Computational Tridimensional Protein Modeling of Cry1Ab19 Toxin from Bacillus thuringiensis BtX-2

Kashyap, S.*, B. D. Singh², and D. V. Amla³

1National Bureau of Agriculturally Important Microorganisms (ICAR), Kusmaur, Kaithauli, Mau Nath Bhanjan-275101, (U. P.) India
2School of Biotechnology, Faculty of Science, Banaras Hindu University, Varanasi 221005, (U. P.) India
3Molecular Biology and Genetic Engineering Division, National Botanical Research Institute, Rana Pratap Marg, P.B. #436, Lucknow 226001, (U. P.) India

Received: June 15, 2011 / Revised: January 17, 2012 / Accepted: February 2, 2012

We report the computational structural simulation of the Cry1Ab19 toxin molecule from B. thuringiensis BtX-2 based on the structure of Cry1Aa1 deduced by x-ray diffraction. Validation results showed that 93.5% of modeled residues are folded in a favorable orientation with a total energy Z-score of -8.32, and the constructed model has an RMSD of only 1.13Å. The major differences in the presented model are longer loop lengths and shortened sheet components. The overall result supports the hierarchical three-domain structural hypothesis of Cry toxins and will help in better understanding the structural variation within the Cry toxin family along with facilitating the design of domain-swapping experiments aimed at improving the toxicity of native toxins.

Keywords: Three-dimensional structure, homology modeling, Cry1Ab19, Bacillus thuringiensis BtX-2, third party annotation

Insecticidal crystal protein produced by the soil bacterium Bacillus thuringiensis (Bt) belongs to a large toxin family and until now has been considered harmless to mammals [17]. Cry1A series toxins are produced as an inactive protoxin within Bt sporangia that, upon ingestion by a susceptible larva, binds to a high-affinity receptor site on the midgut membrane. The bonded toxin disrupts the electrolyte balance across the cell membrane, leading to cell lysis and larval death [8]. The crystal structure of the active toxins in solution has been analyzed for Cry1Aa [9], Cry2A [15], Cry3A [12], Cry3B [5], Cry1Ac [4], Cry4Ba [2], and Cry4Aa [1] by x-ray diffraction studies, whereas Cry11Bb [7], Cry5Aa [14], and Cry5Ba [19] have been predicted by homology modeling methods. In spite of such studies, little effort has been focused on the structural study of Cry1Ab.

Computational modeling of proteins is performed to construct a three-dimensional model using known structural information for one or more related proteins. Prediction of protein structures using bioinformatics tools is becoming ever important as new sequences are pooling up. To some extent, these computational studies can speed up the process of determining experimental structures by molecular replacement phasing of x-ray diffraction studies. A complete understanding of the three-dimensional structure of all of the Cry1 family members is desirable for a comprehensive understanding of the mechanisms governing their toxicity. Therefore, in the present study, we report a model for the structure of the Cry1Ab19 toxin based on the hypothesis of structural similarity with the Cry1Aa toxin. This model supports existing hypotheses of receptor insertion and will further provide an initiation point for domain mutagenesis experiments between Cry1Ab and other toxins for elucidation of possible modes of action.

Sequence alignment between Cry1Ab19 (AAW31761) and Cry1Aa1 (PDB 1ciy A) was performed directly on the HHpred interactive server (http://toolkit.tuebingen.mpg.de/hhpred) to detect homology and predict the structure under global alignment mode, and the end alignment was used in the MODELLER software [16]. Refinement of the developed model was performed on the Summa Lab’s server (http://silvio.cs.uno.edu/proteinrefinementserver/). Linear depiction of three-dimensional structures was generated by PDB Sum (http://www.ebi.ac.uk/pdbsum/). The model was validated on the PROCHECK Web server [11]. Figures and electrostatic potentials were generated with PyMOL, and Ramachandran

*Corresponding author
Phone: +91-547-2530800; Fax: +91-547-2530358; E-mail: sudhanshukshyp@gmail.com

# Supplementary data for this paper are available on-line only at http://jmb.or.kr.
plot assessment was carried out on the RAMPAGE server [13]. Backbones were superimposed and calculation of RMSD was performed on the SuperPose server (http://wishart.biology.ualberta.ca/SuperPose/).

Over existing methods used in the elucidation of toxin structure, molecular modeling has an added advantage due to the use of template information from known biological samples. To determine a structure with molecular modeling, a template sequence having a similarity higher than 40% is required; to this end, we used the HHpred server to screen the PDB database for secondary structure similarity. Subsequently, the structural model of the Cry1Ab19 toxin was developed, using the alignment between Cry1Ab19 and Cry1Aa1 (Smith-Waterman score: 3,337; % identity: 87.9; a.a. overlap: 588; z-score: 4,034.6; E-value: 6.9e-218 for Cry1Aa1) (Fig. 1). Alignment with each Cry1Aa domain was straightforward and there was a clear correspondence of amino acids on the N- and C-terminal sides. Structural

Fig. 1. Amino acid sequence alignment of the Cry1Ab19 endotoxin with Cry1Aa (1ciy:A). The sequential numbers above represent the number of amino acids. The residues highlighted in red represent helices, those in blue, strands, those in green, turns, and those in black, coils. Structural features were generated using SAS software (http://www.ebi.ac.uk/thornton-srv/databases/sas).
comparison of the Cry1Aa toxin with the theoretical model of Cry1Ab19 indicates correspondence to the general model of a Cry protein (Fig. 2). Domain I consists of nine \( \alpha \)-helices and two small \( \beta \)-strands. The identified helices and strands are \( \alpha_1 \) (Pro\(^{68}\)-Ser\(^{81}\)), \( \alpha_{2a} \) (Ala\(^{87}\)-Trp\(^{98}\)), \( \alpha_{2b} \) (Pro\(^{103}\)-Ile\(^{117}\)), \( \alpha_3 \) (Glu\(^{123}\)-Ala\(^{152}\)), \( \alpha_4 \) (Pro\(^{157}\)-Phe\(^{181}\)), \( \alpha_5 \) (Gln\(^{187}\)-Trp\(^{215}\)), \( \alpha_{6} \) (Ala\(^{219}\)-Val\(^{231}\)), \( \alpha_7 \) (Ser\(^{256}\)-Tyr\(^{283}\)), \( \alpha_{7b} \) (Ile\(^{300}\)-Thr\(^{302}\)), \( \beta_0 \) (Pro\(^{304}\)-Asn\(^{308}\)), and \( \beta_1 \) (Ser\(^{316}\)-Ser\(^{323}\)) (Fig. 2, 3A, and 3B). All of the helices in the Cry1Ab19 model were slightly shorter than those in Cry1Aa. The Cry1Ab19 domain I model relates well with data from Gazit et al. [6], who suggested that \( \alpha_4 \) and \( \alpha_5 \) insert into the membrane in an antiparallel manner as a helical hairpin. It is also possible that both helices cross the membrane with their polar sides exposed to the solvent. Kumar and Aronson [10] demonstrated that mutation in the base of helix 3 and the loop between \( \alpha_3 \) and \( \alpha_4 \) cause alterations in the balance of negatively charged residues, which leads to a loss of toxicity. Mutations in helices \( \alpha_2 \) and \( \alpha_6 \) and the surface residues of \( \alpha_3 \) have no important effect on toxicity. Meanwhile, helices \( \alpha_4 \) and \( \alpha_5 \) seem to be very sensitive to mutation. Any mutation aimed to increase amphility in these helices will improve the pore-forming activity of Cry1Ab19-type toxins. This region is considered to be the most variable domain among Cry toxins, and it has been shown to be primarily involved in receptor recognition and determination of specificity.

As with other reported Cry toxin structures, domain II of the Cry1Ab19 toxin consists of three Greek key beta sheets arranged in a beta prism topology. It is composed of 2
helices $\alpha_{8a}$ (Ser$^{442}$-Glu$^{467}$) and $\alpha_{8b}$ (Pro$^{465}$-Gly$^{480}$), and 11 $\beta$-strands, $\beta_2$ (Ile$^{332}$-His$^{341}$), $\beta_3$ (Glu$^{346}$-Ser$^{357}$), $\beta_4$ (Arg$^{382}$-Ala$^{385}$), $\beta_5$ (Tyr$^{392}$-Tyr$^{400}$), $\beta_6$ (Ser$^{414}$-Leu$^{417}$), $\beta_7$ (Thr$^{419}$-Ala$^{422}$), $\beta_8$ (Ala$^{432}$-Tyr$^{434}$), $\beta_9$ (Thr$^{439}$-Asp$^{441}$), $\beta_{10}$ (His$^{461}$-Phe$^{469}$), $\beta_{11}$ (Ala$^{483}$-His$^{490}$) and $\beta_{12}$ (Thr$^{505}$-Pro$^{508}$). Domain III comprises residues 471–608 and is highly conserved, the only important modification being a three-residue deletion between $\beta_{16}$ and $\beta_{17}$. The $\alpha$- and $\beta$-strands in this domain are $\beta_{13a}$ (Thr$^{511}$-Leu$^{513}$), $\beta_{13b}$ (Ser$^{520}$-Val$^{522}$), $\beta_{14}$ (Ile$^{532}$-Arg$^{535}$), $\beta_{15}$ (Gly$^{539}$-Asn$^{542}$), $\beta_{16}$ (Tyr$^{556}$-Ser$^{564}$), $\beta_{17}$ (Leu$^{568}$-Ile$^{571}$), $\beta_{18}$ (Arg$^{577}$-Phe$^{584}$), $\alpha_9$ (Ser$^{609}$-Ser$^{612}$), $\beta_{19}$ (Arg$^{601}$-Gly$^{607}$), $\beta_{20}$ (Ser$^{614}$-His$^{622}$), and $\beta_{21}$ (Val$^{630}$-Pro$^{639}$).

Several studies indicate that site mutations in these residues reduce toxicity and alter channel properties in Cry1Ac [4] and Cry1Aa1 [3, 18].

Stereo checking and structure validation are essential constraints in computational biology. PROCHECK is frequently used for the structural evaluation of proteins in the form of a Ramachandran plot. The recognition of errors was evaluated on the ProSA server, which is

Fig. 4. Evaluation of Cry1Ab19 and comparison with Cry1Aa1.
(A) Evaluation of Cry1Ab19 using the ProSA server. The plot indicates the nearness of the constructed structure with native structures. The $Z$-score of the evaluated model was -8.25, shown as a large dot. (B) Ramachandran plot analysis showing placement of residues in the deduced model (93.5% favorable orientations). Structure residues were considered separately for torsion angles. (C) Superposed 3D backbone structure of Cry1Aa1 (yellow) and Cry1Ab19 (red) showing a high degree of similarity between the reference and the generated model.
employed in the refinement and validation of a computational model in terms of individual residue energy levels. This software indicates the overall model quality by comparing known and x-ray deciphered structures of similar sizes and groups. The assessment is graphically presented in the form of a Z-score (Fig. 4A), which in our case was found to be -8.32, categorizing it within the permissible range (+5 to -15) of native conformational structures. We further extended our validation specifically for proline and glycine residues owing to their inherent ability to influence the overall structural integrity of the molecule utilizing the RAMPAGE server. The graphical output shown in Fig. 4B indicates that most of the residues (93.5%) have ϕ and ψ angles in the core regions and 4.3% are in allowed regions, except for some proline and glycine residues (1.6%), which fall in the outlier region. Most bond lengths, bond angles, and torsion angles are in the range of values expected for a naturally folded protein. Finally, the computed Cry1Ab19 backbones were superimposed on reference Cry1Aa1 (Fig. 4C) for overall evaluation of structural deviations using the SuperPose server, which calculated a low RMSD (1.13Å). In conclusion, the evidence presented here, based on the identification of structurally equivalent Cry1Aa residues in the Cry1Ab19 toxin through homology modeling indicates that, due to high amino acid homology between these two toxins, they share a common tridimensional structure. Cry1Ab19 contains the most variable regions in the loops of domain II, which determine the specificity of these toxins, and the computational model supports the three-domain structural hypothesis of Cry toxins.

Acknowledgments

The authors are grateful to ICAR for the RAship to S.K. Infrastructure facility and encouragement by Director NBAIM are duly acknowledged.

REFERENCES