

Transient local secondary structure in the intrinsically disordered C-term of the Albino3 insertase

Dustin R. Baucom,¹ Mercedes Furr,¹ Vivek Govind Kumar,¹ Patience Okoto,¹ James L. Losey,¹ Ralph L. Henry,² Mahmoud Moradi,^{1,*} Thallapuranam Krishnaswamy Suresh Kumar,^{1,*} and Colin D. Heyes^{1,*}

¹Department of Chemistry and Biochemistry and ²Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas

ABSTRACT Albino3 (Alb3) is an integral membrane protein fundamental to the targeting and insertion of light-harvesting complex (LHC) proteins into the thylakoid membrane. Alb3 contains a stroma-exposed C-terminus (Alb3-Cterm) that is responsible for binding the LHC-loaded transit complex before LHC membrane insertion. Alb3-Cterm has been reported to be intrinsically disordered, but precise mechanistic details underlying how it recognizes and binds to the transit complex are lacking, and the functional roles of its four different motifs have been debated. Using a novel combination of experimental and computational techniques such as single-molecule fluorescence resonance energy transfer, circular dichroism with deconvolution analysis, site-directed mutagenesis, trypsin digestion assays, and all-atom molecular dynamics simulations in conjunction with enhanced sampling techniques, we show that Alb3-Cterm contains transient secondary structure in motifs I and II. The excellent agreement between the experimental and computational data provides a quantitatively consistent picture and allows us to identify a heterogeneous structural ensemble that highlights the local and transient nature of the secondary structure. This structural ensemble was used to predict both the inter-residue distance distributions of single molecules and the apparent unfolding free energy of the transient secondary structure, which were both in excellent agreement with those determined experimentally. We hypothesize that this transient local secondary structure may play an important role in the recognition of Alb3-Cterm for the LHC-loaded transit complex, and these results should provide a framework to better understand protein targeting by the Alb3-Oxa1-YidC family of insertases.

SIGNIFICANCE Intrinsically disordered proteins or regions play key roles in a wide range of biological processes. How intrinsically disordered proteins or regions are able to balance disorder with specificity, which is usually the result of specific structure, is an open but critical question. In this manuscript, using, to our knowledge, a novel combination of experimental and computational techniques, we identify regions of transient local structure in a disordered region of a protein that is critical for targeting light-harvesting proteins to the thylakoid membrane in plants, namely Albino3, which is a member of a wider protein family in bacteria and mitochondria. These results are expected to provide a framework to address how a balance of disorder and transient local structure play key roles in biological processes.

INTRODUCTION

The evolutionarily conserved YidC-Oxa1-Albino3 (Alb3) protein family mediates assembly and insertion of membrane

proteins vital to the processes of energy production in bacteria, mitochondria, and chloroplasts, respectively (1). Located in the inner membrane of bacteria, mitochondria, and the thylakoid membrane of chloroplasts, these insertases have similar functions with the exception of some species-specific differences (1). Each homolog shares a conserved hydrophobic region comprised of five transmembrane domains (TMDs) and a large periplasmic domain between TMDs 1 and 2, with the exception of the six transmembrane domains of YidC in Gram-negative bacteria. In chloroplasts, Alb3 is required for the post-translational integration of LHCPs into the thylakoid membrane (2). Nuclear-encoded, light-harvesting chlorophyll-binding proteins (LHCPs) are imported

Submitted December 31, 2020, and accepted for publication October 13, 2021.

*Correspondence: mmoradi@uark.edu or sthalla@uark.edu or cheyes@uark.edu

Dustin R. Baucom and Mercedes Furr contributed equally to this work.

Dustin R. Baucom's present address is Department of Natural Sciences, Northeastern State University, Tahlequah, Oklahoma, 74464.

Editor: Samrat Mukhopadhyay.

<https://doi.org/10.1016/j.bpj.2021.10.013>

© 2021 Biophysical Society.

from the cytoplasm into the stromal compartment of the chloroplast via a chloroplast targeting sequence (3). Upon entrance into the stroma, the targeting sequence is cleaved, and LHCP binds to a chloroplast signal recognition particle (cpSRP), a heterodimer of cpSRP43 and cpSRP54, to form the transit complex (4). Once at the membrane, via interaction between cpSRP54 and its receptor, cpFtsY, the protein complex then interacts with Alb3 (5,6). Alb3 functions as an insertase and is critical for both the post-translational integration of LHCPs (2) and the cotranslational integration of other chloroplast membrane proteins (7). The C-terminal of Alb3 (Alb3-Cterm) is an intrinsically disordered region of Alb3 that protrudes out into the stroma and facilitates association of the transit complex at the thylakoid membrane through recognition and association with the cpSRP43 subunit (8,9). Upon binding to cpSRP43, Alb3-Cterm gains α -helical structure, as demonstrated in circular dichroism (CD) experiments (9). In addition, Alb3-Cterm binding to cpSRP43 activates GTP hydrolysis by cpSRP54 and cpFtsY (10). An interaction between LHCP and Alb3-Cterm was also observed in *in vitro* binding experiments (11). Furthermore, cpSRP43 binds Alb3-Cterm, whereas cpSRP54 does not (11), suggesting that the interaction between Alb3-Cterm and cpSRP43 facilitates a transfer of LHCP from cpSRP43 to Alb3-Cterm. It was also found that the cpSRP-LHCP complex binds more efficiently to Alb3-Cterm than cpSRP alone (11), suggesting a mechanism in which Alb3 is prevented from being blocked by a cargoless cpSRP through structural rearrangements in cpSRP upon LHCP binding.

Sequence alignments of Alb3 from five different plants have revealed four conserved, positively charged motifs (I–IV) in the stromal C-terminal region (9). In that study, isothermal titration calorimetry (ITC) data suggested that motif I is not important in the binding interaction but that motifs II and IV are required for the interaction with cpSRP43 (9). In a follow-up study by the same group, lysine and arginine were independently mutated to alanine in motifs II and IV, which led to reduced affinity to cpSRP43, although the effect in motif IV was stronger than motif II (12). When the mutations were introduced simultaneously, binding with cpSRP43 was lost completely. This result suggests that specific positively charged residues in motifs II and IV are important for binding cpSRP43 (12). In a separate study, a truncated Alb3 lacking motifs III and IV was shown to be unstable in normal light conditions resulting in $\sim 80\%$ lower amounts of truncated Alb3 present compared with wild-type Alb3 (13). However, when grown under continuous low light conditions, the truncated Alb3-Cterm was able to accumulate and integrate LHCPs with only a minor reduction compared with wild-type, suggesting that motifs I and II and TM5 were sufficient for interaction with cpSRP43 and insertion of LHCP into the membrane, whereas motifs III and IV provided photostability to the protein (13).

However, a key aspect missing from the studies described above is localized structural information in Alb3-Cterm. In this study, we combine biophysical techniques, site-directed mutagenesis, single-molecule fluorescence resonance energy transfer (smFRET) and all-atom-enhanced sampling molecular dynamics (MD) simulations to provide high-resolution localized structural information in Alb3-Cterm. We identify that motifs I and II have a relatively high propensity for secondary structure with smFRET experiments showing a broad peak with shorter inter-residue distance than that of a random coil. In the presence of increasing denaturant concentration, molecules showing high smFRET efficiencies reduce, and those showing lower smFRET efficiencies increase, which are in line with inter-residue distances expected from a random coil, indicating that this local structure can be unfolded. Comparison of smFRET data with independent predictions of structures from enhanced MD simulations highlighted the formation of α -helical transient secondary structure in motifs I and II of Alb3-Cterm with inter-residue distance distributions and apparent unfolding free energies in very good agreement between experiment and simulation. Finally, CD spectra of site-directed mutants of Alb3-Cterm in which helix-breaking residues were substituted for native residues quantified the relative importance of specific residues in forming transient α -helical secondary structure. The results of this study are expected to provide a framework to better understand the role of transient secondary structure within disordered regions of an important membrane-bound insertase during key steps of protein targeting and insertion.

MATERIALS AND METHODS

Cloning, expression, and purification of His-tagged Alb3-Cterm

N-terminally His-tagged c-terminal of PPF1 (Alb3-Cterm in *Pisum sativum*) was cloned into the pQE-80L vector and transformed in BL-21-Star cells. Protein production was induced with 1 mM isopropyl-1-thio-D-galactopyranoside, and the cells were harvested after 3.5 h at 37°C, at which point the absorbance value at 600 nm was 0.6. For the purification of His-tagged Alb3-Cterm, the cells were resuspended in buffer (50 mM HEPES, 150 mM NaCl, 10% (w/v) glycerol, and 0.02% 1-thioglycerol (pH 7.5)). The cells were lysed using ultrasonication, and the cell lysate was separated using ultracentrifugation at 19,000 revolutions per minute (rpm). The supernatant was applied onto an Ni²⁺ Sepharose column (GE Healthcare, Chicago, IL), washed with resuspension buffer, and eluted at 300 mM imidazole (IMZ) using a stepwise IMZ gradient of 10-, 20-, 50-, 100-, 300-, and 500-mM IMZ concentrations. Alb3-Cterm migrates slightly higher than expected, between 25 and 17 kDa. The molecular weight and purity were confirmed by mass spectroscopy (molecular weight = 15.17 kDa). The eluted protein was further purified to homogeneity using a Mono S ion exchange column (Cytiva, Marlborough, MA).

Initial predictions of structured regions

Structural predictions from the amino acid sequence of Alb3-Cterm were made using six different web servers: PONDR, DISOPRED, PrDOS2,

PSIPred, Dynamine, and FoldUnfold (14–19). The primary sequence of wild-type Alb3-Cterm was uploaded to each server, and the output was collected and analyzed to find a region of consensus between all algorithms used.

Enhanced sampling MD simulations

The Loop prediction by Energy-Assisted Protocol algorithm (20) was used to generate an all-atom model of Alb3-Cterm in a fully extended conformation. Modeler (21) was then used to generate 16 different Alb3-Cterm conformations using a Monte Carlo algorithm to minimize the Modeler objective function with different number of iterations. The 16 Modeler conformations were then used to set up 16 independent, well-tempered metadynamics simulations (22) of Alb3-Cterm in an explicit water solvent environment. Systems were solvated in a box of TIP3P waters (23) and 0.15 M NaCl. System size ranged from 23,745 atoms to 557,583 atoms because of the different box sizes. Simulations were performed using the NAMD 2.13 simulation package (24) with the CHARMM36m all-atom additive force field (25,26). Initially, we energy minimized each system for 1000 steps using the conjugate gradient algorithm (27). Then, the systems were equilibrated for 1 ns using a 2-fs time step at 310 K using a Langevin integrator with a damping coefficient of $\gamma = 1 \text{ ps}^{-1}$. The pressure was maintained at 1 atm using the Nose-Hoover Langevin piston method (28). The smoothed cutoff distance for nonbonded interactions was set to 10–12 Å, and long-range electrostatic interactions were computed with the particle mesh Ewald method (29).

After equilibration, well-tempered metadynamics simulations were run for 320 ns (an aggregate of $16 \times 320 = 5120$ ns). The α -variable was used as the collective variable in all metadynamics simulations. This collective variable quantifies the α -helical propensity of a protein or peptide of length N using Eq. 1

$$\alpha = \frac{1}{2} \left(\frac{1}{N-2} \sum_{n=1}^{N-2} f(\theta_n) + \frac{1}{N-4} \sum_{n=1}^{N-4} g(d_n) \right), \quad (1)$$

where θ_n is the angle formed by $C_\alpha^{(n)} - C_\alpha^{(n+1)} - C_\alpha^{(n+2)}$ and d_n is distance between $O^{(n)}$ and $N^{(n+4)}$. The $f(\theta)$ and $g(d)$ are score functions (both ranging from 0 to 1) quantifying the likelihood of θ and d being associated with an α -helix, Eq. 2,

$$f(\theta) = \frac{1 - \left(\frac{\theta - \theta_0}{\delta\theta} \right)^2}{1 - \left(\frac{\theta - \theta_0}{\delta\theta} \right)^4}, \quad g(d) = \frac{1 - \left(\frac{d}{d_0} \right)^6}{1 - \left(\frac{d}{d_0} \right)^8}, \quad (2)$$

where θ_0 and $\delta\theta$ are 88 and 15°, respectively, and d_0 is 3.3 Å. The metadynamics algorithm used a width of 0.01, an initial Hill weight of 0.1, and an update frequency of 1000. The pseudotemperature used for the well-tempered feature was 1200 K (22). Conformations were collected every 4 ps for the explicit simulations.

Secondary structure determination from MD trajectories

Visual Molecular Dynamics (VMD) (30) was used to visualize the Alb3-Cterm conformations and to generate molecular images. The structural identification (STRIDE) algorithm (31) was used to calculate the helical propensity of individual amino acids based on the MD trajectories. All 16 MD trajectories were combined to generate per-residue helical propensity profiles. Overall helical propensity (h) in each metadynamics trajectory (e.g., replica i) was calculated using the weighted average of the

“ α ”-collective variable using the metadynamics-based free energy profile from trajectory $i(F_i(h))$, Eq. 3:

$$\langle h \rangle_i = \frac{\int_0^1 e^{-\frac{F_i(h)}{RT}} h \, dh}{\int_0^1 e^{-\frac{F_i(h)}{RT}} dh} \quad (3)$$

We note that the STRIDE definition of helical structure used in per-residue helical propensities and the α definition of helical structure used in overall helical propensities are related but not the same, and the two quantities are not necessarily expected to have the same average values.

Cloning, expression, and purification of double cysteine His-tagged Alb3-Cterm mutant for smFRET

An endogenous cysteine residue was kept at position 14 in the original sequence, and an additional cysteine residue was introduced at position 52 (S52C). Primers were designed using an Agilent design program and ordered from Integrated DNA Technologies (Coralville, IA). Site-directed mutagenesis was performed using a QuikChange II XL kit (Agilent Technologies, Santa Clara, CA). The primer-encoded point mutation was introduced into the coding sequence of HIS-Alb3-Cterm by polymerase chain reaction. The plasmid was then transformed into DH5 α competent cells, and the construct was sequence verified by Molecular Resource Laboratory, University of Arkansas, Fayetteville, AR, for Medical Sciences, Little Rock, AR. Upon confirmation of the plasmid sequence, the mutant protein was overexpressed in BL-21-Star *Escherichia coli* cells cultured in lysogeny broth at 37°C with agitation at 250 rpm. Overexpression, purification, and quantification methods were the same as described above for His-tagged Alb3-Cterm except that the second chromatographic step was carried out on a size exclusion column. Purity of S52C is depicted by sodium-dodecyl-sulfate-polyacrylamide-gel-electrophoresis (SDS-PAGE) gels stained with Coomassie Brilliant Blue G-250 dye (Thermo Fisher Scientific, Waltham, MA).

Fluorescent labeling and smFRET

The double cysteine Alb3-Cterm mutants and wild-type were both labeled under the same conditions. Alb3-Cterm was incubated in pH 7.5 10 mM phosphate-buffered saline, with 50 μM Alexa Fluor 488 (Alexa488) C5 maleimide and 1 mM Alexa594 C5 maleimide for 30 min at room temperature. Tris (2-carboxyethyl) phosphine hydrochloride treatment was used to inhibit the formation of disulfide bonds. To remove the excess free dye, the labeling mixture was separated on a Bio-Gel P-6DG Gel filtration column (Bio-Rad Laboratories, Hercules, CA).

Labeled protein preparations were diluted to ~ 10 pM for smFRET experiments. Under these conditions, the average time between bursts was over 100 ms, compared with the ~ 1 -ms diffusion time of the protein through the confocal volume and, thus, is statistically highly unlikely to contain more than one molecule at a time. The dilution buffer for the denaturation experiments contained the appropriate concentration of guanidine hydrochloride in pH 7.5 phosphate-buffered saline buffer. smFRET was performed as described in our previous reports (32,33). Briefly, a MicroTime 200 microscope (Picoquant, Berlin, Germany) based on an Olympus IX71 microscope (Olympus Life Science, Center Valley, PA) was used. The protein solutions were exposed to 485 nm of continuous wave excitation at 75 μW power for 3-min intervals, after which the sample aliquot was changed. Fluorescence was collected using a 60 $\times$ water immersion objective with a numerical aperture of 1.3. A 100- μm pinhole was used to block out-of-focus fluorescence. Collected photons were then passed to an FF562-Di03 dichroic mirror (Chroma, McHenry, IL) to split the fluorescence into donor and acceptor channels based on wavelength. Furthermore, band-pass

filters (Chroma) in each channel were used to restrict photons to the respective fluorescent ranges of the donor and acceptor fluorophores. These filters were 520/35 and 620/60 for the donor and acceptor channel, respectively. Single photon avalanche diode detectors (PDM; Microphotonic Devices, Bolzano, Italy) were used to collect donor and acceptor signals in time-tagged time-resolved format.

Single-molecule traces were binned into 1-ms intervals and analyzed using a home-written analysis Python program. First, the donor channel was corrected using the γ -factor that was determined experimentally to be 0.6. This γ -corrected donor value was added to the acceptor channel value, and any burst meeting the threshold of 20 counts/ms was analyzed. Ratiometric fluorescence resonance energy transfer (FRET) efficiencies were calculated for all bursts using the γ -factor and the donor fluorescence leak into the acceptor channel, which was found to be 7%. Direct excitation of the acceptor dye by the 485 nm laser was found to be negligible. All FRET efficiencies for a sample were binned to produce the smFRET histograms.

Computational prediction of smFRET distance distributions

Atomic models of Alexa488 and Alexa594 dyes were attached to a cysteine amino acid and simulated in explicit water for 100 ns. The system size was 7407 atoms for Alexa488 and 5512 atoms for Alexa594. All simulation protocols were similar to those used for Alb3-Cterm simulations as described above. Parameter and topology files were generated using Chemistry at Harvard Macromolecular Mechanics General Force Field (34). Simulations were performed using NAMD 2.13 with simulation parameters similar to those described above for Alb3-Cterm. These dye models replaced the Alb3-Cterm residues of interest (C14 and S52) from the explicit well-tempered metadynamics simulations. For the replacement, S52 was first mutated to C52, and the sampled conformations of both dyes were aligned with both C14 and C52. For each conformation of Alb3-Cterm and the donor and acceptor dyes, two alternative models were generated (donor and acceptor attaching to C14 and C52, respectively, or vice versa). Both the Alb3-Cterm simulation and the dye simulations were advanced 40 ps to generate a new conformation. Because the Alb3-Cterm simulations were longer than the dye simulations, it was necessary to “restart” near the beginning of the dye conformations every 100 ns. Mass center fluorophore-fluorophore distance distributions were then calculated for every conformation. The alignment of separate dye and protein simulation trajectories resulted in some conformations with dye-dye distances that are unrealistically close. To avoid this steric overlap, conformations where the dye-dye distance was less than 5 Å were rejected. This procedure generated more than 136,000 simulated conformations of Alb3-Cterm with dyes attached. The distance data were then binned into histograms and compared with the distance distribution derived from smFRET data (with an experimentally determined R_0 of 57 Å).

Cloning, expression, and purification of His-tagged Alb3-Cterm helix-breaking mutants

Single-point mutants of Alb3-Cterm (A21G, V24G, L26G, and K28G) were cloned into pQE-80L bacterial expression vector, which was used as a template for site-directed mutagenesis. Primers were designed using an Agilent primer design program (Integrated DNA Technologies). Site-directed mutagenesis was performed using a QuikChange II XL kit (Agilent Technologies) followed by polymerase chain reaction. The plasmid was then transformed into DH5 α competent cells. Upon confirmation of the correct plasmid sequences, each mutant was overexpressed in BL-21-Star *E. coli* cells cultured in lysogeny broth at 37°C with agitation at 250 rpm. Overexpressed cells were lysed using ultrasonication and separated from the cell debris using ultracentrifugation at 19,000 rpm. Alb3-Cterm mutants were purified as described above for the wild-type Alb3-Cterm. The purity of the Alb3-Cterm mutant proteins were assessed by 15% SDS-PAGE from

an Ni²⁺ column and subsequent purification using Mono S. Protein bands were visualized by staining the gels with Coomassie Brilliant Blue R-250 dye (Thermo Fisher Scientific). Ovalbumin (A5503) was purchased from Sigma-Aldrich (St. Louis, MO).

CD and intrinsic fluorescence spectrometry

CD measurements were performed on a Jasco J-1500 CD spectrometer (Oklahoma City, OK) equipped with a variable temperature cell holder. Conformational changes in the secondary structure of Alb3-Cterm were monitored in the far-ultraviolet (UV) region between 190 and 250 nm with a protein concentration of 20 μ M in a quartz cuvette with a path length of 1 mm. The scanning speed, band width, and data pitch were set to 50 nm/min and 1.00 and 0.1 nm, respectively. Three scans were taken (within a 1000 HT voltage range) and averaged to obtain the CD spectra.

Deconvolution of protein secondary structure from the CD spectra of wild-type Alb3-Cterm and the single-point mutants was performed using CONTIN, CDSSTR, and SELCON, available with the CDPro program package online (35,36). The CD data from 190 to 250 nm for wild-type Alb3-Cterm and each mutant were input into each of the CDPro algorithms, and the output gave fractional estimations of α -helical content (regular and distorted), β -sheet content (regular and distorted), turns, and disordered regions.

Fluorescence spectra were collected using a Fluorescence Spectrophotometer F-2500 (Hitachi, Tokyo, Japan). Initial intrinsic fluorescence readings of Alb3-Cterm were taken at an excitation of 280 nm, and the emission intensity was recorded within a range from 300 to 450 nm.

Limited trypsin digestion

Limited trypsin digestion of Alb3-Cterm and Alb3-Cterm mutants was performed in HEPES buffer. The initial reaction tube contained 500 μ g of protein and 0.5 μ g of enzyme. The trypsin-containing samples were incubated at room temperature (25°C). Digested samples were removed every 2 min for up to 15 min, and then, the final sample was taken at 20 min. The reaction was stopped via the addition of 10% trichloroacetic acid to each 100 μ L aliquot. Trichloroacetic acid precipitated samples were resolved on a 15% SDS-PAGE gel and subsequently stained using Coomassie Blue (Thermo Fisher Scientific). UN-ScanIT software (Silk Scientific, Orem, UT) was applied to identify the percent of digestion based on the size of the parent band remaining at each time point.

RESULTS

Propensity for local secondary structure in Alb3-Cterm

The first indication of secondary structure in Alb3-Cterm was obtained through a close examination of the far-UV CD spectrum (Fig. 1). This spectrum shows a prominent negative peak centered around 200 nm, indicative of random coil structure (37). However, a small shoulder at 222 nm is also observed (indicated by the arrow in Fig. 1), which suggests the presence of some local helical structure(s) (37). A stable helical structure is typically signified by double minimum at 208 and 222 nm. However, the 208 nm peak is likely buried under the prominent 200-nm ellipticity peak representing random coil behavior. The far UV CD spectrum suggests that the backbone of Alb3-Cterm is predominantly disordered interspersed with weak helical segments.

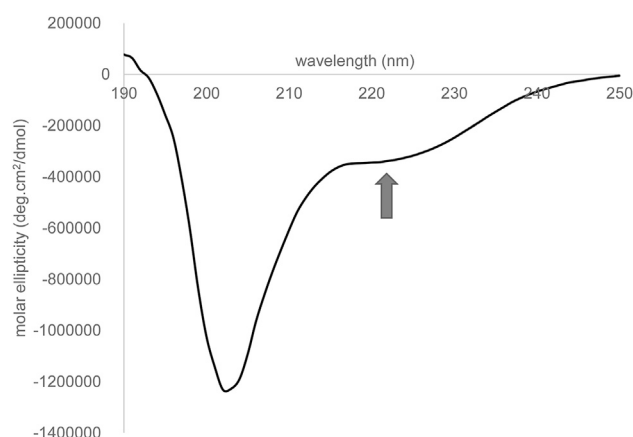


FIGURE 1 Far UV-CD spectrum of Alb3-Cterm. The arrow indicates that a small amount of helical secondary structure may be present.

We then used various common software program servers to predict if secondary structure was possible from the amino acid sequence of Alb3-Cterm. Results from PONDR, PSIPRED, DISOPRED, PrDOS2, Dynamine, and FoldUnfold indicated a consistent prediction of local structure near the N-terminal of Alb3-Cterm (Figs. S1–S5). The complete sequence of Alb3-Cterm is shown in Fig. 2, with the consensus for the predicted ordered region of interest as shown in bold. In this sequence numbering scheme, motif I occurs between residues 15 and 34, motif II occurs between residues 43 and 52, motif III occurs between residues 60 and 77, and motif IV occurs between residues 126 and 136.

Next, we employed an extensive set of all-atom MD simulations in conjunction with state-of-the-art enhanced sampling techniques to characterize the helical propensity of Alb3-Cterm in an explicit aqueous solvent environment, as described in the Materials and methods. Conformations derived from over 5 μ s of enhanced MD simulations were used to calculate the helical propensity of individual Alb3-Cterm amino acid residues (Fig. 3). The regions of highest helical propensity generally agree with the regions predicted by the less-accurate sequence-based secondary structure software used above but provides much more quantitative analysis. Fig. 3 predicts moderate helical propensities (5–20%) between residues 8 and 50, corresponding to

motifs I and II, although some residues in this region drop down close to zero. In addition, moderate helical propensities located near to the C-terminal end of the sequence were also reported by the simulations between residues 75 and 120, which corresponds to residues between motifs III and IV. However, these regions generally have lower helical propensities than those in the 8–50 residue range. The residue-by-residue helical propensities reported in Fig. 3 are based on the STRIDE algorithm (31). Furthermore, overall helical propensity was determined to be $26 \pm 4\%$ from the estimated free energy profile along overall helical propensity, denoted by “ α -collective variable” used in our enhanced sampling algorithms (Fig. 4). We note that the α -collective variable uses a slightly different definition for helical propensity as compared with that used in STRIDE algorithm, as described in the Materials and methods.

smFRET of Alb3-Cterm shows the presence of local secondary structure that can be disrupted by chemical denaturation

By placing an FRET pair of fluorescent dyes at two residues encompassing the predicted helical region, smFRET on freely diffusing Alb3-Cterm was used to estimate the distance between the dye positions in the protein. The top panel in Fig. 5 A shows the smFRET histogram of the Alb3-Cterm under native-like (nondenaturing) conditions. Two populations of FRET efficiencies are evident. The low FRET efficiency peak centered around zero results from statistically labeled donor-only Alb3-Cterm and provides no structural information. The high-FRET peak ($E \approx 0.9$) is from the dual-labeled population. Using this 0.9-FRET-value, an experimentally determined Förster distance for these dyes, R_0 , of 57 Å and applying the equation derived in (38) relating FRET efficiency to end-to-end distance for a disordered protein results in an interdyne distance of 25 Å. Even without taking into account the length of the dye linker, this is shorter than the ~ 42 -Å separation expected for a 38-residue disordered protein as reported for chemically denatured proteins (39) or intrinsically disordered proteins (40,41) (converting R_G to end-to-end distance, R , using $R = \sqrt{6} \times R_G$ (42)), and the disagreement is only worse once the dye linker length is added. This disagreement

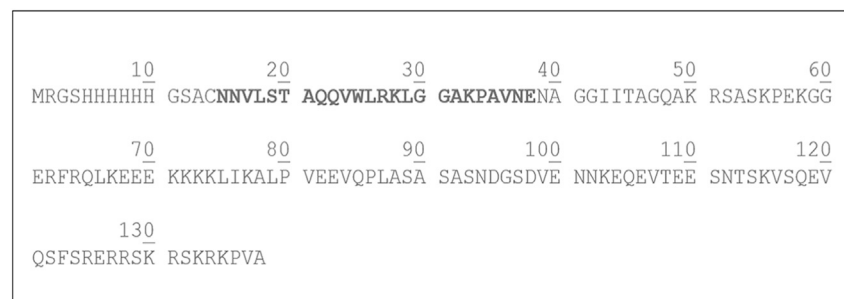


FIGURE 2 Amino acid sequence of Alb3-Cterm. The amino acids highlighted in bold depict the common region of predicted structure. The His-tag is also shown for completeness and included in the numbering scheme. The underlined numbers represent each tenth amino acid residue used in our numbering scheme.

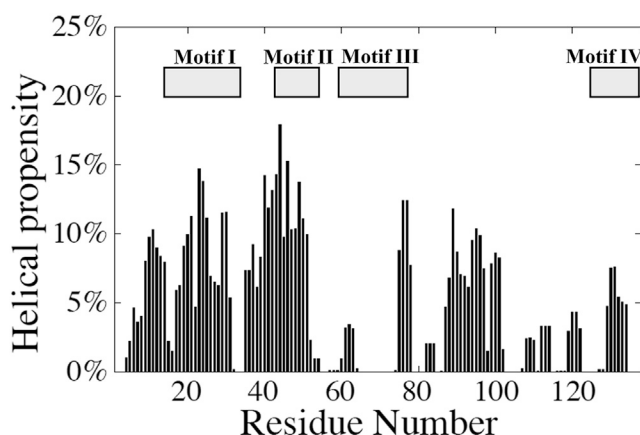


FIGURE 3 Helical propensity per residue from MD simulations. STRIDE algorithm (31) was used to calculate the helical propensity of individual amino acid residues. The region of interest (residues 15–38) has a peak helical propensity of around 15%.

suggests that the protein is more compact or structured than complete disorder or random coil would suggest, in agreement with that expected from Fig. 3. It is also interesting to note the broadness of the high-FRET peak which could indicate that as well as a relatively compact conformation, there is also flexibility and/or disorder in this region. Because the diffusion of the protein through the confocal volume is ~ 1 ms, this would indicate that such transitions between multiple relatively compact states is on timescales slower than 1 ms, indicating that these states are somewhat stable and may further indicate multiple possible sources of secondary structure, as suggested by the various degrees of helical propensity shown in Fig. 3. This point will be verified and discussed in more detail below.

To further investigate the existence of secondary structure in this region, a denaturation titration was performed using guanidine hydrochloride. As shown in Fig. 5 A, as denaturant is titrated in, the high-FRET peak from the original conformation begins to decrease and a new peak at mid-FRET efficiencies appears. In the smFRET histograms, three populations were fitted. The peak at very low FRET efficiencies represents the donor-only labeled population of molecules and was fitted with a lognormal function with parameters corresponding to the smFRET signal of a separately measured donor-only labeled Alb3-Cterm. The high-FRET peak was also fitted to a lognormal function (reversed in its asymmetry), with parameters associated with the FRET-labeled protein under non-denaturing conditions. A Gaussian peak was used to fit the subpopulation of denatured molecules at intermediate FRET. Using these fits, the donor-only peak was removed to highlight the changes in smFRET, as shown in Fig. 5 B, in which the data are overlaid, and the histogram bars are removed for clarity. The disappearance of the high-FRET peak to be replaced with a mid-FRET peak is strongly indicative that some structure was present in normal buffer and that the pro-

tein becomes more disordered in this region as denaturant concentration increases. Furthermore, we observed a denaturant-induced expansion of this completely unfolded state (Fig. 5 C), in agreement with previous reports on expansion of unfolded random coils from globular proteins (38,43,44). The average smFRET of this completely unfolded state is ~ 0.7 , which as described above, using a Förster distance for these dyes, R_0 , of 57 Å, translates to an interdyne distance of 42 Å using the equation derived in (38). This agrees perfectly with that expected for a random coil for chemically denatured proteins (39) or intrinsically disordered proteins (40,41) indicating that this mid-FRET state is indeed a random coil undergoing denaturant-induced expansion.

Additionally, we dialyzed the denaturant from the labeled Alb3-Cterm to determine if the denaturation was reversible (Fig. 5, A and B, *renatured*). The high-FRET peak of the native structure recovered, and the intermediate FRET peak vanished, although a significant fraction of the total number of FRET-labeled molecules was lost, presumably due to aggregation and/or time-dependent proteolytic degradation of the protein.

By comparing the integrated area of the high-FRET peak and intermediate FRET peak fits, the apparent free energy change associated with the “unfolding” process can be determined. For any given concentration of denaturant, the free energy difference between folded and unfolded states, ΔG_{FU} , is given by Eq. 4,

$$\Delta G_{FU} = -RT \ln \frac{U}{F}, \quad (4)$$

where U and F are the fractions of the unfolded and folded states, respectively, R is the ideal gas constant, and T is the temperature. Although this protein is never actually “folded” in the same sense that a globular protein is, we can obtain an approximate free energy of the transient secondary structure by assuming that the apparent free energy difference between the “unfolded” and “folded” states decreases linearly as denaturant increases. Thus, the apparent free energy difference in the Alb3-Cterm under native conditions, ΔG_0 , can be determined according to Eq. 5, as described previously (45,46).

$$\Delta G_{FU} = \Delta G_0 - m[GdnHCl] \quad (5)$$

By plotting the ΔG_{FU} versus denaturant concentration, the y-intercept yields ΔG_0 , and the slope yields the cooperativity, m, for the unfolding process, as previously shown. Fig. 5 D shows this plot and yields a free energy difference between the “folded” and “unfolded” states, $\Delta G_0 = 1.1 \pm 0.1$ kcal/mol and a cooperativity, $m = 0.78 \pm 0.07$ kcal/(mol·M). This small energy difference is consistent with low stability of secondary structure in the region. The small cooperativity value also indicates that the secondary structure is stabilized only slowly with removal of denaturant,

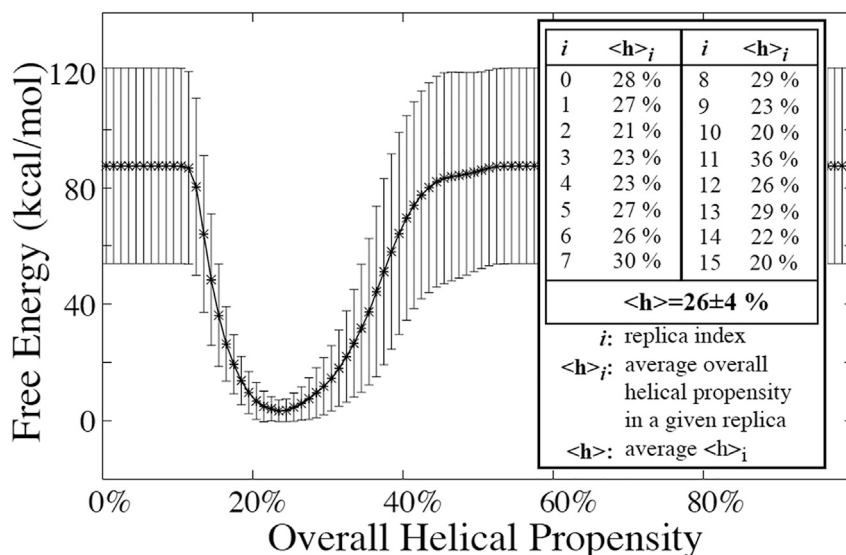


FIGURE 4 Overall helical propensity from MD simulations, as described in [Materials and methods](#). Average overall helical propensities from each of the 16 independent Alb3-Cterm MD simulations are shown inset. The error bars represent the standard deviation based on the 16 independent simulations.

indicating a low propensity for secondary structure and perhaps only transient formation.

It should be highlighted that we use the term “apparent free energy” here because we do not think that this protein is truly described as a two-state model in the classical sense of the unfolding of a globular protein. The broadness of both the high-FRET and the mid-FRET peaks suggest that each of these populations are best described as an ensemble of conformational states rather than a single state, which is also supported by our MD simulations. We use the unfolded and folded terms in [Eq. 4](#) to obtain an estimate of the free energy associated with secondary structure disruption of a somewhat already disordered protein, and so, these numbers should not be directly compared with those of globular proteins. However, to justify the use of [Eq. 4](#), we need to ensure that there is no interconversion between the high-FRET peak and the mid-FRET peak on the 1-ms timescale of the smFRET binning that would affect the shapes and positions of these peaks. If so, it can be assumed that there is a high enough energy barrier between the sets of conformations underlying the high-FRET molecules (containing some secondary structure) and the mid-FRET molecules (consisting of a random coil) that a two-state model can be used to extract an approximate free energy difference between the set of conformations containing transient secondary structure and the random coil. In the [Supporting materials and methods](#), we show that varying the binning time of the smFRET experiment between 1 ms and 200 μ s does not affect the shapes of the smFRET histograms, indicating that there is no interconversion between these sets of conformations on the submicrosecond timescale. The low stability and heterogeneity of the secondary structure observed leads us to use the term “transient secondary structure” to highlight these aspects.

Experimental smFRET and all-atom MD simulations of dye-dye distance distributions are in very close agreement

The relationship between the smFRET efficiency of Alb3-Cterm at 0.0 M denaturant shown in [Fig. 5 B](#) (black curve), and the dye-dye distance was calculated using the Gaussian chain model, as previously described ([38](#)), to provide $E(r)$. Then, the probability of dye-dye distance, $P(r)$, was found using [Eq. 6](#).

$$P(r) = P(E(r)) \frac{dE(r)}{dr} \quad (6)$$

[Fig. 6](#) shows an overlay of the experimental $P(r)$ -values of Alb3 C-term between the dyes at residues 14 and 52 (solid line) and the independently calculated dye-dye distances derived from the enhanced MD simulations (dashed line), highlighting the quantitatively consistent picture between simulation and experiment. The MD-based distance distributions shown in [Fig. 6](#) are generated using 16 separate simulations, and three example conformations that result in a dye-dye distance at the most probable position are shown, which exemplifies the transient nature of the secondary structure (helices shown in purple).

In addition to the dye-dye distance distribution, we also estimated the unfolding apparent free energy associated within residues 14–52 from the STRIDE-based helical propensities ([Fig. 3](#)) using [Eq. 7](#), which is very similar to [Eq. 4](#).

$$\Delta G_0 = -RT \ln \frac{p_U}{p_F} \quad (7)$$

Here, p_F quantifies the probability of any residue in the 14–52 region to be structured, and $p_U = 1 - p_F$. Based on the data shown in [Fig. 3](#), we estimate $\Delta G_0 = 1.5 \pm 0.4$ kcal/mol,

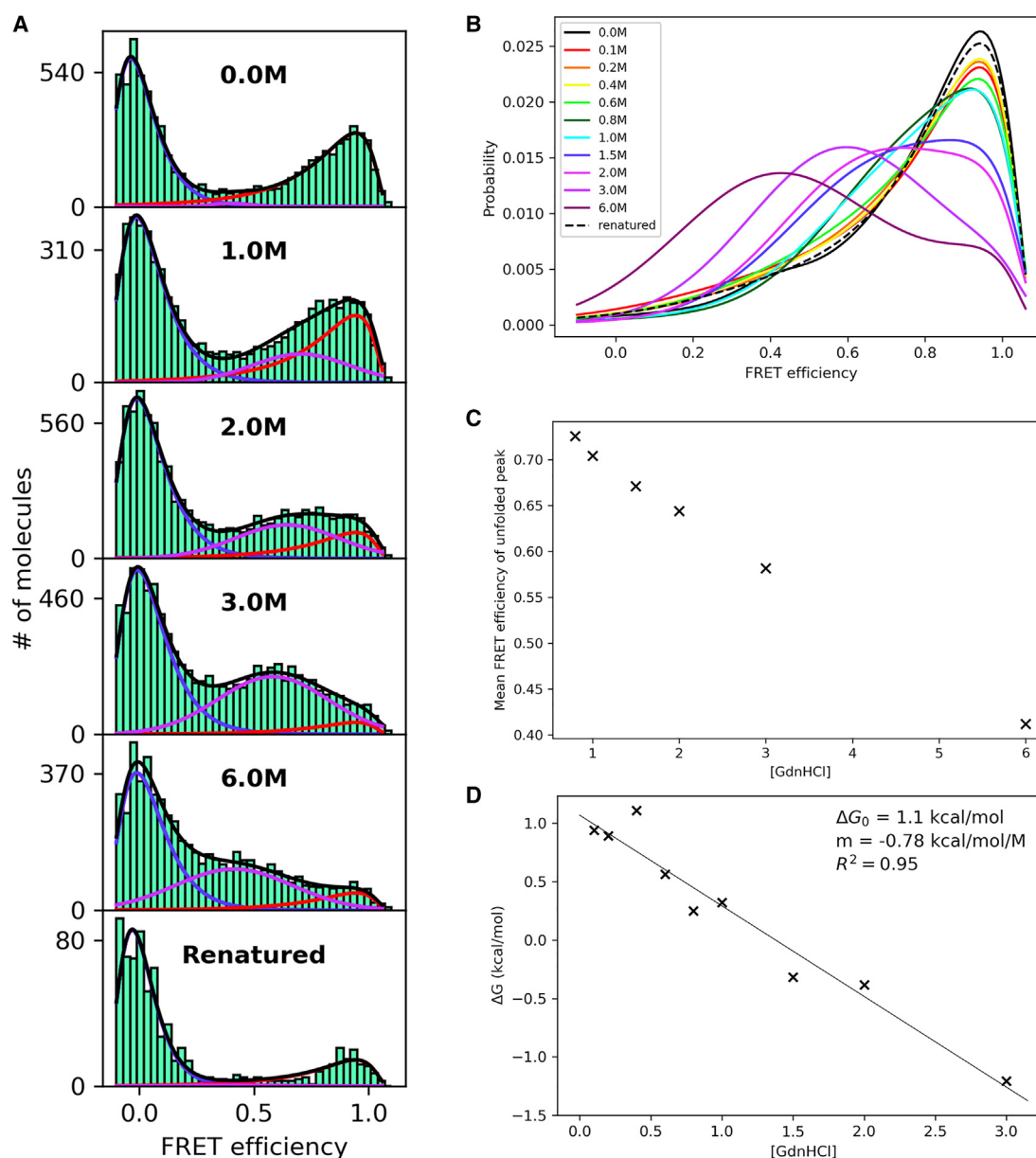


FIGURE 5 (A) Selected smFRET histograms of S52C Alb3-Cterm in increasing guanidine hydrochloride. (B) Fit curves with the donor-only peak removed of the titration smFRET histograms. (C) Mean FRET efficiency of the mid-FRET peak associated with the unfolded state showing denaturant-induced expansion of the random coil. (D) Plot of the unfolding energy versus denaturant concentration to yield ΔG_0 and cooperativity, m . To see this figure in color, go online.

which is in good agreement with the experimental smFRET-based measurement of $\Delta G_0 = 1.1 \pm 0.1$ kcal/mol (Fig. 5 C).

Given that the MD simulations did not use any experimental information from the smFRET experiments above, the agreement between the experimental data and simulations is remarkable and supports our, to our knowledge, novel simulation approach for generating an ensemble of Alb3-Cterm conformations, imparting confidence in our detailed atomic-level descriptions of Alb3-Cterm structural ensembles.

Helix-breaking mutations result in loss of local secondary structure

Four single-point mutations were designed and analyzed with far UV-CD, intrinsic fluorescence, and limited trypsin digestion to probe for changes in structure and stability. Four amino acids, A21, V24, L26, and K28, were selected based on their ability to promote helix formation in the high helical propensity region identified near the N-terminus of Alb3-Cterm, and four mutants were constructed in which each of these residues were separately mutated to

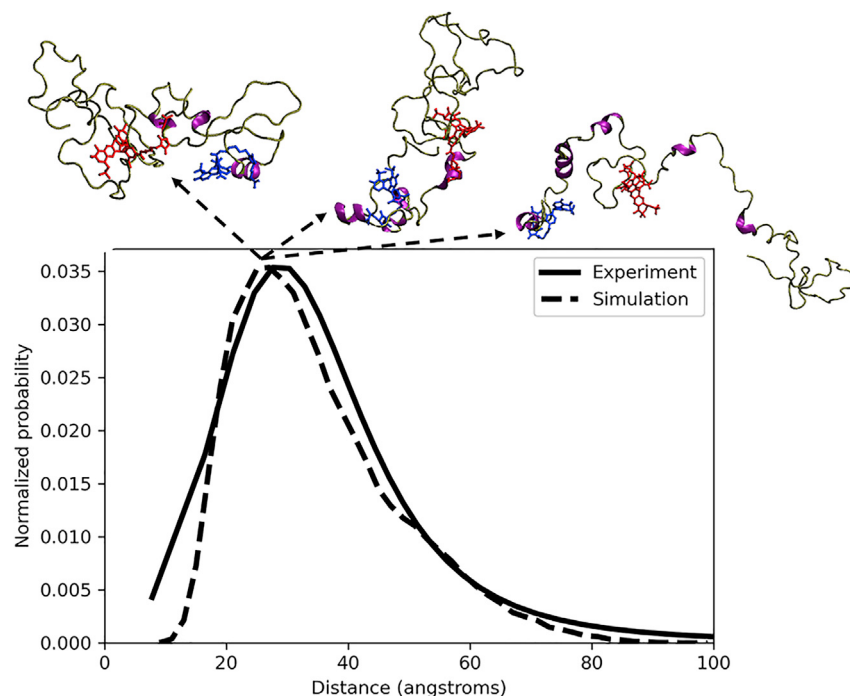


FIGURE 6 Comparison of experimental smFRET, converted into dye-dye distance using the Gaussian chain model as previously described (38), and dye-dye distance distributions from enhanced MD simulations. Three representative converged conformations representing the most probable dye-dye distance are depicted above the distribution plot, with transient helical regions highlighted in purple, unstructured regions shown in green and the dyes shown in blue and red. The transient nature of the helical regions is obvious from their variation from structure to structure. To see this figure in color, go online.

glycine, a known helix-breaking residue (47–49). Although the CD spectrum reports on the whole protein, changes in the CD spectrum upon site-directed mutagenesis of a single residue allows us to determine if any local secondary structure is affected, and thus, we rationalize that the changes in the CD spectrum upon these mutations will allow us to determine the relative local secondary structural composition.

The far UV-CD spectra of wild-type Alb3-Cterm and the four mutants are shown in Fig. 7 A. The A21G, V24G, and L26G mutants each clearly show a weaker shoulder at ~222 nm together, with a narrowing of the peak around 208 nm when compared with wild-type Alb3-Cterm. Negative ellipticity at these wavelengths are generally associated with helical structure, and thus, the data imply at least some loss of helical secondary structure for the A21G, V24G, and L26G mutants compared with wild-type Alb3-Cterm. The K28G mutant shows smaller differences to wild-type Alb3-Cterm but does show a broadening in the shape of the shoulder around 210–230 nm as well as a weaker slope between 190 and 200 nm. This could indicate more complex changes in the helical, sheet, and random coil composition. Detailed deconvolution of the various CD spectra will be discussed in more detail below. In contrast to the CD spectral changes, the intrinsic fluorescence spectra from the lone tryptophan, which is also in motif I, for the wild-type and all four of the single-point mutations showed insignificant shifts in the wavelength of maximal emission or peak shape (Fig. 7 B). The peaks are centered at ~350 nm, which is indicative of a completely solvent-exposed tryptophan, highlighting the local environment of the tryptophan is signifi-

cantly exposed to solvent for wild-type of Alb3-Cterm and all the helix-breaking mutations. Clearly, tertiary structure is not formed to any measurable degree in Alb3-Cterm, which would be expected for a mostly disordered protein with only transient secondary structure.

Quantification of secondary structural content from the CD spectra of wild-type Alb3-Cterm and the four helix-breaking mutants were analyzed using three different deconvolution algorithms in the CDPro program (Table 1). CONTIN/LL (35), CDSSTR (50), and SELCON3 (36) all provide an estimation of protein secondary structure using different fitting approaches (see Supporting materials and methods for more discussion on the different algorithms). Sreerama et al. found that in most cases, CONTIN/LL provided the best agreement to predictions from x-ray structures, with SELCON3 performing slightly better than CDSSTR (36).

From the data in Table 1, three key observations are immediately evident. First, wild-type Alb3-Cterm contains approximately equal amounts of helical, turn, and disordered composition, with very little sheet structure. Second, the disordered fraction is consistently increased in the helix-breaking mutants, although there is also an increase in the amount of regular β -sheet structure as well when compared with wild-type Alb3-Cterm. Third, the helical fractions contain both regular and distorted helices, with distorted helices being more prevalent. The combined fraction of regular and distorted helices was reduced by ~0.15–0.25 in all the helix-breaking mutants when compared with wild-type, with an increase in both the disordered and regular β -sheet fractions. The fraction of turn structure was comparable in wt-Alb3-Cterm and in each of the mutants. Taken

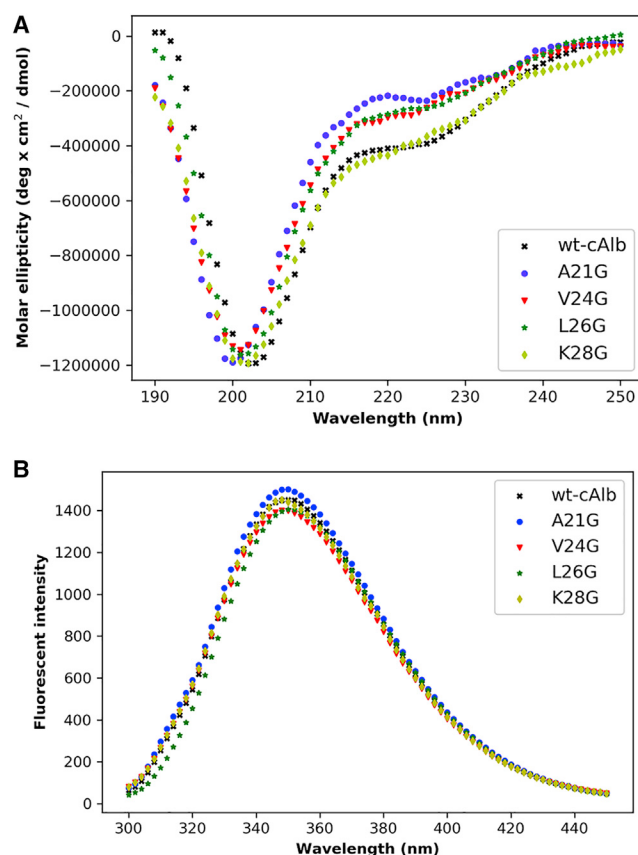


FIGURE 7 The Far-UV CD spectra (A) and intrinsic fluorescence (B) of wt-Alb3-Cterm and the four helix-breaking mutants, A21G, V24G, L26G, and K28G. To see this figure in color, go online.

together, the data suggest that each of the helix-breaking mutants does decrease the helical composition of Alb3-Cterm while also increasing the disorder, but the effects are more complicated than a simple order-disorder transition due to changes in the β -sheet composition. We cannot discriminate at this point if the β -sheet composition is transient or if it is associated with partial aggregation of the proteins upon mutation.

Limited trypsin digestion provides information on the backbone flexibility of proteins. Arginine and lysine residues located in flexible, solvent-accessible regions of proteins are more vulnerable to trypsin cleavage. The results of these digestion assays are shown in Fig. 8, with error bars representing the standard deviation of the stained band intensities from experiments performed in triplicate. wt-Alb3-Cterm contains 46 trypsin cleavage sites, which in conjunction with being predominately unstructured, results in it being highly susceptible to enzymatic cleavage. The rate of digestion is much slower for wt Alb3-Cterm compared with the mutants. After 20 min of incubation with trypsin at 25°C, the wt Alb3-Cterm parent band, migrating at 15 kDa, is reduced to 25% of its original quantity. At the same time point, A21G, K28G, V24G, and L26G are reduced to 8, 5, 0, and 0%, respectively. These results

suggest that the introduction of these single-point mutations led to a significant increase in the backbone flexibility of Alb3-Cterm and, therefore, an overall loss of structural integrity, in agreement with a general loss of secondary structure.

DISCUSSION

In this study, we have examined the complex structural properties of Alb3-Cterm, a predominately intrinsically disordered protein and discovered a region of ~ 30 amino acids long with moderate propensity for secondary structure formation. By comparing single-molecule FRET distributions with predictions from an ensemble of all-atom structures derived from enhanced sampling, all-atom MD simulations, we have uncovered the transient nature of this secondary structure, which rationalizes its low abundance measured in the CD spectrum. We further showed that the secondary structure can be interrupted by chemical denaturant or by using single-point helix-breaking mutations and trypsin digestion assays.

The region in Alb3-Cterm that we have identified here to show transient secondary structure correlates to motifs I and II, which have been argued to play an important role in the integration of LHCP into thylakoid membranes (13). However, other studies have suggested that motif I is not so important but that motifs II and especially IV are (9,12). This debate is also complicated by the fact that different interaction domains on cpSRP43 have been proposed. In one study, the Ankyrin-repeat domain of cpSRP43 was shown to have a strong affinity (~ 200 nM) to Alb3 C-term (8), whereas other studies have shown that the CD2-CD3 domains of cpSRP43 interacts with Alb3-Cterm, albeit with a lower affinity (~ 20 μ M) (9,12). NMR data of Alb3-Cterm consisting of motifs II–IV suggested that CD3 was the primary interaction site for these motifs (12), although removal of the CD3 domain from cpSRP43 only slightly reduced the affinity of cpSRP43 for full-length Alb3-Cterm (53 μ M) (9). Furthermore, it had been previously shown that deletion of the CD3 domain from cpSRP43 does not inhibit LHCP integration into the thylakoid membrane (51). It is possible that the transient secondary structure in motifs I–II observed in this study, and the fact that single-point mutations can inhibit such formation may account for some of these inconsistencies, but additional research may be needed to extract out the exact functional implications. The ability of our, to our knowledge, novel all-atom enhanced MD simulations to provide conformational structural ensembles that show locations of transient secondary structure formation, which was in excellent agreement with experimental smFRET distributions and site-directed mutagenesis assays, is expected to provide an excellent framework to better identify important, specific regions in Alb3-Cterm. These regions will be tested for functional implications both in binding cpSRP subunits and LHCP, as well as their roles

TABLE 1 Estimation of protein secondary structure from far-UV CD spectra of wt-Alb3-Cterm and the mutants and the percent of structured and unordered content was calculated using CONTINLL, CDSSTR, and SELCON3

	Helix (regular)	Helix (distorted)	Sheet (regular)	Sheet (distorted)	Turn	Disordered
Wild-type						
CONTIN/LL	0.149	0.244	0.001	0.072	0.270	0.265
CDSSTR	0.193	0.244	0.076	0.067	0.181	0.254
SELCON3	0.138	0.209	0.013	0.063	0.259	0.329
A21G						
CONTIN/LL	0.037	0.133	0.113	0.093	0.237	0.388
CDSSTR	0.010	0.057	0.165	0.105	0.242	0.414
SELCON3	0.063	0.111	0.132	0.080	0.224	0.409
V24G						
CONTIN/LL	0.038	0.129	0.076	0.079	0.257	0.421
CDSSTR	0.039	0.132	0.126	0.089	0.260	0.356
SELCON3	0.065	0.123	0.099	0.092	0.242	0.409
L26G						
CONTIN/LL	0.065	0.152	0.140	0.083	0.224	0.336
CDSSTR	0.055	0.152	0.110	0.093	0.238	0.353
SELCON3	0.085	0.142	0.119	0.074	0.229	0.347
K28G						
CONTIN/LL	0.046	0.162	0.043	0.082	0.274	0.393
CDSSTR	0.074	0.171	0.117	0.086	0.226	0.324
SELCON3	0.076	0.177	0.020	0.073	0.255	0.387

in subsequent integration of LHCP into thylakoid membranes. Furthermore, because of the similarity in function of the YidC-Oxa1-Alb3 protein family, our results may imply that although YidC and Oxa1 have structured solvent-exposed but flexible regions (52,53) and Alb3 is mostly disordered (9) but shown here to contain transient secondary structure, this class of proteins may have a common structural basis for substrate recognition underlying both co- and post-translational protein targeting.

CONCLUSIONS

We have used, to our knowledge, a novel combination of single-molecule FRET and all-atom MD simulations that take advantage of enhanced sampling techniques to identify local transient secondary structure in the intrinsically disordered C-terminal region of the Alb3 insertase, a key membrane protein for targeting light-harvesting proteins to the thylakoid membrane. The excellent agreement between the experimental and computational interfluorophore distributions of single Alb3-Cterm molecules and the unfolding apparent free energy of the secondary structure allowed us to identify a structural ensemble that highlights the transient and local nature of the secondary structure formation. Single-point helix-breaking mutations, CD spectral deconvolution, and trypsin digestion assays all support the conclusion of local transient secondary structure close to the N-terminus of Alb3-Cterm. As stated in the introduction, there is a great deal of debate as to which of the motifs I–IV are important for cpSRP43 binding and LHCP integration. The observation of transient secondary structure reported here will provide the basis for future studies on the relative importance of such structure in each of the motifs to explain

the basis for the conflicting results, such as examining specific single- and multiple-point helix-breaking mutations in each of the motifs on the binding of cpSRP43 to Alb3 C-term and the integration of LHCP. More generally, these results provide a structural framework to further examine the possible role of local transient secondary structure in otherwise intrinsically disordered regions of proteins that play important roles in protein targeting, such as those of the YidC-Oxa1-Alb3 protein family.

SUPPORTING MATERIAL

Supporting material can be found online at <https://doi.org/10.1016/j.bpj.2021.10.013>.

AUTHOR CONTRIBUTIONS

D.R.B. performed the research, analyzed the data, and wrote the manuscript. M.F. performed the research and analyzed the data. V.G.K. performed the research and analyzed the data. P.O. performed the research and analyzed the data. J.L.L. performed the research and analyzed the data. R.L.H. designed the research and wrote the manuscript. M.M. designed the research and wrote the manuscript. T.K.S.K. designed the research and wrote the manuscript. C.D.H. designed the research and wrote the manuscript.

ACKNOWLEDGMENTS

This research was generously supported by the U.S. Department of Energy (grant no. DE-FG02-01ER15161) and the Arkansas Biosciences Institute. This research is also supported by National Science Foundation grants CHE 1945465 and OAC 1940188. This research is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. This work also used the Extreme Science and Engineering

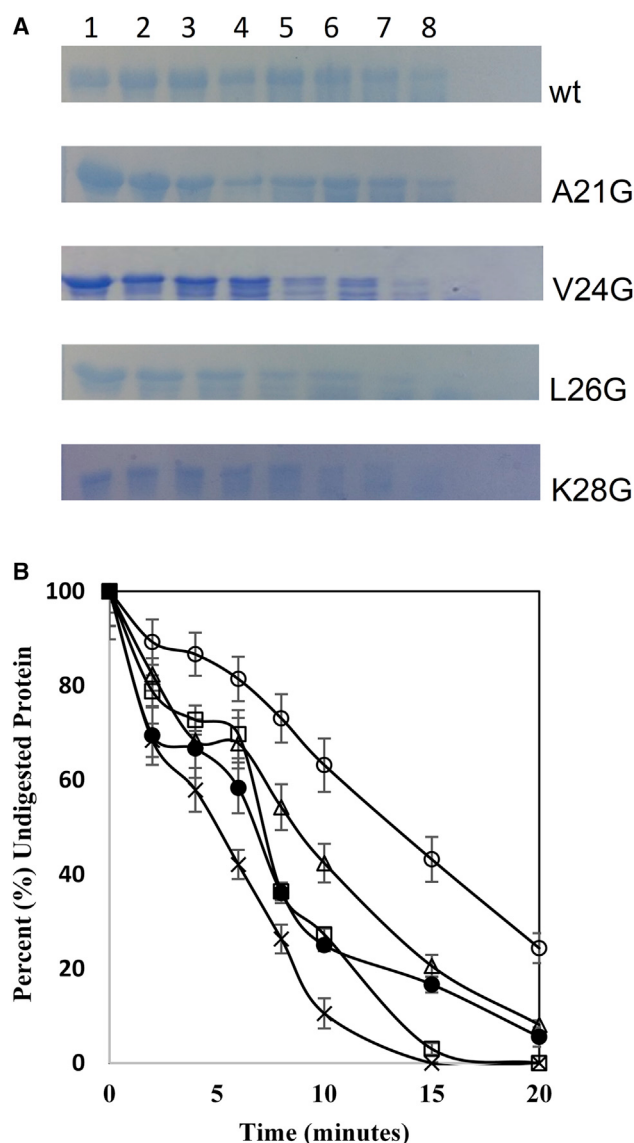


FIGURE 8 (A) SDS-PAGE gels of limited trypsin digestion for Alb3-Cterm and mutants. Shown are the following: lane 1, zero trypsin; lane 2, 2 min; lane 3, 4 min; lane 4, 6 min; lane 5, 8 min; lane 6, 10 min; lane 7, 15 min; and lane 8, 20 min. (B) Densitometric analysis of the SDS-PAGE gels depicting the resistance of wt-Alb3-Cterm (open circle (○)), A21G (open triangle (Δ)), V24G (open square (□)), L26G (cross (x)), and K28G (solid circle (●)) to limited trypsin digestion. To see this figure in color, go online.

Discovery Environment (allocation MCB150129), which is supported by National Science Foundation grant number ACI-1548562.

REFERENCES

- Wang, P., and R. E. Dalbey. 2011. Inserting membrane proteins: the YidC/Oxa1/Alb3 machinery in bacteria, mitochondria, and chloroplasts. *Biochim. Biophys. Acta.* 1808:866–875.
- Moore, M., M. S. Harrison, ..., R. Henry. 2000. Chloroplast Oxa1p homolog albino3 is required for post-translational integration of the light

- harvesting chlorophyll-binding protein into thylakoid membranes. *J. Biol. Chem.* 275:1529–1532.
- Payan, L. A., and K. Cline. 1991. A stromal protein factor maintains the solubility and insertion competence of an imported thylakoid membrane protein. *J. Cell Biol.* 112:603–613.
- DeLille, J., E. C. Peterson, ..., R. Henry. 2000. A novel precursor recognition element facilitates posttranslational binding to the signal recognition particle in chloroplasts. *Proc. Natl. Acad. Sci. USA.* 97:1926–1931.
- Tu, C. J., D. Schuenemann, and N. E. Hoffman. 1999. Chloroplast FtsY, chloroplast signal recognition particle, and GTP are required to reconstitute the soluble phase of light-harvesting chlorophyll protein transport into thylakoid membranes. *J. Biol. Chem.* 274:27219–27224.
- Moore, M., R. L. Goforth, ..., R. Henry. 2003. Functional interaction of chloroplast SRP/FtsY with the ALB3 translocase in thylakoids: substrate not required. *J. Cell Biol.* 162:1245–1254.
- Króliczewski, J., M. Piskozub, ..., B. Króliczewska. 2016. ALB3 insertase mediates cytochrome b6 co-translational import into the thylakoid membrane. *Sci. Rep.* 6:34557.
- Lewis, N. E., N. J. Marty, ..., R. L. Goforth. 2010. A dynamic cpSRP43-Albino3 interaction mediates translocase regulation of chloroplast signal recognition particle (cpSRP)-targeting components. *J. Biol. Chem.* 285:34220–34230.
- Falk, S., S. Ravaut, ..., I. Sinning. 2010. The C terminus of the Alb3 membrane insertase recruits cpSRP43 to the thylakoid membrane. *J. Biol. Chem.* 285:5954–5962.
- Chandrasekar, S., and S. O. Shan. 2017. Anionic phospholipids and the Albino3 translocase activate signal recognition particle-receptor interaction during light-harvesting chlorophyll a/b-binding protein targeting. *J. Biol. Chem.* 292:397–406.
- Bals, T., B. Dünschede, ..., D. Schünemann. 2010. Interplay between the cpSRP pathway components, the substrate LHCP and the translocase Alb3: an in vivo and in vitro study. *FEBS Lett.* 584:4138–4144.
- Horn, A., J. Hennig, ..., I. Sinning. 2015. Structural basis for cpSRP43 chromodomain selectivity and dynamics in Alb3 insertase interaction. *Nat. Commun.* 6:8875.
- Urbischek, M., S. N. von Braun, ..., S. Schwenkert. 2015. The extreme Albino3 (Alb3) C terminus is required for Alb3 stability and function in *Arabidopsis thaliana*. *Planta.* 242:733–746.
- Xue, B., R. L. Dunbrack, ..., V. N. Uversky. 2010. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta.* 1804:996–1010.
- Ward, J. J., L. J. McGuffin, ..., D. T. Jones. 2004. The DISOPRED server for the prediction of protein disorder. *Bioinformatics.* 20:2138–2139.
- Ishida, T., and K. Kinoshita. 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 35:W460–4.
- Jones, D. T., and J. J. Ward. 2003. Prediction of disordered regions in proteins from position specific score matrices. *Proteins.* 53 (Suppl 6):573–578.
- Cilia, E., R. Pancsa, ..., W. F. Vranken. 2013. From protein sequence to dynamics and disorder with DynaMine. *Nat. Commun.* 4:2741.
- Galzitskaya, O. V., S. O. Garbuzynskiy, and M. Y. Lobanov. 2006. FoldUnfold: web server for the prediction of disordered regions in protein chain. *Bioinformatics.* 22:2948–2949.
- Liang, S., C. Zhang, and Y. Zhou. 2014. LEAP: highly accurate prediction of protein loop conformations by integrating coarse-grained sampling and optimized energy scores with all-atom refinement of backbone and side chains. *J. Comput. Chem.* 35:335–341.
- Eswar, N., D. Eramian, ..., A. Sali. 2008. Protein structure modeling with MODELLER. *Structural Proteomics: High-Throughput Methods.* Humana Press, pp. 145–159.
- Barducci, A., G. Bussi, and M. Parrinello. 2008. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* 100:020603.

23. Jorgensen, W. L., J. Chandrasekhar, ..., M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
24. Phillips, J. C., R. Braun, ..., K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.
25. Best, R. B., X. Zhu, ..., A. D. Mackerell, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J. Chem. Theory Comput.* 8:3257–3273.
26. Huang, J., S. Rauscher, ..., A. D. MacKerell. 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods.* 14:71–73.
27. Reid, J. 1971. Large Sparse Sets of Linear Equations. Academic Press, London, NY.
28. Martyna, G. J., D. J. Tobias, and M. L. Klein. 1994. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* 101:4177–4189.
29. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: An $N \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
30. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.
31. Andersen, C. A., A. G. Palmer, ..., B. Rost. 2002. Continuum secondary structure captures protein flexibility. *Structure.* 10:175–184.
32. Henderson, R. C., F. Gao, ..., T. K. Suresh Kumar. 2016. Domain organization in the 54-kDa subunit of the chloroplast signal recognition particle. *Biophys. J.* 111:1151–1162.
33. Gao, F., A. D. Kight, ..., C. D. Heyes. 2015. Regulation of structural dynamics within a signal recognition particle promotes binding of protein targeting substrates. *J. Biol. Chem.* 290:15462–15474.
34. Vanommeslaeghe, K., E. Hatcher, ..., A. D. Mackerell, Jr. 2010. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* 31:671–690.
35. Provencher, S. W. 1982. CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Comput. Phys. Commun.* 27:229–242.
36. Sreerama, N., and R. W. Woody. 2000. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* 287:252–260.
37. Greenfield, N. J. 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* 1:2876–2890.
38. Kuzmenkina, E. V., C. D. Heyes, and G. U. Nienhaus. 2006. Single-molecule FRET study of denaturant induced unfolding of RNase H. *J. Mol. Biol.* 357:313–324.
39. Kohn, J. E., I. S. Millett, ..., K. W. Plaxco. 2004. Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. USA.* 101:12491–12496.
40. Receveur-Brechot, V., and D. Durand. 2012. How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Sci.* 13:55–75.
41. Bernadó, P., and M. Blackledge. 2009. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophys. J.* 97:2839–2845.
42. Flory, P. J. 1953. Principles of Polymer Chemistry. Cornell University Press, Ithaca, NY.
43. Schuler, B., E. A. Lipman, and W. A. Eaton. 2002. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature.* 419:743–747.
44. Kuzmenkina, E. V., C. D. Heyes, and G. U. Nienhaus. 2005. Single-molecule Förster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc. Natl. Acad. Sci. USA.* 102:15471–15476.
45. Amigoulova, E. V., J. Groll, ..., G. U. Nienhaus. 2004. Bio-functionalized polymer surfaces exhibiting minimal interaction towards immobilized proteins. *ChemPhysChem.* 5:552–555.
46. Alonso, D. O., and K. A. Dill. 1991. Solvent denaturation and stabilization of globular proteins. *Biochemistry.* 30:5974–5985.
47. Dong, H., M. Sharma, ..., T. A. Cross. 2012. Glycines: role in α -helical membrane protein structures and a potential indicator of native conformation. *Biochemistry.* 51:4779–4789.
48. Chakrabarty, A., J. A. Schellman, and R. L. Baldwin. 1991. Large differences in the helix propensities of alanine and glycine. *Nature.* 351:586–588.
49. Pace, C. N., and J. M. Scholtz. 1998. A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* 75:422–427.
50. Johnson, W. C. 1999. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins.* 35:307–312.
51. Goforth, R. L., E. C. Peterson, ..., R. L. Henry. 2004. Regulation of the GTPase cycle in post-translational signal recognition particle-based protein targeting involves cpSRP43. *J. Biol. Chem.* 279:43077–43084.
52. Kumazaki, K., S. Chiba, ..., O. Nureki. 2014. Structural basis of Sec-independent membrane protein insertion by YidC. *Nature.* 509:516–520.
53. Jia, L., M. Dienhart, ..., R. A. Stuart. 2003. Yeast Oxa1 interacts with mitochondrial ribosomes: the importance of the C-terminal region of Oxa1. *EMBO J.* 22:6438–6447.