Unfolding bovine α-lactalbumin with T-jump: Characterizing disordered intermediates via time-resolved x-ray solution scattering and molecular dynamics simulations ⓕ

Darren J. Hsu,[1] ⓘD   Denis Leshchev,[1] ⓘD   Irina Kosheleva,[2] ⓘD   Kevin L. Kohlstedt,[1,a)] ⓘD   and Lin X. Chen[1,3,a)] ⓘD

AFFILIATIONS

[1] Department of Chemistry, Northwestern University, Evanston, Illinois 60208, USA
[2] Center for Advanced Radiation Sources, The University of Chicago, Chicago, Illinois 60637, USA
[3] Chemical Sciences and Engineering Division, Argonne National Laboratory, Argonne, Illinois 60439, USA

Note: This paper is part of the JCP Special Collection in Honor of Women in Chemical Physics and Physical Chemistry.
[a)] Authors to whom correspondence should be addressed: kkohlstedt@northwestern.edu and l-chen@northwestern.edu

ABSTRACT

The protein folding process often proceeds through partially folded transient states. Therefore, a structural understanding of these disordered states is crucial for developing mechanistic models of the folding process. Characterization of unfolded states remains challenging due to their disordered nature, and incorporating multiple methods is necessary. Combining the time-resolved x-ray solution scattering (TRXSS) signal with molecular dynamics (MD), we are able to characterize transient partially folded states of bovine α-lactalbumin, a model system widely used for investigation of molten globule states, during its unfolding triggered by a temperature jump. We track the unfolding process between 20 μs and 70 ms and demonstrate that it passes through three distinct kinetic states. The scattering signals associated with these transient species are then analyzed with TRXSS constrained MD simulations to produce protein structures that are compatible with the input signals. Without utilizing any experimentally extracted kinetic information, the constrained MD simulation successfully drove the protein to an intermediate molten globule state; signals for two later disordered states are refined to terminal unfolded states. From our examination of the structural characteristics of these disordered states, we discuss the implications disordered states have on the folding process, especially on the folding pathway. Finally, we discuss the potential applications and limitations of this method.

Published under license by AIP Publishing. https://doi.org/10.1063/5.0039194

INTRODUCTION

Knowledge of disordered structures in proteins sheds light on the conformational space the protein explores during folding. The insights gained from such studies lead to the identification of structural features preserved outside of non-native conformations, such as the most stable components of the protein and, conversely, those prone to disordering.[1,2] New efforts to link disordered states in terms of conserved structural parameters can lead to folding path(s) being described in atomistic detail, providing a more complete picture of the structural dynamics, which aids in diagnosis of diseases and drug design processes.[3–5]

Recently, there has been a major development of structural analysis methods of disordered states, largely stemming from the research on intrinsically disordered proteins.[6] From the experimental side, a multitude of spectroscopic methods have been applied to these systems, such as nuclear magnetic resonance (NMR) and mass spectrometry.[7] On the other hand, theoretical methods such as molecular dynamics (MD) simulations[8,9] and Bayesian inferential structural determination[10] have been able to determine

disordered protein structures. However, challenges remain as protein structures can be diverse during their functions and along their folding/unfolding pathways, which necessitates the incorporation of multiple structural analysis methods to characterize these conformations. Furthermore, some of the disordered conformations are only transiently populated during the folding process, often on the timescale of microseconds.[11–15] Therefore, methods that can capture and kinetically distinguish transient conformations and provide structural information of the protein are needed to investigate these elusive intermediates.

Several experimental methods have been developed and utilized to tackle such a challenge, each with their own advantages. Förster resonance energy transfer can be applied to probe distances of interest between tags installed on the protein system.[16] Environment-sensitive fluorescence from intrinsic tryptophan residue in proteins interrogates the folding pathways.[17] Time-resolved circular dichroism (CD) can be employed to probe secondary and tertiary structures during reactions.[18,19] Hydrogen–deuterium exchange-mass spectrometry is used to identify residues being protected from the solvent during the folding process.[20] Finally, two-dimensional IR spectroscopy tracks the amide vibrations to interpret the folding environment.[21] However, since these methods provide only partial indicators of the protein unfolding process, a method that observes the entire protein is needed to complement these observations.

Time-resolved X-ray solution scattering (TRXSS) is a powerful method to probe the structural evolution of proteins, where difference scattering patterns encode protein secondary and tertiary structural changes as a function of delay time after a reaction trigger that synchronizes protein actions in the ensemble. With the high time-resolution given by pulsed x-rays from synchrotrons (100 ps) and x-ray free electron lasers (femtoseconds), conformational states with distinct kinetics can be separated, and their structures can be characterized.[22–26] TRXSS can be coupled to different triggers to initiate protein structural changes, including optical excitation, ligand dissociation, electron transfer, temperature-jump (T-jump), pH-jump, and reactant concentration-jump.[25–33] In particular, a T-jump that initiates solvent heating and subsequent protein heating using a ns infrared (IR) pulse can be applied to most temperature-sensitive proteins to perturb structural dynamics and to obtain the kinetics of protein structural changes.

As TRXSS experiments on proteins in solution inherently lead to a loss in spatial resolution compared to crystallography due to the orientational averaging of protein structures resulting in one-dimensional data, the scattering intensity I vs momentum transfer q,[34,35] complementary investigations about the protein system should be incorporated to gain atomistic understanding of the processes. MD simulation is one such method that provides detailed descriptions of the protein structure. However, independently TRXSS and MD studies may not give suitable agreement of protein structures, especially when involving disordered states.[36] By coupling experimental data into simulations as constraints, the predicted structures from MD are bounded by the scattering profile of states probed by TRXSS.[37–39] Recently, we have developed a methodology to transform a TRXSS difference signal into an external steering potential while considering the hydration layer effect on the scattering patterns, which is necessary when the (un)folding process involves large changes in the solvent accessible surface area of the

protein system.[40] With this protocol, it is now possible to sample structures that are balanced between a given MD force field and those that are consistent with the scattering data. In addition, the method can be augmented by enhanced sampling methods such as well-tempered metadynamics[41,42] or replica-exchange methods,[43,44] which are necessary for the exploration of the wider conformational space associated with a disordered state.

To demonstrate this combined approach of TRXSS and MD simulation for the structural analyses of transient conformationally disordered states, we selected a model protein system, calcium-bound bovine α-lactalbumin (BLA), as a test case. BLA has been extensively studied especially for its molten globule states[45–49] and therefore richly described with many structural parameters that we can compare against. The molten globule states of BLA are generally described as being compact in size with a radius of gyration (R_g) 10% larger than that of the native state, having a substantial amount of secondary structures, increased backbone fluctuation, and poor side-chain packing.[19,50,51] However, a detailed description of the molten globule structural states is still missing. Specifically, a structural understanding of which α-helices are disordered and the extent of backbone fluctuation remains unanswered.

In this work, we describe a structural analysis of the BLA kinetic unfolded states populated by applying a short T-jump pulse while tracking the intermediate states with TRXSS. From the TRXSS analysis, we investigate the two-state equilibrium unfolding transition and parallel unfolding pathways with the kinetic intermediates after the T-jump. The difference signal associated with each kinetic state on the unfolding pathways is incorporated into our MD simulations. We then present the sampled structures and examine their connection to the molten globule state. Finally, we discuss the applicability and limitations of this combined method and connect it to a broader structural analysis pipeline.

MATERIALS AND METHODS

Experimental details

The BLA T-jump induced unfolding experiment was carried out at the BioCARS 14-ID-B beamline at the Advanced Photon Source (APS) at Argonne National Laboratory.[52] All chemicals were purchased from Sigma-Aldrich and used without further purification. BLA (Sigma-Aldrich L5385) was dissolved at 5 mg/ml in 10 mM tris(hydroxymethyl)aminomethane (TRIS) buffer at pH 7.5 with 2 mM calcium chloride added to ensure complete calcium binding of the protein, which stabilizes the folded state.[47] The solution was filtered with a 0.2 μm syringe filter before use. The sample was withdrawn with a syringe pump into a capillary flow cell, which is mounted on a custom-made, temperature-controlled aluminum nitride (AlN) flow cell holder, as described in detail elsewhere.[29] The holder can be heated with a Peltier heating unit underneath the AlN piece. The T-jump was triggered by a 7-ns (FWHM) IR laser pulse at 1.443 μm, which excited an overtone of O–H stretching mode of water molecules. The protein and solvent structural changes of the sample solution were followed by attenuated polychromatic x-ray pulses with photon energy peaked at 12 keV and 22 μs in length at variable time delays relative to the laser pulse controlled

electronically. The scattering signal was registered on a Rayonix MX340-HS detector, giving a momentum transfer q range from 0.015 Å^{-1} to 2.5 Å^{-1}. The experiments were conducted at three initial temperatures, 60 °C, 65 °C, and 70 °C. Radiation damage was monitored by watching the intensity of the small-angle region; the intensity rises compared to the reference scan when the irradiated protein begins to form aggregates. The flow cell was then cleaned promptly. A detailed description of the experimental setup can be found in previous publications[29,52,53] and the supplementary material.

The TRXSS raw data were subject to masking, polarization correction, azimuthal integration, and outlier rejection to generate a one-dimensional absolute scattering curve for each time delay. The TRXSS difference signal was obtained by subtracting the signal from a positive time delay (x-ray pulses arrives after the laser pulse) with the signal from a negative time delay (ground state where the laser pump arrives after x-ray probing). This difference signal contains the contribution from both protein structural change and solvent heating of the hydration layer. The latter is further subtracted out using a pure solvent heating signal measured on the buffer solution with the same T-jump setup.[29,54] The remaining protein-related signal is used for further kinetic analysis.

Steady-state x-ray solution scattering (XSS) profiles were measured with the same setup at different temperatures from 50 °C to 80 °C to obtain an unfolding curve of BLA. The same was done for pure buffer to provide temperature calibration and enable an estimation of the magnitude of the T-jump. The magnitude of the T-jump was estimated from these data to be 11.5 °C (see the supplementary material).

To convert experimental species-associated difference signals to the input data for the steering MD simulation, three scaling factors were considered. First, the fraction of unfolded protein at equilibrium was calculated according to the two-state sigmoidal unfolding curve and the estimated magnitude of the T-jump (see the section titled Results and Discussion and the supplementary material). Next, since the T-jump will dissipate in several milliseconds, the system may not achieve the equilibrium unfolding fraction. To quantify the degree of unfolding, we scale the experimental difference signals at 5 ms (the largest signal in magnitude) to fit the steady-state difference based on T-jump and obtain the achieved unfolding fraction. Finally, we use the overall population in global analysis (GA) as a normalization factor to scale the curves. The species-associated difference signals were then scaled by these factors to match the scale of steady-state scattering and then input as the signal to the MD simulations. For a detailed procedure, see the supplementary material.

Computational details

X-ray scattering-derived potentials

The harmonic x-ray scattering-derived potential (XSP) that is used to steer the simulation can be derived following our previously published method.[40] In short, the XSP (U^{scat}) has the following form:

$$U^{scat} = \left(\frac{k_\chi}{2}\right)\chi^2 = \sum_q \frac{k_\chi}{2\sigma_q^2}\left\{\Delta S_{exp}(q) - \alpha\left[S_{calc}(q) - S_{ref}(q)\right]\right\}^2, \quad (1)$$

where k_χ is the force constant, σ_q is a weighing factor, ΔS_{exp} is a scaled target difference pattern, α is the excited state fraction that is set to 1, S_{calc} is the scattering pattern given a set of coordinates in a frame, and S_{ref} is the reference (ground state) scattering pattern. ΔS_{exp} is derived by subtracting both the ground state scattering pattern and the solvent response from the excited state pattern. In this work, α is preset to 1 because ΔS_{exp} has been scaled by the degree of unfolding (see the supplementary material), and σ_q corresponds to the standard deviation for each q point, which is also scaled.

Following the Debye formula,[55] S_{calc}, given a set of coordinates $\mathbf{r}$, can be calculated as

$$S_{calc}(q, \mathbf{r}) = \sum_k \sum_j f_k(q, \mathbf{r})f_j(q, \mathbf{r})\frac{\sin(q\,r_{kj})}{q\,r_{kj}}, \quad (2)$$

where $S_{calc}(q, \mathbf{r})$ is the scattering intensity, j and k are atom indices, $f(q, \mathbf{r})$ are atomic form factors that depend on exposure to solvent, and r_{kj} is the distance between atoms j and k.

For the form factors, we used the following formula:[56,57]

$$f_i(q, \mathbf{r}) = f_i^v(q) - C_1(q)f_i^s(q) + c_2\,s_i(\mathbf{r})f^w(q), \quad (3)$$

where $f_i^v(q)$ is the Waasmaier–Kirfel atomic form factor *in vacuo* for atom i,[58] $C_1(q)f_i^s(q)$ is the excluded volume factor with

$$C_1(q) = c_1^3 \exp\left(-\frac{\left(\frac{4\pi}{3}\right)^{2/3}q^2 r_m^2(c_1^2 - 1)}{4\pi}\right), \quad (4)$$

where c_1 is an adjustable parameter accounting for solvent volume exclusion, r_m is the average atomic radius set at 1.62 Å, and c_2 is the solvation shell parameter that alters the contribution of the hydration shell effect term in Eq. (3). In that term, s_i is the fraction of solvent accessible surface area (SASA) per protein atom and $f^w(q)$ is the form factor for water. We refer interested readers to Ref. 40 for treatment of the form factors.

To calculate forces $\mathbf{F}_k^{scat}$ acting on the atoms, $-\nabla U^{scat}$ is taken with some simplification,

$$\mathbf{F}_k^{scat} = 2k_\chi\alpha\sum_q \frac{1}{\sigma_q^2}\left\{\Delta S_{exp}(q) - \alpha\left[S_{calc}(q) - S_{ref}(q)\right]\right\}$$
$$\times \sum_j\left\{f_k(q)f_j(q)\left[\cos(qr_{kj}) - \frac{\sin(qr_{kj})}{qr_{kj}}\right]\right\}\frac{\mathbf{r}_{kj}}{r_{kj}^2}. \quad (5)$$

Finally, to reduce the fluctuation due to S_{calc} in $\mathbf{F}_k^{scat}$, an exponential averaging method is applied,[38]

$$S_{calc}^{EA}(t) = N_t^{-1}\sum_{k=0}^{n} \exp(-k\delta t/\tau)S_{calc}(t - k\delta t), \quad (6)$$

where n is the number of snapshots, spaced by updated interval δt, N_t is the resulting normalization constant $\sum_{k=0}^{n} \exp(-k\delta t/\tau)$, and τ is the characteristic memory length. In this work, δt was set at 50 steps (100 fs) and τ was set at 5000 steps (10 ps). Between

$\tau < t < 2\tau$, $\mathbf{F}_k^{scat}$ in Eq. (5) is scaled by $[1 - \cos(\pi(t - \tau)/\tau)]/2$, which transitions from 0 to 1 over the course of τ to gradually apply S_{calc} in Eq. (5) to the force. Interested readers are referred to the source code and manual, which are publicly available on GitHub (github.com/darrenjhsu/XSNAMD).

Molecular dynamics simulations

The MD simulations were carried out using NAMD 2.11 including the external force module described above.[59] For all simulations, the time step was set to 2 fs, while the particle-mesh Ewald method was used to calculate the full electrostatic interactions.[60] The NPT ensemble was maintained using Langevin dynamics and a Nosé–Hoover barostat (P = 1 atm) with a decay time constant of τ = 100 fs. Bonds between hydrogens and heavy atoms were constrained with the SHAKE algorithm.[61]

The crystal structure of calcium-bound BLA (PDB entry: 1F6S[62]) was protonated with the CHARMM36[63] force field and solvated with sTIP3P water.[64] The four disulfide bonds are created using the patch included in the force field.[63] The periodic cubic water box was padded by 20 Å on all dimensions around the protein, resulting in about 14 000 water molecules. Calcium and chloride ions were added to give a 2 mM calcium concentration while maintaining electroneutrality.

The energy of the simulated system was minimized for 5000 steps and equilibrated at 56 °C for 20 ns. The snapshots of BLA structures in last 5 ns of the trajectory were taken as the equilibrium ensemble. The average scattering pattern from this ensemble was matched against experimental steady-state XSS S_{exp} at 60 °C by minimizing the objective function,

$$\chi^2(\kappa, c_1, c_2) = \frac{1}{N_q} \sum_{i=1}^{N_q} \left(\frac{\kappa S_{exp}(q_i) - S_{calc}(q_i, c_1, c_2, \rho)}{\kappa \sigma_{exp}(q_i)} \right)^2, \quad (7)$$

fitted against κ, c_1, and c_2. In the above expression, N_q is the number of q points, κ is a scaling factor, and σ_{exp} is the experimental standard deviation of data. ΔS_{exp} is then scaled by κ as well. The solvent density, ρ, was set to the water density at 56 °C (see the supplementary material). The difference between two temperatures (56 °C in MD and 60 °C in experiment) is calculated using the difference in the melting temperatures, defined as the temperature where half of the protein population is unfolded, in this work and the literature (see the section titled Results and Discussion).[47] We used the room temperature (20 °C) and the melting temperatures (74 °C vs 69 °C) to create a reference line to correct our experimentally set temperatures to that of the simulations. The XSP [Eq. (1)] based on each of the difference scattering patterns between the folded and each unfolded state (see the section titled Results and Discussion) with an initial energy of about 100 kcal/mol was then enabled for 10 ns to prepare the system in a local minimum on the potential energy surface, as determined by the fluctuation of χ^2 values (see the supplementary material). For this simulation, ρ was adjusted to the water density at 67 °C.

From this point, for each of the three intermediate states, **I**, **U₁**, and **U₂**, as determined from GA and described below in the section titled Results and Discussion, a well-tempered metadynamics (WTmetaD) simulation was launched with the TRXSS signal associated with that state as the restraint to sample the relevant structures. WTmetaD applies a history-dependent potential to the simulation system, effectively discouraging the system from visiting the same location in the conformational space constructed by the selected collective variables (CV) and encouraging exploration of new conformations. The "well-tempered" part of this method refers to the down-scaling of added potential based on existing energy deposition at the same location of the conformational space, which improves convergence.[41,42] The XSP level is increased by increasing k_χ until the calculated difference signal achieves reasonable agreement with the input signal. Among the three kinetically distinct states, the initial XSP is set at 200 kcal/mol for **U₁** and **U₂** states but is raised to 3600 kcal/mol for the **I** state as lower XSP values do not produce a better fit to the data. In general, the fit to data as a function of initial XSP roughly resembles steps. The two collective variables for the WTmetaD calculations are the α-helical content score and R_g with a range from 0 to 1 in steps of 0.01 for the former and 13 Å–35 Å in 0.1 Å bins for the latter.[65] The α-helical content scoring function looks at angles between C_α atoms in neighboring residues and contacts of O and N atoms four residues away. The choice of collective variables is guided by observing the rather coil-rich region around residues 70 in the BLA crystal structure, which leads to the presumption that R_g and α-helical content should have low correlation. The WTmetaD was run with an initial energy deposition rate of 0.2 kcal mol^{-1} ps^{-1} and a bias temperature ΔT = 3600 K. The convergence was determined by the decay in mean deposited energy in 2 ns windows during simulations (see the supplementary material). Based on the convergence curves, the simulation for the **I** state is terminated at 100 ns and those for **U₁** and **U₂** states at 250 ns.

RESULTS AND DISCUSSION

Steady-state structural analysis

X-ray solution scattering

We begin with the examination of the steady-state BLA structure over the temperature range of 50 °C–80 °C. The steady-state XSS patterns of BLA at different temperatures are shown in Fig. 1. As the temperature increases, the scattering intensity below 0.05 Å^{-1} increases, while that in the higher q region decreases. The temperature-dependent scattering patterns can be modeled as a two-state transition, where the population vs temperature can be fitted with a sigmoid function (see the inset of Fig. 1 and the supplementary material). From the resulting fully folded and unfolded state scattering patterns, we obtained an initial R_g of 15.41 ± 0.06 Å for the folded state, 20.32 ± 0.02 Å for the unfolded state, and a melting temperature of 74 ± 2 °C. The initial R_g matches well with earlier SAXS studies.[50,66] Our derived melting temperature is 5 °C higher than the reported 69 °C from an earlier work,[47] most likely due to imperfect thermal contact between the flow cell and the capillary. This deviation is corrected in the MD simulations, as detailed in the section titled Materials and Methods.

Equilibrium MD

In addition to examining the steady-state XSS profile of BLA, we also calculate the equilibrium MD structures. Equilibrium MD
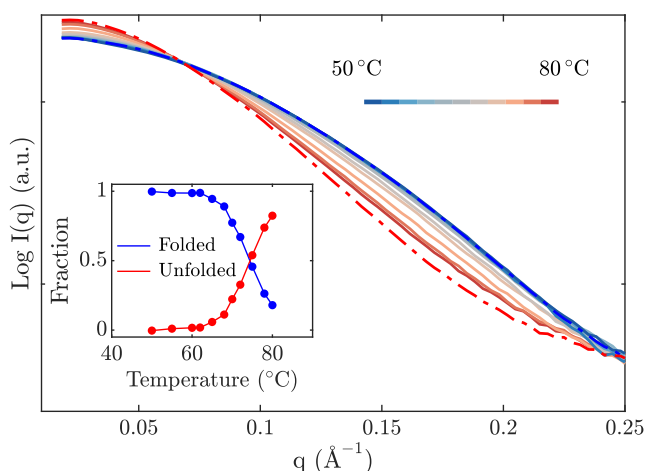
FIG. 1. Steady-state XSS pattern $I(q)$ of BLA for a series of temperature scans between 50 °C and 80 °C. The scattering patterns for fully folded and unfolded states are shown in blue and red dashed-dotted lines, respectively. (Inset) The fraction of folded (blue circles) and unfolded BLA (red circles) using a two-state transition model for the fit.

simulations were carried out for 20 ns at 56 °C during which the protein stayed in a stable conformation with R_g = 13.9 ± 0.1 Å and an α-helix score of 0.508 ± 0.006. The R_g value calculated here is slightly smaller than that determined by the steady-state XSS experiment because only the protein atoms are included in the calculation, whereas the hydration shell also contributes to the R_g extracted from the XSS analysis.

We calculated the average secondary structure content for each state as a function of residue indices in equilibrium MD using the `timeline` code in VMD (Visual Molecular Dynamics) [Fig. 2(a)].[67] For a visual reference, the secondary structures are shown color-coded using the crystal structure [Fig. 2(b)]. In the equilibrium ensemble, there are four well-defined α-helices, residues 5–12 (αI), 23–34 (αII), 85–100 (αIII), and 102–107 (αIV), as well as one β-sheet, residues 41–56 (βI) (see Fig. 1).

The extent of residue solvent exposure is shown in Fig. 2(c) and is again mapped onto the crystal structure in Fig. 2(d). Except for αIV, all other secondary structures are buried in the core of the protein. The location of these buried residues is consistent with the protected residues shown in a previous NMR study.[19]
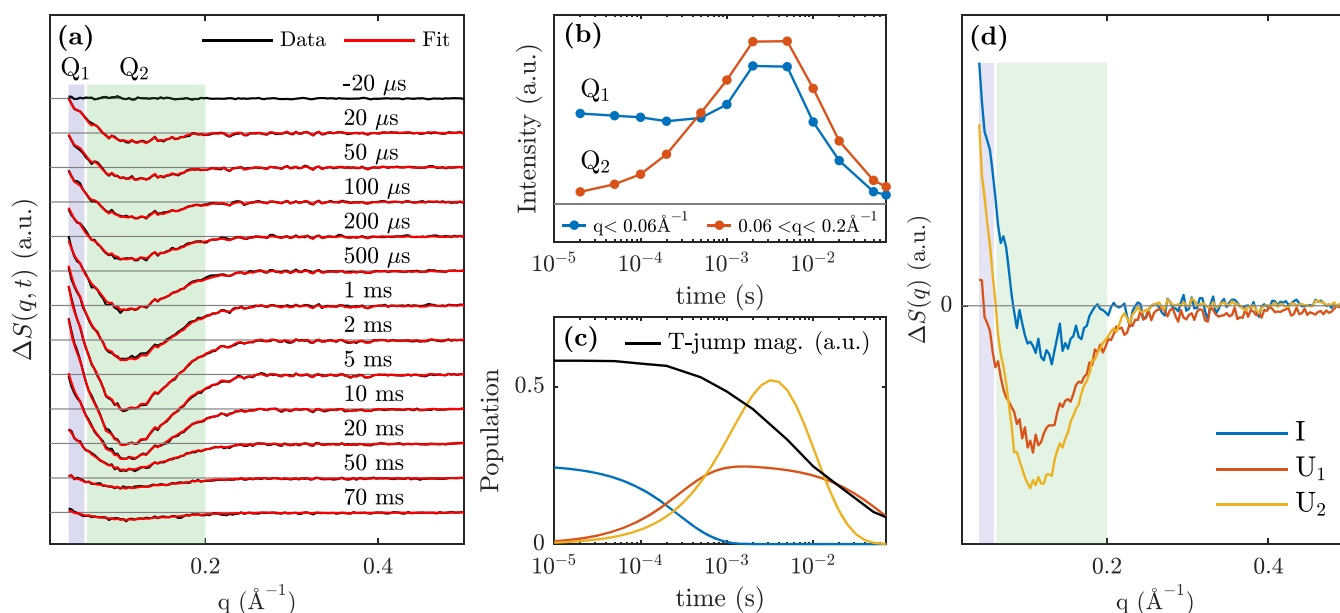
Time-resolved structural analysis

T-jump TRXSS

We now consider the structural response of BLA upon a T-jump perturbation using TRXSS. The structural responses of BLA after a 11.5 °C T-jump from initial temperatures of 60 °C, 65 °C, and 70 °C were individually collected from 20 μs to 70 ms delay following the T-jump. The T-jump magnitude is estimated by comparing the time-resolved solvent response signal between 2.2 Å^{-1} < q < 2.5 Å^{-1} to steady-state XSS difference signals of pure buffer at different temperatures relative to the signal at the initial temperatures (see the supplementary material). The compiled experimental

results done at an initial temperature of 60 °C are shown in Fig. 3, and the results at 65 °C and 70 °C are qualitatively similar and can be found in the supplementary material. The difference signal $\Delta S(q, t)$ at select time delays with respect to the T-jump laser pump is shown in Fig. 3(a). The ground state patterns are collected with a negative time delay (-20 μs), where the x-ray probe arrives before the T-jump laser pump. The solvent response is extracted from scattering signals from the pure buffer after applying the T-jump laser pulse (see the supplementary material and Figs. S2–S4 for details on data processing). Two characteristic q ranges, $q < 0.06$ Å^{-1} (Q_1) and 0.06 Å^{-1} < q < 0.2 Å^{-1} (Q_2), are shown in blue and green shades in Fig. 3(a), respectively. The integrated magnitude of ΔS in these two regions is shown in Fig. 3(b). At the earliest time delay (20 μs), there is already a significant amount of positive signal in the Q_1 region, which grows only slightly toward later time delays. In contrast, the
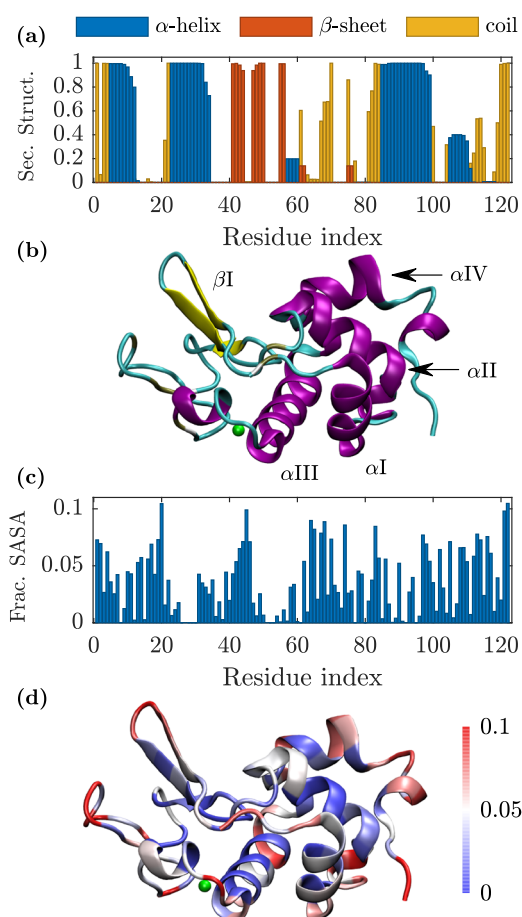
FIG. 2. Equilibrium ensemble properties of BLA. (a) The secondary structure content across each residue. (b) The BLA crystal structure labeled with secondary structure notations used in the text. (c) The fraction of the solvent accessible surface area (SASA) as a function of residue index. (d) The fraction of SASA mapped onto the crystal structure backbone. Residues more exposed to solvent are colored in red, while those buried are colored in blue. In (b) and (d), a green dot denotes the bound calcium ion.

FIG. 3. Time-resolved structural analyses of BLA. (a) Difference signals $\Delta S(q)$ (black) at representative time delays after T-jump and the GA fits to them (red). The two characteristic q ranges, Q_1 and Q_2, defined in the text, are shaded in blue and green [also shaded in (d)]. (b) Integration of (absolute) difference signals as a function of time in Q_1 and Q_2 to show the different time evolution in these regions. (c) Population of the three intermediates (I: blue; U_1: red; and U_2: yellow) derived from the kinetic traces using GA. The T-jump magnitude is included in black. (d) ΔS for each intermediate derived from GA. Legends and colors are shared between (c) and (d).

signal in Q_2 is small initially but becomes significantly larger in magnitude at longer time delays. If there were only one intermediate state, the difference signals in these regions would have been proportional. Therefore, these findings indicate that several intermediates contribute to the signal.

To reveal the intermediate states in terms of their difference scattering patterns, GA was utilized based on the information from singular value decomposition (SVD) analysis of the difference data up to $q < 0.8$ Å^{-1}, which returned three significant components judging by the lag 1 autocorrelation values, indicating the same number of intermediates. In short, GA decomposes the data into a set of time-independent species-associated difference scattering curves that evolve only in population following the supplied physical kinetic model. We tried multiple kinetic models that can be constructed with the three states, including a sequential model and two parallel pathway models, and judged the results based on both goodness of fit and the resulting species-associated difference scattering patterns (see the supplementary material). The best fitting kinetic model contains three structures in two parallel processes: (1) a three-state sequential pathway (folded → I → U_1) and (2) a two-state transition pathway (folded → U_2). The selection of this model is largely guided by the slight difference in the signal growth rate of Q_1 and Q_2 between 100 μs and 1 ms [Fig. 3(b)], which indicates that two states are populated on that timescale with different kinetics. The resulting species-associated kinetic traces for the 60 °C dataset are shown in Fig. 3(c), and their respective difference scattering patterns are shown in Fig. 3(d). The two-state transition has a rising time of 2.2 ms and a decay time of 40 ms, while the three-state transition involves a rapidly (<20 μs) populated I state that

transitions with a lifetime of 920 μs to the U_2 state, decaying with a time constant of 8.6 ms. In both pathways, the decay time is coupled to the solvent cooling after a T-jump, typically starting after about a few ms, making the interpretation difficult.[29,30,68] While the signals share the positive feature in the Q_1 region and the dip between Q_2, the magnitudes of these features in each state are different, therefore creating three distinct species-associated difference signals. The other datasets with the initial temperature at 65 °C and 70 °C were fitted with the same kinetic model and returned similar results. These fits can be found in the supplementary material.

XS–well-tempered metadynamics (WTmetaD)

We now calculate the XSP based on the TRXSS difference signals [Eq. (1)] and utilize it in WTmetaD simulations. For each of the three kinetic states, a WTmetaD simulation combined with an XSP was run. All simulations reached convergence at the end as shown by the decay in the mean magnitude of Gaussian energy depositions over simulation time (see the supplementary material). The converged free energy surfaces (FESs) are depicted in Fig. 4. The FES for the I state (green) has a deep valley at around R_g = 16 Å, with the most populated conformational space at R_g = 16.1 Å and α-helix score = 0.40. The U_1 state (blue) exhibits a much broader distribution of R_g and α-helix score, with a minimum at R_g = 22.0 Å and α-helix score = 0.33. Finally, the FES of the U_2 state shows two local minima, one at R_g = 19.7 Å and α-helix score = 0.37 and the other at R_g = 21.5 Å and α-helix score = 0.32.

The FESs clearly show that the I state is an intermediate, while the U_1 and U_2 states are terminal unfolded states with higher R_g
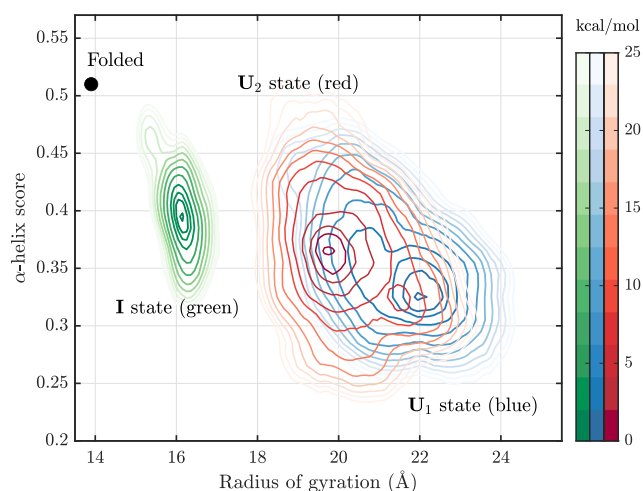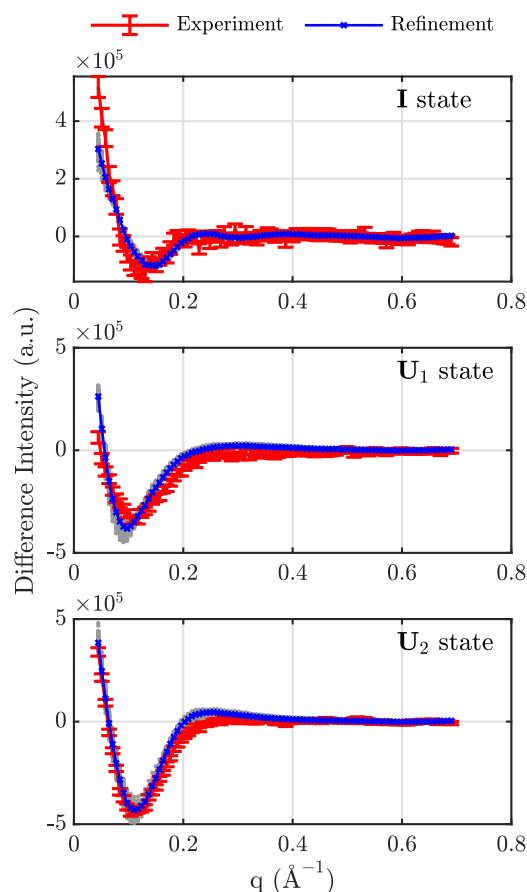
FIG. 4. FESs for each intermediate **I** (green), **U₁** (blue), and **U₂** (red) of the well-tempered metadynamics with XSP enabled. Energy minimum of each FES is set to zero. For comparison, the folded state sampled in equilibrium run is also included (black dot).

FIG. 5. Signal refinement for all three protein disordered states. Experiment difference signals (red) are shown with error bars. The thermal averages of computed difference scattering signal $\langle \Delta S(q) \rangle$ (blue) are from the individual scattering patterns (gray dots), which show a high level of self-consistency.

values and lower α-helix scores. These results are consistent with the kinetic model applied to the signal decomposition in the TRXSS experiments even though the MD simulation did not *a priori* utilize the kinetic model in the experimental data analyses (see the above results). The R_g values for the two terminal unfolded states, determined using atomic coordinates, also match closely with the unfolded R_g obtained from steady-state XSS experiments. These findings provide reassurance that the WTmetaD is indeed sampling relevant conformations.

To assess the agreement between the refinement results and the input difference signals, we calculated the thermal average of the simulated difference curves $\langle \Delta S(q) \rangle$ for each of the three disordered states, as shown in Fig. 5 along with the input signals. For all three states, the refined signals (blue) do not completely fit the experimental signals (red). In the **I** state, the main deviation is at the extreme low-q region, where the refined difference signal is lower than the input difference signal. In addition, the main dip in the refined signal shifts to a slightly higher q value. For the **U₁** state, the refined signal is higher than the input at the extreme low q, the minimum of the dip is at a lower q, and the refined signal recovers to zero faster. Finally, the refined signal of the **U₂** state shows some deviation around $q = 0.2$ Å⁻¹. The refined signals, however, match better with the derived species-associated difference patterns using the 65 °C and 70 °C datasets, which indicates that the refinement is indeed successful (see the supplementary material and Fig. S18). It is important to note that the x-ray scattering-guided MD (XSMD) refinement strategy does not aim at finding structures that give completely matching signals but rather samples structures striking a balance between constraints of the MD force field and the input signals.[69]

In order to characterize the changes in the secondary structure and SASA, the thermally averaged values are calculated based on the FES sampling. The differences of SASA (ΔSASA) and the secondary structure (ΔSec. Struct.) in the three states **I**, **U₁**, and **U₂** from those of the equilibrium ensemble are shown in Fig. 6 and also mapped on to the crystal structures for visual guidance. For the **I** state, the average SASA increased from 72 nm² in the ground state ensemble to 95 nm². The most significant gain in SASA is in a segment with residues 90–110, and a slight gain is also observed in residues 1–40 [Fig. 6(a), upper panel]. These regions correspond to αI, αIII, and αIV, which lost a significant portion of the helix character [Fig. 6(a), lower panel]. However, in **I**, its R_g is about 16.1 Å, only slightly above the ground state $R_g = 13.9$ Å. It shows that the **I** state is disordered in terms of secondary structures but rather compact in its overall size, similar to the description of a molten globule state.[50]

For the **U₁** and **U₂** states, the SASA drastically increased to 126 nm² and 122 nm², respectively. While residues 100–110 are also more exposed to solvent as in the **I** state, exposure of residues 30–60 accounts more for the positive SASA change, which includes βI and the nearby backbone that connects to αII [Figs. 6(b) and
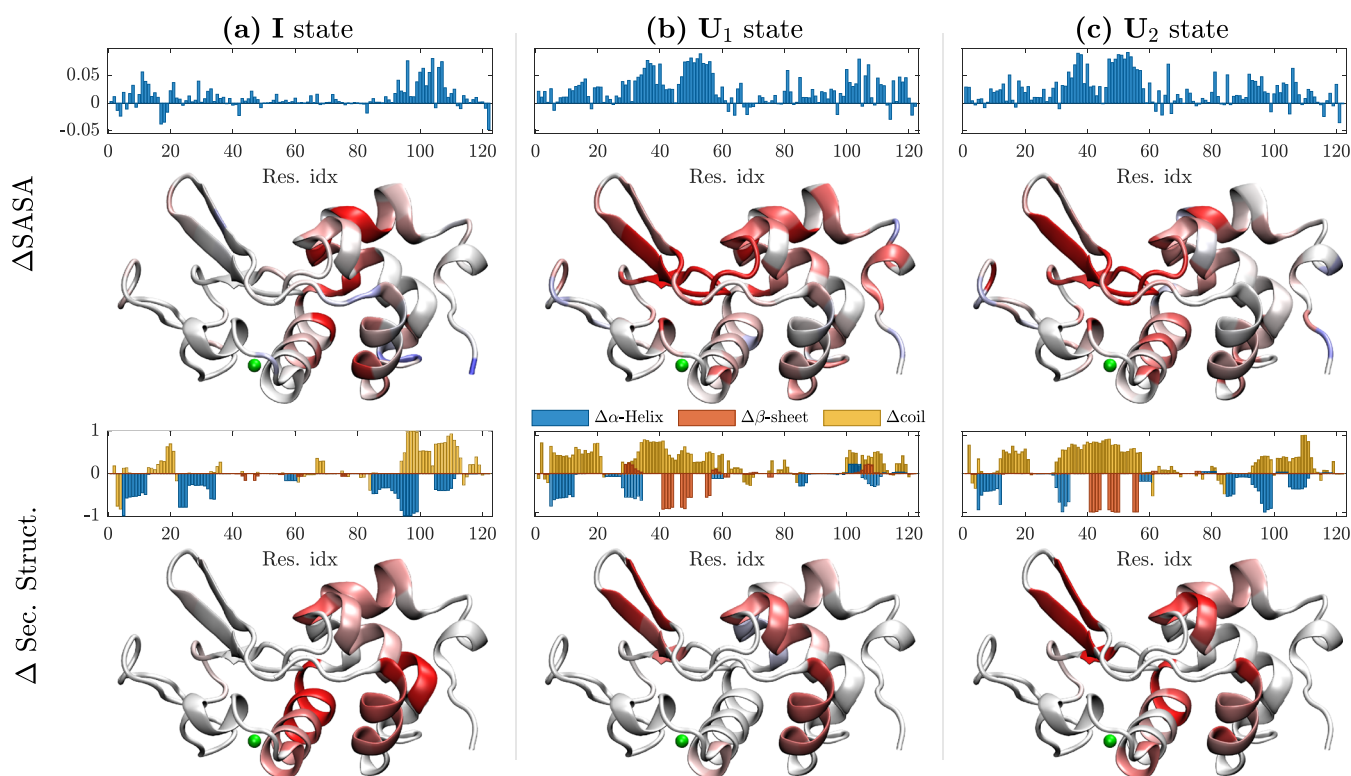
FIG. 6. Changes of residue-level solvent accessible surface area ΔSASA (upper panels) and secondary structure content ΔSec. Struct. (lower panels) with mappings onto the crystal structures showing the magnitude for each intermediate structure **I** (a), **U$_1$** (b), and **U$_2$** (c). For ΔSASA, the bar charts show changes in the fraction of SASA for each residue. For each protein structure, more exposed residues are shown in increasingly pronounced red shade. For changes in secondary structures, residues that formed α-helices or β-sheets are colored in blue, and those unfolded are colored in red. Changes in random coil are not mapped onto crystal structures. For all protein images, a green dot denotes the bound calcium ion.
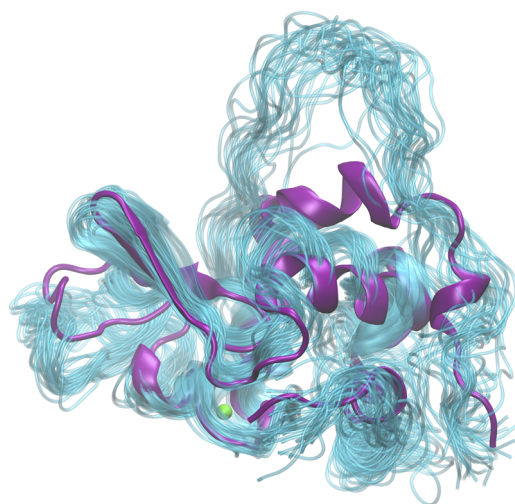
FIG. 7. The molten globule structures (cyan) sampled from the **I** state in the free energy basin defined in Fig. 4. Only the structures in the basin that are <1.5 kcal/mol are rendered. Each structure is aligned by Cα to the crystal structure (purple).

6(c), upper panels]. It is consistent with that the β-sheet in these two states completely disappears and that this region shows a large gain in the random coil structure content [Figs. 6(b) and 6(c), lower panels]. We can be sure that, in both states, the effect of incorporating the TRXSS signal causes the β-sheet to unfold completely and become more exposed to solvent. An important difference between **U$_1$** and **U$_2$** states is the higher solvent exposure of αIII in **U$_2$**, which can be seen in the backbone structure map [Figs. 6(b) and 6(c), upper panels]. It suggests that the protein expanded from the core in the **U$_2$** state, as these residues are among the most buried in the equilibrium ensemble [Fig. 2(d)]. On the contrary, the protein gained surface area through surface residue exposure in the **U$_1$** state.

BLA molten globule state as tracked by TRXSS and MD

As elucidated from the MD studies above, the **I** state can be categorized as a compact molten globule state. We now resolve further the structural ensemble of the **I** state. Some characteristics of this state include a compact size (R_g = 16.1 Å) and a slightly lower α-helical score of 0.40 compared to 0.51 in the ground state (Fig. 4)

but higher than the terminal unfolded states U_1 and U_2. The R_g is consistent with the 15.7 Å reported by Dolgikh and co-workers[50] for the molten globule state, as determined by SAXS. The structures relevant to the molten globule state can be extracted and examined through our MD methodology. The sampled structures in the free energy basin less than 1.5 kcal/mol are aligned to and overlaid on the ground state, as shown in Fig. 7. The overall protein topology is roughly conserved as in the ground state, but most α-helices are somewhat disordered. Despite the disordering, only αIV significantly extends out and forms a random coil. It is in line with similar studies done on the partially unfolded human α-lactalbumin,[2] where the low resolution reconstruction of SAXS data suggests that the C-terminal residues are likely disordered and form a tail.

The extension of αIV does not lead to an expansion of the nearby secondary structures, αII and αIII, as the residues in these components still resemble helices, and they can be aligned well onto the crystal structure. While in Fig. 6(a), lower panel, most α-helices are not categorized as helices, the α-helical score is rather high for the **I** state. It is due to that `timeline` categorizes the residues either as a part of a helix or not, but the α-helical score is a continuous quantity that estimates the similarity of a residue's configuration to that in a helix. Therefore, while the `timeline` disqualifies most residues in the **I** state as helices, these residues do have characteristics of helical arrangement, as shown in Fig. 7. It follows that there is likely significant incomplete side-chain packing that results in perturbed α-helices. This perturbation is clearly due to the input TRXSS signal, which yielded longer distances between atoms on the side-chain. Overall, using the TRXSS signal to refine our MD simulations allowed us to gain insight into the disordered structures of the molten globule state.

Implications for the BLA folding process

If we consider the unfolding process initiated in this experiment to be the reverse of the folding process,[19] then the unfolding process in our experiment (e.g., molten globule state formation from the native state in Fig. 4) will be the last process in the folding pathway during the formation of the native structure. Similarly, the later expansion process in the unfolding will be mirrored to an initial collapse into the molten globule state. Our experiment and refinement on the structures therefore give consistent results with the literature description of the folding process.[19] Comparing the finding from the steady-state study, where only one unfolded state is observed, to that in the time-resolved study, where two unfolded states are necessary to explain the data evolution; it shows that the two unfolded states may have the similar thermodynamic property, but their kinetic accessibility may be modulated by the existence of the **I** state.

We can now start to describe the folding process with our sampled structures. The protein in its terminal unfolded states still has a considerable amount of α-helices while being greatly expanded in R_g. In particular, the αII and αIII seem to suffer less loss in the structure. This is consistent with the reports of residual far-UV CD signals, linked to secondary structures, after the protein has passed its melting point determined by the near-UV CD, which reflects tertiary structures.[47] Vanderheeren and co-workers argued that this disparity stems from the binding of calcium near residues 80 (Fig. 6, green dots in molecular graphics). In our refinement, it seems that the

residues near the calcium binding site (residues 82–87) are indeed fairly intact in both U_1 and U_2 states, especially residues 85–91. The Cys91 is linked through a disulfide bond to the Cys73 on the random coil region, which may have contributed to the creation of a framework that stabilized the residues in between. The same, however, cannot be said on the **I** state, where the αIII helix is more disordered (but not totally unfolded) compared to the U_1 and U_2 states. It seems that the gain in enthalpy from a non-specific contact in the compact state is countered by the loss of secondary structures and conformational entropy. This view is consistent with the work by Forge and co-workers.[19] It follows that forming the native contact may be the main barrier between the molten globule state and the native state.[70] Considering the interchangeable nature of the U_1 and U_2 states as seen in the steady-state data, and difference in SASA between them, it seems that the deciding factor of whether the protein assumes a molten globule state during folding is the presence of a solvent inaccessible core. If the protein is exposed to the solvent from the core, as in αIII of the U_2 state [Fig. 6(c)], then the protein residues could have more space to adjust for the native contact within individual helices. The protein can then bypass the molten globule state and form the native structure directly. Overall, the reciprocal-space TRXSS data can be explained by these sampled structures whose properties also match the existing descriptions of the folding process. Through the analyses, we are able to obtain the atomistically detailed structures, which allow for closer looks of the folding dynamics.

Applications and limitations of XSMD

The XSMD approach presented in this work was able to drive the protein to structures that fit the data reasonably well. In addition, without utilizing kinetic information determined from the experimental data, this method successfully revealed the putative unfolding pathways. We show that it is possible to utilize the TRXSS signal to find possible intermediate structural ensembles, and we expect that many protein (un)folding systems can benefit from this method, provided that the TRXSS signal points to a rather homogeneous ensemble. Potential applicable systems include insulin dissociation[29,71] or photoactive receptors.[27] However, since the possibility of the heterogeneous unfolded state ensemble cannot be ruled out, using ensemble-averaged and maximum-entropy methods may allow for more effective fitting of the data.[72] We expect our method to be fully compatible with these more complex methods with minimal modifications of the implementation.[69]

The present method can be seen as a way of making a Bayesian inference on the prior distribution of the sampled structures based on the supplied MD force field, resulting in a posterior distribution of the refined structure using the fit of the input scattering signal as the likelihood. One property of the Bayesian inference is that the posterior distribution can be the prior of the next inference. In other words, the structures that are sampled can be subject to further screening with other experimental data. For example, it can be incorporated into the Bayesian inferential structural determination method as another constraint.[10] However, further estimations of uncertainties in various parameters in this method are necessary. The excited state fraction α will inherit some uncertainty from data analysis routines. The fitting parameters c_1 and c_2 were assumed to stay the same for computational convenience. However, these values

may change due to the perturbation in solvent conditions, especially during rapid-mixing experiments.[73,74] One way to estimate the credible intervals for these parameters could be to also refine them during the MD simulations. In addition, the information acquired here can thus be connected to a central structural analysis pipeline to gain a more comprehensive understanding of protein dynamics in the disordered states.

CONCLUSION

We have conducted time-resolved x-ray solution scattering experiments to probe the bovine α-lactalbumin unfolding process in which three kinetically distinct states are revealed. We formulated the state-associated difference scattering signals as x-ray scattering-derived potentials to steer the well-tempered metadynamics simulations to refine the structures for each state. The structures of the three states were found to be a molten globule state, $\mathbf{I}$, and two terminal unfolded states, $\mathbf{U_1}$ and $\mathbf{U_2}$. The sampled structures of the $\mathbf{I}$ state appear to have a similar overall topology of the folded native protein but less condensed secondary structures, while the other two states are much more conformationally disordered and have expanded sizes. These findings allow us to extract structural information at a higher accuracy than using TRXSS or MD simulation alone, especially for conformationally disordered protein structures. The results from this work can significantly help to decipher protein folding processes with an atomistically detailed description.

SUPPLEMENTARY MATERIAL

See the supplementary material for details about (1) XSS data collection, (2) TRXSS data reduction, (3) steady-state XSS reduction and analysis, (4) TRXSS solvent signal subtraction, (5) global analysis of the TRXSS signal, (6) the estimation of the T-jump magnitude, (7) the estimation of the excited state fraction, (8) preparation of the species-associated difference signal for MD simulations, (9) convergence of XSMD simulations, (10) convergence of WTmetaD simulations, (11) comparison between refined difference and species-associated differences, and (12) calculating the SASA change for each residue.

ACKNOWLEDGMENTS

DATA AVAILABILITY

The data and simulation trajectories that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]C. Bracken, L. M. Iakoucheva, P. R. Romero, and A. K. Dunker, "Combining prediction, computation and experiment for the characterization of protein disorder," Curr. Opin. Struct. Biol. **14**, 570–576 (2004).

[2]J. Ho CS, A. Rydström, M. S. S. Manimekalai, C. Svanborg, and G. Grüber, "Low resolution solution structure of HAMLET and the importance of its alpha-domains in tumoricidal activity," PLoS One **7**, e53051 (2012).

[3]T. Mittag and J. D. Forman-Kay, "Atomic-level characterization of disordered protein ensembles," Curr. Opin. Struct. Biol. **17**, 3–14 (2007).

[4]N. I. Brodie, K. I. Popov, E. V. Petrotchenko, N. V. Dokholyan, and C. H. Borchers, "Conformational ensemble of native α-synuclein in solution as determined by short-distance crosslinking constraint-guided discrete molecular dynamics simulations," PLoS Comput. Biol. **15**, e1006859 (2019).

[5]I. M. Ilie, D. Nayar, W. K. den Otter, N. F. A. van der Vegt, and W. J. Briels, "Intrinsic conformational preferences and interactions in α-synuclein fibrils: Insights from molecular dynamics simulations," J. Chem. Theory Comput. **14**, 3298–3310 (2018).

[6]V. N. Uversky, "Introduction to intrinsically disordered proteins (IDPs)," Chem. Rev. **114**, 6557–6560 (2014).

[7]A. Schramm, C. Bignon, S. Brocca, R. Grandori, C. Santambrogio, and S. Longhi, "An arsenal of methods for the experimental characterization of intrinsically disordered proteins—How to choose and combine them?," Arch. Biochem. Biophys. **676**, 108055 (2019).

[8]K. Lindorff-Larsen, N. Trbovic, P. Maragakis, S. Piana, and D. E. Shaw, "Structure and dynamics of an unfolded protein examined by molecular dynamics simulation," J. Am. Chem. Soc. **134**, 3787–3791 (2012).

[9]C. Camilloni and M. Vendruscolo, "Statistical mechanics of the denatured state of a protein using replica—Averaged metadynamics," J. Am. Chem. Soc. **136**, 8982–8991 (2014).

[10]J. Lincoff, M. Haghighatlari, M. Krzeminski, J. M. C. Teixeira, G.-N. W. Gomes, C. C. Gradinaru, J. D. Forman-Kay, and T. Head-Gordon, "Extended experimental inferential structure determination method in determining the structural ensembles of disordered protein states," Chem. Commun. **3**, 74 (2020).

[11]C. M. Jones, E. R. Henry, Y. Hu, C. K. Chan, S. D. Luck, A. Bhuyan, H. Roder, J. Hofrichter, and W. A. Eaton, "Fast events in protein folding initiated by nanosecond laser photolysis," Proc. Natl. Acad. Sci. U. S. A. **90**, 11860–11864 (1993).

[12]S. Piana, K. Lindorff-Larsen, and D. E. Shaw, "Atomic-level description of ubiquitin folding," Proc. Natl. Acad. Sci. U. S. A. **110**, 5915–5920 (2013).

[13]A. B. Stickrath, M. W. Mara, J. V. Lockard, M. R. Harpham, J. Huang, X. Zhang, K. Attenkofer, and L. X. Chen, "Detailed transient heme structures of Mb-CO in solution after CO dissociation: An X-ray transient absorption spectroscopic study," J. Phys. Chem. B **117**, 4705–4712 (2013).

[14]H. Gelman and M. Gruebele, "Fast protein folding kinetics," Q. Rev. Biophys. **47**, 95–142 (2014).

[15]V. Muñoz and M. Cerminara, "When fast is better: Protein folding fundamentals and mechanisms from ultrafast approaches," Biochem. J. **473**, 2545–2559 (2016).

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

[17]K. Schlepckow, J. Wirmer, A. Bachmann, T. Kiefhaber, and H. Schwalbe, "Conserved folding pathways of α-lactalbumin and lysozyme revealed by kinetic CD,

fluorescence, NMR, and interrupted refolding experiments," J. Mol. Biol. **378**, 686–698 (2008).

[18]E. Chen, M. J. Wood, A. L. Fink, and D. S. Kliger, "Time-resolved circular dichroism studies of protein folding intermediates of cytochrome *c*," Biochemistry **37**, 5589–5598 (1998).

[19]V. Forge, R. T. Wijesinha, J. Balbach, K. Brew, C. V. Robinson, C. Redfield, and C. M. Dobson, "Rapid collapse and slow structural reorganisation during the refolding of bovine alpha-lactalbumin," J. Mol. Biol. **288**, 673–688 (1999).

[20]L. Konermann, J. Pan, and Y.-H. Liu, "Hydrogen exchange mass spectrometry for studying protein structure and dynamics," Chem. Soc. Rev. **40**, 1224–1234 (2011).

[21]X.-X. Zhang, K. C. Jones, A. Fitzpatrick, C. S. Peng, C.-J. Feng, C. R. Baiz, and A. Tokmakoff, "Studying protein–protein binding through T-jump induced dissociation: Transient 2D IR spectroscopy of insulin dimer," J. Phys. Chem. B **120**, 5134–5145 (2016).

[22]H. Takala, A. Björling, O. Berntsson, H. Lehtivuori, S. Niebling, M. Hoernke, I. Kosheleva, R. Henning, A. Menzel, J. A. Ihalainen, and S. Westenhoff, "Signal amplification and transduction in phytochrome photosensors," Nature **509**, 245–248 (2014).

[23]J. G. Kim, T. W. Kim, J. Kim, and H. Ihee, "Protein structural dynamics revealed by time-resolved X-ray solution scattering," Acc. Chem. Res. **48**, 2200–2208 (2015).

[24]J. Kim, K. H. Kim, K. Y. Oang, J. H. Lee, K. Hong, H. Cho, N. Huse, R. W. Schoenlein, T. K. Kim, and H. Ihee, "Tracking reaction dynamics in solution by pump–probe X-ray absorption spectroscopy and X-ray liquidography (solution scattering)," Chem. Commun. **52**, 3734–3749 (2016).

[25]D. Rimmerman, D. Leshchev, D. J. Hsu, J. Hong, B. Abraham, R. Henning, I. Kosheleva, and L. X. Chen, "Probing cytochrome *c* folding transitions upon phototriggered environmental perturbations using time-resolved X-ray scattering," J. Phys. Chem. B **122**, 5218–5224 (2018).

[26]T. W. Kim, S. J. Lee, J. Jo, J. G. Kim, H. Ki, C. W. Kim, K. H. Cho, J. Choi, J. H. Lee, M. Wulff, Y. M. Rhee, and H. Ihee, "Protein folding from heterogeneous unfolded state revealed by time-resolved X-ray solution scattering," Proc. Natl. Acad. Sci. U. S. A. **117**, 14996–15005 (2020).

[27]H. S. Cho, F. Schotte, N. Dashdorj, J. Kyndt, R. Henning, and P. A. Anfinrud, "Picosecond photobiology: Watching a signaling protein function in real time via time-resolved small- and wide-angle X-ray scattering," J. Am. Chem. Soc. **138**, 8815–8823 (2016).

[28]D. J. Hsu, D. Leshchev, D. Rimmerman, J. Hong, M. S. Kelley, I. Kosheleva, X. Zhang, and L. X. Chen, "X-ray snapshots reveal conformational influence on active site ligation during metalloprotein folding," Chem. Sci. **10**, 9788–9800 (2019).

[29]D. Rimmerman, D. Leshchev, D. J. Hsu, J. Hong, I. Kosheleva, and L. X. Chen, "Direct observation of insulin association dynamics with time-resolved X-ray scattering," J. Phys. Chem. Lett. **8**, 4413–4418 (2017).

[30]H. S. Cho, F. Schotte, V. Stadnytskyi, A. DiChiara, R. Henning, and P. Anfinrud, "Dynamics of quaternary structure transitions in R-state carbonmonoxyhemoglobin unveiled in time-resolved X-ray scattering patterns following a temperature jump," J. Phys. Chem. B **122**, 11488–11496 (2018).

[31]I. Josts, S. Niebling, Y. Gao, M. Levantino, H. Tidow, and D. Monteiro, "Photocage-initiated time-resolved solution X-ray scattering investigation of protein dimerization," IUCrJ **5**, 667–672 (2018).

[32]D. Rimmerman, D. Leshchev, D. J. Hsu, J. Hong, B. Abraham, R. Henning, I. Kosheleva, and L. X. Chen, "Revealing fast structural dynamics in pH-responsive peptides with time-resolved X-ray scattering," J. Phys. Chem. B **123**, 2016–2021 (2019).

[33]L. Henry, M. R. Panman, L. Isaksson, E. Claesson, I. Kosheleva, R. Henning, S. Westenhoff, and O. Berntsson, "Real-time tracking of protein unfolding with time-resolved X-ray solution scattering," Struct. Dyn. **7**, 054702 (2020).

[34]D. I. Svergun, "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing," Biophys. J. **76**, 2879–2886 (1999).

[35]T. D. Grant, "*Ab initio* electron density determination directly from solution scattering data," Nat. Methods **15**, 191 (2018).

[36]D. Kimanius, I. Pettersson, G. Schluckebier, E. Lindahl, and M. Andersson, "SAXS-guided metadynamics," J. Chem. Theory Comput. **11**, 3491–3498 (2015).

[37]A. Björling, S. Niebling, M. Marcellini, D. Van Der Spoel, and S. Westenhoff, "Deciphering solution scattering data with experimentally guided molecular dynamics simulations," J. Chem. Theory Comput. **11**, 780–787 (2015).

[38]P.-c. Chen and J. S. Hub, "Interpretation of solution X-ray scattering by explicit-solvent molecular dynamics," Biophys. J. **108**, 2573–2584 (2015).

[39]C. Paissoni, A. Jussupow, and C. Camilloni, "Martini bead form factors for nucleic acids and their application in the refinement of protein–nucleic acid complexes against SAXS data," J. Appl. Crystallogr. **52**, 394–402 (2019).

[40]D. J. Hsu, D. Leshchev, I. Kosheleva, K. L. Kohlstedt, and L. X. Chen, "Integrating solvation shell structure in experimentally driven molecular dynamics using X-ray solution scattering data," J. Chem. Phys. **152**, 204115 (2020).

[41]A. Laio and M. Parrinello, "Escaping free-energy minima," Proc. Natl. Acad. Sci. U. S. A. **99**, 12562–12566 (2002).

[42]A. Barducci, G. Bussi, and M. Parrinello, "Well-tempered metadynamics: A smoothly converging and tunable free-energy method," Phys. Rev. Lett. **100**, 020603 (2008).

[43]Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," Chem. Phys. Lett. **314**, 141–151 (1999).

[44]Y. M. Rhee and V. S. Pande, "Multiplexed-replica exchange molecular dynamics method for protein folding simulation," Biophys. J. **84**, 775–786 (2003).

[45]K. Kuwajima and S. Sugai, "Equilibrium and kinetics of the thermal unfolding of alpha-lactalbumin. The relation to its folding mechanism," Biophys. Chem. **8**, 247–254 (1978).

[46]M. Ikeguchi, K. Kuwajima, M. Mitani, and S. Sugai, "Evidence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: A comparative study of the folding reactions of α-lactalbumin and lysozyme," Biochemistry **25**, 6965–6972 (1986).

[47]G. Vanderheeren and I. Hanssens, "Thermal unfolding of bovine alpha-lactalbumin. Comparison of circular dichroism with hydrophobicity measurements," J. Biol. Chem. **269**, 7090–7094 (1994).

[48]L. C. Wu, Z.-y. Peng, and P. S. Kim, "Bipartite structure of the α-lactalbumin molten globule," Nat. Struct. Mol. Biol. **2**, 281–286 (1995).

[49]M. Arai and K. Kuwajima, "Rapid formation of a molten globule intermediate in refolding of α-lactalbumin," Folding Des. **1**, 275–287 (1996).

[50]D. A. Dolgikh, L. V. Abaturov, I. A. Bolotina, E. V. Brazhnikov, V. E. Bychkova, R. I. Gilmanshin, Y. O. Lebedev, G. V. Semisotnov, E. I. Tiktopulo, and O. B. Ptitsyn, "Compact state of a protein molecule with pronounced small-scale mobility: Bovine α-lactalbumin," Eur. Biophys. J. **13**, 109–121 (1985).

[51]S. Chakraborty, V. Ittah, P. Bai, L. Luo, E. Haas, and Z.-y. Peng, "Structure and dynamics of the α-lactalbumin molten globule: Fluorescence studies using proteins containing a single tryptophan residue," Biochemistry **40**, 7228–7238 (2001).

[52]T. Graber, S. Anderson, H. Brewer, Y.-S. Chen, H. S. Cho, N. Dashdorj, R. W. Henning, I. Kosheleva, G. Macha, M. Meron, R. Pahl, Z. Ren, S. Ruan, F. Schotte, V. Šrajer, P. J. Viccaro, P. Westferro, P. Anfinrud, and K. Moffat, "BioCARS: A synchrotron resource for time-resolved X-ray science," J. Synchrotron Radiat. **18**, 658–670 (2011).

[53]M. Cammarata, M. Levantino, F. Schotte, P. A. Anfinrud, F. Ewald, J. Choi, A. Cupane, M. Wulff, and H. Ihee, "Tracking the structural dynamics of proteins in solution using time-resolved wide-angle X-ray scattering," Nat. Methods **5**, 881–886 (2008).

[54]M. Cammarata, M. Lorenc, T. K. Kim, J. H. Lee, Q. Y. Kong, E. Pontecorvo, M. Lo Russo, G. Schirò, A. Cupane, M. Wulff, and H. Ihee, "Impulsive solvent heating probed by picosecond X-ray diffraction," J. Chem. Phys. **124**, 124504 (2006).

[55]P. Debye, "Zerstreuung von Röntgenstrahlen," Ann. Phys. **351**, 809–823 (1915).

[56]D. Svergun, C. Barberato, and M. H. J. Koch, "CRYSOL—A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates," J. Appl. Crystallogr. **28**, 768–773 (1995).

[57]D. Schneidman-Duhovny, M. Hammel, J. A. Tainer, and A. Sali, "Accurate SAXS profile computation and its assessment by contrast variation experiments," Biophys. J. **105**, 962–974 (2013).

[58]D. Waasmaier and A. Kirfel, "New analytical scattering-factor functions for free atoms and ions," Acta Crystallogr., Sect. A: Found. Crystallogr. **51**, 416–431 (1995).

[59]J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, "Scalable molecular dynamics with NAMD," J. Comput. Chem. **26**, 1781–1802 (2005).

[60]U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh Ewald method," J. Chem. Phys. **103**, 8577–8593 (1995).

[61]J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes," J. Comput. Phys. **23**, 327–341 (1977).

[62]E. D. Chrysina, K. Brew, and K. R. Acharya, "Crystal structures of apo- and holo-bovine α-lactalbumin at 2.2-Å resolution reveal an effect of calcium on inter-lobe interactions," J. Biol. Chem. **275**, 37021–37029 (2000).

[63]J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. De Groot, H. Grubmüller, and A. D. Mackerell, "CHARMM36m: An improved force field for folded and intrinsically disordered proteins," Nat. Methods **14**, 71–73 (2016).

[64]S. Boonstra, P. R. Onck, and E. van der Giessen, "CHARMM TIP3P water model suppresses peptide folding by solvating the unfolded state," J. Phys. Chem. B **120**, 3692–3698 (2016).

[65]G. Fiorin, M. L. Klein, and J. Hénin, "Using collective variables to drive molecular dynamics simulations," Mol. Phys. **111**, 3345–3362 (2013).

[66]M. Arai, K. Ito, T. Inobe, M. Nakao, K. Maki, K. Kamagata, H. Kihara, Y. Amemiya, and K. Kuwajima, "Fast compaction of alpha-lactalbumin during folding studied by stopped-flow X-ray scattering," J. Mol. Biol. **321**, 121–132 (2002).

[67]W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," J. Mol. Graphics **14**, 33–38 (1996).

[68]M. C. Thompson, B. A. Barad, A. M. Wolff, H. Sun Cho, F. Schotte, D. M. C. Schwarz, P. Anfinrud, and J. S. Fraser, "Temperature-jump solution X-ray scattering reveals distinct motions in a dynamic enzyme," Nat. Chem. **11**, 1058–1066 (2019).

[69]G. Hummer and J. Köfinger, "Bayesian ensemble refinement by replica simulations and reweighting," J. Chem. Phys. **143**, 243150 (2015).

[70]R. B. Best, G. Hummer, and W. A. Eaton, "Native contacts determine protein folding mechanisms in atomistic simulations," Proc. Natl. Acad. Sci. U. S. A. **110**, 17874–17879 (2013).

[71]P. Banerjee, S. Mondal, and B. Bagchi, "Effect of ethanol on insulin dimer dissociation," J. Chem. Phys. **150**, 084902 (2019).

[72]M. R. Hermann and J. S. Hub, "SAXS-restrained ensemble simulations of intrinsically disordered proteins with commitment to the principle of maximum entropy," J. Chem. Theory Comput. **15**, 5103–5115 (2019).

[73]A. Yaghmur, P. Laggner, B. Sartori, and M. Rappolt, "Calcium triggered L_α-H_2 phase transition monitored by combined rapid mixing and time-resolved synchrotron SAXS," PLoS One **3**, e2072 (2008).

[74]B. Angelov, A. Angelova, M. Drechsler, and S. Lesieur, "Rapid mixing stopped-flow small-angle X-ray scattering study of lipoplex formation at beamline ID02@ESRF," J. Surf. Invest.: X-Ray, Synchrotron Neutron Tech. **9**, 105–110 (2015).