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

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

Botulinum Research Center and Institute of Advanced Sciences, North Dartmouth, MA
University of California, Berkley, CA; University of Massachusetts, Lowell, MA

Objective

- In the absence of solution structure, our aim is to model the solution structure of tLCA (truncated Light Chain).
- Find out the answer of the following questions
  a) Is LCA is partially disordered protein?
  b) Is the folded structure of LCA in solution different from the crystal structure?
  c) What are the conformational features of LCA in solution at the minima of free energy landscape?

Introduction

- A critical step in developing an inhibitor against a protein's function requires a comprehensive knowledge of its solution conformational states and the molecular basis of its function.
- In general, a crystal structure can address this issue, but the static view becomes inadequate for systems where thermal fluctuation and/or large functionally relevant conformational states might be populated.
- Perhaps due to considerable inherent flexibility of Botulinum toxin light chain A (LCA), crystallization proved to be a formidable task. This also leads to the apparent difference between the solution structure and the crystal structure.
- The solution structure might also explain how two different sequence motifs, one located at the N-terminus and the other close to the C-terminus of LCA can functionally interact better with their different biological partners at the intracellular surface of the plasma membrane, than in the crystal structure, leading to increases longevity of the enzymatic activity.
- Experimental data supporting the high potency, longevity, and exclusive site selectivity characteristics of LCA strengthen the hypothesis that this molecule has a different conformational features in solution that 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 exact length of LCA in the functional state.
- 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.
- 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.
- The difficulty in developing a biologically consistent inhibitor against BoNT is due in part to the structurally flexible pre-molten globule-like structure (PRIME) the endopeptidase adopts under optimum enzyme activity conditions.
- 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.
- The crystal structure of only truncated LCA (424 a.a.) is quite likely that the conformational states of the PRIME and molten globule states are at variance with the crystal structure.
- Therefore it is important to describe the ensemble structures and characterize the dynamic properties of the enzyme in solution, where it exists in a manifold of conformational subpopulations.
- The role of dynamic structure alterations of this protein becomes more intriguing in view of the extreme specificity of the endopeptidase for its substrate and its potential utility in designing specific antidotes against botulism threats.
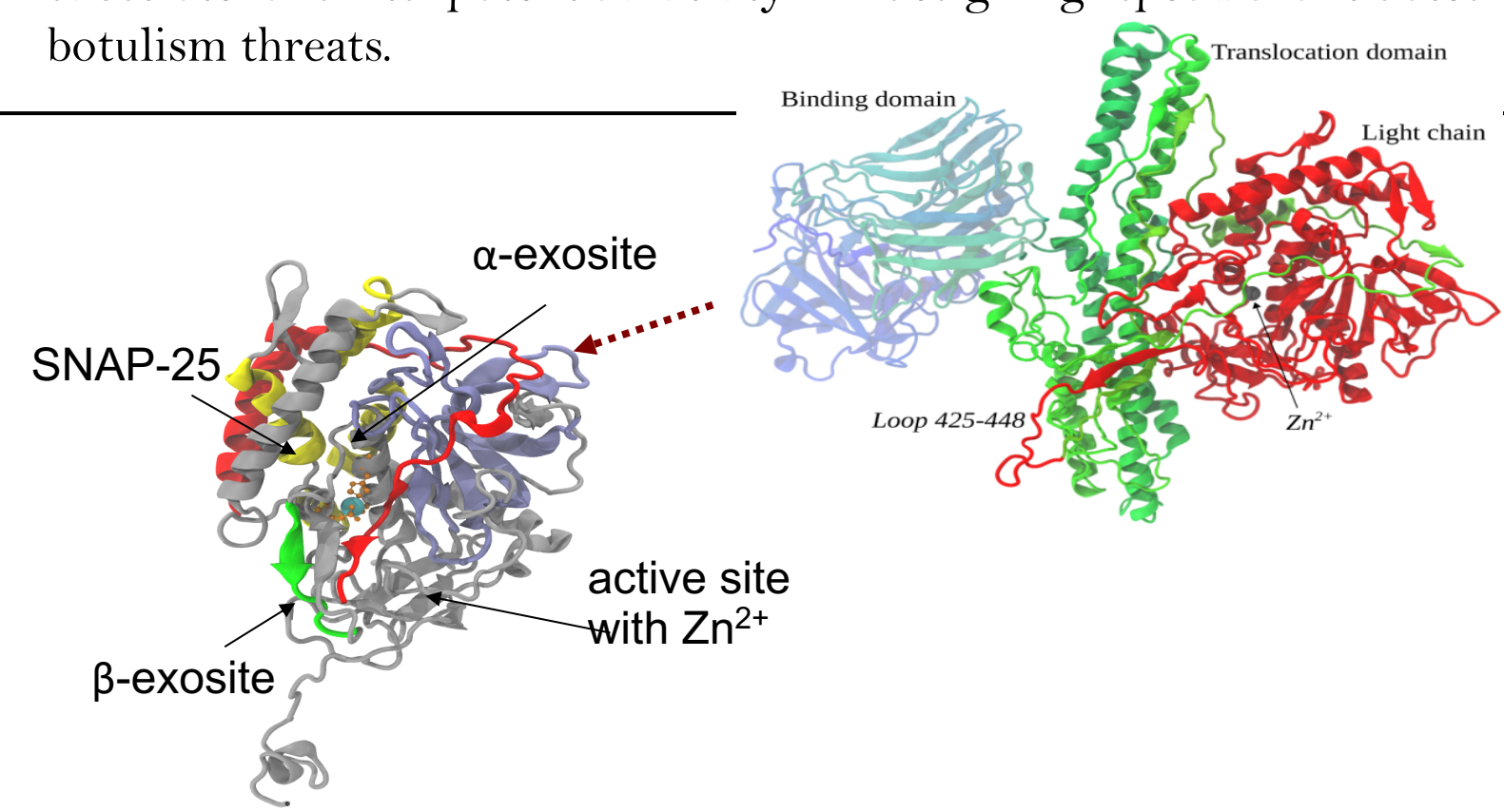
Figure 1: Right figure: Crystal structure of botulinum toxin A (PDB ID: 3BTA). Left figure: a- exosite (substrate binding) and ß-exosite (substrate recognition). tLCA used to model solution structures. Phe425-Lys448 was synthesized in silico, which is energy minimized to obtained an equilibrium structure. LC (Met1-Lys448) was extracted from equilibrium structure.

Our Approach

Step I: Using tLCA as a model protein. We used 0.8 and 1.6 mg/ml aqueous solutions of tLCA in SAXS experiments at 25 and 37 ºC.
Step II: Crystal structure of tLCA used in molecular modelling *in silico*, to access the 3D spatial distribution of protein atoms in different solution conformation.
Step III: Compared the results of SAXS experiments, namely the distribution of atomic pair distance P(r), with the theoretical estimate of the same quantity, $P_{th}(r)$, accessible in molecule modelling *in silico*.
Step IV: Contrasted the solution structure (experimental and theoretical) with published X-ray structures available only for the tLCA. In a sense, P(r) is a molecule ruler that can be used to explore the tertiary structure and shape of proteins.
Step-V: Finally, nonlinear regression in conjunction with a generalized additive model to reconstruct theoretically the distribution of atomic pair distance P(r) using the structure output from MD simulations.

Results
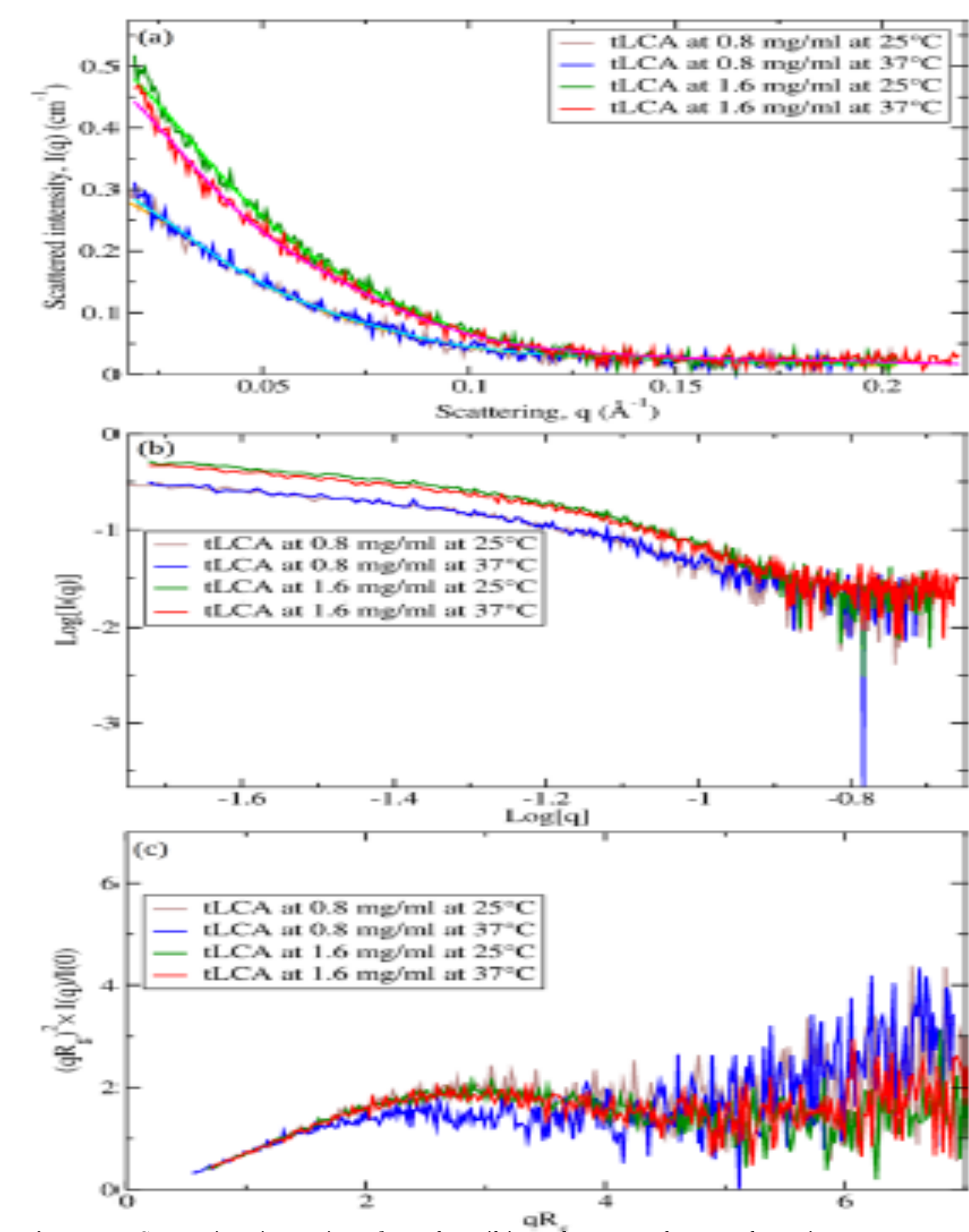
Figure 2: Scattering intensity plots describing the secondary and tertiary structure of BoNT/A LC in solution: Shown for 0.8 and 1.6 mg/mL solutions of tLCA at 25 and 37 °C are the curves of scattering intensity I(q) vs q (panel a), Log[I(q)] vs Log[q] (for UV-CD; panel b), and (qRg)2 × (I(q)/I(0)) vs qRg (near UV-CD; panel c). The spectral contribution from the buffer solution was subtracted from the combined (protein plus buffer solution) spectra.
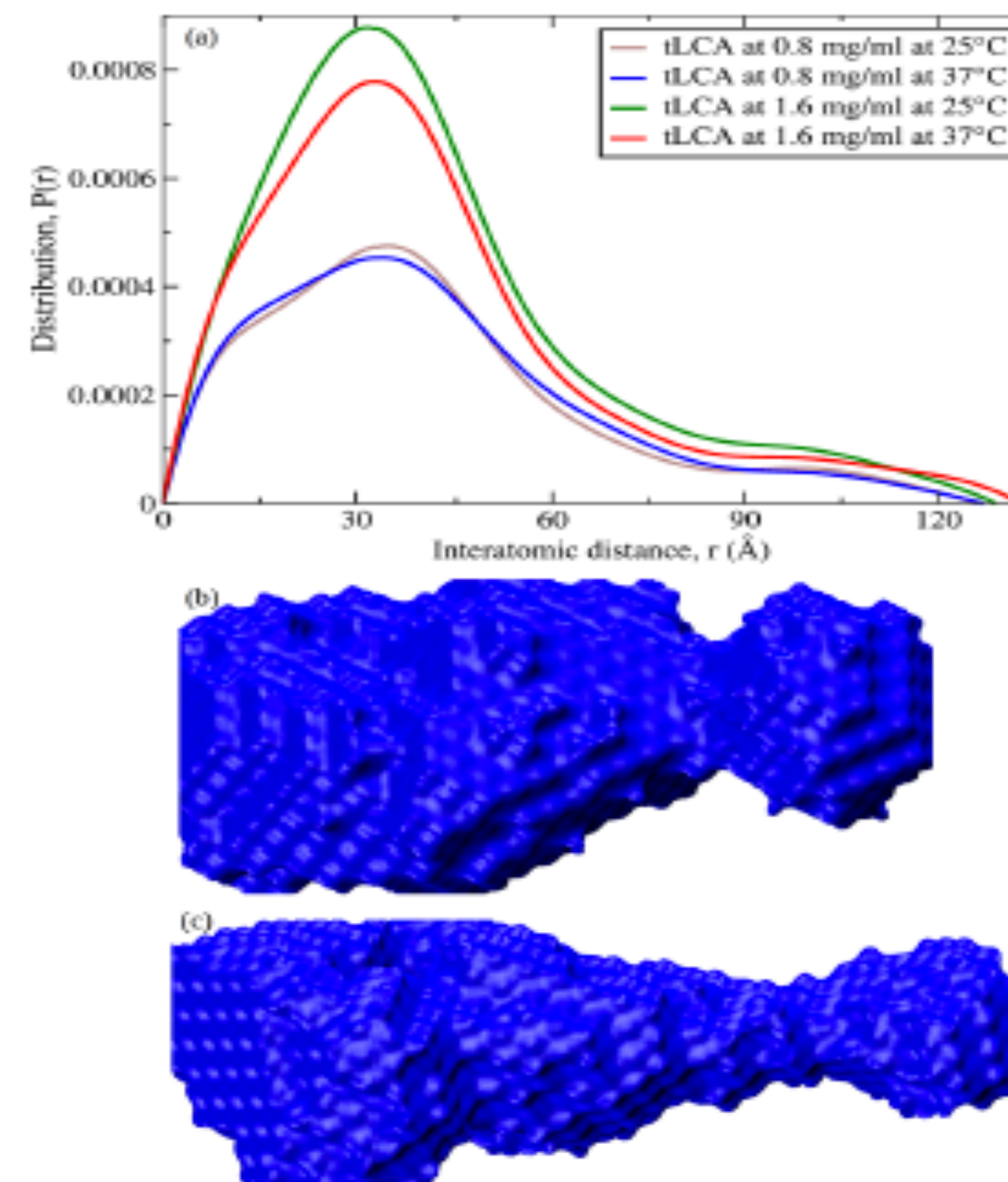
Figure 3: Distribution P(r) and average 3D shape of BoNT/A LC in solution: Displayed are the profiles of un-normalized probability distribution of atomic pair distances P(r) (panel a) and results of ab initio model calculations (in QuickSurf representation) for tLCA at 25 °C (panel b) and 37 °C (panel c). In panel a, the curves of P(r) for 1.6 mg/mL solution are shifted up along the y-axis for clarity.

Theoretical reconstruction of P(r)
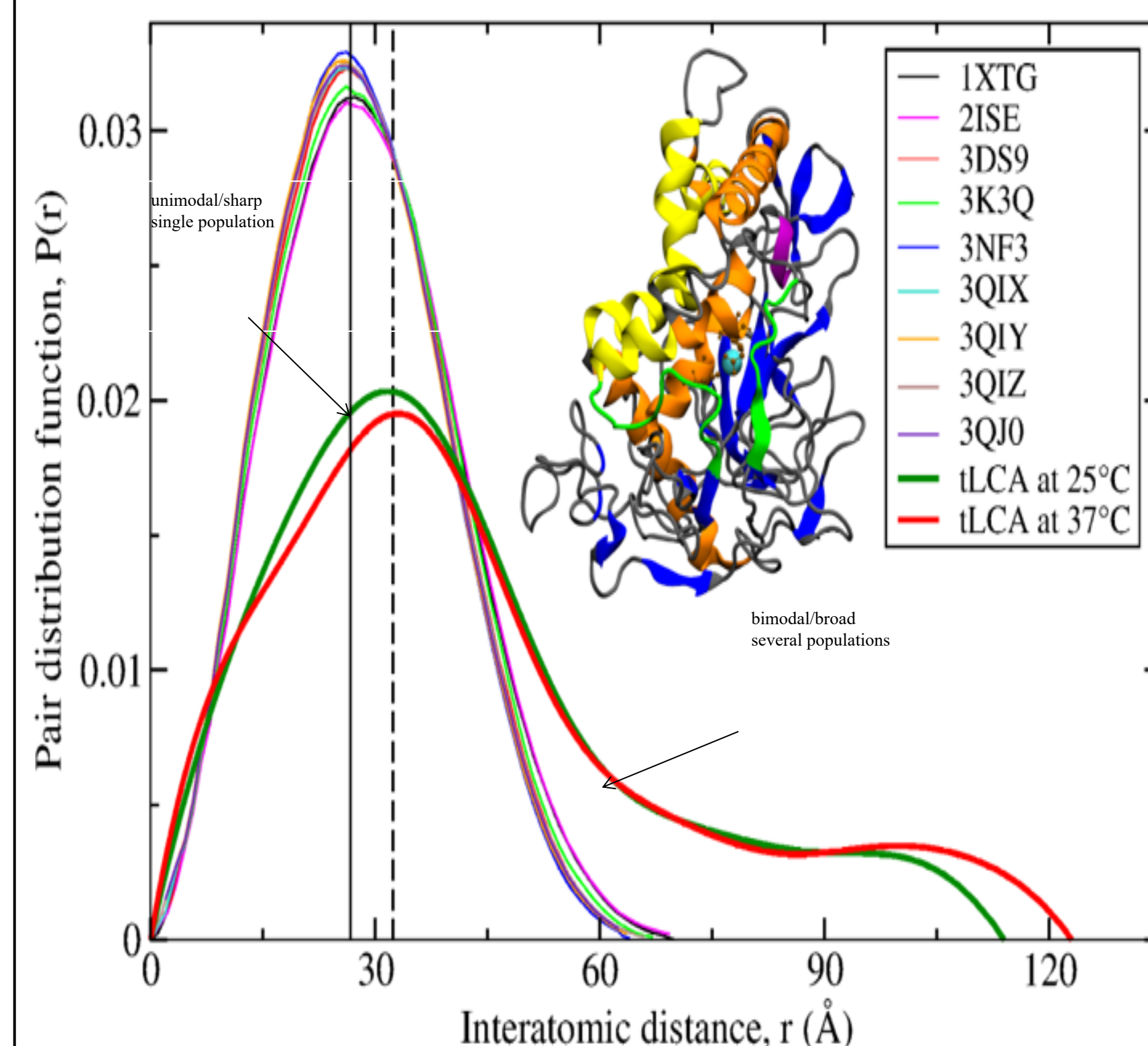
Figure 4: Distribution P(r) for crystal structures of tLCA and its fragments: Profiles of the distributions of atomic pair distances P(r) for the crystal structures of BoNT/A LC molecule (PDB entries are listed in the graph) are compared with the experimental distribution P(r) for the 1.6 mg/mL solution 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 ∼ 26−27 Å for the crystal structures and the first maximum (first mode) at r ∼ 32−33 A' for the experimental distributions.
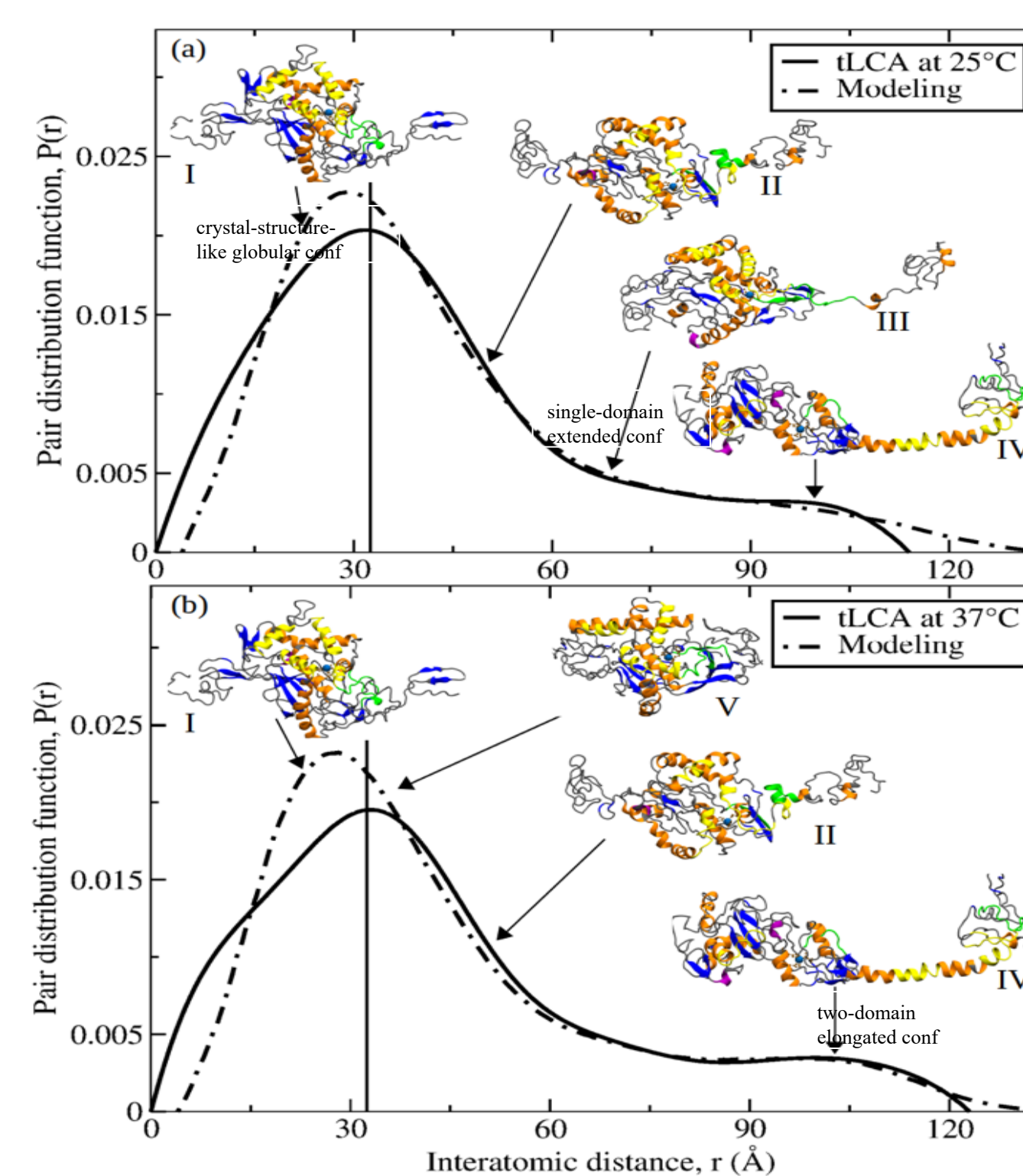
Figure 5: Theoretical reconstruction of distribution P(r) and interpretation of SAXS spectra: Superposed are profiles of the normalized distributions of atomic pair distances P(r) for tLCA obtained experimentally (SAXS) and theoretically (MD simulations and machine learning) at 25 °C (panel a) and 37 °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.

Important Findings

1) Analysis of Kratky plots suggests that tLCA is not a globular protein, rather, it is partially unfolded.
2) The distribution of atomic pair distance was found to be bimodal, with the highest maximum of probability mass around ∼ 33 -35 A (first mode) and the lowest maximum around ∼100 -105 A (second mode).
3) By contrast, for all the crystal structure of tLCA resolved to date the distribution of atomic pair distribution P(r) is unimodal and sharply peaked around ∼ 25 A.
4) Quantitative analysis of the force-extension curves generated in silico revealed that a large number of partial unfolding transition occur in tLCA that are characterized by low ∼50 − 150 pN unfolding forces in the 20 − 40 nm range of molecular elongation, indicating the free energy landscape of tLCA unfolding is a collection of multiple energy minima for the native state.
5) This also indicates that the intermediate states are separated by low energy barriers, which makes tLCA unstable to mechanical and thermal factors in the cell.
6) Molecular modelling 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.
7) No single conformation could account for bimodal distribution. Additionally, all the crystal structure P(r) is significantly lower than first mode of SAXS derived P(r), indicating globular conformation is not the desired conformation for LCA in solution.
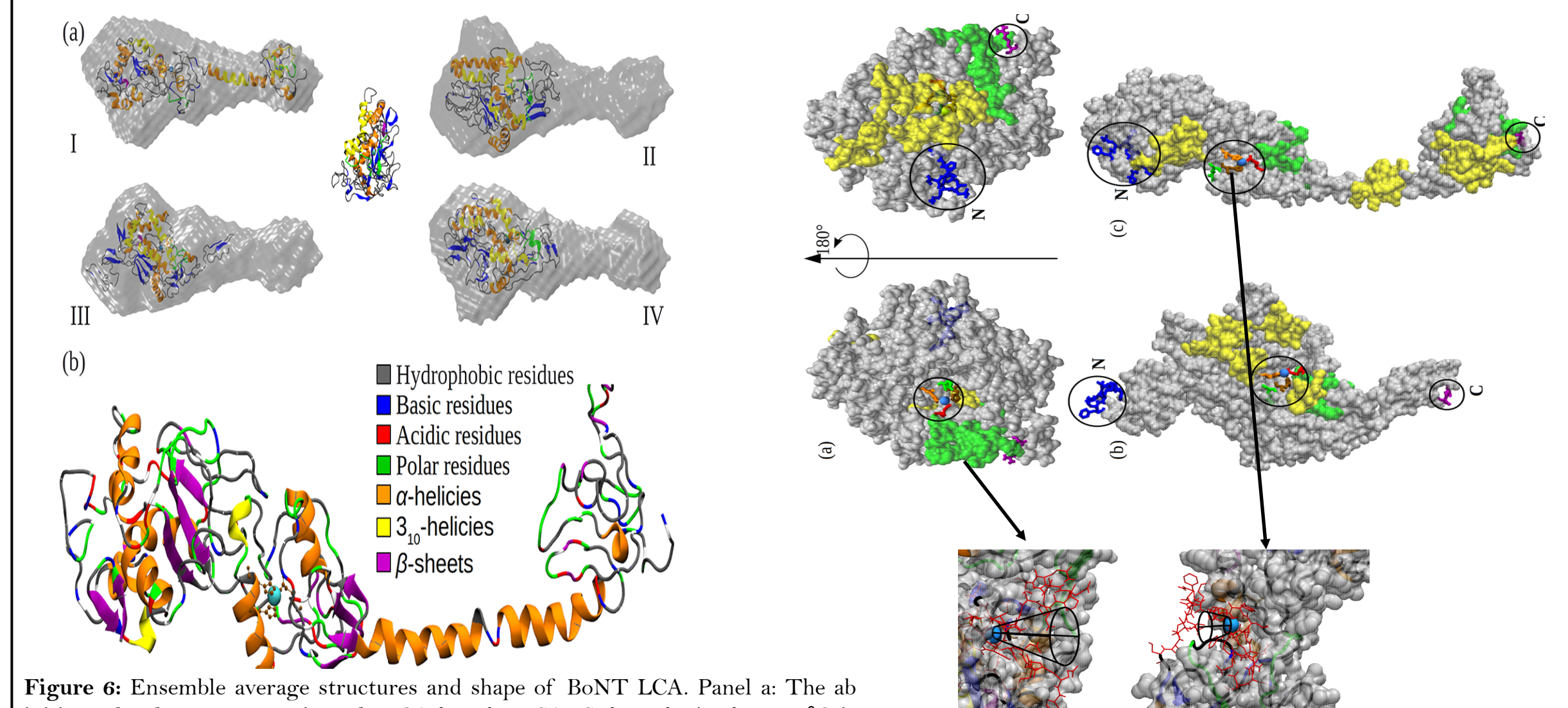
Figure 6: Ensemble average structures and shape of BoNT LCA. Panel a: The ab initio molecular reconstruction of tLCA based on SAXS data obtained at 37 °C is directly superimposed with the ensemble average solution structures I−IV resolved in silico. Panel b: Magnified structure I showing 40-residue flexible 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.

Figure 7: Catalytic site of BoNT LCA: Compared are different configurations of the active site of tLCA in the crystal structure 1XTG (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.

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 poison known to mankind.
- Our findings suggest that the native solution state of BoNT/A protein is not defined by its crystal structures..
- LCA 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.
- The present analysis further validates the existence of the PRIME conformation, which accounts for the optimal activity of this unique enzyme..
- This study will be very helpful in developing effective new countermeasures against botulism.

Acknowledgement