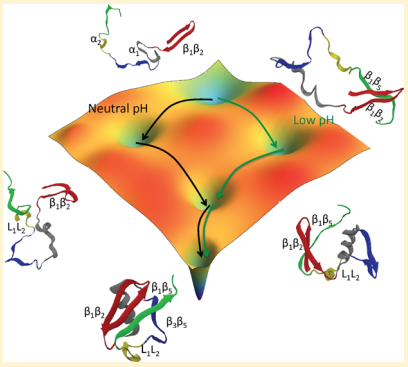
Govardhan Reddy*,† and D. Thirumalai‡

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

‡Biophysics Program, Institute for Physical Science and Technology, and Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742, United States

Supporting Information

ABSTRACT: Folding of Ubiquitin (Ub), a functionally important protein found in eukaryotic organisms, is investigated at low and neutral pH at different temperatures using simulations of the coarse-grained self-organized-polymer model with side chains (SOP-SC). The melting temperatures (T_m's), identified with the peaks in the heat capacity curves, decrease as pH decreases, in qualitative agreement with experiments. The calculated radius of gyration, showing dramatic variations with pH, is in excellent agreement with scattering experiments. At T_m, Ub folds in a two-state manner at low and neutral pH. Clustering analysis of the conformations sampled in equilibrium folding trajectories at T_m, with multiple transitions between the folded and unfolded states, shows a network of metastable states connecting the native and unfolded states. At low and neutral pH, Ub folds with high probability through a preferred set of conformations resulting in a pH-dependent dominant folding pathway. Folding kinetics reveal that Ub assembly at low pH occurs by multiple pathways involving a combination of nucleation-collapse and diffusion collision mechanism. The mechanism by which Ub folds is dictated by the stability of the key secondary structural elements responsible for establishing long-range contacts and collapse of Ub. Nucleation collapse mechanism holds if the stability of these elements are marginal, as would be the case at elevated temperatures. If the lifetimes associated with these structured microdomains are on the order of hundreds of microseconds, then Ub folding follows the diffusion−collision mechanism with intermediates, many of which coincide with those found in equilibrium. Folding at neutral pH is a sequential process with a populated intermediate resembling that sampled at equilibrium. The transition state structures, obtained using a P_{fold} analysis, are homogeneous and globular with most of the secondary and tertiary structures being native-like. Many of our findings for both the thermodynamics and kinetics of folding are not only in agreement with experiments but also provide missing details not resolvable in standard experiments. The key prediction that folding mechanism varies dramatically with pH is amenable to experimental tests.



INTRODUCTION

Major advances in experiments[1,2] and theory,[3−9] and creation of coarse-grained models rooted in theory[10−14] have produced a comprehensive framework for quantitatively describing the way single domain proteins fold. More recently, technical advances have made it possible to generate long (nearly milliseconds in some instances) folding trajectories using atomically detailed simulations in water for several small proteins.[15,16] These developments have ushered a new era in protein folding in which it is imperative to develop theoretical and computational models so that detailed comparisons with experiments can be made.[17] Many researchers assume that this task requires all atom simulations using empirical force fields. An alternative is to develop coarse-grained (CG) models, which have proven to have exceptional predictive power not only in the study of self-assembly of proteins but also in understanding larger complexes and biological machines.[10,11,13] We use this approach, which we contend is powerful, not to mention computationally tractable, to simulate the folding of Ubiquitin (Ub) at low and neutral pH.

The importance of the 76-residue Ub, a regulatory protein present in eukaryotic organisms, can hardly be overstated. Ub is involved in a large number of functions ranging from transcriptional regulation to protein degradation and executes these functions by attaching (ubiquitinating) to a number of substrate proteins with great structural diversity. Depending on the function, monoubiquitation[18] and polyubiquitination,[19] both of which are post-translational modifications, have been characterized. In addition to the intrinsic interest in the folding of this small protein, recent studies have established a link between the stability, dynamics, and function of Ub.[20] The monomeric native fold of Ub has five β-strands, a long α-helix, a 3_{10} helix woven together by a complex topology (Figure 1A). The C_α contact-map (Supporting Information Figure S1) illustrates

Special Issue: Biman Bagchi Festschrift

Received: April 10, 2015
Revised: July 1, 2015
Published: July 1, 2015

DOI: 10.1021/acs.jpcb.5b03471
J. Phys. Chem. B 2015, 119, 11358−11370

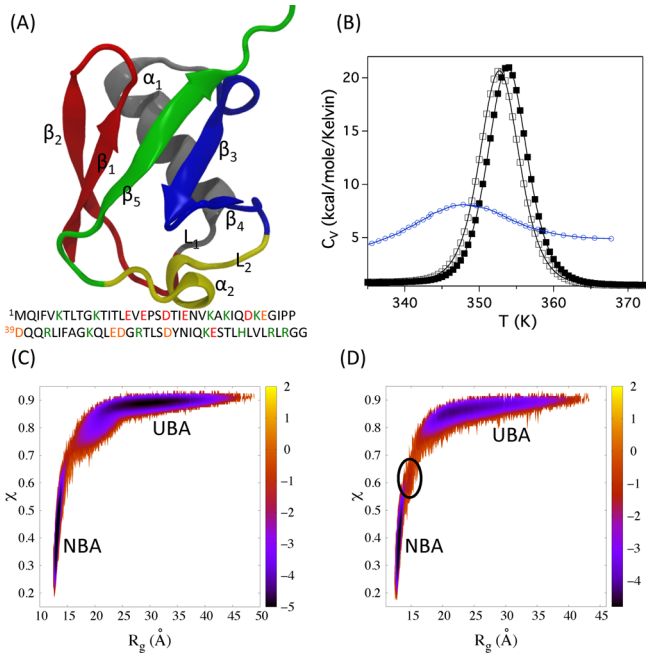


Figure 1. Thermodynamics of Ub folding. (a) Ribbon diagram[58] of the crystal structure of Ub (PDB ID: 1UBQ). The strands are labeled β_1, β_2, β_3, β_4, and β_5, helix is α_1, 3_{10} helix is α_2, and the interacting loops are L_1 and L_2. The sequence of Ub is given below. Positively charged residues are in green, and negatively charged residues are in red. (B) Heat Capacity C_v as a function of temperature T at low pH (empty squares) and neutral pH (solid squares). For comparison we show the experimental data[44] in blue for heat capacity at pH = 3.0. Although the calculated T_m deviates only by a few degrees from experiments the simulations do not capture the measured peak height and the width of the heat capacity curve. (C) The free energy surface of Ub at low pH and T_m = 353 K projected onto the radius of gyration, R_g, and structural overlap parameter, χ, shows two-state behavior. (D) Same as panel C except the free energy profiles correspond to folding at neutral pH and T_m = 354 K. The circle highlights a high-energy intermediate.

prominent interactions between the residues of $\beta_1\beta_2$ hairpin, and long-range contacts involving strands β_1 and β_5, and β_3 and β_5, and the residues in the loops L_1 and L_2. Contacts between residues that are distant along the sequence (especially strands β_1 and β_5, β_3 and β_5, and loops L_1 and L_2) makes for high contact order, which could be the reason for the complex folding kinetics for a moderate sized protein. The overall folding itself can be accurately estimated using $\tau_F \approx \tau_0 \exp(\sqrt{(N)})$,[21] where $\tau_0 \approx 1\ \mu$s. For the 76-residue Ub, $\tau_F \approx 6$ ms, which agrees remarkably well with experiments.[22,23] However, regardless of how accurate theoretical estimates of τ_F may be, this by itself does not provide insight into the molecular underpinnings of the folding process, which require simulations.

Here, we explore Ub folding using the self-organized polymer model[24] with emphasis on native interactions, which has been used to study not only protein folding[25] but also several other complex processes ranging from bacterial transcription initiation,[26] response of microtubule to force,[27,28] and indentation of virus particles.[29] The model emphasizes native interactions based on the structure of the folded state. The role in non-native interactions, which has been discussed extensively (see below for additional discussions), has been shown to be much less dominant in determining the folding of well designed proteins.[14,30,31] Various aspects of Ub folding have been explored using both atomistic and CG simulations.[32−42] Previous CG simulations, based on C_α representation without consideration of electrostatic effects have elucidated the slow dynamics in monomeric Ub at low temperatures[40,43] and revealed a change in the folding mechanism as the temperature is lowered.[40] Because Ub folding thermodynamics depends dramatically on pH,[44] it is crucial to consider electrostatic interactions. Using the SOP model with side chains (SOP-SC) including charge effects, we provide a quantitative description of the thermodynamics and kinetics of folding as a function of pH, which we mimic by modifying the electrostatic interactions. The simulations capture the thermodynamics of Ub folding qualitatively and predict, for the first time, the dimensions in the unfolded state accurately. Interestingly, we predict that although there is a network of connected states linking the folded and unfolded states implying multiple folding pathways, there is a dominant folding path along which Ub self-assembles underscoring the need for probabilistic description of the folding process. The dominant path is found to change dramatically with pH. Our results for folding thermodynamics and kinetics are in semiquantitative agreement with a number of experiments, thus establishing that CG models can capture the physics of protein folding.

METHODS

Self Organized Polymer-Side Chain (SOP-SC) Model. We model the polypeptide chain using the coarse-grained self organized polymer side chain model (SOP-SC)[24] in which each residue is represented using two interaction centers, one for the backbone atoms and the other for the side chains (SCs). Out of the 76 residues in Ub, 23 are charged (Figure 1A and Figure S1), which we include in the SOP-SC model (see below). The centers of the interaction centers, representing the backbone atoms and the side chain atoms, are at the C_α atom position of the residue, and the center of mass of the side chain, respectively. The SCs interact via a residue-dependent Betancourt-Thirumalai statistical potential.[45] At low pH, the acidic residues are protonated, which minimizes the effect of electrostatic interactions on the folding of Ub. To simulate folding at 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[46] (Protein Data Bank (PDB) ID: 1UBQ).

The energy of a conformation in the SOP-SC models is a sum of bonded (B) and nonbonded (NB) interactions. The interaction between a pair of covalently connected beads (two successive C_α atoms or SC connected to a C_α atom) is represented by Finite Extensible Nonlinear Elastic (FENE) potential. The nonbonded interactions are a sum of native (N) and non-native (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.

The force-field in the SOP-SC model is

$$E_{TOT} = E_B + E_{NB}^{N} + E_{NB}^{NN} + \lambda E^{el} \quad (1)$$

The FENE potential, E_B, between covalently linked beads is given by

$$E_B = -\sum_{i=1}^{N_B} \frac{k}{2} R_o^2 \log\left(1 - \frac{(r_i - r_{cry,i})^2}{R_o^2}\right) \quad (2)$$

DOI: 10.1021/acs.jpcb.5b03471
J. Phys. Chem. B 2015, 119, 11358−11370

where N_B is the total number of bonds between the beads in the coarse grained model of the polypeptide chain. For Ub, $N_B = 151$.

The nonbonded native interactions, E_{NB}^{N}, in eq 1 is

$$E_{NB}^{N} = \sum_{i=1}^{N_N^{bb}} \epsilon_h^{bb}\left[\left(\frac{r_{cry,i}}{r_i}\right)^{12} - 2\left(\frac{r_{cry,i}}{r_i}\right)^{6}\right] + \sum_{i=1}^{N_N^{bs}} \epsilon_h^{bs}\left[\left(\frac{r_{cry,i}}{r_i}\right)^{12} - 2\left(\frac{r_{cry,i}}{r_i}\right)^{6}\right] + \sum_{i=1}^{N_N^{ss}} 0.5(\epsilon_i^{ss} - 0.7)(300.0k_B)\left[\left(\frac{r_{cry,i}}{r_i}\right)^{12} - 2\left(\frac{r_{cry,i}}{r_i}\right)^{6}\right] \quad (3)$$

where $N_N^{bb}(= 177)$, $N_N^{bs}(= 486)$, and $N_N^{ss}(= 204)$ are the numbers of backbone−backbone, backbone-side chain, side chain−side chain native interactions, respectively, k_B is the Boltzmann constant, r_i is the distance between the ith pair of residues, and $r_{cry,i}$ is the corresponding distance in the crystal structure. The numbers in the parentheses are for Ub. The strength of interaction between the pair of side chain beads i, ϵ_i^{ss}, is taken from the Betancourt−Thirumalai statistical potential.[45] The values of ϵ_h^{bb} and ϵ_h^{bs} are the same as the ones used in our previous studies on the folding of GFP.[25] Thus, the crucial E_{NB}^{N}, which determines protein stability, is transferable.

The non-native interactions, E_{NB}^{NN}, in eq 1 is taken to be

$$E_{NB}^{NN} = \sum_{i=1}^{N_{NN}} \epsilon_l\left(\frac{\sigma_i}{r_i}\right)^{6} + \sum_{i=1}^{N_{ang}} \epsilon_l\left(\frac{\sigma_i}{r_i}\right)^{6} \quad (4)$$

where $N_{NN}(= 10\,159$ in Ub) is the total number of non-native interactions, $N_{ang}(= 224$ in Ub) is the number of pair of beads separated by two bonds in the SOP-SC model, σ_i is the sum of the radii of the ith pair of residues. The radii for side chains of amino acids are given in Table S2. The values of the interaction parameters used in the energy function are given in Table S1 in the SI.

Because Ub has 23 charged residues (Figure 1A), we expect electrostatic interactions to be important at neutral pH. The last term in eq 1 accounts for electrostatic effects, which are modeled using the screened Coulomb potential,

$$E^{el} = \sum_{i=1}^{N_c-1} \sum_{j=(i+1)}^{N_c} \frac{q_i q_j \exp(-\kappa r_{ij})}{\epsilon r_{ij}} \quad (5)$$

where N_c is the number of charged residues, q_i and q_j are the charges on the side chains of the ith and jth residues respectively, κ is the inverse Debye length, and r_{ij} is the distance between interaction centers located at the centers of mass of side chains i and j. If charges are present on two bonded residues, then electrostatic interactions between these residues is ignored. The value of q_i, measured in units of electron charge, is +1 for positively charged residue and is −1 for negatively charged residues. In implicit solvent simulations of proteins a range of dielectric constants with values from 2 to 20 are typically used.[47] We used a value of $10\epsilon_o$ (ϵ_o is vacuum permittivity), which gave a reasonable radius of gyration of the protein in the unfolded state. We calculated κ assuming that a monovalent salt of 10 mM is present in the solution. The parameter λ in eq 1 is intended to account for pH effects. At neutral pH, $\lambda = 1.0$. In acidic pH, E^{el} is not as relevant, and hence we set $\lambda = 0$. At low pH, the charges on the negatively charged residues are neutralized. Because the positively charged residues no longer can engage in salt bridges, the polypeptide chain should swell leading to an increase in R_g. The residues bearing the positively charged residues would be hydrated, resulting in a reduction in the value of the effective charge (q_i). As a result, electrostatic static repulsion between the like charges would be softened. Given that these charges are well separated in Ub, it stands to reason that interactions between positive charges would not be as strong in the unfolded state as might be naively estimated based on Coulomb's law. So to a first approximation, we neglected the small repulsive interaction, as it is likely to be a perturbation to the hydrophobic interaction. This approximation is not inconsistent with experiments showing that a mutant of S6 in which 16 charged residues were replaced folded (albeit with altered properties)[48] leading the authors to argue that charge interactions must not be paramount to folding.

The parameters in the SOP-SC energy function are given in the Supporting Information (SI).

The rationale for using native-centric CG models to decipher the folding mechanisms of proteins can be traced to several previous computational and theoretical[3,4] studies. The earliest studies using lattice models[30,31] showed that, for well designed proteins, non-native interactions are likely relevant only during the initial stages of folding resulting in the collapse of the polypeptide chains.[31] These findings were also corroborated in certain atomic detailed simulations in implicit solvents.[49] It was also shown that the conformations in the transition state ensemble contain predominantly native interactions,[31,50] with non-native interactions forming with small probability.[31] More recently, analyses based on atomically detailed and C_α-Go model simulations for the distribution of the fraction of native contacts in the transition path were found to be similar.[14] These studies reinforce the notion that on time scales exceeding the collapse time, it is likely that only native interactions direct protein folding.

Simulations and Data Analysis. Following our earlier studies, we used[25,51] low-friction Langevin dynamics simulations[52] to obtain the thermodynamic properties, and Brownian dynamics simulations[53] to simulate the folding kinetics (see SI for details).

We used the structural overlap function[54] $\chi = 1 - \frac{1}{N_{tot}}\sum_{i=1}^{N_{tot}} \Theta(\delta - |r_i - r_i^0|)$ to distinguish between different populated states of the protein. Here, N_{tot} (= 11 026) is the number of pairs of interaction centers in the SOP-SC model of Ub assuming that the interaction centers are separated by at least two bonds, r_i is the distance between the ith pair of beads, r_i^0 is the corresponding distance in the folded state, Θ is the Heaviside step function, and δ = 2 Å. Examples of folding trajectories at neutral and acidic pH along with the distribution of χ are displayed in Figure S2 in the SI. Using χ as an order parameter, we calculated the fraction of molecules in the folded and unfolded basins as a function of temperature T (see SI for details). The radius of gyration R_g is calculated using $R_g = (1/2N^2)(\sum_{i,j} \vec{r}_{ij}^{\,2})^{1/2}$.

Identifying the Folding Network. In order to determine the network of connected states during Ub folding, we clustered the conformations of Ub using a structural metric based on the distribution of reciprocal of interatomic distances (DRID)[55] and leader-like clustering algorithm.[56,57] To evaluate the DRID metric, two sets of atoms are required. The first is a set of n centroids, and the second is a set of atoms N_{atom}. The centers of the 74 backbone sites of the SOP-SC model (out of the 76

DOI: 10.1021/acs.jpcb.5b03471
J. Phys. Chem. B 2015, 119, 11358−11370

backbone beads, the two termini backbone sites are omitted) are used for both the centroid set (n = 74), and the distance evaluation set (N_{atom} = 74). For each individual centroid i, three moments of distribution of reciprocal distances (μ_i, ν_i, ξ_i) are used to describe the features of atomic distances in a particular conformation. Hence, a conformation is described by a DRID vector of 222 (= 3 × n) components. The geometric distance ρ, between two conformations described by the DRID vectors (μ_i, ν_i, ξ_i) and (μ_i', ν_i', ξ_i') is obtained by $\rho = ((1/3n)\sum_{i=1}^{n}[(\mu_i - \mu_i')^2 + (\nu_i - \nu_i')^2 + (\xi_i - \xi_i')^2])^{1/2}$. The moments of distribution of the reciprocal distances for centroid i (μ_i, ν_i, ξ_i) are

$$\mu_i = \frac{1}{N_{atom} - 1 - nb_i}\sum_{j}^{*}(1/d_{ij}) \quad (6a)$$

$$\nu_i = \left[\frac{1}{N_{atom} - 1 - nb_i}\sum_{j}^{*}(1/d_{ij} - \mu_i)^2\right]^{1/2} \quad (6b)$$

$$\xi_i = \left[\frac{1}{N_{atom} - 1 - nb_i}\sum_{j}^{*}(1/d_{ij} - \mu_i)^3\right]^{1/3} \quad (6c)$$

where nb_i is the number of atoms in the distance evaluation set bonded to the centroid i, the summation is over the set of atoms assigned for distance evaluation, and the asterisk in eq 6a indicates that the atoms bonded to the centroid i are omitted in the summation. Two conformations are assigned to different clusters if the geometric distance ρ between them is greater than a certain cutoff value.

To identify a suitable cutoff value of ρ for clustering the conformations, the number of clusters N_{clust} as a function of different cutoff values of ρ is calculated (Figure S3). The N_{clust} increases exponentially as the cutoff value for ρ is decreased (Figure S3). For the Ub clusters in low pH (Figure 4), we use a cutoff value ρ = 0.0055 Å^{-1} because for this value eight clusters with probabilities greater than 0.01 exist. Based on the secondary structural content in the eight clusters, we further coarse-grained them into five clusters (Figure S4). The cumulative probability of observing a conformation in one of the five clusters exceeds 0.98, which means most of the sampled conformations can be uniquely assigned to one of the major clusters. A similar procedure is used to cluster Ub conformations in neutral pH (see Figure 5). The clustering analyses were performed for conformations sampled both at equilibrium and during the folding process.

RESULTS AND DISCUSSION

pH-Dependent Heat Capacity. The heat capacity, $C_v(T)$ [= $(\langle E^2\rangle - \langle E\rangle^2)/k_BT^2$; $\langle E(T)\rangle$, and k_B are the average internal energy and Boltzmann constant respectively], as a function of temperature, T, shows that Ub folds in low and neutral pH in an apparent two-state manner (Figure 1B). The melting temperature of Ub in low (neutral) pH is $T_m \approx$ 353 K (354 K) . These values are in reasonable agreement with experiments,[44,59] which show that T_m varies approximately from 320 K to >360 K depending on pH. The computed heat capacity curves are only in qualitative agreement with calorimetric data.[44] The dramatic changes in the pH-dependent T_m are not quantitatively reproduced. Most importantly, the full width of $C_v(T)$ at half-maximum, which changes from ≈18 K at pH = 2 to about ≈10 K in pH = 4 in experiments, is ≈5 K in simulations (Figure 1B). The calorimetric enthalpy[51,60] estimated from the specific heat data in low (neutral) pH is 142.1(142.5) kcal/mol. These values are approximately double the experimental values.[44] Interestingly, all atom simulations[41] greatly underestimate the calorimetric enthalpy. In general, it is difficult to accurately compute heat capacity curves using simulations with any empirical force field. In light of this observation, we consider the agreement between simulations and experiments using the same force field (meeting the transferable criterion) as in the our previous reports[25,51] on SH3 and GFP as reasonable.

Temperature-Dependent Ordering of the NBA and Secondary Structural Elements (SSEs). The temperature dependence of the fraction of Ub in the NBA, $f_{NBA}(T)$, shows a cooperative transition to the folded state (Figure 2A). The melting temperature, determined using $f_{NBA}(T_m)$ = 0.5, coincides with the peak position of $C_v(T)$ (Figure 1B). To

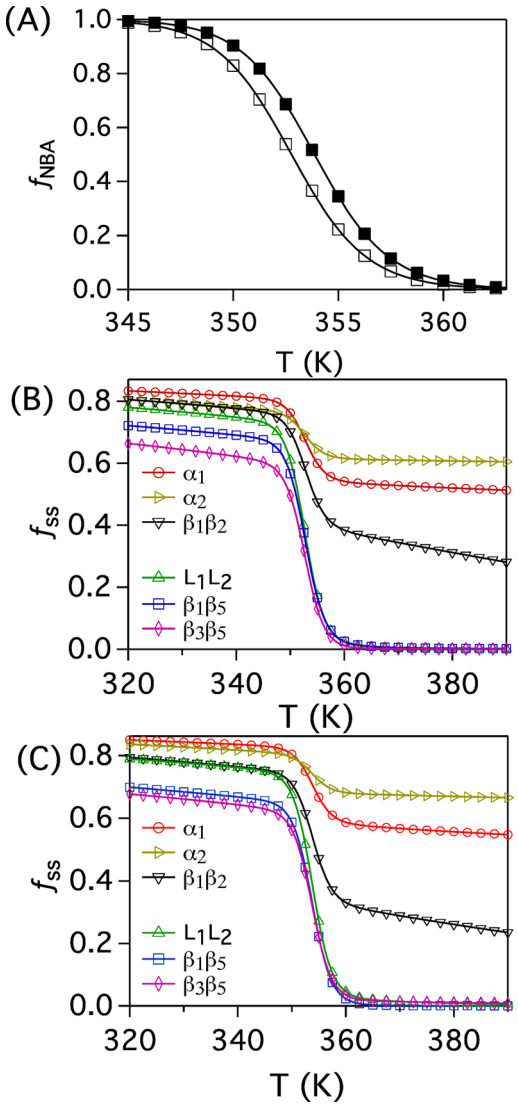


Figure 2. (A) Fraction of the protein folded, f_{NBA} in low pH (empty squares) and neutral pH (solid squares) as a function of temperature, T. Fraction of the various secondary structural elements, f_{ss} as a function of temperature T in (B) low pH and (C) neutral pH. The secondary structures α_1, α_2, and $\beta_1\beta_2$ (Figure 1A) stabilized by local contacts do not completely unfold even at temperatures above T_m. Ub unfolds upon rupture of nonlocal tertiary interactions involving $\beta_1\beta_5$, $\beta_3\beta_5$, and L_1L_2.

DOI: 10.1021/acs.jpcb.5b03471
J. Phys. Chem. B 2015, 119, 11358−11370

dissect the ordering of the secondary structural elements, SSEs (= $\beta_1\,\beta_2$, $\beta_1\,\beta_5$, $\beta_3\,\beta_5$, L_1L_2, α_1, α_2), we computed $f_{ss} = \langle N_{ss}\rangle / N_{ss}^{o}$, where $\langle N_{ss}\rangle$ is the average number of native contacts present in SSE at T, and N_{ss}^{o} is the total number of such contacts in the coarse-grained PDB structure. The secondary structural elements $\beta_1\beta_5$, $\beta_3\beta_5$, and L_1L_2, which are stabilized by nonlocal contacts are absent in the unfolded Ub. The rupture of these contacts is primarily responsible for the unfolding of the protein at $T > T_m$ in both low and neutral pH (Figure 2). In contrast, α_1, α_2, and $\beta_1\beta_2$, stabilized by local contacts, persist even at $T > T_m$ and ≈50% of the contacts between the residues stabilizing these structures are present even at $T \approx 400$ K (Figure 2). Interestingly, snapshots from atomically detailed simulations[41] at $T > T_m$ also show (see Figure 2 in ref 41) persistence of helical structures.

pH-Dependence of the Radius of Gyration (R_g). Plots of average R_g as a function of T show that the dimension of Ub at $T > \approx 355$ K in neutral pH is considerably more compact than in acidic pH (Figure 3A). The unfolded state R_g distribution at

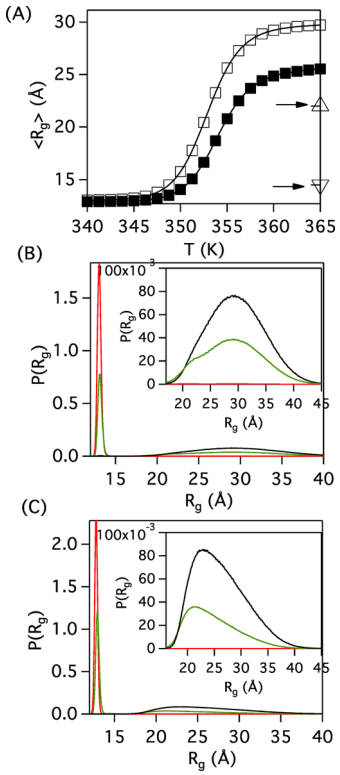


Figure 3. (A) R_g as a function of T for low pH (empty squares) and neutral pH (solid squares) calculated from SOP-SC simulations. Arrows correspond to $\langle R_g\rangle$ of Ub in the unfolded state from atomistic simulations[41] using modified CHARMM22 (black inverted triangle) and OPLS force fields[64] (triangle), which are ≈14.5 Å and ≈22 Å, respectively. Probability distribution of R_g (B) low pH at T = 343 K (red), T = 353 K (green), T = 363 K (black), and (C) neutral pH for T = 325 K (red), T = 353 K (green), T = 375 K (black). Insets in panels B and C shows R_g distribution in the unfolded state.

neutral and acidic pH shows conformations with R_g in the range $20 \lesssim R_g \lesssim 40$ with the peak of the probability distribution at (≈ 30 Å and ≈23 Å for high pH and neutral pH, respectively; see inset in Figure 3B,C). The average values of R_g in the unfolded state at high temperatures are ≈30 Å and ≈25 Å at low and neutral pH, respectively. These values are in excellent agreement with experiments,[61−63] which report a mean R_g of Ub at pH 2.5 and 7.0 are ≈32 Å and ≈26 Å, respectively. In neutral pH, dominated by attractive electrostatic interactions, Ub samples compact conformations as the centers of mass distance between the secondary structural elements $\beta_1-\beta_5$, $\beta_3-\beta_5$, L_1-L_2 are in proximity, thus explaining the compactness (Figure S5). In acidic pH, the interaction between these SSEs are not nearly as strong resulting in expansion of the polypeptide chain (Figure S5).

The $\langle R_g\rangle$ of the unfolded state of Ub computed from the probability distribution obtained from atomistic simulations with a modified CHARMM22 potential is $\langle R_g\rangle \approx 14.5$ Å (Piana et al.,[41] (Figure S1)) compares poorly with the experiments (Figure 3A). Although Ub samples conformations with R_g in the range $12 \lesssim R_g \lesssim 33$ in these simulations,[41] the peak of the probability distribution is between 12−13 Å, resulting in $\langle R_g\rangle \approx 14.5$ Å. In contrast, simulations[64] using the OPLS force field show that R_g of the unfolded protein is in the range $12 \lesssim R_g \lesssim 40$, with an estimated $\langle R_g\rangle \approx 21-22$ Å showing that even the sizes unfolded states of proteins cannot be computed unambiguously using all-atom empirical force fields.[65] More recently, it has been shown that atomic description produces unusually compact unfolded states.[66] Hydrogen exchange experiments[66] confirm that the current atomistic force fields sample compact unfolded conformations with persistent native-like secondary structure due to excessive intramolecular hydrogen bonding. It should be noted that recent computations[67,68] suggest that by tuning the protein−water interactions[67] or by using a variant of a water model generated by adjusting the dispersion interactions[68] one can alter the dimensions of the unfolded or intrinsically disordered proteins, providing reasonable R_g values in better agreement with the experiments. It remains to be ascertained whether these fixes also produce less compact structures for proteins with native states.

Hint of a High Energy Intermediate in the Free Energy Surface at Neutral pH. The folding trajectories in Figure S2 show that at T_m, Ub makes a number of cooperative transitions between the native basin of attraction (NBA) and unfolded basin of attraction (UBA). In acidic pH, such transitions between NBA and UBA in a two-state manner (Figure 1C). On the other hand, in neutral pH, the NBA → UBA involves an intermediate (Figure 1D), although its presence is not evident in the specific heat plot (Figure 1B). Using these folding trajectories, we constructed the free energy surface, $\Delta G(R_g, \chi) = -k_B T_m \ln(P(R_g,\chi))$, where $P(R_g,\chi)$ is the joint probability distribution of R_g and χ at T_m. The $\Delta G(R_g, \chi)$ profiles display two major basins (NBA and UBA) in low and neutral pH conditions (Figure 1C,D). These two basins are separated by a barrier (Figure 1C,D). In neutral pH the free energy surface has an additional high energy basin, which can be associated with an intermediate (Figure 1D). The shoulder, corresponding to the intermediate, is on the NBA side. Below we show that the shoulder corresponds to the metastable states sampled by Ub in the UBA (see below).

The $\Delta G(R_g, \chi)$ profiles show that acquisition of the native state occurs only after substantial compaction of the polypeptide chain (Figure 1C,D). In both neutral and acidic pH the value of χ, even after a large decrease in R_g, is relatively high. We infer that folding only commences after populating an ensemble of minimum energy compact structures,[69] which has been explicitly demonstrated for Ub folding using single molecule pulling experiments.[70]

Network of Connected States at T_m. In order to assess the complexity of the folding thermodynamics of Ub, we performed a clustering analysis (see Methods) at the melting temperatures using a 37 μs trajectory at acidic pH and a 5 μs trajectory in neutral pH (Figure S2). Even though at T_m, the protein in low

DOI: 10.1021/acs.jpcb.5b03471
J. Phys. Chem. B 2015, 119, 11358−11370

and neutral pH appears to fold in a two-state like manner (Figure 1C,D), the clustering analysis reveals that Ub samples prominent metastable states where it adopts secondary and tertiary structures to varying degrees (Figures 4 and 5).

Of the five prominent clusters identified in acidic pH, three are metastable states labeled MS1−3 in Figure 4. In the MS1 state,

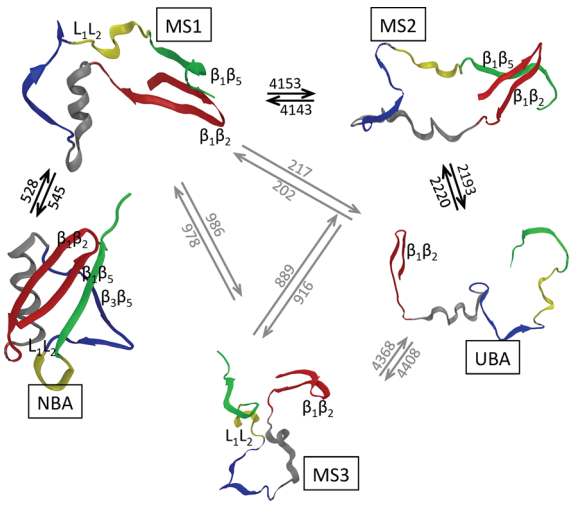


Figure 4. Network of connected states obtained using a clustering analysis of the folding trajectories in low pH at $T \approx T_m$ (= 353 K). There are three metastable clusters labeled MS1−3, which are considered equilibrium intermediates. A representative conformation with the secondary structural elements from each cluster is shown. Dark arrows correspond to the dominant pathway and the gray arrows show possible subdominant routes connecting NBA and UBA. Interestingly, MS3 is not connected to NBA. The numbers show the number of transitions between the clusters in the 37 μs trajectory.

the SSEs $\beta_1\beta_2$, $\beta_1\beta_5$,and L_1L_2 are formed, whereas in MS2 only the $\beta_1\beta_2$ hairpin and $\beta_1\beta_5$ are present. In the MS3 state, the hairpin $\beta_1\beta_2$ and L_1L_2 contacts are present. The network of connected states shows that the dominant thermodynamic pathway for assembly of Ub is NBA ⇆ MS1 ⇆ MS2 ⇆ UBA. Although, conformations belonging to MS1 and MS2 are sampled in the folding trajectories, globally Ub appears as a two-state folder because the lifetimes of the MS1 and MS2 states are small at T_m. (Figures 1B and S2).

In neutral pH, the reversible pathway between UBA and NBA involves MS1 and MS3 states (Figure 5). Electrostatic interactions destabilize the MS2 state, and hence it is not sampled in the folding pathways. The MS1 state has a long enough lifetime in neutral pH that it is discernible as a high energy intermediate in the free energy surface in Figure 1D. The dominant Ub folding pathway in neutral pH connects the states NBA ⇆ MS1 ⇆ MS3 ⇆ UBA, which is different from the dominant pathway in low pH. This is the first indication that the folding mechanism depends on pH, which we demonstrate below using kinetic simulations. The MS3, state, which is not a part of the low pH dominant folding pathway (Figure 4), is stabilized in neutral pH by favorable interactions among the charged residues at the interface of the SSEs L_1, L_2, and α_1 (Figure S1). Interactions associated with these structural elements play a key role in Ub folding close to neutral pH (see below).

The thermodynamic pathway can be compared to experiments. Based on experiments[71,72] at pH 7.5 and Ψ-analysis, two folding pathways for Ub were inferred. In the major folding

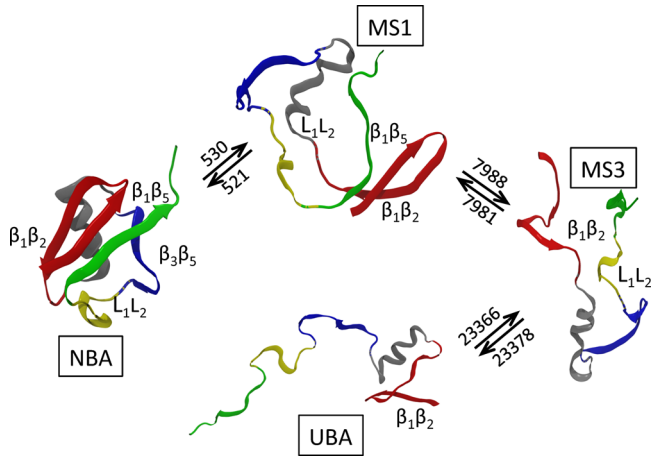


Figure 5. Network of connected states obtained using a clustering analysis of the folding trajectories in neutral pH at $T \approx T_m$ (= 355 K) reveal four clusters. In addition to the NBA and UBA, there is a intermediate cluster, MS1, and a metastable cluster MS3. A representative conformation with the secondary structural elements from each cluster is shown. The numbers on the arrows show the number of transitions between the clusters in the 5 μs trajectory.

pathway, the hairpin $\beta_1\beta_2$ forms, then the helix α_1 stacks onto $\beta_1\beta_2$ forming tertiary contacts. Subsequently, sheet β_5 interacts with β_1 forming the $\beta_1\beta_5$ contacts. The dominant pathway inferred from experiments on the mutant F45W agrees partially with our predictions in neutral pH. Our analysis (Figure 5) reveals that tertiary contacts exist only between $\beta_1\beta_2$ strands. We find that the contacts between α_1 and $\beta_1\beta_2$ alone are not stable. We suggest that the next step in the assembly is the formation of contacts between α_1, L_1, and L_2 due to the charged residues (Figure S1), giving rise to transient population of the MS3 state. Finally, sheet β_5 interacts with β_1 enabling the formation of the rest of the tertiary contacts. In such compact intermediates where L_1L_2 and $\beta_1\beta_5$ contacts are formed, we also observe interactions between α_1 and $\beta_1\beta_2$ in some of the Ub folding trajectories in neutral pH conditions (see below and structure I3 in Figure 8).

The dominant pathway identified using simulations in low pH (Figure 4) agrees with the minor pathway inferred from the experiments[71,72] at pH 7.5. In this pathway, the hairpin β_1 β_2 forms first, then the sheet β_5 interacts with β_1 forming the $\beta_1\beta_5$ contacts (Figure 4, clusters MS1 and MS2). Subsequently, helix α_1 forms tertiary contacts with sheets β_1, β_2, and β_5. The UBA cluster is similar to the structure *U*1 in the atomistic simulations[41] where the contacts between the $\beta_1\beta_2$ sheets are present. The conformations of Ub in the MS1 and MS2 states with contacts between $\beta_1\beta_2$, $\beta_1\beta_5$, and L_1L_2 are similar to the cluster U1′ in the atomistic simulations.

Temperature and pH Dependence of Folding Mechanism. We generated at least 50 folding Brownian dynamics simulation trajectories in low and neutral pH at two temperatures to probe the dynamics of Ub folding. The simulations are initiated from an ensemble of unfolded conformations generated at temperatures $T > T_m$. Regardless of the temperature or pH the SSEs, α_1, α_2, and $\beta_1\beta_2$ hairpin are the first structural elements to form. The variations in the self-assembly of Ub occur only in the subsequent stages.

Low pH. At T = 300 K Ub folds along four pathways, which we illustrate using one of the folding trajectories (Figure 6A and 7). In all the pathways, the SSEs α_1, α_2, and the hairpin $\beta_1\beta_2$ are always present due to their stability. Subsequently there is a

DOI: 10.1021/acs.jpcb.5b03471
J. Phys. Chem. B 2015, 119, 11358−11370

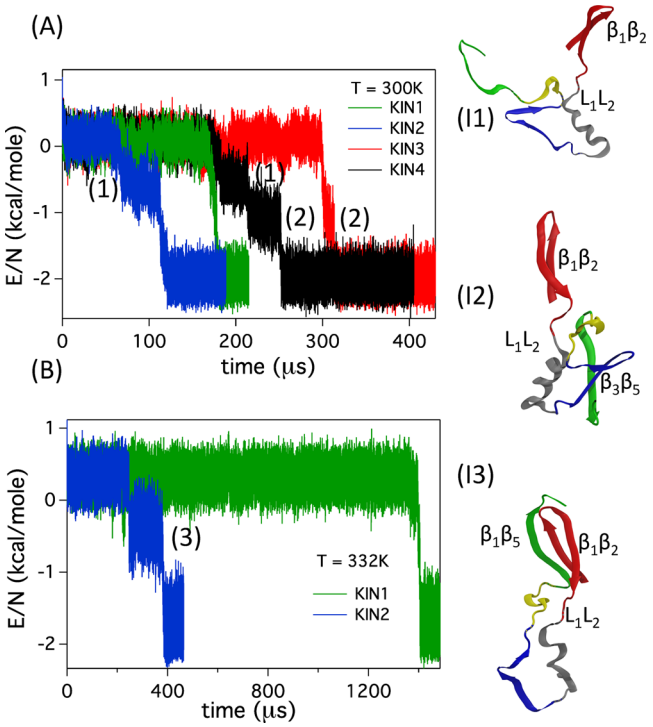


Figure 6. Ub folding kinetics in low pH. The folding pathways inferred from change in energy per residue as a function of time, t at (A) T = 300 K and (B) T = 332 K. Two kinetic intermediates I1 and I2 are identified in the folding pathways at T = 300 K, and the intermediate I3 is populated at T = 332 K. Representative structures of the kinetic intermediates are on the right.

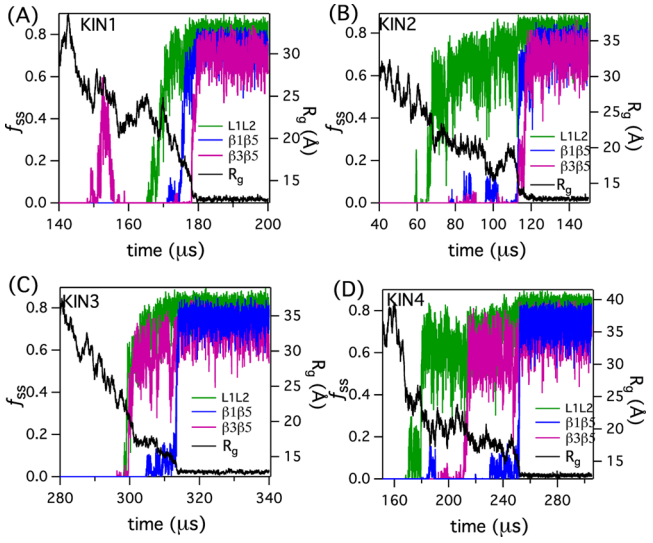


Figure 7. Fraction of native contacts in various secondary structural elements ($\beta_1\beta_5$, $\beta_3\beta_5$, and L_1L_2), f_{ss} and R_g as a function of time in low pH folding trajectories at T = 300 K. The plots in four panels are for trajectories labeled KIN1, KIN2, KIN3, and KIN4 in Figure 6A. The green, blue and magenta colors correspond to secondary structures L_1L_2, $\beta_1\beta_5$, and $\beta_3\beta_5$, respectively. Radius of gyration, R_g as a function of time is shown in black.

bifurcation in the folding pathways. Using energy per residue as a reporter of folding, we find that in the pathways KIN2 and KIN3, one intermediate is populated prior to reaching the folded state, whereas there is no persistent intermediate in KIN1 (Figure 6). The observed bifurcation in the folding pathways, where one route involves a direct UBA → NBA transition, while in the rest NBA is reached in stages, is the hallmark of the kinetic partitioning mechanism (KPM).[73−75] The two intermediates are structurally different (Figure 6A, I1 and I2). In KIN4, both the intermediates (Figure 6, I1 and I2) are sampled, whereas folding in KIN2, only I1 is sampled and I2 is accessed in KIN3. I1 is stabilized by the secondary structural elements $\beta_1\beta_2$ and L_1L_2, whereas I2 is stabilized by the contacts between $\beta_1\beta_2$, L_1L_2, and $\beta_3\beta_5$ (Figure 7). There are similarities between I1 and the MS3 state (Figure 4) identified in the equilibrium simulations.

In KIN1, where Ub appears to fold in a two-state-like manner (Figure 6A), the contacts stabilizing the secondary structures L_1L_2, $\beta_1\beta_5$, and $\beta_3\beta_5$ form almost simultaneously (Figure 7A). They assemble successively separated by a time ≈15 μs (Figure 7A), which is only a fraction of the first passage time. The interactions among the nonlocal SSEs needed to stabilize the compact states, and collapse of Ub measure by decrease in $R_g(t)$ occur nearly simultaneously (Figure 7A). Folding along this pathway thus follows the nucleation-collapse (NC) mechanism.

In KIN2 and KIN4, the L_1L_2 contacts form ahead of $\beta_1\beta_5$ and $\beta_3\beta_5$ leading to I1 (Figure 7B,D). The kinetic intermediate I2 is formed when contacts L_1L_2 and $\beta_3\beta_5$ are established simultaneously or successively as observed in KIN3 and KIN4, respectively (Figure 7C,D). In the other pathways intermediates with well-defined structures form (Figure 7B,C). These figures also show that $R_g(t)$ decrease continues even after some of the nonlocal SSEs form. The route to the native state in KIN4 is via both the intermediates found in KIN2 and KIN3 (Figure 7D). The presence of a nearly direct transition to the native state in KIN1 and folding through the intermediates in the other pathways is in accord with the kinetic partitioning mechanism (KPM).

The assembly of native Ub along the KIN2, KIN3, and KIN4 pathways at T = 300 K can be rationalized by the diffusion-collision mechanism (DCM).[76] In the nascent stages of folding microdomains (for example $\beta_1\beta_2$ and $\beta_1\beta_5$ sheet in I3) form which diffuse freely. Subsequently, some of these collide and coalesce to form the kinetic intermediates with lifetimes on the order of ≈100 μs. In the final stages of folding, the rest of the secondary structural elements collide with the core of the protein structure, and coalesce to form the native structure (see below for further discussion).

At the higher temperature T = 332 K Ub folds by the KPM (Figure 6B). In KIN1, the protein folds in a two-state manner, and in KIN2 a single intermediate (Figure 6B) whose structure is different from the ones observed at T = 300 K is populated. This intermediate,[40,41] stabilized by the SSEs $\beta_1\beta_2$, L_1L_2, and $\beta_1\beta_5$ (Figure 6, I3), is similar to the MS1 state at T_m, thus providing evidence that Ub samples the equilibrium structures in the process of folding (see below for additional discussion). By analyzing the formation of individual secondary structures (Figure S6) we find that at the higher temperature, the contacts that stabilize L_1L_2 leading to S1 (MS3 cluster) are unstable (Figure S6) in contrast to what is observed at T = 300 K (Figure 7B,D). The L_1L_2 and $\beta_1\beta_5$ contacts are stabilized simultaneously at T = 332 K (Figure S6) for time scales on the order of ≈100 μs leading to I3. At higher temperatures the contacts between the β-sheets $\beta_1\beta_5$ are important in stabilizing the intermediate, as these are the end-to-end contacts in Ub. These interactions minimize the conformational fluctuations leading to lifetimes that are long enough for S3 to form (Figure 6B). At T = 332 K, a two-state-like folding pathway is observed when $\beta_3\beta_5$ contacts immediately form after structuring of L_1L_2 and $\beta_1\beta_5$ contacts (Figure S6A).

DOI: 10.1021/acs.jpcb.5b03471
J. Phys. Chem. B 2015, 119, 11358−11370

Neutral pH. Folding at neutral pH, where electrostatic interactions play a major role, is dramatically different. In this case, both at T = 335 K and 300 K a well-populated intermediate is observed (MS1 in Figure 5, and I2 in Figure 8). In particular,

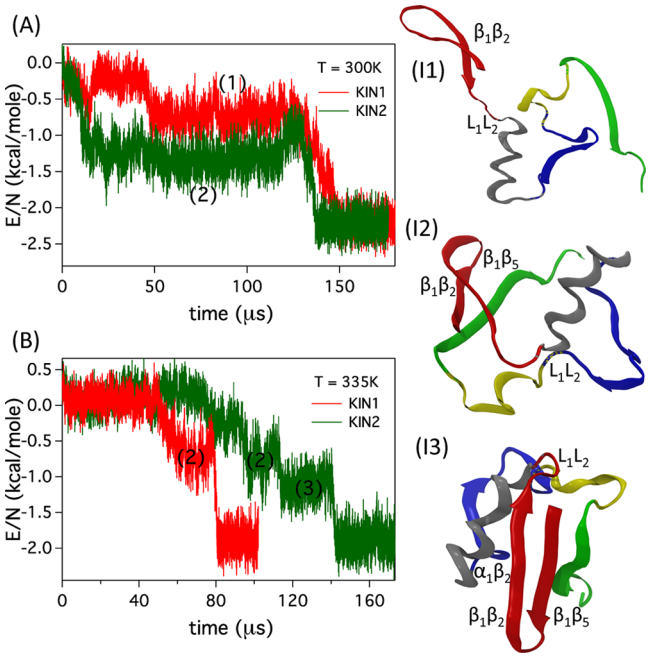


Figure 8. Folding kinetics in neutral pH at T = 300 K and 335 K. Three intermediates are populated in the folding trajectories and representative structures of the intermediates labeled I1, I2, and I3 are shown on the right.

due to the electrostatic interactions between residues in L_1 and L_2, the L_1L_2 loop formation precedes the formation of the I2 intermediate (Figure S7). The assembly of Ub occurs by the DCM where preformed microdomains collide. The intermediate I2 can further become compact due to favorable electrostatic interactions between the helix α_1 and β_2 strand leading to I3 (Figure 8 and S8). This intermediate is not observed in high pH folding. The final stage of compaction is delayed until the microdomains adopt native-like topology (Figure S7).

Comparison with Experiments. It should be borne in mind that there are discrepancies in the interpretations of the different experimental data, which makes it difficult to make direct comparisons with a single experiment. With this caveat, we note that our findings, which are by and large consistent with experiments,[22,23,77−88] provide a complete picture of the structures sampled by Ub during folding. However, experiments have only characterized a subset of the predicted intermediates. In accord with the present findings, experiments inferred that Ub folds through an intermediate at low temperatures or mildly denaturing conditions or when mutations slow down folding. A common characteristic in all the intermediates, regardless of pH or temperature, is that the $\beta_1\,\beta_2$ hairpin is stable for which there is substantial experimental evidence.[23,79,86] The I1 intermediate found here rationalizes experimental studies,[79,86,89] which provide evidence for a stable α_1 helix and unstable β_3, β_4, and β_5 strands. In addition protein vivisection suggested[90] an intermediate structure of Ub ubiquitin in which the small β_4 (Figure 1A) strand is unstructured (Figures 6 and 8B). Taken together we conclude that our simulations provide a complete structure of the populated intermediates filling in gaps in experimental studies.

The environment-dependent complex folding pathways are captured in Figure 10. The folding mechanisms, involving characterization of the network of connected intermediate structures and transitions between them, are vastly different if the external conditions are altered. The most general characteristic is that folding is a stochastic process in which assembly occurs by multiple pathways. Depending on the conditions, the flux through these pathways can be altered and, as demonstrated here, one can even use a single dominant pathway for folding. Validation of our predictions requires experiments probing folding kinetics as a function of pH and temperature.

Coincidence of Equilibrium and Kinetic Intermediates. The structures of the I1 kinetic intermediates in low pH at T = 300 K, and I3 at T = 332 K (Figure 6) are similar to those observed in the MS1 and MS3 states (Figure 4). The dominant folding/unfolding pathway identified at T_m (Figure 4) is very similar to the KIN2 folding pathway at T = 335 K. In neutral pH, the dominant folding pathway (Figure 5) is identical to the folding pathways at both T = 300 K and 335 K.

The coincidence of equilibrium and kinetic intermediates is not without precedence. A series of insightful NMR experiments have established that, during the folding of apomyoglobin, a kinetic intermediate is populated[91] that has the same structure as the one characterized at equilibrium.[92] Our study leads to the prediction that the network of states accessed kinetically are also found in folding trajectories at equilibrium. This prediction can be tested using NMR experiments as a function of pH just as was done for apomyoglobin.

Identification of the Transition State Ensemble (TSE) Using P_{fold}. The transition state structures of Ub are identified from the folding trajectory at T_m (Figure S2). Putative transition state structures are picked from the saddle-point region of the free energy projected onto the variables E and χ using the conditions -50.0 kcal/mol < E < -40.0 kcal/mol and 0.66 < χ < 0.68. Starting from these structures, we calculated the commitment probability, P_{fold},[93] of reaching the NBA. The set of structures with $P_{fold} \approx 0.4-0.6$ is identified as the transition state ensemble (TSE) (Figures 9C and S9). To our knowledge, this is the first demonstration that TSE has been quantitatively identified for Ub without any prejudice about the underlying reaction coordinate.

The average TSE structure is globular with most of the SSEs and tertiary contacts intact as in the folded state, and they are fairly homogeneous (Figure 9A), supporting the conclusions based on Ψ-value analysis[71,72] and all atom simulations.[41] The average χ, an estimate of the contact-order in the TSE, is approximately 0.67, in agreement with reported values for various proteins.[94] The $\beta_1\beta_2$ hairpin and α_1 are fully structured in the TSE, which is not surprising given their thermodynamic stability (Figure 3B). Compared to the folded structure, the contacts between the β-sheets $\beta_3\beta_5$ are absent in the TSE, although the strands β_3 and β_5 are fully structured (Figure 9B). The formation of a compact TSE in which majority of the SSE and tertiary interactions are consolidated further supports that at $T \approx T_m$ Ub folds by the NC mechanism.

The TSE structures identified in the simulations are in reasonable agreement with the inferences drawn from the Φ-value analysis,[95] and are in better agreement with the Ψ-value analysis[72,96] and T-jump infrared spectroscopy experiments.[23] Based on these studies,[72,95,96] it is suggested that the N-terminal part of the protein, helix α_1 and sheet $\beta_1\beta_2$, are ordered in the

DOI: 10.1021/acs.jpcb.5b03471
J. Phys. Chem. B 2015, 119, 11358−11370

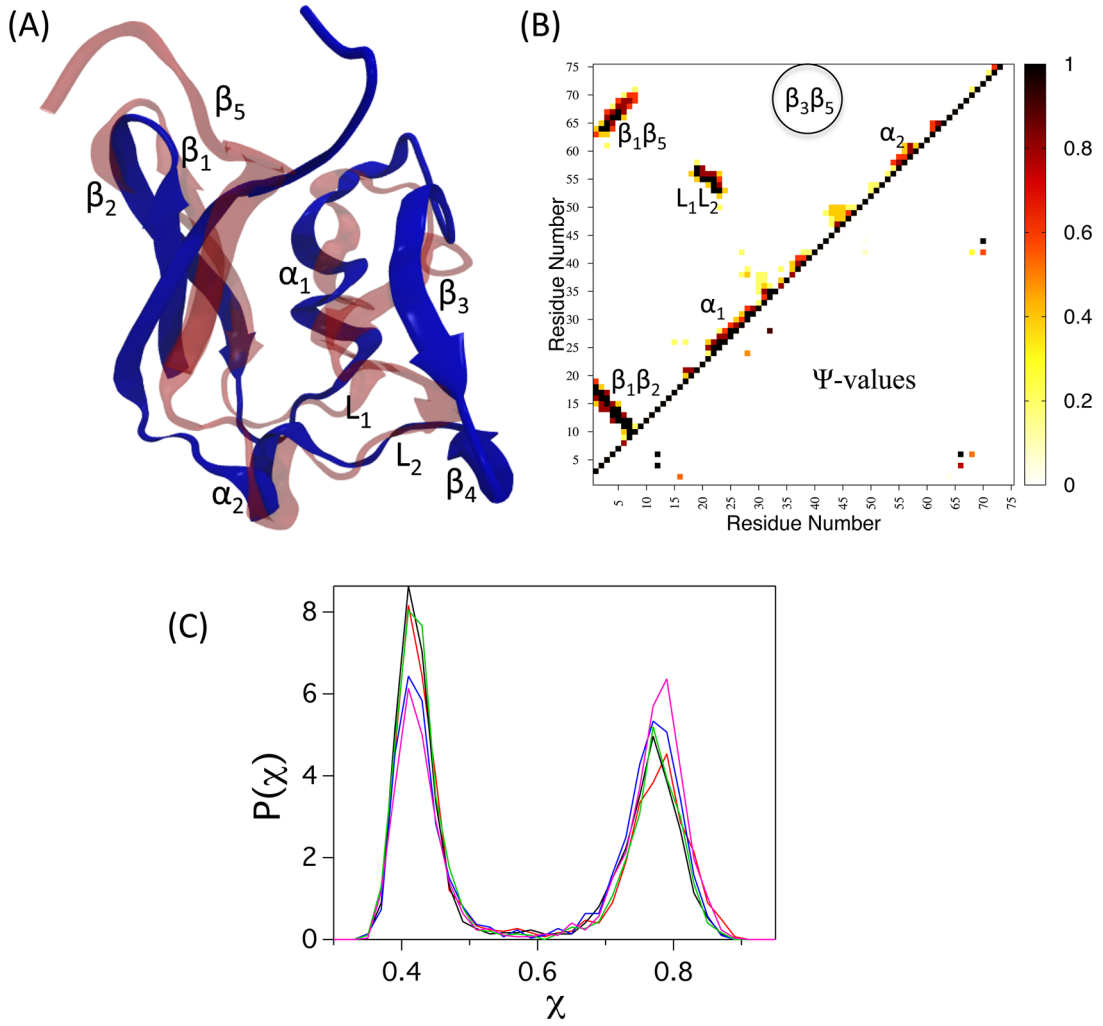


Figure 9. (A) Two superimposed representative structures from the transition state ensemble at low pH. (B) The upper diagonal of the plot shows C_α contact-map of the transition state ensemble. There are no interactions between $\beta_3\beta_5$ (shown in circle). The lower diagonal of the plot shows experimental[71] Ψ-values. (C) The distribution of the final structural overlap parameter, χ, of at least 500 simulation trajectories spawned from the transition state structures. Data is shown for five different structures. The distribution shows that roughly half of these trajectories go to the folded basin and the other half reach unfolded basin ($p_{fold} \approx 0.5$).

TSE, which agrees with our simulations (Figure 9A). However, the experiments disagree among each other on the TSE structure in the C-terminal region of the protein. The Φ-value analysis[95] suggests that the C-terminal region of the protein is unfolded, while the Ψ-value analysis[72,96] and T-jump infrared spectroscopy experiments[23] infer the opposite. According to these experiments, the transition state is extensively ordered with structure comprising the four β-strands and the α-helix. Our simulations show that the β_5 does make contacts with β_1 in the TSE in agreement with the experiments based on Ψ-value analysis. In addition, the p_{fold} analysis shows that the C-terminus α-helix is at least partially structured.

A picture of the TSE using all-atom MD simulations in water and projection onto a one-dimensional reaction coordinate was proposed.[41] Using the dynamics in the projected coordinate, they computed Φ-values for only hydrophobic residues using certain (untested) assumptions. The trends (not absolute values) in experiments and simulations are similar.[41] On this basis, they asserted that the structures in the barrier region in the one-dimensional coordinate is the TSE. Because of the completely different methods and the models used (our TSE is most appropriate for acidic pH) in the two studies it is difficult to directly compare the p_{fold}-based determination of the TSE with the one from.[41] Nevertheless, in both the studies, TSEs are homogeneous, compact, and native-like.

Relevance of Non-native Interactions. We provide generic arguments showing that non-native interactions ought to play only a subdominant role in the folding of evolved small proteins that ostensibly fold in an apparent two-state manner. In order to keep the arguments simple, let us assume that non-native interactions largely affect the unfolded state. There is anecdotal evidence that this is the case in a mutant of NTL9.[97] We write the free energy difference between the unfolded states containing non-native (NN) interactions and one described using only native (N) interactions as $\Delta\Delta G_U = \Delta H_U - T(\Delta S_U)$, where ΔH_U (ΔS_U) is the enthalpy (entropy) changes between the NN and N models of the unfolded state. Typically, but not always, we expect that NN interactions ought to stabilize the unfolded state compared to the unfolded state described by the N model. However, $\Delta H_U < 0$ also implies $T(\Delta S_U)$ will be negative because certain conformations formed by favorable NN interactions are disallowed in the native interaction dominated model. Thus, the sign of $\Delta\Delta G_U$ is determined by the magnitude of $T(\Delta S_U)$, which cannot be too large to negate ΔH_U. If this were

DOI: 10.1021/acs.jpcb.5b03471
J. Phys. Chem. B 2015, 119, 11358−11370

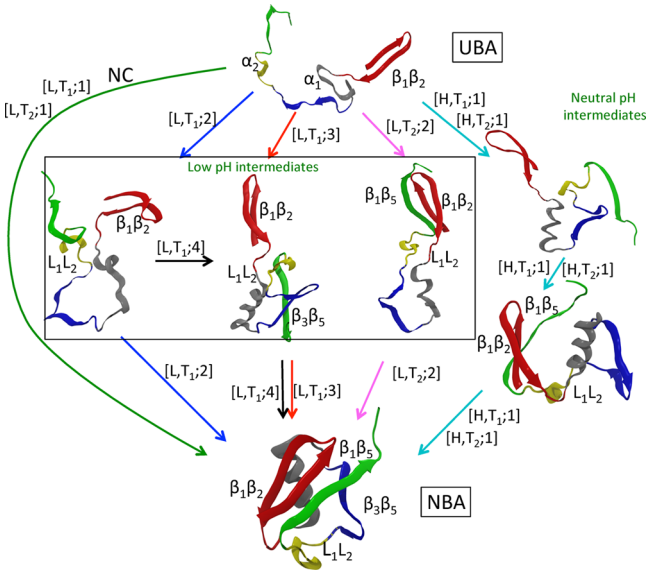


Figure 10. Complex folding pH and temperature-dependent folding pathways for Ub. The first index in $[L,T_1;\alpha]$ denotes low pH, the second is temperature, and α = 1, 2, 3, 4 labels the kinetic pathways. A similar interpretation with $T_2 > T_1$ holds for $[H,T_2;\alpha]$ with H standing for neutral pH. In the unfolded state secondary structures stabilized by local contacts (α_1, α_2, $\beta_1\beta_2$) are always present. The transition from UBA to NBA in $[L,T_1,1]$ and $[L,T_2,1]$ (green arrow) is described by the nucleation-collapse (NC) mechanism. Blue, red, and black arrows routes to the NBA from UBA at T_1 through well-defined intermediates whose structures are displayed. The high-temperature route through an intermediate is shown by lavender arrows. The folding pathway at neutral pH at both temperatures is in aqua blue. Three of the four intermediates (the left two and the one on the right) are also populated in equilibrium folding trajectories.

the case, then NN interactions alone would stabilize the folded states to a greater extent than N interactions, which is unlikely. So we will assume that $|\Delta H_U/T(\Delta S_U)| > 1$.

Because hydrophobic interactions between small hydrophobic species is entropic in origin, we expect that non-native interactions are most relevant when a salt bridge, not present in the folded state, can form in the unfolded state, and hence may be important in Ub folding around neutral pH. It is clear that $|\Delta\Delta G_U/\Delta G_{NU}| \ll 1$, where ΔG_{NU} is the stability of the native state with respect to the unfolded state. If this inequality is violated, then the protein would not fold! Thus, it follows that NN interactions are most likely to be perturbation and not a dominant determinant of the folding thermodynamics. This conclusion is increasingly valid for proteins whose native states are highly stable. A corollary of this argument is that folding rates, and in most cases unfolding rates as well, are unlikely to change significantly (less than an order of magnitude). Our previous works on lattice models, which treated N and NN interactions on equal footing, provide illustrations of the arguments provided here.[31]

Do the arguments given above imply that NN interactions[98] are not relevant at all? We discuss two examples suggesting that favorable NN interactions may affect the stability and kinetics of folding. (1) Mutations of surface exposed charged residues in Fyn SH3 domain, a small protein (≈56 residues) showed the folding rates, with respect to the WT, increases by a factor of ≈3 for the E46K mutant and by a factor of ≈8 for mutant for the E46K-E11K-D16K-H21K-N30K (Fyn5) mutant (see Table 1 in ref 99). The unfolding rate decreases only by a factor of 2 for these mutants, which can be taken to mean that the effects of the dramatic mutations largely affect the unfolded states. Even though these mutations, especially the Fyn5 construct, are drastic, the effect on the folding kinetics is modest, and the factor of 8 increase can be accounted for by $\approx 2k_BT$ change in the barrier height, which roughly mirrors the enhanced stability of the Fyn5 mutant (see Table 1 in ref 99). The relatively small changes (less than a factor of 10 even in Fyn5 mutant) in the rates are in accord with the arguments given above. To explain these changes, coarse-grained simulations were performed using C_α representation of the protein. We do not believe the explanation based on these simulations is adequate for two reasons. First, these simulations included only NN interactions between charged residues using the Debye−Huckel potential. The parameters of the NN electrostatic interactions are very different from those used for native electrostatic interactions. In other words, NN and N interactions are not treated on *equal footing*. Second, the inferences that non-native interactions might be present in the transition state were made based on free energy profiles computed using the fraction of native contacts without the benefit of P_{fold} analysis. (2) In a more compelling case, it has been shown in a number of NTL9 mutants that the unfolded state may be stabilized by a non-native salt bridge.[97] The most dramatic change occurs in experiments at pH 5.5 in the K12 M mutant (a rather large perturbation) in which the free energy of the mutant is enhanced by $\approx 3k_BT$ (a 50% increase) relative to the WT, with all other mutants exhibiting less than k_BT change (see Table 1 in ref 97). Apparently, in the WT a salt bridge forms between D8K12 in the unfolded state, which is clearly abolished in the mutant, thus increasing the free energy of the unfolded state. The result is the K12 M is more stable than the WT. These results were explained by using CG simulations in which Debye−Huckel potential was used with NN interactions only between charged residues.[100] In this study, NN and N electrostatic interactions were treated on equal footing. The findings corroborate the experimental observations.

We draw two conclusions for the discussion presented here. First, only in dramatically altered sequences are NN interactions significant in affecting the stability. Second, these changes involving charged residues can be taken into account within CG models by slightly altering the strength of hydrophobic interactions. After all large perturbations in both Fyn SH3 and NTL9 can modulate both electrostatic as well hydrophobic interactions because one expects changes in hydration in the unfolded states as a consequence of these mutations. Based on the current evidence from simulations,[14,30,31] we conclude that generically non-native interactions are likely to be *only a perturbation* and *not* a dominant factor in the folding of small single domain proteins. This conclusion is in accord with the arguments given above and theoretical considerations.[4]

CONCLUSIONS

In summary, using coarse-grained models and molecular dynamics simulations we have dissected the folding of Ub as a function of temperature at acidic and neutral pH. The major findings in this work are as follows: (1) We quantitively predict the pH-dependent changes in the radius of gyration of Ub. The values of mean R_g at high temperatures are in excellent agreement with experiments. (2) A significant prediction of our study is that the folding pathways can be dramatically altered by changing pH. The major pathway at low pH resembles the minor pathway at neutral pH. The structures of some of the intermediates and transition states, which are only indirectly inferred from

DOI: 10.1021/acs.jpcb.5b03471
J. Phys. Chem. B 2015, 119, 11358−11370

experiments, are fully resolved. (3) Our work also highlights the balance between the number of local and nonlocal contacts determines the folding mechanisms of protein folding in general.[101] In the context of Ub, the secondary structural elements stabilized by local contacts form in the early stages of the folding process. Only subsequently and after considerable compaction, do secondary structures, stabilized by nonlocal contacts, form. The formation of these nonlocal contacts determine the folding rates and are strongly influenced by the folding conditions. (4) Although there are dominant folding pathways under all conditions for folding in general, and Ub in particular, self-assembly can occur by alternate subdominant routes. Thus, the assembly mechanism of proteins should be described in probabilistic terms—a notion that appears naturally in the statistical mechanical description of folding.[4,21] In accord with this general principle, we find that the folding mechanism is complex especially at acidic pH. Under these conditions, a fraction of molecules folds by a nucleation-collapse mechanism, whereas in others long-lived metastable intermediates are populated prior to collapse and the formation of the native-state. This finding is in line with the KPM,[73] which is now firmly established for a number of proteins.[74,75] At neutral pH, Ub folds by a sequential mechanism in which local SSEs first form. Subsequently an intermediate stabilized by long-range contacts (L_1 L_2 and β_1 β_5) is populated prior to the formation of the native-state.

An important enterprise in molecular simulations is to benchmark force fields, which should be done by comparing simulations and experiments. Minimally such comparisons should include specific heat profiles, dependence of the dimensions (R_gs) of the protein as a function of temperature and denaturants, and time dependent changes in R_g and other measurable properties probing the kinetics of self-assembly. We hasten to add that it is almost impossible to calculate accurately (nor should one attempt such computations) material-dependent properties (specific heat being one example) using simulations with *ad hoc* empirical force-fields including CG models or atomically detailed models. For purposes of direct comparisons with experiments, it is prudent to create transferable CG models[24,102] by benchmarking against experiments. The transferable CG force-field we have created has been remarkably successful in semiquantitatively reproducing many experimental quantities for srcSH3[51] and GFP[25] as a function of denaturants. Such simulations are currently beyond the scope of atomic detailed simulations because of a lack of reasonable force fields for denaturants and the sheer size of GFP.

Despite the ability to reproduce experimental measurements and make testable predictions for a large number of proteins using coarse-grained models, they have obvious limitations. The absence of explicit inclusion of the solvent, which has an impact on the fluctuations of the unfolded state, makes it difficult to quantitively reproduce the measured heat capacity curves. Finally the knowledge of the native structure needed in these simulations can be legitimately criticized. Despite these reservations, the potential utility of coarse-grained models in protein and RNA folding is substantial.[10] Most importantly, such simulations can be carried out using standard desktop computers.

■ ASSOCIATED CONTENT

Supporting Information

Description of the simulation methods; Table S1−S2; Figures S1−S9. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b03471.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +91-80-22933533. Fax: +91-80-23601310 E-mail: greddy@sscu.iisc.ernet.in.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are indebted to George Makhadatze for pointing out an error in the simulated heat capacity curves in an earlier version. We thank William Eaton, Stefano Piana, Eugene Shaknovich, and Tobin Sosnick for valuable comments. We also acknowledge discussions with Koby Levy and Dan Raleigh on non-native electrostatic interactions in NTL9. GR acknowledges startup grant from Indian Institute of Science-Bangalore. DT acknowledges a grant from the National Science Foundation through grant CHE 1361946. 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.

■ DEDICATION

We wish to dedicate this article to Biman Bagchi who has made pioneering contributions to a number of areas using judicious use of theory and computations.

■ REFERENCES

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

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

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

(4) Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 167−195.

(5) Dill, K.; Chan, H. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **1997**, *4*, 10−1.

(6) 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.

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

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

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

(10) Hyeon, C.; Thirumalai, D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat. Commun.* **2011**, *2*, 487.

(11) Whitford, P. C.; Sanbonmatsu, K. Y.; Onuchic, J. N. Biomolecular dynamics: order-disorder transitions and energy landscapes. *Rep. Prog. Phys.* **2012**, *75*, 076601.

(12) Tozzini, V. Minimalist models for proteins: a comparative analysis. *Q. Rev. Biophys.* **2010**, *43*, 333−371.

(13) Vicatos, S.; Rychkova, A.; Mukherjee, S.; Warshel, A. An effective Coarse-grained model for biological simulations: Recent refinements and validations. *Proteins: Struct., Funct., Genet.* **2014**, *82*, 1168−1185.

(14) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 17874−17879.
(15) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341−346.
(16) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How fast-folding proteins fold. *Science* **2011**, *334*, 517−520.
(17) Thirumalai, D.; Liu, Z.; O'Brien, E. P.; Reddy, G. Protein folding: From theory to practice. *Curr. Opin. Struct. Biol.* **2013**, *23*, 22−29.
(18) Hicke, L. Protein regulation by monoubiquitin. *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 195−201.
(19) Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **2009**, *78*, 477−513.
(20) Lee, S. Y.; Pullen, L.; Virgil, D. J.; Castaneda, C. A.; Abeykoon, D.; Bolon, D. N. A.; Fushman, D. Alanine Scan of Core Positions in Ubiquitin Reveals Links between Dynamics, Stability, and Function. *J. Mol. Biol.* **2014**, *426*, 1377−1389.
(21) Thirumalai, D. From Minimal Models to Real Proteins: Time Scales for Protein Folding Kinetics. *J. Phys. I* **1995**, *5*, 1457−1467.
(22) 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.
(23) 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.
(24) 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.
(25) Reddy, G.; Liu, Z.; Thirumalai, D. Denaturant-dependent folding of GFP. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17832−17838.
(26) Chen, J.; Darst, S. A.; Thirumalai, D. Promoter melting triggered by bacterial RNA polymerase occurs in three steps. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 12523−12528.
(27) Theisen, K. E.; Zhmurov, A.; Newberry, M. E.; Barsegov, V.; Dima, R. I. Multiscale Modeling of the Nanomechanics of Microtubule Protofilaments. *J. Phys. Chem. B* **2012**, *116*, 8545−8555.
(28) Kononova, O.; Kholodov, Y.; Theisen, K. E.; Marx, K. A.; Dima, R. I.; Ataullakhanov, F. I.; Grishchuk, E. L.; Barsegov, V. Tubulin Bond Energies and Microtubule Biomechanics Determined from Nano-indentation in Silico. *J. Am. Chem. Soc.* **2014**, *136*, 17036−17045.
(29) Kononova, O.; Snijder, J.; Brasch, M.; Cornelissen, J.; Dima, R. I.; Marx, K. A.; Wuite, G. J. L.; Roos, W. H.; Barsegov, V. Structural Transitions and Energy Landscape for Cowpea Chlorotic Mottle Virus Capsid Mechanics from Nanomanipulation in Vitro and in Silico. *Biophys. J.* **2013**, *105*, 1893−1903.
(30) Camacho, C.; Thirumalai, D. Modeling the role of disulfide bonds in protein folding: Entropic barriers and pathways. *Proteins: Struct., Funct., Genet.* **1995**, *22*, 27−40.
(31) Klimov, D.; Thirumalai, D. Multiple protein folding nuclei and the transition state ensemble in two state proteins. *Proteins: Struct., Funct., Genet.* **2001**, *43*, 465−475.
(32) Fernandez, A.; Colubri, A.; Berry, R. S. Three-body correlations in protein folding: the origin of cooperativity. *Phys. A* **2002**, *307*, 235−259.
(33) Marianayagam, N.; Jackson, S. The folding pathway of ubiquitin from all-atom molecular dynamics simulations. *Biophys. Chem.* **2004**, *111*, 159−171.
(34) 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.
(35) Alonso, D.; Daggett, V. Molecular dynamics simulations of hydrophobic collapse of ubiquitin. *Protein Sci.* **1998**, *7*, 860−874.
(36) Irback, A.; Mitternacht, S. Thermal versus mechanical unfolding of ubiquitin. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 759−766.
(37) Sorensen, J.; Head-Gordon, T. Toward minimalist models of larger proteins: A ubiquitin-like protein. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 368−379.
(38) Dastidar, S.; Mukhopadhyay, C. Unfolding dynamics of the protein ubiquitin: Insight from simulation. *Phys. Rev. E* **2005**, *72*, 051928.
(39) 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.
(40) Zhang, J.; Qin, M.; Wang, W. Multiple folding mechanisms of protein ubiquitin. *Proteins: Struct., Funct., Genet.* **2005**, *59*, 565−579.
(41) 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.
(42) Mandal, M.; Mukhopadhyay, C. Microsecond molecular dynamics simulation of guanidinium chloride induced unfolding of ubiquitin. *Phys. Chem. Chem. Phys.* **2014**, *16*, 21706−21716.
(43) Sorenson, J. M.; Head-Gordon, T. Toward Minimalist Models of Larger Proteins: a Ubiquitin-likeProtein. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 368−379.
(44) Wintrode, P.; Makhatadze, G.; Privalov, P. Thermodynamics of ubiquitin unfolding. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 246−253.
(45) 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. *Protein Sci.* **1999**, *8*, 361−369.
(46) Vijay-kumar, S.; Bugg, C.; Cook, W. Structure of ubiquitin refined at 1.8 A resolution. *J. Mol. Biol.* **1987**, *194*, 531−544.
(47) Fogolari, F.; Brigo, A.; Molinari, H. Protocol for MM/PBSA molecular dynamics simulations of proteins. *Biophys. J.* **2003**, *85*, 159−166.
(48) Kurnik, M.; Hedberg, L.; Danielsson, J.; Oliveberg, M. Folding without charges. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 5705−5710.
(49) Cardenas, A.; Elber, R. Kinetics of cytochrome C folding: Atomically detailed simulations. *Proteins: Struct., Funct., Genet.* **2003**, *51*, 245−257.
(50) Li, L.; Mirny, L.; Shakhnovich, E. Kinetics, thermodynamics and evolution of non-native interactions in a protein folding nucleus. *Nat. Struct. Biol.* **2000**, *7*, 336−342.
(51) 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.
(52) Veitshans, T.; Klimov, D.; Thirumalai, D. Protein folding kinetics: Timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Folding Des.* **1997**, *2*, 1−22.
(53) Ermak, D. L.; Mccammon, J. A. Brownian Dynamics with Hydrodynamic Interactions. *J. Chem. Phys.* **1978**, *69*, 1352−1360.
(54) Guo, Z.; Thirumalai, D. Kinetics and Thermodynamics of Folding of a de novo Designed four Helix Bundle. *J. Mol. Biol.* **1996**, *263*, 323−343.
(55) Zhou, T.; Caflisch, A. Distribution of reciprocal of interatomic distances: A fast structural metric. *J. Chem. Theory Comput.* **2012**, *8*, 2930−2937.
(56) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* **2007**, *23*, 2625−2627.
(57) Späth, H. *Cluster Analysis Algorithms for Data Reduction and Classification of Objects*; Ellis Horwood Series in Computers and Their Applications; Ellis Horwood: Chichester, U.K., 1980.
(58) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(59) Ibarra-Molero, B.; Loladze, V.; Makhatadze, G.; Sanchez-Ruiz, J. Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry* **1999**, *38*, 8138−8149.
(60) Zhou, Y.; Hall, C. K.; Karplus, M. The calorimetric criterion for a two-state process revisited. *Protein Sci.* **1999**, *8*, 1064−1074.
(61) Huang, J.-r.; Gabel, F.; Jensen, M. R.; Grzesiek, S.; Blackledge, M. Sequence-Specific Mapping of the Interaction between Urea and

DOI: 10.1021/acs.jpcb.5b03471
J. Phys. Chem. B 2015, 119, 11358−11370

Unfolded Ubiquitin from Ensemble Analysis of NMR and Small Angle Scattering Data. *J. Am. Chem. Soc.* **2012**, *134*, 4429−4436.
(62) 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.
(63) 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.
(64) 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.
(65) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98−105.
(66) Skinner, J. J.; Yu, W.; Gichana, E. K.; Baxa, M. C.; Hinshaw, J. R.; Freed, K. F.; Sosnick, T. R. Benchmarking all-atom simulations using hydrogen exchange. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15975−15980.
(67) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113−5124.
(68) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.
(69) Camacho, C. J.; Thirumalai, D. Minimum energy compact structures of random sequences of heteropolymers. *Phys. Rev. Lett.* **1993**, *71*, 2505−2508.
(70) Garcia-Manyes, S.; Dougan, L.; Badilla, C. L.; Brujic, J.; Fernandez, J. M. Direct observation of an ensemble of stable collapsed states in the mechanical folding of ubiquitin. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 10534−10539.
(71) 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.
(72) 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.
(73) Guo, Z. Y.; Thirumalai, D. Kinetics of Protein-Folding - Nucleation Mechanism, Time Scales, and Pathways. *Biopolymers* **1995**, *36*, 83−102.
(74) Peng, Q.; Li, H. Atomic force microscopy reveals parallel mechanical unfolding pathways of T4 lysozyme: Evidence for a kinetic partitioning mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 1885−1890.
(75) Stigler, J.; Ziegler, F.; Gieseke, A.; Gebhardt, J. C. M.; Rief, M. The Complex Folding Network of Single Calmodulin Molecules. *Science* **2011**, *334*, 512−516.
(76) Karplus, M.; Weaver, D. Diffusion-Collision model for protein folding. *Biopolymers* **1979**, *18*, 1421−1437.
(77) Vallee-Belisle, A.; Michnick, S. W. Visualizing transient protein-folding intermediates by tryptophan-scanning mutagenesis. *Nat. Struct. Mol. Biol.* **2012**, *19*, 731.
(78) 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.
(79) 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.
(80) 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.
(81) 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.
(82) 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.
(83) Went, H.; Benitez-Cardoza, C.; Jackson, S. Is an intermediate state populated on the folding pathway of ubiquitin? *FEBS Lett.* **2004**, *567*, 333−338.
(84) 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.
(85) 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.
(86) Cordier, F.; Grzesiek, S. Temperature-dependence properties as studied by of protein hydrogen bond high-resolution NMR. *J. Mol. Biol.* **2002**, *317*, 739−752.
(87) Khorasanizadeh, S.; Peters, I.; Butt, T.; Roder, H. Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* **1993**, *32*, 7054−7063.
(88) 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.
(89) 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.
(90) Zheng, Z.; Sosnick, T. R. Protein Vivisection Reveals Elusive Intermediates in Folding. *J. Mol. Biol.* **2010**, *397*, 777−788.
(91) Jennings, P. A.; Wright, P. E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **1993**, *262*, 892−896.
(92) Hughson, F. M.; Wright, P. E.; Baldwin, R. L. Structural characterization of a partly folded apomyoglobin intermediate. *Science* **1990**, *249*, 1544−1548.
(93) Du, R.; Pande, V.; Grosberg, A.; Tanaka, T.; Shakhnovich, E. On the transition coordinate for protein folding. *J. Chem. Phys.* **1998**, *108*, 334−350.
(94) Paci, E.; Lindorff-Larsen, K.; Dobson, C.; Karplus, M.; Vendruscolo, M. Transition state contact orders correlate with protein folding rates. *J. Mol. Biol.* **2005**, *352*, 495−500.
(95) Went, H.; Jackson, S. Ubiquitin folds through a highly polarized transition state. *Protein Eng., Des. Sel.* **2005**, *18*, 229−237.
(96) Sosnick, T.; Dothager, R.; Krantz, B. Differences in the folding transition state of ubiquitin indicated by phi and psi analyses. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 17377−17382.
(97) Cho, J.; Raleigh, D. Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins. *J. Mol. Biol.* **2005**, *353*, 174−185.
(98) Chen, T.; Song, J.; Chan, H. S. Theoretical perspectives on nonnative interactions and intrinsic disorder in protein folding and binding. *Curr. Opin. Struct. Biol.* **2015**, *30*, 32−42.
(99) Zarrine-Afsar, A.; Zhang, Z.; Schweiker, K. L.; Makhatadze, G. I.; Davidson, A. R.; Chan, H. S. Kinetic consequences of native state optimization of surface-exposed electrostatic interactions in the Fyn SH3 domain. *Proteins: Struct., Funct., Genet.* **2012**, *80*, 858−870.
(100) Azia, A.; Levy, Y. Nonnative Electrostatic Interactions Can Modulate Protein Folding: Molecular Dynamics with a Grain of Salt. *J. Mol. Biol.* **2009**, *393*, 527−542.
(101) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. Impact of Local and Non-local Interactions on Thermodynamics and Kinetics of Protein Folding. *J. Mol. Biol.* **1995**, *252*, 460−471.
(102) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing. *J. Phys. Chem. B* **2012**, *116*, 8494−8503.

DOI: 10.1021/acs.jpcb.5b03471
J. Phys. Chem. B 2015, 119, 11358−11370