
Fabian Dingfelder,† Stephan Benke,† Daniel Nettels,‡ and Benjamin Schuler*†‡

†Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
‡Department of Physics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

ABSTRACT: The 303-residue cytolytic toxin ClyA forms a stable α-helical monomer. In the presence of detergents or membranes, however, the protein makes a large conformational transition to the protomer state, which is competent for assembly into a dodecameric cytolytic pore. In this study, we map the structure of the ClyA monomer during denaturant-induced unfolding with single-molecule Förster resonance energy transfer (FRET) spectroscopy. To this end, we probe intramolecular distances of six different segments of ClyA by placing donor and acceptor fluorophores at corresponding positions along the chain. We identify an intermediate state that contains the folded core consisting of three of the α-helices that make up the helical bundle present in the structure of both the monomer and the protomer, but with the C- and N-terminal helices unfolded, in accord with the secondary structure content estimated from circular dichroism (CD) spectroscopy. The existence of this intermediate is likely to be a consequence of the structural bistability underlying the biological function of ClyA: The terminal helices are part of the largest rearrangements during protomer formation, and the local differences in stability we detect may prime the protein for the required conformational transition.

INTRODUCTION

During the past two decades, our understanding of the elementary processes of protein folding has improved tremendously. These advances are largely due to a focus on single-domain proteins that typically exhibit fast folding and can be well described by two-state models that enable a quantitative thermodynamic and kinetic analysis.1,2 By dissecting such systems into their secondary structure building blocks, combined with methods from ultrafast spectroscopy,3 other experimental techniques, and the continued advances in simulation and theory,3−10 we now have a remarkably detailed picture of many protein folding processes from picoseconds to milliseconds. These developments in conceptual understanding and experimental tools now enable us to revisit questions that had been the focus of protein folding research in its early days, where most of the model systems were relatively large, multidomain, and often oligomeric proteins.7 Already 40 years ago, it was clear that many of these proteins populate folding intermediates that could be detected experimentally,9 and many of the ideas and methods used to investigate biomolecules to this day were established for identifying and characterizing such intermediates. However, we can now take advantage of methods such as single-molecule spectroscopy,9−16 which were not available at that time.

Folding intermediates have been observed under a variety of conditions, such as low pH or at moderate concentrations of denaturant.7,6,17,18 Alternatively, equilibrium intermediate states can be populated by introducing mutations that destabilize the native state or by the removal of ligands. For example, CD spectroscopy of cytochrome c, β-lactamase, and apomyoglobin revealed an acid-denatured state around pH 2, which folds to an intermediate state upon further reduction in pH. The characteristics of this compact state were a high content of secondary structure (as identified by far-UV CD) and few tertiary interactions (near-UV CD), a constellation that is often referred to as a molten globules state.18,19 In this and many other cases, however, obtaining more detailed structural information on intermediate states from ensemble techniques is often challenging, especially under conditions where the native state, the intermediate, and the unfolded state coexist. A technique that allows structural information on intermediates to be obtained even at atomic resolution is nuclear magnetic resonance (NMR) spectroscopy. For example, the structure of the intermediate of a small four-helix bundle protein that is sparsely populated at equilibrium (2−3%) could be solved by NMR.20 However, obtaining high-resolution information from NMR becomes increasingly challenging for larger proteins. Another method that enables folding intermediates to be mapped is single-molecule fluorescence spectroscopy, especially in combination with FRET.9,13,21,22 A particular benefit of this technique is the ability to resolve individual subpopulations, so that the
intermediate populations can be separated from the unfolded or native state. Furthermore, single-molecule FRET is a versatile approach for the investigation of protein kinetics and dynamics\(^ {28,41}\) and the properties of conformationally heterogeneous systems, such as unfolded and intrinsically disordered proteins\(^ {25,26}\) or misfolded states.\(^ {28,41}\) Even though the focus of most single-molecule FRET experiments has been the folding of small proteins, the approach has clearly been shown to have the potential for resolving structural and kinetic heterogeneity in the folding of multidomain and oligomeric proteins.\(^ {25–30}\)

In this study, we map the structure of an equilibrium intermediate of the large protein cytolysin A (ClyA, 34 kDa) with single-molecule FRET spectroscopy by placing donor–acceptor pairs into different positions to monitor distance changes in different parts of the protein. ClyA is a pore-forming bacterial toxin that is expressed as a highly α-helical monomer. In the presence of membranes or detergents, the protein undergoes a very large structural transformation into the protomer conformation (Figure 1), which then assembles into the cytolytic pore.\(^ {30}\) The ClyA monomer is a bundle of four long helices (αA, αB, αC, and αF) flanked by a shorter C-terminal helix (αG).\(^ {39}\) Membrane binding is thought to be triggered by its interaction with the “β-tongue”, a short hydrophobic β-hairpin at one end of the protein (Figure 1a).\(^ {32}\) Upon protomer formation, the amphipathic N-terminal helix αA largely displaces from the helix bundle of the monomer and flips upward by 180° to become an extension of helix αB.\(^ {32}\) Concomitantly, large parts of the hydrophobic core are repacked, and the arrangement and extent of the other helices change such that more than half of all residues are affected by the reorganization of the molecule. ClyA thus represents a remarkable bistable system that can interconvert between two very different conformations. The kinetic mechanism of the monomer-to-protomer conversion has previously been investigated, which revealed that the process involves an expanded intermediate state.\(^ {30}\) As for other such metamorphic proteins,\(^ {28,41}\) the existence of two stable conformational states for the same polypeptide sequence suggests a delicate balance of interactions,\(^ {28,41}\) which for a large protein such as ClyA may be expected to lead to detectable differences in conformational stability across the protein. Here we thus investigate ClyA for the presence of equilibrium intermediates that may be related to the conformational change essential for its biological function.

## METHODS

**Protein Expression, Purification, and Labeling.** ClyA was expressed, purified, and labeled as described previously.\(^ {30}\) Briefly, six double-cysteine variants (A2C/Q56C, K107C/N157C, A187C/K290C, Q56C/E252C, T2C/E252C, and T2C/V303C) were produced by site-directed mutagenesis of the pClyA vector coding for N-terminally His\(_6\)-tagged ClyA\(^ {33}\) using the QuickChange protocol (Stratagene). Protein expression was carried out in *E. coli* Tuner DE3 (Merck Millipore) at 20 °C for 12 h. The double-cysteine variants as well as wt-ClyA were purified via immobilized metal affinity chromatography (IMAC) with a nickel-nitrilotriacetic acid (NTA) resin (Thermo Fischer).

wt-ClyA for the CD experiments was buffer-exchanged to 10 mM sodium phosphate (pH 7.4) on a HiPrep desalting 26/100 column (GE Healthcare) and further purified by anion exchange chromatography on a HighPrep Q FF 16/100 (GE Healthcare) applying an elution gradient of 0–1 M NaCl.\(^ {30}\)

The protein was then denatured by adding guanidinium chloride (GdmCl) to a final concentration of ~6 M and concentrated to 230 μM protein. For final purification, size exclusion chromatography on a Superdex 200 Increase 10/300 column (GE Healthcare) was performed, yielding a final ClyA concentration of 50 μM.

Following nickel chelate affinity chromatography, the double-cysteine variants for labeling were concentrated to ~400 μM and reduced by adding Dithiothreitol (DTT) to a final concentration of 10 mM. The reducing agent was removed by a HiTrap desalting column (GE Healthcare) using a buffer containing 50 mM sodium phosphate and 150 mM NaCl (pH 7.3). For donor labeling, Alexa Fluor 488 C5 maleimide (Invitrogen) dissolved in dimethyl sulfoxide (DMSO) was added in a dye-to-protein molar ratio of 1:0.7 with the protein concentration being ~100–300 μM. The reaction was carried out at room temperature for 2 h and then quenched by the addition of DDT to a final concentration of 50 mM. After removal of free dye and buffer exchange to

![Figure 1](image-url)
10 mM sodium phosphate (pH 7.3) via a HiTrap desalting column (GE Healthcare), the singly labeled protein was separated from unlabeled and doubly labeled protein by anion exchange chromatography, using a Mono Q GL anion exchange column (GE Healthcare). The protein was concentrated to \( \sim 100 \mu M \), and Alexa Fluor S94 C5 maleimide (Invitrogen), dissolved in DMSO, was added in 3-fold molar excess. The labeling reaction was carried out overnight on ice and quenched by the addition of DTT to a final concentration of 50 mM. Unreacted dye was removed by a HiTrap desalting column, followed by anion-exchange chromatography as in the first labeling step. The masses of all doubly labeled variants were confirmed by electrospray ionization mass spectrometry.

**CD Spectroscopy.** Equilibrium spectra (Figure 3a) were recorded on a Jasco J715 CD spectrometer at 293 K. The 50 \( \mu M \) ClyA stock solution in 4 M GdmCl was diluted 11-fold with buffer (25 mM sodium phosphate, 75 mM NaCl, pH 7.0) containing different amounts of GdmCl. The samples were incubated for at least 1 h, and data were recorded in a 1 mm path length cuvette in the range of 200–250 nm. The scanning speed was 10 nm/min, and 15 spectra were averaged. For the measurement in 4 M GdmCl, a 10-\( \mu M \) sample was measured in 0.5 mm path length cuvette, and 20 accumulations were averaged.

**Singular Value Decomposition (SVD).** SVD was used to determine the number of components necessary to account for the changes in the CD spectra as a function of denaturant concentration.\(^{42}\) CD spectra measured at \( N \) different denaturant concentrations were combined into an \( M \times N \) matrix \( H \), where \( M \) is the number of wavelengths. This matrix can be decomposed into the product of three matrices: \( H = U S V^T \). The orthogonal \( M \times M \) matrix \( U \) contains information about the shape of the CD spectra, with the columns of \( U \) being the basis vectors (Figure 3c). The diagonal matrix \( S \) contains the singular values sorted by magnitude as diagonal elements (inset Figure 3c). The columns of the orthonormal matrix \( V \) contain information regarding the change in signal amplitude as a function of denaturant concentration (amplitude vectors). The first two amplitude vectors were fitted globally with a two-state model to obtain the unfolding midpoint denaturant concentration, with the midpoint as a shared fit parameter (Figure 3d).

**Single-Molecule Measurements.** Single-molecule measurements were performed on a Micro Time 200 confocal instrument (PicoQuant) at 295 K, as described previously.\(^{42}\) In brief, pulsed interleaved excitation (PIL) was applied to remove the contribution of molecules lacking an active acceptor dye. The fluorescence photons emitted after donor or acceptor excitation were collected via an UplanApo 60/1.2W objective (Olympus) and distributed onto four single-photon avalanche detectors (r-SPAD, PicoQuant), first according to polarization (polarization cube, PicoQuant) and then according to color (595 DCXR dichroic mirror, Chroma). Prior to detection, photons were filtered by E252/50 M or HQ650/100 band-pass filters (Chroma) for donor and acceptor channels, respectively. The protein stock solution of the different ClyA variants in 4 M GdmCl was diluted at least 100-fold with buffer (25 mM sodium phosphate, 75 mM NaCl, 140 mM \( \beta \)-mercaptoethanol, 0.001\% (w/v) \( \text{Tween 20, pH 7.0} \)) containing different amounts of GdmCl, resulting in sample concentrations of \( \sim 40–125 \mu M \). The \( \text{Tween 20} \) concentration used to reduce surface adhesion is below the critical micelle concentration (5.9 \( \times 10^{-5} \) M; 0.007\% (w/v)) and does not induce the conversion to the protomer conformation.\(^ {30}\) The samples were incubated for at least 2 h for equilibration prior to measurements. Measurements were conducted in custom-made quartz cuvettes that were plasma-activated (FEMTO, Diener) and pretreated at least for 15 min with 0.1 mg/mL poly(lysine)-graft-poly(ethylene glycol) (PLL/PEG) (PLL(20)-g-[3.5]-PEG(2), Susos) to minimize surface adhesion of ClyA molecules. Measurements were carried out for 30 min (Figure 4).

**Single-Molecule FRET Data Analysis.** FRET data analysis was carried out as described previously.\(^ {32}\) The numbers of detected donor and acceptor photons were corrected for the different quantum yields of the dyes, different detection efficiencies, crosstalk, acceptor direct excitation, and background.\(^ {45}\) Contiguous photons detected after donor excitation pulses with interphoton times of less than 100 \( \mu s \) were combined into one fluorescence burst. Bursts were considered for data analysis if more than 50 photons were identified. Bursts with a stoichiometry ratio \( S > 0.7 \) were omitted. For each burst, the transfer efficiency was calculated according to \( E = n_A/(n_A+n_D) \), with \( n_A \) and \( n_D \) being the numbers of corrected acceptor and donor photons detected after donor excitation, respectively. The resulting transfer efficiencies were binned in a histogram.

To extract the relative fractions of the native, intermediate, and unfolded populations for each variant, the histograms were fitted with a sum of appropriate peak functions, one for each subpopulation.\(^ {42}\) The width of each subpopulation peak was a shared fit parameter throughout each titration series, while the peak positions and amplitudes were allowed to be different in each histogram (Figure 4). The uncertainty in the peak populations was estimated by increasing and decreasing the width of one of the populations by 15\% for the fit of the transfer efficiency histograms, with the widths of the other populations remaining unchanged.\(^ {32}\) This procedure was iteratively applied to all populations contributing to the transfer efficiency histogram, and the lowest and highest values of the resulting populations are reported as error bars (Figure 5).

**Dual-Focus Fluorescence Correlation Spectroscopy (2F-FCS).** 2F-FCS measurements were performed on a confocal single-molecule instrument (MT200, PicoQuant), with two foci (interfocal distance of 469 nm, calibrated as previously described\(^ {46}\)) produced by passing two collinear, alternatingly pulsed laser beams of orthogonal polarization through a differential interference contrast prism (U-DICTHC, Olympus). The light sources were a SC-450-4 supercontinuum laser (Fianium, UK) with a zS82/15 band-pass filter, resulting in a wavelength of 582 ± 15 nm, and a Solea supercontinuum laser (PicoQuant, Germany) with wavelength-selected output at 585 ± 5 nm. The repetition rate of each laser was 20 MHz, and the laser power was 30 \( \mu W \) (measured at the back aperture of the objective). Fluorescence emission, collected by the objective, was focused on a 150-\( \mu m \) pinhole before being distributed by a 50/50 beam splitter onto two single-photon avalanche photodiodes (r-SPAD 50, PicoQuant) after passing a bandpass filter (ET525/50M, Chroma). The photon arrival times were recorded by two channels of a HydraHarp 400 counting module (PicoQuant). Fluorescence intensity auto- and cross-correlation curves from the data of each focus were fitted globally as described by Pieper et al.\(^ {32}\) From the resulting translational diffusion coefficient, \( D \), the Stokes radius was calculated according to the Stokes–Einstein equation: \( R_S = \) .
Calculation of Mean Transfer Efficiencies from Interdye-Distance Dynamics in a Potential of Mean Force. The spatial distribution of sterically accessible dye positions, taking into account the flexible linkers, was modeled using the software FRET Positioning and Screening (FPS)\(^5^\) (Figure 2). As input, the structure of the ClyA monomer (PDB code: 1QOY) and the dimensions of the fluorescent dyes (Alexa Fluor 488 C5 maleimide and Alexa Fluor 594 C5 maleimide) and the linker lengths and widths used were 16 and 4.5 Å (Alexa Fluor 488 C5 maleimide) as well as 18 and 4.5 Å (Alexa Fluor 594 C5 maleimide). From the output, an expected interdye-distance efficiency, we took into account rapid fluctuations of the interdye distance that are expected to occur on time scales similar to the fluorescence lifetime of the donor (~4 ns in the absence of the acceptor).

We modeled the dynamics as diffusive motion in the potential of mean force given by the distance distribution. Following the theory and notation of Gopich and Szabo\(^5^\) we briefly describe the procedure: Very generally, the following relation between the excited state of the donor dye holds

\[
\langle E \rangle = 1 - \tau_D / \tau_0
\]

where \(\tau_D = 1/k_D\) is the mean fluorescence lifetime of the donor in the absence of an acceptor. \(\langle E \rangle\) is given by

\[
\langle E \rangle = \int_0^\infty p_D^e(t) dt
\]

Here, \(p_D^e(t)\) is the survival probability of the excited state of the donor \(D^e\), which formally can be written as:

\[
p_D^e(t) = \left\{ \exp\left(-k_D t - \int_0^t k_D \left(\frac{R_0}{r(t)}\right)^6 dt\right) \right\}_{r(t)}
\]

where \(\langle \cdot \rangle_{r(t)}\) denotes averaging over all possible paths of distance fluctuations, \(r(t)\). A related quantity is the population density, \(p_D^e(r,t)\), of \(D^e\), which is given by the solution of the Smoluchowski equation

\[
\frac{dp_D^e(r,t)}{dt} = \left[ \mathcal{L} - k_D - k_f \left(\frac{R_0}{r}\right)^6 \right] p_D^e(r,t)
\]

of the diffusion operator

\[
\mathcal{L} = D \frac{d^2}{dr^2} \left[p_e^r(r) \frac{d}{dr} (p_e^r(r))^{-1}\right]
\]

where \(D\) is the effective diffusion coefficient. The Smoluchowski equation needs to be solved with the initial condition \(p_D^e(r,t) = 0\) at \(t = 0\) and that the doner was excited and that the interphoton time is much longer than the reconfiguration time of the interdye distance. From the solution we obtain the survival probability after integration over \(r\):

\[
\langle E \rangle = \int_0^\infty p_D^e(r,t) dr
\]

For the calculations in practice, we discretized the Smoluchowski equation with respect to \(r\) on a grid of \(N = 50\) distances defined on an interval \((r_i, r_i)\) outside of which \(p_e^r(r)\) is zero to good approximation. We obtain \(i = 1, \ldots, N\) rate equations of the form

\[
\frac{dp_D^e(r_i,t)}{dt} = D \left[ \frac{p_D^e(r_{i-1},t) - p_D^e(r_{i+1},t)}{\Delta} - \frac{p_D^e(r_{i+1/2},t) - p_D^e(r_{i-1/2},t)}{\Delta} \right] - k_D(1 + (R_0/r)^6) p_D^e(r_i,t)
\]

where \(r_i = r_{i-1} + (i-1)\Delta\) with \(\Delta = (r_{i-1} - r_{i-1})/(N-1)\). We further defined \(p_D^e(r_{i+1},t) = p_D^e(N+1/2,t)\) and \(p_D^e(r_{i-1/2},t) = p_D^e(N-1/2,t)\). In the last formula, \(i\) is not limited to integers. At the boundaries, we set \(p_D^e(N+1) = p_D^e(N-1) = 0\). The effective diffusion coefficient \(D\) is the sum of the diffusion coefficients of the individual dyes including their linkers attached to the protein. These values are difficult to measure directly. Here we assume that \(D\) is similar to the diffusion coefficient of a free dye, which is a reasonable approximation based on molecular dynamics simulations.\(^5\) The experimentally determined diffusion coefficient of Alexa Fluor 488 was used \(435 \mu m^2/s\). To estimate error bounds (Figure 2c), \(D\) was varied from a value corresponding to distance relaxation much slower than the fluorescence lifetime (static regime) to twice the value of the diffusion coefficient of the free dye, which corresponds to the largest physically meaningful value. To account for the two possible labeling permutations, the distance distributions and mean transfer efficiencies were calculated for both permutations and were found to be in good agreement.

To test the free rotational mobility of the fluorescent dyes, we determined fluorescence polarization anisotropies for the six different variants from the single-molecule data and found values between 0.12 and 0.14 for variants 2/56, 187/290, 56/252, 2/252, and 2/303, suggesting rapid orientational averaging of the dyes and thus an orientational factor of \(k^2 \approx 2/3\). For variant 107/157, we observed a donor anisotropy of
of 0.25. In this case, reduced orientational averaging may contribute to the discrepancy between experimentally observed values and those calculated based on the accessible volumes. The donor fluorescence anisotropies decrease with increasing denaturant concentration, with values between 0.03 to 0.06 in 4 M GdmCl. Only for variant 107/157, we still observe an increased anisotropy of 0.16. As expected, the assumption of free and rapid orientational averaging is thus even more likely to hold under denaturing conditions.

**Equilibrium Three-State Model.** The three-state model used for fitting the relative populations of the native (N), intermediate (I), and unfolded state (U) (Figure 5) is given by

\[ N \leftrightarrow I \leftrightarrow U \]

with the equilibrium constants \( K_1 = c_1/c_2 \) and \( K_2 = c_2/c_3 \). The fractions of the populations are then described by

\[ f_N = (1 + K_1 + K_2)^{-1}, \quad f_I = (K_1 + 1 + K_2)^{-1}, \quad f_U = (K_1^{-1}K_2^{-1} + 1)^{-1}. \]

The equilibrium constants depend on denaturant concentration \( c_{\text{GdmCl}} \) according to

\[ K_{i=1,2} = \exp(-m_i(c_{\text{GdmCl}} - c_{\text{m}})/RT), \]

with \( R \) being the ideal gas constant, \( T \) being the temperature, \( c_{\text{m}} \) being the midpoint concentrations, and \( m_i \) being the equilibrium parameters.\(^2,55\) Fit parameters were the \( m \)-values and the midpoint concentrations.

## RESULTS

**Mapping the Structure of ClyA with Single-Molecule FRET.** For mapping the structural changes during unfolding of ClyA, six fluorescently labeled variants of the 303-residue protein were designed for single-molecule FRET based on the crystal structure of the soluble ClyA monomer (PDB code: 1QOY\(^39\)) and the protomer (taken as one subunit of the pore, PDB code: 2WCD\(^12\)) (Figure 1a and b). By placing the FRET dyes into different positions, different parts of the structure with very different segment lengths (ranging from 50 to 301 residues) were probed (Figure 2). To minimize effects on conformational stability, the residues exchanged to cysteines for fluorescence labeling were chosen in a way that they are solvent-exposed both in the monomer and protomer conformations (Figure 1).

First, we recorded transfer efficiency histograms with confocal single-molecule detection of freely diffusing sample molecules for all ClyA variants under native conditions. The measured mean transfer efficiencies (\( E \)) were compared to the values calculated based on the crystal structure (Figure 2). The accessible volumes of the dyes, considering their flexible linkers, were modeled as described by Seidel and co-workers\(^51,56\) and are depicted in blue and orange for Alexa Fluor 488 (donor) and Alexa Fluor 594 (acceptor), respectively in Figure 2b. The measured and calculated values of \( \langle E \rangle \) for the variants are compared in Figure 2c. The calculations were performed assuming different time scales for interdye-distance dynamics with respect to the fluorescence lifetime of the donor. We note that the dynamics have a detectable effect on the resulting transfer efficiencies, especially if the distance between the attachment points of the linkers is similar to or smaller than the linker lengths (see Methods). Donor fluorescence anisotropies were determined for all variants to test for free rotational mobility of the dyes. Only for variant 107/157, we find an anisotropy in excess of 0.2, potentially contributing to the discrepancy between the measured and calculated mean transfer efficiency (see Methods). The last five C-terminal residues of ClyA are not resolved in the crystal structure of the monomer and are presumably unstructured. Therefore, for the calculation of variant A2C/V303C, we added the length of five times 0.38 nm to the contour length of the C-terminal dye linker (bottom panel in Figure 2b), which leads to an increased uncertainty in the \( E \) estimate. The larger width of the transfer efficiency peak of this variant compared to the others further indicates an additional distance heterogeneity persistent on a millisecond time scale. Altogether, the distance values from the measurements and those based on the crystal structure are in reasonable agreement and illustrate the level of accuracy.

\[ \begin{align*}
\text{Figure 2. Mapping of the native state of ClyA with single molecule FRET.} \\
(a) & \text{Transfer efficiency histograms of the six different variants investigated, fitted with appropriate peak functions (lines) to obtain mean transfer efficiencies, } \langle E_{\text{meas}} \rangle. \\
(b) & \text{Crystal structure of ClyA (PDB code: 1QOY\(^39\)) with the accessible volumes of the fluorophores depicted in blue and orange, for the donor and acceptor dyes, respectively. The part of the polypeptide chain that is probed in each variant is colored in magenta. (c) Comparison of measured and calculated mean transfer efficiency values based on the structure. Interdye-distance dynamics are modeled as diffusive process in a potential of mean force (see text). The error range for } \langle E_{\text{calc}} \rangle \text{ results from the limiting cases of fast diffusion (upper bound) or no diffusion of the dyes (lower bound) with respect to the fluorescence lifetime (see Methods for details). There might be additional contributions to the error on } \langle E_{\text{calc}} \rangle, \text{ originating, e.g., from residual interactions of the dyes with the protein surface. However, those contributions are difficult to quantify and are thus not included here. Note that the five C-terminal residues are not resolved in the crystal structure, so the length of the flexible dye linker was increased for calculating the expected mean transfer efficiency for variant 2/303 (\*).}
\end{align*} \]
attainable in the single-molecule FRET experiments of this system.

Secondary Structure Changes from CD Spectroscopy.
To quantify the secondary structure content of ClyA as a function of denaturant concentration, equilibrium CD spectra of unlabeled wt-ClyA were recorded (Figure 3a). Starting from 50 μM ClyA in 4 M GdmCl, the protein was diluted 11-fold with buffers containing various concentrations of the denaturant GdmCl and equilibrated. From the CD spectrum recorded at the lowest GdmCl concentration (0.36 M), where the protein is folded, the secondary structure content was quantified using the K2d algorithm (provided by the Web server DICHRWEB), which is based on a reference database of spectra of proteins with known crystal structure. The estimated helicity was 81%, within error in agreement with a helical content of 77% determined from the crystal structure. Above ∼4 M GdmCl, all helical signal is lost, indicating complete unfolding.

To analyze the unfolding transition, we first plotted the normalized mean-residue molar ellipticity at 222 nm as a function of denaturant concentration (Figure 3b). The resulting data show a single transition with surprisingly low cooperativity for a protein of this size, fitting the transition with an equilibrium two-state model results in a midpoint concentration of 2.29 ± 0.03 M GdmCl. The ellipticity at 222 nm does not reveal direct evidence for the existence of an equilibrium intermediate. Hence, to use the information from the entire spectral range, we quantify the number of components required to describe the data set using SVD. At least three singular values above the noise were identified (Figure 3c), indicative of the existence of more than two states. However, the third and fourth basis vectors contain substantially higher levels of noise than the first two (Figure 3c), and their amplitude vectors are dominated by noise (Figure 3d), whereas the amplitude vectors of the first two components resemble the transition obtained from the single-wavelength analysis (Figure 3b). Fitting the first two amplitude vectors globally with a two-state model, using the midpoint concentration as shared fit parameter, yielded an unfolding midpoint of 2.25 ± 0.10 M GdmCl, in agreement with the single-wavelength analysis. Overall, the results obtained from CD spectroscopy provide hints for a deviation from simple two-state behavior but remain rather inconclusive regarding the properties of a potential unfolding intermediate.

Mapping the Intermediate State of ClyA with Single-Molecule FRET. To test the existence of an intermediate in more detail, transfer efficiency histograms of the six FRET-labeled ClyA variants were recorded at various denaturant concentrations (Figure 4). The protein stock solutions in 4 M GdmCl were diluted at least 100-fold with the respective GdmCl buffer. The samples were then incubated for at least 2 h to equilibrate. The shape of transfer efficiency histograms did not change over the time course of a measurement (Figure S1), confirming completed equilibration. In all cases, the transition proceeds from the folded state under fully native conditions to a single unfolded population at 4 M GdmCl, in agreement with the loss of CD signal at this denaturant concentration (Figure 3b). However, in contrast to the CD results, the existence of an intermediate state becomes obvious in the single-molecule experiments by the appearance of an additional population that is most prominent in the range between about 1 and 2.5 M GdmCl (Figure 4), especially for the variants 187/290, 56/252, 2/252, and 2/303. Interestingly, the patterns of transfer efficiency peaks are different for the different variants, both in terms of the number of peaks and the denaturant concentrations where they are populated. What can we infer about the structure of the intermediate from the single-molecule FRET measurements?

ClyA variant 2/56 (Figure 4a), which probes the N-terminal helix αA, shows a pronounced transition from the low transfer efficiency expected for the folded helix (Figure 2, first row) to a population with ⟨E⟩ ≈ 0.4 at a GdmCl concentration of ∼1.25 M. No clearly separated third peak is observed at ∼2.5 M, but some peak broadening is detected, indicating the existence of two overlapping subpopulations with similar transfer efficiencies. The mean interdye distance of the intermediate is thus close to that of the unfolded state for these labeling positions, suggesting that the N-terminal helix αA is unfolded in the intermediate. The small shift in ⟨E⟩ at ∼2.5 M GdmCl suggests that part of the segment 2−56 remains structured in the intermediate state (possibly involving the second helical segment of αA, αA2, after Pro36 and the following short 3α helix, see Figure 1a) and only unfolds in the second transition along with the rest of the protein.

In contrast, ClyA 107/157, which probes helix αC, reveals a very different behavior. For this variant, an intermediate population is not detectable, and apparent two-state behavior is observed, with a single transition at ∼2.25 M GdmCl (Figure 4b), implying that helix αC remains folded in the intermediate. This hypothesis is supported by the similar behavior of ClyA 56/252 (Figure 4d), which probes the segment encompassing helices αB−αF, i.e., the core of the protein without the N- and C-terminal helices. In this case, the mean transfer efficiency in
the intermediate is only slightly increased compared to the native state, suggesting that helices αB-αF remain largely structured. Only above ~2 M GdmCl do we observe the transition to the low transfer efficiency population expected for a fully unfolded sequence with a dye separation of 200 residues.

Variant 187/290 probes the segment including helices αF and αG. Interestingly, in this case the transfer efficiency of the intermediate is lower than for the fully unfolded state,
indicating a very large interdye distance compared to both the native and the unfolded state (Figure 4c). Since we already know from variant S6/252 that helix αF remains largely intact in the intermediate state, we conclude that only the C-terminal helix αG unfolds in the first transition. The combination of an extended and still stably folded helix αF with an unfolded helix αG would explain the large distance between residues 187 and 290, which decreases upon unfolding of helix αF in the second transition to the fully unfolded state.

With variant 2/252, the full polypeptide chain with the exception of the C-terminal helix αG is probed (Figure 4e). In the first transition, the intermediate population appears at (E) \( \approx 0.5 \). This value is in accord with helix αA being largely unfolded in the intermediate (as suggested based on variant 2/ S6) and the core helices αB-αF retaining their structure and arrangement (as suggested by variant 56/252). Finally, in the terminally labeled variant 2/303 (Figure 4f), the intermediate appears at an efficiency of \( \approx 0.67 \), presumably corresponding to helices αG and αA1 being unfolded. The mean transfer efficiency of the native state population shifts continuously toward lower transfer efficiency upon increasing denaturant concentrations, which may indicate noncooperative unfolding or denaturant-induced expansion of the C-terminal 12-residue segment that follows helix αG.

To quantify the relative populations of the native, intermediate, and unfolded states as a function of denaturant concentration, the transfer efficiency histograms were fitted with appropriate peak functions where only the width was assumed to be constant for each population. Figure 5 shows the relative fractions of the three subpopulations for all six variants from the resulting peak areas (Figure 4). Overall, the intermediate state population is observed consistently between \( \sim 1 \) M and \( \sim 2.5 \) M GdmCl. The transition from the intermediate to the unfolded state occurs between \( \sim 2 \) and \( 2.5 \) M GdmCl for the different variants (Figure 5 and Figure S2) and coincides approximately with the unfolding midpoint concentration determined by CD spectroscopy (Figure 3b and d), confirming that the intermediate is still highly helical. At 1.75 M GdmCl, where the intermediate is populated almost exclusively according to the single-molecule experiments, it contains about 75% of the helicity of the native state (Figure 3b), in keeping with all helices except αA and αG being folded.

The data were fitted with an equilibrium three-state model (see Methods). Only for variant 107/157 (Figure 5b), the intermediate is not apparent, so a two-state model was used for the fit. In most cases, the data are reasonably well described by the equilibrium models. For variants 2/252 and 2/303, the unfolded state population exhibits a plateau between 1.25 and 2.25 M GdmCl, which is not well described by the fit (Figure 5e, f). A possible explanation is that the low transfer efficiency population in Figure 4e, f in this GdmCl concentration range may not correspond to the unfolded state but to a second intermediate conformation, which is in slow exchange with the first one, possibly linked to the unfolding/undocking of helix αA2. We note that the introduction of cysteine residues and labels at different positions of the polypeptide chain is likely to affect the stability of native state and intermediates to slightly different extents, which complicates a precise quantitative comparison of stabilities.

**Chain Dynamics in the Intermediate State.** Further evidence for the unfolding of helix αA in the intermediate comes from nanosecond FCS (nsFCS) at 1.75 M GdmCl for variants 2/56 and S6/252, which probes the distance dynamics between donor and acceptor dyes (Figure 6). Besides the expected anticorrelated signal for very short lag times due to photon antibunching, ClyA 2/56 shows an additional anticorrelated component in the donor autocorrelation is likely to originate from rotational dynamics, as supported by the asymmetry of the signal for positive and negative lag times. A possible explanation is that the low transfer efficiency population in Figure 4e, f in this GdmCl concentration range may not correspond to the unfolded state but to a second intermediate conformation, which is in slow exchange with the first one, possibly linked to the unfolding/undocking of helix αA2. We note that the introduction of cysteine residues and labels at different positions of the polypeptide chain is likely to affect the stability of native state and intermediates to slightly different extents, which complicates a precise quantitative comparison of stabilities.
Finally, the dimensions of the intermediate were investigated by 2f-FCS and compared to those of the native and unfolded states (Figure S3). The Stokes radius of the intermediate was found to be $3.5 \pm 0.4$ nm, not much greater than the value observed for the native state ($3.2 \pm 0.3$ nm), in agreement with the presence of a folded helical core. Only upon complete unfolding does the Stokes radius increase substantially, reaching $5.9 \pm 0.6$ nm at 4 M GdmCl.

### Discussion and Conclusion

Altogether, single-molecule FRET experiments unequivocally reveal the presence of an equilibrium unfolding intermediate for the cytolytic pore toxin ClyA. From the analysis of the unfolding experiments monitored by single-molecule FRET of six variants probing different protein segments, combined with CD, nsFCS, and 2f-FCS, the following picture emerges for the unfolding of ClyA (Figure 7): In the absence of denaturant, only the native state (N) is populated. Addition of moderate concentrations of GdmCl ($\sim 1$–2.5 M) leads to the formation of an intermediate state (I) in which the C- and N-terminal helices are unfolded. Increasing the denaturant concentration further leads to global unfolding of the protein (U).

![Figure 7](image)

**Figure 7.** Cartoon representation of the states involved in equilibrium folding of ClyA. In the absence of denaturant, only the native state (N) is populated. In the presence of $\sim 1$–3 M GdmCl, an intermediate state (I) is formed, in which the C- and N-terminal helices are unfolded. Increasing the denaturant concentration further leads to global unfolding of the protein (U).

As part of this rearrangement, the N-terminal helix $\alpha$A is translocated within the structure and rotated by almost 180° (Figure 1). This major conformational change requires a substantial amount of flexibility in the structure, presumably making helix $\alpha$A prone to unfolding in the presence of relatively low concentrations of denaturant. Although helix $\alpha$G is not affected as much during protomer conformation, it is shortened from both ends and loses its tight interactions with $\alpha$A, likely to increase its tendency to unfold.

These local differences in stability are also reflected by a structural analysis of the differences between monomer and protomer conformation of ClyA. A map of side chain interactions allows us to distinguish contacts that are unique to monomer or protomer from those that are common to both conformations (Figure 1c). Unaltered interactions (displayed in black in Figure 1c) are primarily located at the interface between helices $\alpha$B and $\alpha$C. Interestingly, some contacts of helix $\alpha$F with $\alpha$G and $\alpha$C are present in both conformations, even though the helical packing between helices $\alpha$F and $\alpha$C differs in the monomer and the protomer. Similarly, an alignment of the monomer and protomer structures illustrates the parts of the protein that are structurally most similar in both conformations (Figure 1c, inset). Helices $\alpha$B and $\alpha$C are hardly affected by the conformational transition, reflecting the stable core of ClyA. The largest change is apparent for helix $\alpha$A, as expected considering its large translocation during protomer formation.

The conformational bistability of ClyA may thus be encoded by differences in local stability across the protein that facilitate the rearrangement of parts of the structure, while others remain stably folded (Figure 1c). It would be interesting to use recent advances in coarse-grained models for the conformational changes of ClyA and for interpreting single-molecule FRET experiments to develop a more detailed structural description of the intermediate. Such coarse-grained simulations have been previously applied to resolve the conformational switching of lymphotactin, another protein that can adopt two native states. It was shown that the transition involves a partly unfolded state that is stabilized by common contacts of both folds.

Could the equilibrium intermediate observed here be related to the kinetic intermediate observed previously in time-resolved single-molecule and CD measurements of ClyA? In that work, an off-pathway intermediate was identified that is transiently populated during the interconversion from monomer to protomer upon interaction with the membrane-mimicking detergent DDM. However, the kinetic intermediate exhibited a very low transfer efficiency for variant 56/252 used in the time-resolve measurements, indicating that it is largely unfolded and can therefore not resemble the equilibrium intermediate observed here. It is conceivable, however, that an additional kinetic intermediate similar to the equilibrium intermediate identified here may exist, since its transfer efficiency for variant 56/252 is between those of the monomer and the protomer and may thus have eluded detection in the kinetics. Using the different variants employed here for the equilibrium measurements in the DDM-induced conversion kinetics may help to settle this question.
ASSOCIATED CONTENT

Supporting Information
The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.8b07026.

Transfer efficiency histograms of different variants; dimensions of the native, intermediate, and unfolded state determined with dual-focus FCS (PDF)

AUTHOR INFORMATION

Corresponding Author
*Phone: +41 44 63 5535. E-mail: schuler@bioc.uzh.ch.

ORCID
Benjamin Schuler: 0000-0002-5970-4251

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Rudi Glockshuber and Daniel Roderer for the pClyA plasmid and expression protocols and the Functional Genomics Center Zurich for expert mass spectrometry analysis. This work was supported by the Swiss National Science Foundation.

REFERENCES

(37) Vaidyanathan, M. S.; Sathyanarayana, P.; Maiti, P. K.; Visweswariah, S. S.; Ayappa, K. G. Lysis dynamics and membrane
Correlation Spectroscopy from Picoseconds to Seconds. Peptide Dynamics Investigated with Single-Molecule FRET and 175