

Botulinum Endopeptidase: SAXS Experiments and MD Simulations Reveal Extended Solution Structures That Account for Its Biochemical Properties

Published as part of *The Journal of Physical Chemistry virtual special issue "Peter J. Rossky Festschrift"*.

Raj Kumar, Farkhad Maksudov, Olga Kononova, Kenneth A. Marx, Valeri Barsegov,*
and Bal Ram Singh*

 Cite This: *J. Phys. Chem. B* 2020, 124, 5801–5812

 Read Online

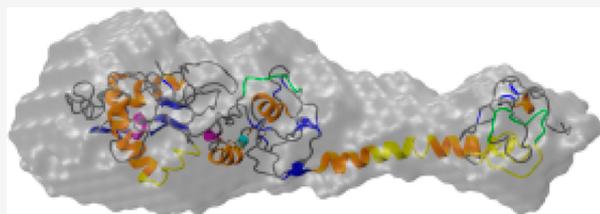
ACCESS |

 Metrics & More

 Article Recommendations

 Supporting Information

ABSTRACT: Development of antidotes against botulism requires understanding of the enzymatically active conformations of Botulinum neurotoxin serotype A (BoNT/A) light chain (LCA). We performed small angle X-ray scattering (SAXS) to characterize the solution structures of truncated light chain (tLCA). The 34–37 Å radius of gyration of tLCA was 1.5-times greater than the averaged 22–23-Å radius from the crystal structures. The bimodal distribution of interatomic distances $P(r)$ indicated the two-domain tLCA structure with 129–133 Å size, and Kratky plots indicated the tLCA partial unfolding in the 25–37 °C temperature range. To interpret these data, we employed molecular dynamics simulations and machine learning. Excellent agreement between experimental and theoretical $P(r)$ profiles helped to resolve conformational subpopulations of tLCA in solution. Partial unfolding of the C-terminal portion of tLCA (residues 339–425) results in formation of extended conformations with the larger globular domain (residues 2–298) and the smaller unstructured domain (339–425). The catalytic domain, buried 20 Å-deep inside the crystal structure, becomes accessible in extended solution conformations (8–9 Å deep). The C- and N-termini containing different functional sequence motifs are maximally separated in the extended conformations. Our results offer physical insights into the molecular basis of BoNT/A function and stress the importance of reversible unfolding–refolding transitions and hydrophobic interactions.



1. INTRODUCTION

Each of the seven serotypes (A–G) of botulinum neurotoxin (BoNT) blocks the release of acetylcholine at the nerve–muscle junctions, leading to severe flaccid muscle paralysis.^{1,2} Although botulinum neurotoxin serotype A (BoNT/A) is the most potent (poisonous) molecule known to mankind,³ it is most popularly known for its therapeutic effect at very low doses.^{3,4} The molecule, particularly the light chain domain (LC), is known to survive and be effective in host cells for long time periods (months), an unusually robust property for exogenous proteins.^{5–7} Despite many decades of ongoing research, there is little understanding of the molecular-level basis behind this unusual stability and high effectiveness. Several studies have demonstrated that crystal structures alone cannot account for these phenomena, suggesting that the properties of the BoNT/A molecule might be rooted in a highly flexible and dynamic solution structure.^{8–11} Furthermore, while each of the seven BoNT serotypes contains an identical Zn²⁺-binding motif in their active sites, their substrates are either entirely different proteins, or the cleavage site on the substrate is different for each serotype. Each of the BoNT endopeptidases is extremely selective for substrate, and each has an exclusive cleavage site. Exclusive selectivity and

specificity are key to the endopeptidase actions of BoNT. Moreover, only entire substrate proteins or their large polypeptide segments are recognized to include exosites for effective cleavage by BoNT endopeptidases.¹² All the BoNT endopeptidases recognize one or more of the three SNARE proteins involved in the exocytosis. Therefore, an understanding of the optimally active conformational of BoNT is an important step toward discerning the comprehensive mechanism of BoNT action. Here, we examine the solution structural variability of BoNT/A, in order to understand its intracellular stability and range of catalytic mechanisms.

A critical step in developing an inhibitor against a protein's function requires a comprehensive knowledge of its active solution conformational states and the molecular basis of its function. In general, a crystal structure can address this issue,

Received: March 30, 2020

Revised: June 15, 2020

Published: June 16, 2020



but the static view becomes inadequate for systems where significant thermal fluctuation and/or large functionally relevant conformational states might be populated. To achieve crystallization to the solid state involves crystal packing forces that might lead to distortions of the biologically relevant solution active conformation(s). Difficulty in crystallization might also result from the protein system requiring too large an energy input to pack into the crystal unit cell. This is likely the case here, where our previous experimental studies⁸ revealed that, for the LCA, crystallization proved to be a formidable task despite multiple efforts by researchers, perhaps due to the significant flexibility of the C-terminal region being difficult to overcome. The apparent difference between the solution structure and the crystal structure is likely one of the main reasons that a correct explanation of enzymatic mechanism or design of effective inhibitors has been lacking.^{8,10,11,13,14} The solution structure might also help explain how two different sequence motifs, one located at the N-terminus¹⁵ and the other close to the C-terminus¹⁶ of BoNT/A LC (LCA), can functionally interact better with their different biological partners (SNAP-25 substrate and septin oligomers, respectively) at the intracellular surface of the plasma membrane, than in the crystal structure, leading to increased longevity of the enzymatic activity. Experimental data supporting the high potency, longevity, and exclusive site selectivity characteristics of LCA^{8–11} strengthen the hypothesis that this molecule has a different conformation in solution than in the available crystal structures. Hence, a thorough solution structural characterization of LCA will provide the necessary information critical to understanding its toxicity, as well as its mechanism of therapeutic activity.

The LCA used in most laboratories are the truncated forms of the 448-residue full length polypeptide. However, there is no clarity as to the exact length of LCA in the functional state. Although the BoNT/A gene encodes for 448 residues, published data cites lengths of 437 and 444 residues.^{4,17–19} The crystal structure (3BTA) potentially suggests a length of 448 residues; however, validated crystal structure have missing coordinates for residues 432–449.¹⁹ There is no definitive information about the length of an active enzyme inside the cell, considered by most investigators to be mainly LC-425 and LC-424.^{20,21} Although these forms have enzymatic activity, their interactions with potential inhibitors and with peptide substrate and whole protein SNAP-25 substrate are very different from those for the full-length LCA protein.^{4,20,21} This may result from dynamic structural differences in their conformations. The difficulty in developing a biologically consistent inhibitor against BoNT is due in part to the structurally flexible premolten globule-like structure (PRIME) the endopeptidase adopts under optimum enzyme activity conditions.^{8–11} There are reports of intracellular chemical modifications of LCs which might also play a role in their activity and survival, but the structure and dynamics are overarching features in mediating the effect.²² The PRIME conformation is increasingly being recognized as an important dynamic structural property for the development of inhibitors.^{9–11,23–25}

The crystal structure of only truncated LCA (tLCA) (obtained by removing 24 residues at the C-terminus) has been solved by X-ray crystallography.^{12,26} However, as argued above, it is quite likely that the conformational states of the PRIME and molten globule (MG) states are at variance with the crystal structure. Therefore, it is important to adequately

describe the ensemble structures and characterize the dynamic properties of the enzyme in solution, where it exists in a manifold of conformational subpopulations. In this regard, our observation of the PRIME conformation of LCA with optimal enzymatic activity is significant.⁹ The role of dynamic structure alterations of this protein becomes more intriguing in view of the extreme specificity of the endopeptidase for its substrate, SNAP-25, and its potential utility in designing specific antidotes against botulism threats.

Since a crystal structure of full-length LCA is not available, we, therefore, used tLCA to model the solution structure. In this study, we examine the structural aspects of LCA in truncated form to answer the following questions: (i) Is LCA a partially disordered protein? (ii) Is the folded structure of LCA in solution different from the crystal structure? (iii) What are the conformational features of LCA in solution at the minima of free energy landscape. To answer these questions, we have employed small-angle X-ray scattering (SAXS) combined with molecular dynamic (MD) simulations and machine learning to develop models of solution structures of tLCA. This approach has allowed us to visualize changes in the solution molecular shape of the BoNT endopeptidase, outlining the greater accessibility of the active site (which according to the crystal structure remains located in a ~ 20 Å deep crevice) and changes in the exosites. More importantly, this approach helped to resolve the ensemble of possible structures which might be populated in the energy minima of this enzyme. SAXS analysis suggested significant conformational changes that occur in solution. Kratky plots of SAXS suggested that tLCA is not a globular protein; rather, it is a partially unfolded protein. Both SAXS and MD simulations indicated that the tLCA native structure is not a single structure; rather, it is an ensemble of structures at the minima of free energy landscape. We also present a model to explain the possible role of the missing C-terminus residues in overall folding dynamics and catalysis. We find that tLCA has an extended conformation in solution relative to the crystal structure. The comparative extended molecular shape derived from SAXS at 25 and 37 °C not only confirms the PRIME conformation of the optimally active enzyme at 37 °C, but also reveals active site exposure and suggests a possible role of the 24-residue C-terminal peptide segment in the enzymatic activity. Furthermore, the extended shape allows the two different previously characterized functional sequence motifs located in the N-terminus and C-terminus regions to better interact with their biological ligands in the crowded environment at the intracellular surface of the plasma membrane.

2. EXPERIMENTAL AND THEORETICAL METHODS

2.1. Purification of Truncated BoNT/A LC. tLCA were purified according to the method described previously.²⁷ *Escherichia coli* culture of recombinant tLCA was obtained from Dr. Swaminathan's lab (Brookhaven National Laboratory, New York). Proteins were purified in phosphate buffer (10 mM sodium phosphate pH 7.5) containing 300 mM NaCl. Prior to conducting experiments, all light chains were dialyzed in 10 mM phosphate buffer at pH 7.4 containing 150 mM NaCl, 1 mM DTT and 10% glycerol. The protein concentration was determined spectrophotometrically by comparison with the molar extinction coefficient of 0.83 $\text{mg}^{-1} \text{mL cm}^{-1}$ measured at $\lambda = 2800$ Å.¹¹

2.2. SAXS Data Acquisition and Analysis. The SAXS experiments were carried out at the MacCHESS (Macro-

molecular X-ray Science at the Cornell High Energy Synchrotron Source) facility at Cornell University. SAXS data were collected for protein in a phosphate buffer solution at 25 and 37 °C. Buffer subtraction was adjusted for the excluded volume of the protein. The predicted scattering intensity at $q = 0 \text{ \AA}^{-1}$ and the radius of gyration R_g were determined by Guinier analysis²⁸ and compared for two protein concentrations (0.8 and 1.6 mg/mL) to determine interparticle interaction and aggregation. The values of R_g were derived using the Guinier approximation, $I(q) = I(0) \exp[q^2 R_g^2/3]$, where $I(q)$ is the scattered intensity and $I(0)$ is the forward scattered intensity. $I(q)$ and $I(0)$ are inferred respectively from the slope and the intercept of the linear fit of $\ln[I(q)]$ vs q at low q values. For a reliable straight line fit we used the values of q from the range $0.0239 \leq q \leq 0.2222$ such that $qR_g < 1.3$. The distribution of atomic pair distances $P(r)$ contains information about the interatomic distances of a molecule for a given molecular conformation. This enables one to calculate R_g using the entire scattering curve. For example, R_g can be estimated from the second moment of $P(r)$. The distribution of atomic pair distances $P(r)$ also reveals the maximum dimension D_{max} of the molecule, defined as the point r at which $P(r) = 0$. The distributions $P(r)$ were calculated using the Fourier inversion of the scattering intensity $I(q)$ using the GNOM package (RAW software). A range of D_{max} was tested, and the final choice of D_{max} was based on the following three criteria: (1) $P(r)$ should be positive for all values of r ; (2) R_g evaluated with the GNOM package should agree with that obtained using the Guinier analysis; and (3) the curve of $P(r)$ should remain stable when the value of D_{max} is increased beyond an estimated molecular length.

2.3. Three-Dimensional Structure Calculation. The SAXS data were processed using RAW software²⁹ to determine the values of R_g and D_{max} and to obtain the distribution $P(r)$ for each protein conformation. Because $P(r)$ provides a distance profile of all pairs of atoms in a molecule, this measure characterizes the overall molecular shape. This information was used to calculate *ab initio* dummy atom models with DAMMIF program (BioXTAS RAW software^{29,30}), which also evaluates the quality of the models based on agreement between the predicted and experimental X-ray scattering. The quality of dummy atom models was further assessed by evaluating the agreement between replicate models and was quantified by the mean normalized spatial discrepancy (mean NSD).³¹ The mean NSD values were found to be 0.761 and 0.707 for the 25 and 37 °C models, respectively. This indicates very good agreement between individual models, which were then averaged. The average models were superimposed with the previously determined crystal structure using the i-tasser³² (iterative threading assembly refinement) and VMD³³ programs.

2.4. Computational Molecular Modeling. **2.4.1. Atomic Structural Model.** In all-atomic molecular modeling of BoNT/A LC, we used the resolved crystal structure of truncated BoNT/A light chain from the Protein Data Bank (PDB entry: 1XTG), which includes the resolved amino acid residues 2–425.

2.4.2. All-Atom MD Simulations. The *in silico* atomic model was constructed using the PDB structure 1XTG. We used the GROMACS 2020 software package^{34,35} and CHARMM27 force field^{36–38} to carry out the MD simulations in explicit water. The initial protein structure was solvated with at least

1.2 nm of SPC/E³⁹ water. Steepest descent minimization was performed to remove possible inconsistencies in the crystal structure (e.g., steric clashes). Electrostatics was treated using the particle-mesh Ewald (PME) method⁴⁰ with short-range electrostatic and van der Waals cutoffs set at 1.2 nm. The temperature then was coupled with the velocity-rescaled 300 K thermostat.⁴¹ The Parrinello–Rahman barostat⁴² was used in the equilibration step to set the pressure to 1 bar, and LINCS was used to constrain all bond lengths.⁴³ We carried out several 0.5- μ s production equilibrium simulation runs using the GPU-accelerated version of the GROMACS 2020 software, using a 2 fs time step. In a pilot study, we used our own software with the Solvent Accessible Surface Area (SASA) model of implicit solvation in conjunction with the CHARMM19 unified hydrogen force field.⁴⁴ In these MD simulations in implicit water, we used the CHARMM program⁴⁵ to create the atomic model and energy-minimize the model with the steepest descent algorithm.⁴⁵ The obtained system was gradually heated to 25 °C during 300 ps and equilibrated at this temperature for 5 ns. Langevin equations of motion were propagated with a 1 fs time step. These all-atom MD simulations in implicit water were accelerated on graphics processing units (GPUs).^{46,47}

2.4.3. Pulling Simulations. In steered MD (SMD) simulations, the pulling force was applied to the N- or C-terminus of tLCA through a virtual spring (mimicking the cantilever tip) attached to the C_α -atom of the pulled residue: Pro2 (N-terminus) or Arg425 (C-terminus). The other end of the molecule was constrained: the C_α -atom of Arg425 or Pro2, respectively. The virtual cantilever tip was moving with velocity $v_f = 10^4 \mu\text{m/s}$, mimicking the cantilever motion in single-molecule experiments. These dynamic force-ramp measurements *in silico* were carried out using a time-dependent pulling force to unfold the tLCA molecule, $f(t) = \kappa(v_f t - \Delta x)$, where $\kappa = 100 \text{ pN/nm}$ is the cantilever spring constant and Δx is the displacement of the tagged residue from its initial position.

2.5. Theoretical Reconstruction of SAXS Spectra.

2.5.1. Generating Ensemble of Structures. The output from the SMD simulations was used to construct and analyze the force (F)–extension (X) curves for the tLCA molecule; see Figure S1 in the Supporting Information. The FX -curves were used to select a total of 13 partially unfolded conformations of tLCA molecule (Figure S1), which were equilibrated at 25 °C for 0.5 μ s. During equilibration, some of the structures underwent partial refolding, which enabled us to generate a total of 60 structures of the tLCA molecule.

2.5.2. Structure Analysis. To avoid selecting conformations with a large structure overlap, each pair of these 60 structures was first analyzed using the structure overlap function, $\xi = 1/2N(N-1) \sum_{i \neq j=1}^N \Theta(|r_{ij} - r_{ij}^0| - \beta r_{ij}^0)$. This is a distance-dependent measure of structural similarity between a given conformation and a reference state. Here, r_{ij} and r_{ij}^0 are the interparticle distances between the C_α -atoms in i th and j th residues in a given structure and in the reference structure, respectively ($\beta = 0.2$ is the tolerance for the distance change), and $\Theta(x)$ is the Heaviside step function (N is the total number of atoms). Structures with low overlap (less than $\xi = 0.5$) were used in theoretical reconstruction of the distribution of atomic pair distances $P_{th}(r)$, for which a total of 25 structures were selected. The other structures were discarded.

2.6. Machine Learning Based Reconstruction of Distribution $P(r)$. We employed nonlinear regression in conjunction with a generalized additive model to reconstruct

theoretically the distribution of atomic pair distances $P(r)$ using the structure output from MD simulations. The step-by-step implementation of the machine learning algorithm for theoretical reconstruction of $P(r)$ is the following. *Step 1*: Assign random weights (populations) w_i to each distribution of interatomic distances p_i , which represents the i -th conformation in the ensemble of $n = 25$ conformations. *Step 2*: Form weighted superposition $P_{th}(r) = w_1p_1(r) + w_2p_2(r) + \dots + w_np_n(r)$ to obtain the theoretical distribution $P_{th}(r)$, $i = 1, 2, \dots, n$. *Step 3*: Calculate the mean squared error (MSE) penalty function defined as $MSE = 1/m \sum_{j=1}^m (P_j(r_j) - P_{th_j}(r_j))^2$, where P_j and P_{th_j} are the values of experimental and theoretical distributions corresponding to the atomic pair distance r_j , $j = 1, 2, \dots, m$ ($m = 171$ is total number of data points in the experimental or theoretical distribution). *Step 4*: Minimize the MSE by varying the populations (i.e., regression coefficients) w_i for all $i = 1, 2, \dots, n$ and using the stochastic gradient descent algorithm. *Step 5*: Repeat *step 4* while the MSE continues changing (decreasing). *Step 6*: Stop when the MSE has reached the threshold value (3×10^{-6}). Identify conformations with small population ≤ 0.001 , and remove the corresponding structures. *Step 7*: Repeat *steps 4–6* until a small number of structures (3–5 structures) is obtained, for which the total population is $\sum_{i=1}^k w_i = 0.99$.

3. RESULTS

3.1. SAXS Analysis of tLCA Conformations. The conformations of tLCA in 0.8 and 1.6 mg/mL solutions were investigated using SAXS at 25 and 37 °C (Figure 1, parts a and b). The overall quality of scattering data was assessed by calculating qR_g and correlation coefficient r^2 used in Guinier plots. We found that $qR_g < 1.3$, which limits the maximum value of scattering vector q . It also helps to ensure that estimated parameters are within 10% of their true values.^{28,48} The correlation coefficient r^2 provides a goodness of fit of the straight line used to calculate Guinier parameters. Both qR_g and r^2 were within an acceptable range (Table 1). Representative curves of the scattering intensity $I(q)$, displayed in Figure 1a, provide the following ranges for the scattering vector (see Table 1): for 0.8 mg/mL tLCA, $q = 0.0239–0.221$ (at 25 °C) and $0.0257–0.2118$ (at 37 °C); for 1.6 mg/mL tLCA, $q = 0.0239–0.2167$ (at 25 °C) and $0.0281–0.219$ (at 37 °C).

We further analyzed the Guinier profiles, i.e. the log–log plots of $I(q)$ vs q . The Guinier profiles were found to be linear for both concentrations (Figure 1b), indicating that tLCA in solution is in a monodisperse form.²⁸ R_g , which characterizes the overall size of the molecule, was determined by analyzing the Guinier plots with the GNOM package;⁴⁹ R_g determined with the Guinier approximation was found to fit well the linear plots (as indicated by r^2 value; Table 1). A nonlinear Guinier plot is an indicator of the following: poor sample quality, interparticle effects, or polydispersity. Therefore, the linear log–log profiles of $I(q)$ vs q suggested that tLCA solutions at 25 and 37 °C are monodisperse. The correlation coefficients r^2 for all samples was found to be close to 1, thus indicating that the determined values of R_g were statistically significant.

To evaluate the statistical significance of the data (R_g), we used the coefficient of correlation χ^2 . When χ^2 is close to 1, this indicates the goodness of fit of the statistical model to the experimental data. From Guinier plots, $R_g = 34.25$ Å for 0.8 mg/mL tLCA at 25 °C and 35.00 Å at 37 °C. Doubling the concentration to 1.6 mg/mL resulted in a small increase of R_g to 35.17 Å at 25 °C and to 36.90 Å at 37 °C (Table 1). The

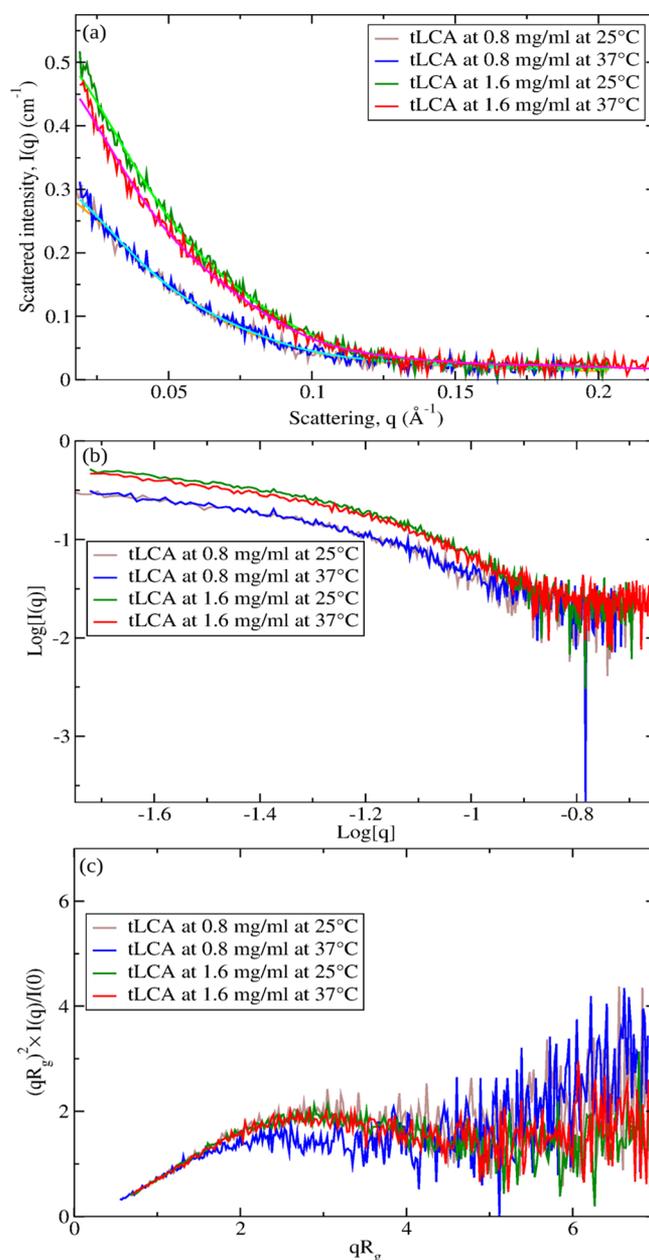


Figure 1. Scattering intensity plots describing the secondary and tertiary structure of BoNT/A LC in solution: Shown for 0.8 and 1.6 mg/mL solutions of tLCA at 25 and 37 °C are the curves of scattering intensity $I(q)$ vs q (panel a), $\text{Log}[I(q)]$ vs $\text{Log}[q]$ (far UV-CD; panel b), and $(qR_g)^2 \times I(q)/I(0)$ vs qR_g (near UV-CD; panel c). The spectral contribution from the buffer solution was subtracted from the combined (protein plus buffer solution) spectra as described in [Experimental and Theoretical Methods](#).

GNOM calculations revealed only slightly lower values of R_g in the $\sim 34–35$ Å range for 0.8 mg/mL and in the $33.9–34.14$ Å range for 1.6 mg/mL tLCA (Table 1). Importantly, these values are at odds with a much smaller radius of gyration $R_g = 22–23$ Å determined for the crystal structures of tLCA (the inset to Figure 3). This is clear evidence that the solution structures of tLCA are more extended compared to the crystal structure. Next, we analyzed Kratky plots, which profile the dependence of the product $(qR_g)^2 I(q)/I(0)$ on qR_g . Interestingly, the Kratky plots for different values of concentration and temperature show a distinct peak, but to the right of the peak

Table 1. Molecular Parameters of BoNT/A LC from SAXS Experiments^a

concentration (mg/mL)	temperature (°C)	q_{min} (Å ⁻¹)	q_{max} (Å ⁻¹)	R_g^b (Å)	qR_g^c	r^2	D_{max}^d (Å)	R_g^e (Å)	total estimate	χ^2
0.8	25	0.0239	0.221	34.25	0.61–1.28	0.906	126	34.03 ± 0.31	0.5774	1.15
0.8	37	0.0257	0.2118	35.0	0.64–1.26	0.891	131	34.95 ± 0.32	0.6639	1.058
1.6	25	0.0239	0.2167	35.17	0.66–1.25	0.947	129	33.9 ± 0.54	0.6746	1.085
1.6	37	0.0281	0.2191	36.9	0.7–1.26	0.939	133	34.14 ± 0.48	0.6158	1.246

^aShown for each protein concentration 0.8 and 1.6 mg/mL and temperature 25 and 37 °C are the minimum and maximum values of scattering vector q_{min} and q_{max} , the radius of gyration R_g , qR_g , the correlation coefficient r^2 , the longest atomic pair distance D_{max} , the total estimate, and the coefficient of correlation χ^2 . ^bValues of R_g are obtained from the Guinier plot. ^cValues of R_g are obtained from the GNOM plot. ^dValues of D_{max} are obtained from the GNOM plot. ^eRange of values of the product qR_g between the minimum and the maximum values.

the scattering intensity does not decay to zero (Figure 1c), which would have been indicative of a globular fold. Hence, the results of analysis of Kratky plots also indicate that in the 25–37 °C temperature range the tLCA molecule in 0.8–1.6 mg/mL aqueous solution does not form a globular folded structure; rather, the tLCA molecule forms extended conformations.

3.2. Atomic Pair Distance Distribution and *Ab Initio* Modeling. Intrigued by the results from Guinier and Kratky plots, we analyzed the distribution of atomic pair distances $P(r)$ using the GNOM package.⁴⁹ The profiles of unnormalized distribution $P(r)$ for 0.8 and 1.6 mg/mL solutions of tLCA at 25 and 37 °C temperature are shown in Figure 2a.

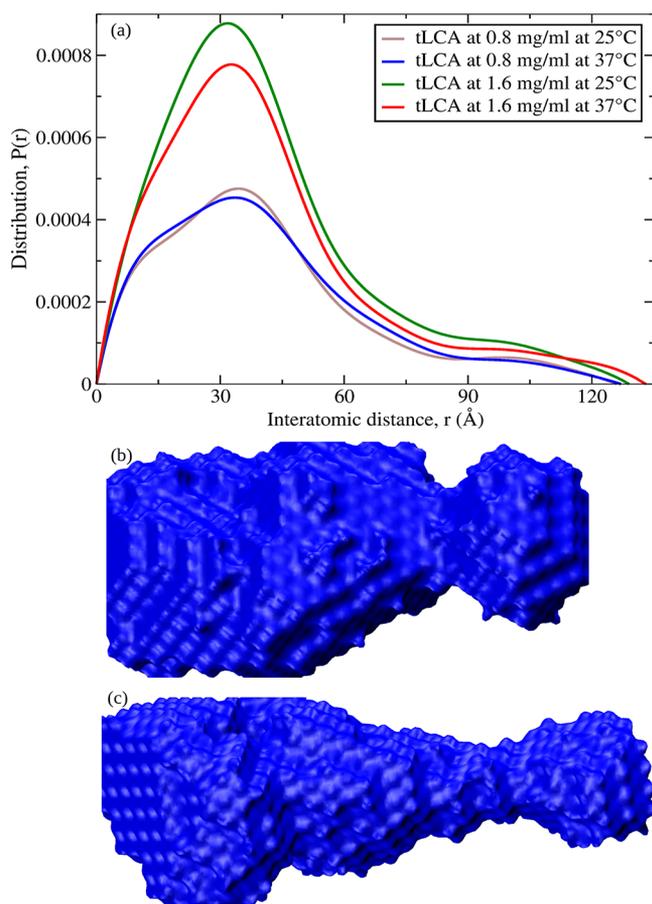


Figure 2. Distribution $P(r)$ and average 3D shape of BoNT/A LC in solution: Displayed are the profiles of un-normalized probability distribution of atomic pair distances $P(r)$ (panel a) and results of *ab initio* model calculations (in QuickSurf representation) for tLCA at 25 °C (panel b) and 37 °C (panel c). In panel a, the curves of $P(r)$ for 1.6 mg/mL solution are shifted up along the y -axis for clarity.

Surprisingly, $P(r)$ was found to be bimodal, suggesting the predominance of a multidomain ensemble-average conformation of the tLCA molecule. This feature could also result from dynamic equilibrium between the populations of compact and expanded structures of tLCA. The larger first peak (first mode) of the probability density mass corresponds to the shorter $D_1 \sim 32$ – 33 Å interatomic distance, whereas the smaller second peak (second mode) corresponds to the longer $D_2 \sim 95$ – 100 Å interatomic distance (Figure 2a). The profiles of $P(r)$ for 37 °C are slightly shifted to the right suggesting less compact (more extended) conformations at higher temperature. Analysis of $P(r)$ revealed that the largest interatomic distance is 129 Å at 25 °C and 133 Å at 37 °C (Table 1). Of note, these are much larger distances compared to the largest 71–72 Å distance for the crystal structures.

Next, we carried out *ab initio* dummy-atom modeling³⁰ of the tLCA molecule using the bimodal form of the distribution $P(r)$ at 25 and 37 °C. This enabled us to construct the unique solutions for the ensemble-averaged three-dimensional (3D) shape of tLCA, in order to characterize its quaternary structure. The results obtained for 25 and 37 °C are displayed, respectively, in parts b and c of Figure 2, which show the elongated shape of the tLCA molecule. These findings are further supported by low values of the mean normalized spatial discrepancy, 0.761 for 25 °C vs 0.707 for 37 °C.⁵⁰ At 25 °C, the ensemble-average tLCA molecule has the two-domain structure with a larger globular domain at one end and a smaller domain at the other end. At 37 °C, the multidomain tLCA molecule elongates further, and the central part of the molecule expands (Figure 2, parts b and c). These findings are entirely consistent with the bimodal form of the distribution $P(r)$ at 25 °C with a broad shoulder at 37 °C (Figure 2a).

3.3. Comparing Distributions $P(r)$ for Crystal and Solution Structures of tLCA. The results obtained thus far strongly suggest the existence of an extended multidomain tertiary structure of the tLCA molecule in solution. Yet, the experimental SAXS data do not provide any information whether the bimodal form of the distribution $P(r)$ is due to a single population (state) of tLCA conformations or due to several populations (e.g., globular and extended states) coexisting in equilibrium. To address this point and to obtain atomic-level structural insights for tLCA in aqueous solution, we turned to computational molecular modeling.

First, we analyzed the crystal structures of the tLCA molecule from the Protein Data Bank (PDB). There are a total of nine different crystal structures of tLCA and its fragments resolved to date (see Table 2). We compared the distributions $P(r)$ corresponding to these crystal structures with the curves of $P(r)$ for the 1.6 mg/mL solution of tLCA at 25 and 37 °C. The results displayed in Figure 3 show major differences. Indeed, the distributions $P(r)$ for all 9 crystal

Table 2. Crystal Structures of BoNT/A LC Available from the Protein Data Bank^a

structure	source	resolved residues	missing residues	PBD entry
BoNT/A complexed with SNAP-25	Mark et al., 2004	2–425	none	1XTG
BoNT/A WT crystal form A	Burnett et al., 2007	2–422	none	2ISE
BoNT/A with octapeptide II inhibitor	Zuniga et al., 2008	3–417	1, 2, 28, 29, 200–209, 250–253	3DS9
BoNT/A catalytic domain complexed with Llama Aa1	Dong et al., 2010	2–250	0, 1, 200–210	3K3Q
BoNT/A LC with JTH-NB-7239 peptide	Zuniga et al., 2010	1–417	27, 28, 60, 199–210, 247–255, 307, 419–425	3NF3
BoNT/A LC with zinc	Thompson et al., 2011	2–420	1, 27, 57, 200–210, 247–256, 421–430	3QIX
BoNT/A LC complexed with hydroxamate-based inhibitor PT-1	Thompson et al., 2011	2–421	1, 63, 64, 199–210, 244–256, 325, 326, 422–430	3QIY
BoNT/A LC complexed with hydroxamate-based inhibitor PT-2	Thompson et al., 2011	2–421	1, 199–208, 245–255, 422–430	3QIZ
BoNT/A LC complexed with hydroxamate-based inhibitor PT-3	Thompson et al., 2011	2–420	1, 199–209, 244–256, 421–430	3QJ0

^aShown for each structure are the publication, the lists of resolved and missing residues (not resolved in the PDB structure), and the PDB code.

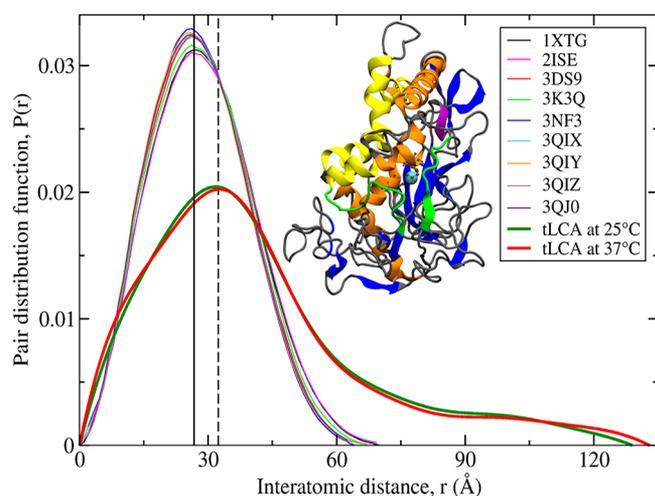


Figure 3. Distribution $P(r)$ for crystal structures of tLCA and its fragments: Profiles of the distributions of atomic pair distances $P(r)$ for the crystal structures of BoNT/A LC molecule (PDB entries are listed in the graph; see also Table 2) are compared with the experimental distribution $P(r)$ for the 1.6 mg/mL solution of tLCA at 25 and 37 °C temperature from SAXS experiments (color denotation is given in the graph). Vertical solid and dashed lines correspond to the maximum of $P(r)$ at $r \sim 26$ – 27 Å for the crystal structures and the first maximum (first mode) at $r \sim 32$ – 33 Å for the experimental distributions. Crystal structure of tLCA (PDB entry is 1XTG) displayed in the inset shows α -helices (in orange color), β -sheets (in blue), 3_{10} -helices (in purple), catalytic domain (in brown) with the zinc ion (blue ball), and random coil and turns (in gray). The α -exosite (in yellow) is the interface at the junction of four α -helices: α -helix 1 (residues 102–113), α -helix 2 (residues 310–321), α -helix 3 (residues 335–348), and α -helix 4 (residues 351–358). The β -exosite (in green) contains loop 250 (residues 242–259) and loop 370 (residues 359–370).

structures are unimodal and sharp, whereas the distributions for solution structures are bimodal and broad. In addition, the peak of probability density mass for the unimodal distribution for crystal structures is shifted by ~ 6 Å to the left from the location of the first mode for the bimodal distribution for solution conformations (Figure 3). Hence, all nine crystal structures of tLCA differ substantially from the solution structures in the 25–37 °C temperature range.

3.4. Ensemble of Solution Conformations of tLCA.

Quantitative analysis of the distributions of atomic pair distances $P(r)$ for crystal structures and solution structures

pointed to the existence of extended conformations of tLCA in aqueous solution with the interatomic distances exceeding 120 Å (Figures 2 and 3). Very long molecular extensions typically correspond to molecular conformations in which the N- and C-termini are spatially well separated. Furthermore, two separate sequence motifs, one at each terminus in LCA, were experimentally characterized and found to function simultaneously with different ligands. At the C-terminus, a dileucine motif containing residues Leu428 and Leu429 binds various members of the septin protein family.¹⁶ At the N-terminus, residues Pro2–Phe8 play a role in the binding to membrane bound SNAP-25 substrate⁵¹ (see Discussion). Therefore, it is advantageous from a functional viewpoint that these features are spatially well separated. For these reasons, we employed the steered MD simulations to generate an ensemble of extended conformations.

We used the most complete crystal structure of tLCA (PDB entry: 1XTG) and steered MD (SMD) simulations to generate computationally partially unfolded conformations of the tLCA molecule (see Experimental and Theoretical Methods). Using the force–extension profiles, depicted in Figure S1 in the Supporting Information, we selected the structures that correspond to (i) the minima numbered 1–6 for pulling at the N-terminus while constraining the C-terminus of tLCA molecule and (ii) the minima numbered 1–7 for pulling at the C-terminus while constraining the N-terminus of tLCA (Experimental and Theoretical Methods). These 13 partially unfolded conformations were then energy-minimized and equilibrated at 25 °C, which resulted in a total of 60 structures for tLCA. An example of one ~ 800 ns long MD equilibration run with the secondary structure alterations (at ~ 60 ns) and partial refolding (at ~ 190 ns) is displayed in Supporting Information, Figure S2. Analysis of structure correlations helped us to eliminate similar structures (with structure overlap $>50\%$) and to reduce the total number of structures to 25 (see Experimental and Theoretical Methods).

3.5. Theoretical Reconstruction of Distribution $P(r)$.

We used all 25 structures of tLCA to model the experimental distribution $P(r)$ at 25 and 37 °C. Selection of the most relevant structures with their weights was performed with the nonlinear regression approach used in machine learning (see Experimental and Theoretical Methods). For each of $i = 1, 2, \dots, n$ ($n = 25$) structures, the distribution of atomic pair distances $p_i(r)$ was calculated. The theoretical distribution $P_{th}(r)$ was constructed using a weighted superposition (additive model), $P_{th}(r) = w_1 p_1(r) + w_2 p_2(r) + \dots + w_n p_n(r)$,

where w_i is the statistical weight (population) for the i th structure in the ensemble. To perform the numerical fit of the theoretical distribution $P_{th}(r)$ to its experimental counterpart $P(r)$, we used nonlinear regression to minimize the mean squared error (MSE). Populations (regression coefficients) w_1, w_2, \dots, w_n were adjusted using the stochastic gradient descent algorithm (Experimental and Theoretical Methods). By discarding the structures with small weights <0.01 we identified the four most important conformations which account for 99% of all solution conformations of the tLCA molecule at 25 °C with $D_{max} = 129$ Å and at 37 °C with $D_{max} = 133$ Å. The theoretical profiles $P_{th}(r)$ are compared with the curves of $P(r)$ from SAXS experiments in Figure 4, which also shows the conformations of tLCA populated at 25 and 37 °C.

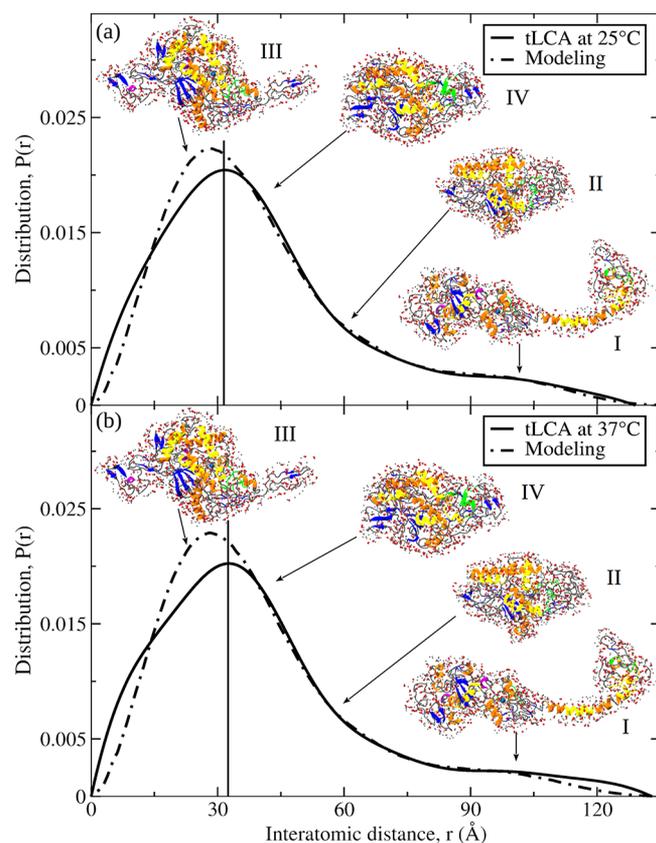


Figure 4. Theoretical reconstruction of distribution $P(r)$ and interpretation of SAXS spectra: Superposed are profiles of the normalized distributions of atomic pair distances $P(r)$ for tLCA obtained experimentally (SAXS) and theoretically (MD simulations and machine learning) at 25 °C (panel a) and 37 °C (panel b). The snapshots of tLCA structures generated *in silico* numbered I, II, III, and IV (shown with water molecules from the first solvation shell) correspond to the most important molecular conformations. Color denotations and structure assignments are the same as in the inset to Figure 3.

The theoretical distribution $P_{th}(r)$ captures the main features of its experimental counterpart $P(r)$. For 25 °C, the following was observed: (i) the similar width of the modes, i.e., ~ 36 Å (theory) vs ~ 41 Å (experiment) for the first mode and ~ 28 Å (theory) vs ~ 32 Å (experiment) for the second mode; (ii) locations of the modes, i.e., ~ 28 Å (theory) vs ~ 32 Å (experiment) for the first mode and ~ 100 Å (theory) vs ~ 100 Å (experiment) for the second mode; and (iii) the shoulder

between 45 and 75 Å distance connecting the modes. Similar agreement is obtained for 37 °C. Yet, $P_{th}(r)$ disagrees with $P(r)$ in the tail regions, e.g., in the rate of decrease to zero. Nevertheless, given the large size and high complexity of the tLCA structure and a limited number of conformations sampled in the MD simulations, the agreement between experiment and theory is very good.

The equilibrium populations of tLCA structures in solution came to $w_I = 0.36$, $w_{II} = 0.33$, $w_{III} = 0.22$, $w_{IV} = 0.09$ (for 25 °C); and $w_I = 0.32$, $w_{II} = 0.30$, $w_{III} = 0.33$, $w_{IV} = 0.05$ (for 37 °C). The first mode of $P(r)$ (with the higher peak) corresponds to the globular conformations III and IV, which are similar to the crystal structure of tLCA. The second mode of $P(r)$ (with the lower peak) is due to extended two-domain conformation I which is very different from the crystal structure of tLCA in that they contain the partially unfolded C-terminal portion of the tLCA molecule. The shoulder separating the modes corresponds to structure II. These conclusions are supported by the structure similarity (see Experimental and Theoretical Methods) between the solution conformations III, IV and the crystal structure 1XTG (Figures 3 and 4). The globular conformations III and IV have the largest structure overlap $\xi_{III} = 0.57$ and $\xi_{IV} = 0.56$ with the crystal structure, whereas the extended conformation I has the smallest overlap $\xi_I = 0.23$ with the crystal structure. For conformation II, the structure overlaps $\xi_{II} = 0.53$, is between the largest (0.57) and smallest (0.23) values.

4. DISCUSSION

The molecular basis of the biochemical and biophysical properties of botulinum neurotoxin serotype A (BoNT/A) can only be partly understood by analyzing the static crystal structures. The exceptionally high potency of this endopeptidase molecule, along with spurious solubility issues due to flexibility of the C-terminal domain that have proved to be difficult to overcome in crystallization assays, suggest the interplay in solution between molecular flexibility, foldability, and enzymatic functions. Indeed, to be highly effective, the molecule needs to be flexible enough to be able to adapt its structure in different cellular environments, yet stable enough to resist structural alterations in its active form(s). The role of dynamic structure alterations of LCA in solution becomes more intriguing in view of its high specificity for the substrate, SNAP-25, one of the SNARE proteins. LCA catalyzes the cleavage of SNAP-25, which is involved in the exocytotic machinery involved in neurotransmitter release from nerve terminals. In addition to high substrate selectivity, LCA has an entirely exclusive cleavage site despite sharing a nearly identical Zn^{2+} binding site with other (six) serotypes, which are comparatively less potent. Unlike other enzymes, LCA employs a longer substrate for recognition and activity, and it also remains active in the molten globule state. No other group of proteases are known to have such high specificity and selectivity and to retain high activity in the molten globule state.⁹ In addition to the molten globule, the preeminent molten globule enzyme (PRIME) conformation of LCA has been demonstrated to play an important role in LCA endopeptidase activity.^{8,9} Hence, to illuminate the mechanism of action of LCA and to identify potential inhibitors of this selective binding, it is important to understand the protein folding dynamics and conformational transitions of this molecule in aqueous solution.

X-ray crystallography provides valuable structural information about globular proteins, but these experiments cannot access the conformational dynamics of disordered proteins in solution. Based on X-ray data, many effective enzyme inhibitors have been developed and reported; yet, to date, no progress has been made in the development of inhibitors for the BoNT/A system. One explanation for the lack of progress is that the X-ray data for LCA is incomplete. Indeed, the full length LCA molecule has 444⁴ amino acid residues, whereas the crystal structures of tLCA has only 425 residues. Comparative analysis of the catalytic activity of LCA vs tLCA demonstrates the important role of the 25-residue long C-terminal portion of LCA in catalytic activity,⁴ which also agrees with the results of prior studies.²¹ However, a more plausible explanation for the lack of progress in the development of inhibitors for BoNT/A is that the available crystal structures do not represent the biologically active conformation, and that the solution structures are far more relevant to explaining BoNT/A function.

The crystal structure of full-length LCA is not yet solved. The crystal structure of tLCA (PDB entry 1XTG) is the closest model to represent its structural properties. The published CD spectra for LCA and tLCA were found to be similar and characteristic of folded proteins with reduced ellipticity for LCA,⁴ thus suggesting more α -helical content for tLCA. In addition, the catalytic activity of tLCA was found to be lower than that for LCA,⁴ which points to the role of missing residues in the structure and function of this enzyme. Our approach of using tLCA as a model system has enabled us to directly compare the results of SAXS measurements, namely the distribution of atomic pair distances $P(r)$, with the theoretical estimate of the same quantity $P_{th}(r)$ accessible in molecular modeling *in silico*. This approach also allows us to contrast the solution structure (experimental as well as theoretical) with published X-ray structures available only for the tLCA. In a sense, $P(r)$ is a molecular ruler that can be used to explore the tertiary structure and shape of proteins. Therefore, we used 0.8 and 1.6 mg/mL aqueous solutions of tLCA in SAXS experiments at 25 and 37 °C, as well as the crystal structure of tLCA molecule in molecular modeling *in silico*, to access the 3D spatial distribution of protein atoms in different solution conformations.

Analysis of the Kratky plots suggests that tLCA is not a globular protein; rather, it is partially unfolded in the 25–37 °C temperature range (Figure 1). Furthermore, the distribution of atomic pair distances $P(r)$ was found to be bimodal, with the highest maximum of probability mass around ~ 33 – 35 Å (first mode) and the lowest maximum around ~ 100 – 105 Å (second mode); see Figure 2. This correlates very well with the elongated two-domain ensemble-average solution structures of tLCA from *ab initio* model calculations (Figure 2). The highest first mode (lowest second mode) corresponds to the larger domain (smaller domain) in the structure. By contrast, for all the crystal structures of tLCA resolved to date (Table 2), the distribution of atomic pair distances $P(r)$ is unimodal and sharply peaked around ~ 25 Å distance (Figure 3). Furthermore, the average 129–133 Å size of solution structures (from SAXS experiments) is roughly twice the size of the crystal structures (65–70 Å; see Figures 2 and 3). Hence, the crystal structures of tLCA (Figure 3; Table 2) alone cannot account for the solution structure.

Although the SAXS data have enabled us to resolve the solution ensemble-average tertiary structure of tLCA, SAXS

measurements could not answer the following question: whether all tLCA molecules adopt the same conformation or there exist several interconverting subpopulations of different solution conformations of tLCA molecules that yield the same average structure? Therefore, to provide structural insights into the conformational transitions occurring in LCA molecules in solution, we turned to computational molecular modeling. The large size and overwhelming complexity of the tLCA molecule prevented us from using advanced sampling techniques. For this reason, we employed the all-atom MD simulations of forced unfolding of the crystal structure of tLCA to generate an entire manifold of extended tLCA conformations. Quantitative analysis of the force–extension curves generated *in silico* revealed that a large number of partial unfolding transitions occur in tLCA that are characterized by low ~ 50 – 150 pN unfolding forces (see peak forces in Figure S1) in the 20–40 nm range of molecular elongation. Invoking the free energy landscape perspective, this means that the free-energy landscape of tLCA unfolding is a collection of multiple energy minima for the native state and intermediate states separated by low energy barriers, which makes tLCA unstable to mechanical and thermal factors inside the cell. Furthermore, the unfolding forces were lower (with respect to the baseline) when tLCA was stretched at the C-terminus compared to the N-terminus (Figure S1), which implies that the C-terminal portion of tLCA is less stable than the N-terminal part.

Molecular modeling of the minimum-energy conformations of tLCA have enabled us to generate an entire ensemble of tLCA conformations, including the crystal-structure-like globular conformations, the single-domain extended conformations and the more elongated two-domain structures. Using these conformations and a simple machine learning approach to nonlinear regression, we carried out theoretical reconstruction of the entire distribution of atomic pair distances. By performing a numerical fit of the theoretical distributions $P_{th}(r)$ to their experimental counterparts $P(r)$, we were able to identify four conformations that are most pertinent to the solution structure of tLCA (Figure 4; snapshots I–IV). Structure alignment of these conformations with the *ab initio* structure reconstituted using the experimental SAXS data is displayed in Figure 5. Because no single conformation could account for the bimodal shape of the experimental distribution $P(r)$, we conclude that the overall 3D structure of tLCA is not represented by a single conformation, but rather, by an equilibrium ensemble of conformations of tLCA in solution (see Figure 5).

While conformations II and IV have tertiary structures close to the crystal structures (Figures 3 and 4) and correspond to the larger globular portion in the *ab initio* structure (Figure 5a), conformations III and I have very different more elongated tertiary structures, allowing them to determine the second smaller mostly unstructured domain in the *ab initio* structure (Figure 5a). For example, structure I contains the 40-residue long flexible α -helical connector Lys299–Phe338, which links the 86-residue long largely unstructured domain Asp339–Arg425 with the globular domain Pro2–Asn298 (Figure 5b). The α -helical connector (Lys299–Phe338) contains α -helix 2 (residues 310–321) and a smaller portion of α -helix 3 (residues 335–348). The 86-residue unstructured domain (Asp339–Arg425) contains amino acids from α -helix 3, α -helix 4 (residues 351–358), and loop 370 (residues 359–370) from the crystal structures (see Figure 3 caption). The unstructured domain is formed by long loops and turns and has only a short

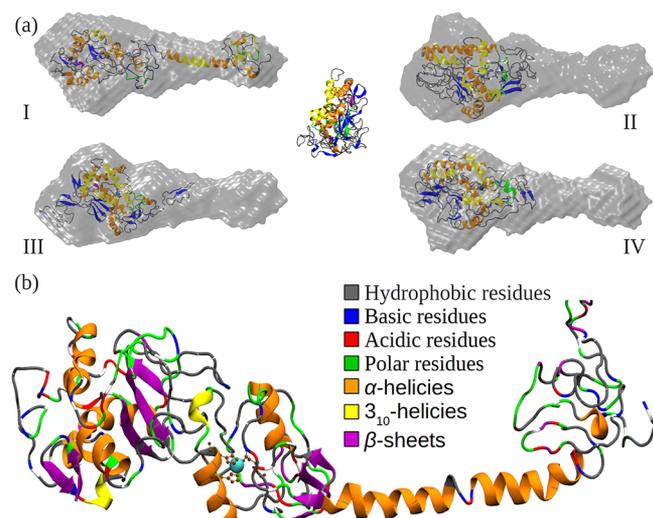


Figure 5. Ensemble average structures and shape of BoNT LCA. Panel a: The *ab initio* molecular reconstruction of tLCA based on SAXS data obtained at 37 °C is directly superimposed with the ensemble average solution structures I–IV resolved *in silico* (see Figure 4). Panel b: Magnified structure I showing 40-residue flexible α -helical connector (Lys299–Phe338), which contains α -helix 2 (residues 310–321) and a small portion of α -helix 3 (residues 335–348). The linker connects the globular domain (Pro2–Asn298) to the 86-residue long unstructured domain (Asp339–Arg425), which contains a portion of α -helix 3, α -helix 4 (residues 351–358) and loop 370 (residues 359–370). Color denotation for hydrophobic, polar, basic, and acidic amino acids is displayed in the graph.

helix Thr414–Lys417. Molecular modeling shows that the α -helical connector is capable of bending around residues Phe317–Lys318–Glu319, whereby the unstructured domain (Asp339–Arg425) attaches itself to the globular domain of tLCA (Pro2–Asn298) on the microsecond time scale (see structure I and snapshots 1–3 in Figure S2).

The results obtained also indicate that hydrophobic interactions are important in the initial recognition and binding of a substrate molecule by LCA. Both recognition and binding processes might be greatly enhanced by partial unfolding of LCA, during which cryptic hydrophobic residues become exposed. Not coincidentally, low 100–150 pN unfolding forces corresponding to a 5–40 nm magnitude extension mark the disruption of weak residue–residue contacts between hydrophobic residues (Figure S1). For example, of 86 residues forming the unstructured domain (Asp339–Arg425) in structure I, 36 amino acids are hydrophobic. These include the following patches of hydrophobic residues: Leu341–Tyr342, Met344–Leu345, Ile348–Tyr349, Phe354–Val355, Phe357–Phe358, Phe366–Leu367, Val373–Phe374, Ile378–Val379, Ile386–Tyr387, Phe390–Leu392, Leu397–Ala399, and Leu422–Val423. In the 40-residue long α -helical connector (Lys299–Phe338), there are 16 hydrophobic amino acids: Ala300, Ile303, Val304, Ala308, Leu310, Tyr312, Met313, Val316, Phe317, Tyr321, Leu322, Leu323, Phe331, Val333, Leu336, and Phe338. Therefore, the reversible unfolding–refolding transitions in LCA might comprise an important part of the molecular mechanism by which the enzyme recognizes the substrate in solution. This conclusion is supported by the optimal activity of LCA at 37 °C temperature, at which it is more unfolded as compared to its structure at lower 25 °C temperature.⁹ This finding is also supported by the results of our previous study of the urea

concentration dependence of LCA activity and molecular properties.¹⁰ Specifically, enzymatic activity of LCA was retained to ~50% in 5 M urea and, surprisingly, to ~100% in 7 M urea. The all-atom MD simulations of LCA denaturation in 5 M urea indicated a significantly extended length of the tLCA molecule compared to the structure in 0 M urea, and the enzymatically active structure in 7 M urea was found to be even more extended in length.¹⁰

The existence of several conformational subpopulations in the LCA folding–unfolding dynamics can also be understood by considering the size of its specific ligands (e.g., SNAP25 peptide; 206 amino acids) and the 18–20 Å depth of the active site in the crystal structure (calculated as the distance from the solvent accessible surface to the zinc ion). For a substrate or inhibitor molecule to access the active site of LCA, the molecule has to unfold by populating more extended conformations. Our MD simulation results clearly show that the C-terminal portion of tLCA is highly flexible with low unfolding energy cost; therefore, it can provide the mechanism for reversible unfolding–refolding to enable the LCA molecule to expose its active site. For example, as shown comparatively in Figure 6, in solution structure I the active site measured

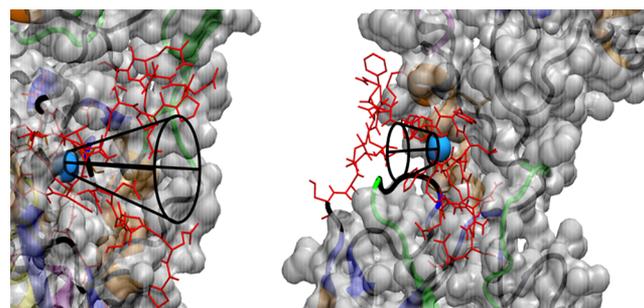


Figure 6. Catalytic site of BoNT LCA: Compared are different configurations of the active site of tLCA in the crystal structure IXTG (left panel) and in structure I displayed in Figure 5b (right panel). Atoms of the residues in the active rim are connected by thin red lines representing the covalent bonds. Color denotation is same as in the inset to Figure 3 with the addition of random coils and turns (in black) and solvent accessible surface (in transparent gray). The depth of the active site pocket is approximated by the black cones.

from the zinc atom is buried only 8–9 Å from the solvent accessible surface, which is more than a 2-fold reduction of the depth of the active site found in the crystal structure. This conclusion is also supported by the significant reduction in the enzymatic activity of tLCA, which emphasizes the role in catalysis of the missing 20-residue long C-terminal portion of LCA.⁴

In addition to possessing improved active site exposure to substrate, the extended forms of unfolded tLCA found in this study also serve to illustrate an additional advantageous feature of this flexible structure, one that places the N-terminus near the maximum distance from the C-terminus (~117 Å in structure I) compared to their close proximity in the crystal structure (~48 Å), as shown in Figure 7. Two separate sequence motifs, one found at each terminus, have been characterized to carry out separate functions in LCA. At the C-terminus, an acidic dileucine motif (containing residues Leu428, Leu429) recruits and binds various members of the septin protein family and causes these cytoskeleton–plasma membrane linking species to sequester and localize at the

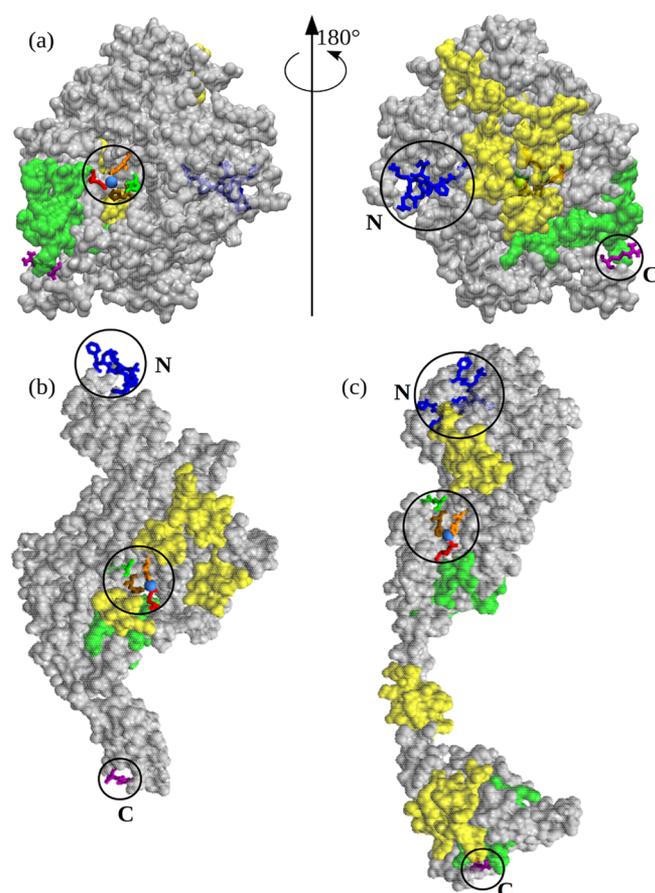


Figure 7. Comparison of crystal and solution structures of BoNT/LCA. Panel a: The crystal structure of tLCA molecule (PDB entry: 1XTG) showing residues Pro2-Phe8 (in blue color), and Arg425 (in purple); residues His223, Glu224, His227, and Glu262 in the active site are shown in brown, green, orange, and red, respectively. Panels b and c display partially unfolded extended solution structures III and I, respectively, reconstructed using computational molecular modeling (see Figures 4 and 5). The structures show locations of the N-terminus and the C-terminus (circled) with the N-terminal residues Pro2-Phe8 (in blue) and C-terminal residue Arg425 (in purple). Residues in the α - and β -exosites are shown, respectively, in yellow and green color; the solvent accessible surface is shown in transparent gray.

membrane surface (in LCA the dileucine motif is four amino acids longer than in our tLCA structure shown in Figure 7 ending at Arg425). This dileucine motif, validated by mutation studies (Leu428Ala and Leu429Ala), binds with and forms septin oligomeric ring and fiber structures with bound LCAs, colocalizing to form punctate complexes that are involved in protein trafficking functions.¹⁶ This sequestered septin bound environment for LCA was demonstrated to be a requirement, responsible for its long-lived intracellular proteolytic activity and neuroparalytic effects. The second LCA sequence motif at the N-terminus Pro2-Phe8 (the N-terminal 7 amino acids of tLCA in this study), was found to be a factor in the binding to membrane bound SNAP-25 substrate (via substrate residues 80–110 and 180–197), causing colocalization with LCA at the membrane.⁵¹ Since both C- and N- termini contain sequence motifs that function simultaneously with different ligand systems, it is, therefore, highly advantageous that these features exist spatially well separated on flexible structures. Extended solution structures, such as structures III (110 Å) and I (132

Å) in Figure 7, allow maximum steric flexibility for dynamic behavior and interactions with molecular partners in the crowded molecular environment at the membrane surface, a feature not present in the close proximity of these sequence motifs in the C- and N- termini found in the tLCA crystal structure.

5. CONCLUSIONS

In this combined experimental and computational study, we have provided insights into the solution structure of LCA and its relationship to the biological function of BoNT/A, the most toxic poison known to mankind. These structures and a wealth of biochemical evidence from prior studies^{8–11} together strongly suggest our major conclusion, that the native solution state of BoNT/A protein is not defined by its crystal structures (Table 2); rather, it exists as an ensemble of several interconverting conformational isoforms. Although in our analysis we used the tLCA molecule, these conclusions are also, and perhaps more so, applicable to full length LCA, which is known to be more flexible in solution.⁴ The present analysis further validates the existence of the PRIME conformation (i.e., extended conformation of LCA at 37 °C), which accounts for the optimal activity of this unique enzyme.⁹ Structural insights at the atomic level of detail we have gathered in this study for BoNT/A under physiological conditions can now be used by researchers to help rationally design novel molecular inhibitors, thereby developing effective new countermeasures against botulism.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.0c02817>.

Figure S1, which shows the results of simulations of forced-induced unfolding of BoNT/A light chain, and Figure S2, which shows the secondary structure propensities for the two-domain structure I. (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Valeri Barsegov – Department of Chemistry, University of Massachusetts, Lowell, Massachusetts 01854, United States; orcid.org/0000-0003-1994-3917; Email: Valeri_Barsegov@uml.edu

Bal Ram Singh – Botulinum Research Center, Institute of Advanced Sciences, Dartmouth, Massachusetts 02747, United States; Email: bsingh@umassd.edu

Authors

Raj Kumar – Botulinum Research Center, Institute of Advanced Sciences, Dartmouth, Massachusetts 02747, United States

Farkhad Maksudov – Department of Chemistry, University of Massachusetts, Lowell, Massachusetts 01854, United States

Olga Kononova – Department of Chemistry, University of Massachusetts, Lowell, Massachusetts 01854, United States; orcid.org/0000-0001-9267-312X

Kenneth A. Marx – Department of Chemistry, University of Massachusetts, Lowell, Massachusetts 01854, United States

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jpcb.0c02817>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by DOD (Grant W81XWH-14-C-0070) to B.R.S. and by the Infinity Foundation (V.B.).

REFERENCES

- (1) Simpson, L. L. Molecular Pharmacology of Botulinum Toxin and Tetanus Toxin. *Annu. Rev. Pharmacol. Toxicol.* **1986**, *26*, 427–453.
- (2) Arnon, S. S.; Schechter, R.; Inglesby, T. V.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Fine, A. D.; Hauer, J.; Layton, M.; Lillibridge, S.; Osterholm, M. T.; O'Toole, T.; Parker, G.; Perl, T. M.; Russell, P. K.; Swerdlow, D. L.; Tonat, K.; et al. Botulinum Toxin as a Biological Weapon: Medical and Public Health Management. *J. Am. Med. Assoc.* **2001**, *285*, 1059–1070.
- (3) Kumar, R.; Dhaliwal, H. P.; Kukreja, R. V.; Singh, B. R. The Botulinum Toxin as a Therapeutic Agent: Molecular Structure and Mechanism of Action in Motor and Sensory Systems. *Semin. Neurol.* **2016**, *36*, 010–019.
- (4) Feltrup, T. M.; Patel, K.; Kumar, R.; Cai, S.; Singh, B. R. A Novel Role of C-Terminus in Introducing a Functionally Flexible Structure Critical for the Biological Activity of Botulinum Neurotoxin. *Sci. Rep.* **2018**, *8*, 1–13.
- (5) De Paiva, A.; Meunier, F. A.; Molgó, J.; Aoki, K. R.; Dolly, J. O. Functional Repair of Motor Endplates after Botulinum Neurotoxin Type A Poisoning: Biphasic Switch of Synaptic Activity between Nerve Sprouts and Their Parent Terminals. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 3200–3205.
- (6) Keller, J. E.; Neale, E. A.; Oyler, G.; Adler, M. Persistence of Botulinum Neurotoxin Action in Cultured Spinal Cord Cells. *FEBS Lett.* **1999**, *456*, 137–142.
- (7) Adler, M.; Keller, J. E.; Sheridan, R. E.; Deshpande, S. S. Persistence of Botulinum Neurotoxin A Demonstrated by Sequential Administration of Serotypes A and E in Rat EDL Muscle. *Toxicon* **2001**, *39*, 233–243.
- (8) Cai, S.; Singh, B. R. Role of the Disulfide Cleavage Induced Molten Globule State of Type A Botulinum Neurotoxin in Its Endopeptidase Activity. *Biochemistry* **2001**, *40*, 15327–15333.
- (9) Kukreja, R.; Singh, B. R. Biologically Active Novel Conformational State of Botulinum, the Most Poisonous Poison. *J. Biol. Chem.* **2005**, *280*, 39346–39352.
- (10) Kumar, R.; Kukreja, R. V.; Li, L.; Zhmurov, A.; Kononova, O.; Cai, S.; Ahmed, S. A.; Barsegov, V.; Singh, B. R. Botulinum Neurotoxin: Unique Folding of Enzyme Domain of the Most-Poisonous Poison. *J. Biomol. Struct. Dyn.* **2014**, *32*, 804–815.
- (11) Kumar, R.; Kukreja, R. V.; Cai, S.; Singh, B. R. Differential Role of Molten Globule and Protein Folding in Distinguishing Unique Features of Botulinum Neurotoxin. *Biochim. Biophys. Acta, Proteins Proteomics* **2014**, *1844*, 1145–1152.
- (12) Breidenbach, M. A.; Brunger, A. T. Substrate Recognition Strategy for Botulinum Neurotoxin Serotype A. *Nature* **2004**, *432*, 925–929.
- (13) Li, L.; Singh, B. R. Role of Zinc Binding in Type A Botulinum Neurotoxin Light Chain's Toxic Structure. *Biochemistry* **2000**, *39* (34), 10581–10586.
- (14) Patel, K.; Cai, S.; Singh, B. R. Current Strategies for Designing Antidotes against Botulinum Neurotoxins. *Expert Opin. Drug Discovery* **2014**, *9* (3), 319–333.
- (15) Kononova, O.; Litvinov, R. I.; Zhmurov, A.; Alekseenko, A.; Cheng, C. H.; Agarwal, S.; Marx, K. A.; Weisel, J. W.; Barsegov, V. Molecular Mechanisms, Thermodynamics, and Dissociation Kinetics of Knob-Hole Interactions in Fibrin. *J. Biol. Chem.* **2013**, *288*, 22681–22692.
- (16) Vagin, O.; Tokhtaeva, E.; Garay, P. E.; Souda, P.; Bassilian, S.; Whitelegge, J. P.; Lewis, R.; Sachs, G.; Wheeler, L.; Aoki, R.; Fernandez-Salas, E. Recruitment of Septin Cytoskeletal Proteins by Botulinum Toxin A Protease Determines Its Remarkable Stability. *J. Cell Sci.* **2014**, *127*, 3294–3308.
- (17) DasGupta, B. R.; Dekleva, M. L. Botulinum Neurotoxin Type A: Sequence of Amino Acids at the N-Terminus and around the Nicking Site. *Biochimie* **1990**, *72*, 661–664.
- (18) Kriegstein, K. G.; DasGupta, B. R.; Henschen, A. H. Covalent Structure of Botulinum Neurotoxin Type A: Location of Sulfhydryl Groups, and Disulfide Bridges and Identification of C-Termini of Light and Heavy Chains. *J. Protein Chem.* **1994**, *13*, 49–57.
- (19) Lacy, D. B.; Tepp, W.; Cohen, A. C.; DasGupta, B. R.; Stevens, R. C. Crystal Structure of Botulinum Neurotoxin Type A and Implications for Toxicity. *Nat. Struct. Biol.* **1998**, *5*, 898–902.
- (20) Baldwin, M. R.; Bradshaw, M.; Johnson, E. A.; Barbieri, J. T. The C-Terminus of Botulinum Neurotoxin Type A Light Chain Contributes to Solubility, Catalysis, and Stability. *Protein Expression Purif.* **2004**, *37*, 187–195.
- (21) Mizanur, R. M.; Frasca, V.; Swaminathan, S.; Bavari, S.; Webb, R.; Smith, L. A.; Ahmed, S. A. The C Terminus of the Catalytic Domain of Type A Botulinum Neurotoxin May Facilitate Product Release from the Active Site. *J. Biol. Chem.* **2013**, *288*, 24223–24233.
- (22) Tsai, Y. C.; Maditz, R.; Kuo, C. L.; Fishman, P. S.; Shoemaker, C. B.; Oyler, G. A.; Weissman, A. M. Targeting Botulinum Neurotoxin Persistence by the Ubiquitin-Proteasome System. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16554–16559.
- (23) Arndt, J. W.; Jacobson, M. J.; Abola, E. E.; Forsyth, C. M.; Tepp, W. H.; Marks, J. D.; Johnson, E. A.; Stevens, R. C. A Structural Perspective of the Sequence Variability Within Botulinum Neurotoxin Subtypes A1–A4. *J. Mol. Biol.* **2006**, *362*, 733–742.
- (24) Burnett, J. C.; Ruthel, G.; Stegmann, C. M.; Panchal, R. G.; Nguyen, T. L.; Hermone, A. R.; Stafford, R. G.; Lane, D. J.; Kenny, T. A.; McGrath, C. F.; Wipf, P.; Stahl, A. M.; Schmidt, J. J.; Gussio, R.; Brunger, A. T.; Bavari, S. Inhibition of Metalloprotease Botulinum Serotype A from a Pseudo-Peptide Binding Mode to a Small Molecule That Is Active in Primary Neurons. *J. Biol. Chem.* **2007**, *282*, 5004–5014.
- (25) Toth, S. I.; Smith, L. A.; Ahmed, S. A. Extreme Sensitivity of Botulinum Neurotoxin Domains towards Mild Agitation. *J. Pharm. Sci.* **2009**, *98*, 3302–3311.
- (26) Segelke, B.; Knapp, M.; Kadkhodayan, S.; Balhorn, R.; Rupp, B. Crystal Structure of Clostridium Botulinum Neurotoxin Protease in a Product-Bound State: Evidence for Noncanonical Zinc Protease Activity. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6888–6893.
- (27) Kumaran, D.; Rawat, R.; Ahmed, S. A.; Swaminathan, S. Substrate Binding Mode and Its Implication on Drug Design for Botulinum Neurotoxin A. *PLoS Pathog.* **2008**, *4*, e1000165.
- (28) Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. X-Ray Solution Scattering (SAXS) Combined with Crystallography and Computation: Defining Accurate Macromolecular Structures, Conformations and Assemblies in Solution. *Q. Rev. Biophys.* **2007**, *40*, 191–285.
- (29) Hopkins, J. B.; Gillilan, R. E.; Skou, S. BioXTAS RAW: Improvements to a Free Open-Source Program for Small-Angle X-Ray Scattering Data Reduction and Analysis. *J. Appl. Crystallogr.* **2017**, *50*, 1545–1553.
- (30) Franke, D.; Svergun, D. I. DAMMIF, a Program for Rapid Ab-Initio Shape Determination in Small-Angle Scattering. *J. Appl. Crystallogr.* **2009**, *42*, 342–346.
- (31) Kozin, M. B.; Svergun, D. I. Automated Matching of High- and Low-Resolution Structural Models. *J. Appl. Crystallogr.* **2001**, *34*, 33–41.
- (32) Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: A Unified Platform for Automated Protein Structure and Function Prediction. *Nat. Protoc.* **2010**, *5*, 725–738.
- (33) Humphrey, W.; Dalke, A.; Schulten, K. Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (34) van der Spoel, L. A. H. *GROMACS 2020 Source Code*; 2020.
- (35) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. Gromacs: High Performance Molecular

Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(36) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulation. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(37) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(38) Lindahl, E.; Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B. Implementation of the Charmm Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* **2010**, *6*, 459–466.

(39) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(40) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(41) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(42) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(43) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.

(44) Ferrara, P.; Apostolakis, J.; Caflisch, A. Evaluation of a Fast Implicit Solvent Model for Molecular Dynamics Simulations. *Proteins: Struct., Funct., Genet.* **2002**, *46*, 24–33.

(45) Brooks, B. R.; Brooks, C.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: Molecular Dynamics Simulation Package. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(46) Zhmurov, A.; Kononova, O.; Litvinov, R. I.; Dima, R. I.; Barsegov, V.; Weisel, J. W. Mechanical Transition from α -Helical Coiled Coils to β -Sheets in Fibrin(Ogen). *J. Am. Chem. Soc.* **2012**, *134*, 20396–20402.

(47) Kononova, O.; Litvinov, R. I.; Blokhin, D. S.; Klochkov, V. V.; Weisel, J. W.; Bennett, J. S.; Barsegov, V. Mechanistic Basis for the Binding of RGD- and AGDV-Peptides to the Platelet Integrin $\text{AIIb}\beta 3$. *Biochemistry* **2017**, *56*, 1932–1942.

(48) Feigin, L. A.; Svergun, D. I.; Taylor, G. W. *Structure Analysis by Small-Angle X-Ray and Neutron Scattering* **1987**, DOI: 10.1007/978-1-4757-6624-0.

(49) Svergun, D. I. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J. Appl. Crystallogr.* **1992**, *25*, 495–503.

(50) Volkov, V. V.; Svergun, D. I. Uniqueness of Ab Initio Shape Determination in Small-Angle Scattering. *J. Appl. Crystallogr.* **2003**, *36*, 860–864.

(51) Chen, S.; Barbieri, J. T. Association of Botulinum Neurotoxin Serotype a Light Chain with Plasma Membrane-Bound SNAP-25. *J. Biol. Chem.* **2011**, *286*, 15067–15072.