J Mol Model (2013) 19:5301–5316
DOI 10.1007/s00894-013-2010-x

ORIGINAL PAPER

Structural analysis and molecular dynamics simulations of novel δ-endotoxin *Cry*1Id from *Bacillus thuringiensis* to pave the way for development of novel fusion proteins against insect pests of crops

Budheswar Dehury · Mousumi Sahu · Jagajjit Sahu · Kishore Sarma · Priyabrata Sen · Mahendra K. Modi · Madhumita Barooah · Manabendra Dutta Choudhury

Received: 30 May 2013 / Accepted: 9 September 2013 / Published online: 24 October 2013

Abstract The theoretical three-dimensional structure of a novel δ-endotoxin Cry1Id (81 kDa) belonging to Cry1I class, toxic to many of the lepidopteran pests has been investigated through comparative modeling. Molecular dynamics (MD) simulations was carried out to characterize its structural and dynamical features at 10 ns in explicit solvent using the GROMACS version 4.5.4. Finally the simulated model was validated by the SAVES, WHAT IF, MetaMQAP, ProQ, ModFOLD and MolProbity servers. Despite low sequence identity with its structural homologs, Cry1Id not only resembles the previously reported Cry structures but also shares the common five conserved blocks of amino acid residues. Although the domain II of Cry1Id superpose well with its closest structural homolog Cry8Ea1, variation of amino acids and length in the apical loop2 of domain II was observed. In this work, we have hypothesized that the variations in apical loop2 might be the sole factor for providing variable surface accessibility to Cry1Id protein that could be important in receptor recognition. MD simulation showed the proposed endotoxin retains its stable conformation in aqueous solution. The result from this study is expected to aid in the development hybrid Cry proteins and new potent fusion proteins with novel specificities against different insect pests for improved pest management of crop plants.

Keywords Cry1Id · Endotoxin · Fusion protein · Hybrid cry protein · Molecular dynamics simulation

Introduction

The endospore forming gram-positive bacterium *Bacillus thuringiensis* (Bt) produces crystalline (Cry) protein inclusions (termed as δ-endotoxins) with natural insecticidal effect on the insect pests, mites and nematodes of the order Lepidoptera, Coleoptera, Diptera, Hymenoptera, Homoptera, Orthoptera and Mallophaga etc. [1, 2]. To date several Cry proteins have been isolated and characterized from different strains *Bacillus thuringiensis* with activity against different insect pests. The sprayable formulations containing Cry proteins are used as a dynamic element in the field of insect pest management [3]. These Cry proteins produced by Bt is considered as potential resource as an alternative to synthetic chemical pesticides to control major insect pests in plants [4, 5]. Furthermore, the members of the crystalline δ-toxin family are widely used in biopesticide formulations and also in generation of transgenic crops for insect control.

The insecticidal activity of the bacterium Bt is mainly attributed to the crystal proteins encoded by the Cry genes. These toxins are named crystal (Cry) proteins because of their

Electronic supplementary material The online version of this article (doi:10.1007/s00894-013-2010-x) contains supplementary material, which is available to authorized users.

B. Dehury · M. Sahu · J. Sahu · K. Sarma · P. Sen · M. K. Modi · M. Barooah (✉)
Department of Agricultural Biotechnology, Assam Agricultural University, Jorhat 785013, Assam, India
e-mail: m17barooah@yahoo.co.in

B. Dehury · M. D. Choudhury
Department of Life Science and Bioinformatics, Assam University, Silchar 788011, Assam, India

abilities to auto-crystallize in the bacterial cytoplasm. Diverse Cry proteins have different insecticidal specificities; mostly dependent on the genes which encode the Cry proteins. The Cry proteins exist in the primitive stage as inactive pro-toxin form which latter converted into mature cytotoxic endotoxin by the action of certain kinds of insect midgut proteases [6]. Eventually the transformation of Cry proteins from inactive form (crystal pro-toxin) to cytotoxic toxin follows a multifaceted process. The mode of action of the Cry proteins generally follows the ingestion by a susceptible animal; activated by the gut protease where they bind to specific receptors on the midgut epithelium, leading to toxins oligomerization, membrane insertion and finally resulting in the pore structure formation [7]. This activation process appears to involve a sequential series of proteolytic cleavages, starting at the C-terminus and proceeding toward the N-terminus until the protease-stable toxin is generated [8]. During the process of sporulation the bacterium Bt synthesizes cytoplasmic inclusions that contain one or more Cry proteins as inactive protoxins [2]. Ingestion of crystal proteins by insect larvae results in the conversion of protoxins into active toxins at alkaline pH. The activated toxins bind to insect-specific receptors present on the plasma membrane of the midgut epithelial cells and create transmembrane pores, leading to disruption of ionic balance, cell lysis and death of insects [7, 9, 10]. Due to the high selectivity and effectiveness of these toxins, the introduction of Cry genes into plants for generation of Bt crops has considerably increased in recent years [11].

Most of Cry proteins encoded and expressed by a variety of *B. thuringiensis* isolates, exhibit significant similarity in three dimensional (3D) structure and mode of action despite having considerable difference at sequence level and in the target specificity [5]. The primary sequence and three-dimensional structural analysis of Cry proteins has provided substantial insight into their structure-function activity. Although a remarkable difference in their insecticidal specificities exists among the Cry proteins, they share a common folding pattern as well as in their domain architecture. Generally most of the Cry proteins are comprised of three domains: a seven- helix-bundle domain (DI), a three-antiparallel-β-sheet domain (DII), and a β-sandwich domain (DIII). The domain I of Cry proteins are mostly helical in nature forming an α-helical bundle which is thought to be involved in membrane insertion and pore formation [12–14]. In contrast, domain II is composed of three antiparallel β-sheets which form a "Greek key" topology where the β-sheets are arranged in such a manner to form a β-prism fold. Furthermore, domain II contains the surface-exposed loops which are considered as the most probable candidates for receptor binding [15, 16]. Domain III consists of two twisted antiparallel β-sheets, forming a β-sandwich with a jelly roll topology. In addition domain III is treated as a multifunctional domain which performs a number of key roles in the biochemistry, structural integrity, membrane penetration, ion channel function and a major determinant of receptor binding [2]. Complete multiple sequence alignment among the members of Cry protein family has revealed five highly conserved blocks in the N-terminal half [17, 18]. Based on spatial structural superposition of the known Cry proteins, the five blocks were found to be lying in the center of individual domain or interface of three domains, implying their putative involvement in inter-domain contacts, flexibility and balance of the overall stability of Cry proteins [19].

The Cry genes have been classified as Cry1 to Cry40, cyt1, and cyt2 and are ranked according to their homology [20]. So far only seven structures of Cry toxins from Bt namely Cry1Aa (PDB ID: 1CIY) [21], Cry2Aa (PDB ID: 1I5P) [22], Cry3Aa (PDB ID: 1DLC) [13], Cry3Bb (PDB ID: 1JI6) [23], Cry4Aa (PDB ID: 2C9K) [15], Cry4Ba (PDB ID: 1 W99) [24], Cry8Ea (PDB ID: 3 EB7) [19] have been determined by X-ray crystallographic methods.

Among the family of *Cry* genes, the Cry1 sub family of proteins were extensively studied and have been used in lepidopteran insect pest management of various crops. Generally, the crystals are composed of pro-toxins of approximately 130 kDa, but Cry1I-type genes are usually silent genes capable of encoding a protein of about 81 kDa in *B. thuringiensis* strains [2, 25–27]. Among the diverse classes of Cry family proteins, the Cry1I class has been grouped in to six subclasses ranging from Cry1Ia to Cry1If. These members within the subclass display several unique features in terms of their mechanism of action and specificity toward various receptors. Although the crystal structure Cry1Aa (133 kDa) belonging to Cry1I group has been reported, but Cry1Id (81 KDa) which belongs to the same Cry1I class (shares only 43.0 % sequence identity with Cry1Aa) has not been studied yet. Moreover, the Cry1 toxins have been extensively used in studies of lepidopteron insect control but have attracted less attention and fewer efforts have been focused on Cry1Id member's structural studies. In addition, it was ascertained that the three dimensional structure of the novel Cry protein, Cry1Id was not available in the protein data bank, it is imperative to have three-dimensional structural information to understand the structure-function behavior of Cry1Id and mechanisms underlying toxicity. In the present study, the theoretical three-dimensional structure of the novel Cry protein, Cry1Id (Cry1I-subgroup) which is toxic to many lepidopteran pests was obtained by comparative modeling and a 10 ns molecular dynamics (MD) simulations was carried out to understand its structural and dynamical features. Structure-function study of Cry1Id with respect to its closest structural homologue Cry8Ea1 was extensively studied and the most probable mechanism of action of this very Cry has been proposed. The findings from this study will abet in the development hybrid Cry proteins and new potent fusion proteins with novel specificities against different insect pests for improved pest management of crop plants.

Materials and methods

Sequence retrieval and analysis

The reviewed full length amino acid sequence of Cry1Id protein (UniProtKB ID: Q9XDL1) [25] was obtained from UniProtKB database of ExPaSy [28]. The full length of the toxin protein was comprised of 719 amino acids (core protein had 593 residues; 54–646). The IntroProScan tool was used to infer the protein family, super family and the domain arrangement within the protein. Conserved domains of the Cry1Id were explored by using the following databases: Pfam (http://pfam.sanger.ac.uk/) [29], Simple Modular Architecture Research (SMART) tool (http://smart.embl-heidelberg.de) [30] and Conserved Domain Database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) [31]. Primary structural analysis from the amino acid sequence of Cry1Id toxin was performed *via* Protparam tool of ExPaSy proteomic server, while secondary structure of Cry1Id was predicted by CONCORD (http://helios.princeton.edu/CONCORD/) [32] web server.

Molecular homology modeling

Sequence comprising the domains (DI-DII-DIII) of Cry1Id was used to build up the 3-D structures using the comparative protein modeling method of Modeller9v11 [33]. For search of suitable templates, DELTABLAST (Domain Enhanced Lookup Time Accelerated BLAST) [34] search tool was used against Protein Data Bank (PDB) (http://www.rcsb.org/). DELTABLAST was preferred against normal BLASTP because the retrieval accuracy and sensitivity toward protein analysis is more in the case of DELTABLAST than normal BLAST [35]. To ensure the correctness in the template identification process, apart from DELTABLAST, Cry1Id was subjected to various meta-servers like 3D Jury [36], Pcons.net [37], GeneSilico [38] and Geno3D [39] to find reliable templates. In addition, the protein threading approach implemented by I-TASSER [40] and protein fold recognition server Phyre version 2.0 (Protein Homology/analogY Recognition Engine) [41] were also used to determine the best templates in terms of fold recognition. Considering the suitable templates obtained from DELTA-BLAST and various meta-servers search, three-dimensional model building of the Cry1Id protein was performed using multi-template approach using four crystal structures of Cry proteins (PDB ID: 3 EB7, 1CIY, 1DLC and 1JI6) of *Bacillus thuringiensis* as the most appropriate templates. The target-template alignment was performed using ClustalX [42] and manually corrected to avoid any error in the model building process. Easy Sequencing in Postscript version 2.2 (ESPript) (http://www.ipbs.fr/ESPript) [43] was used to display the target-template alignment. Based on the target-template alignment 20 different 3D models of Cry1Id were generated through Modeller.

Table 1 Templates considered for homology modeling of *Cry*1ID toxin protein of *Bacillus thuringiensis*

Template (PDB ID)/ (No of amino acids)	% Identity	Positives	% Query coverage	E-value	Resolution (A°)
3 EB7 (589)	49.0	67.0	98.0	0.00	2.3
1CIY (590)	43.0	59.0	98.0	0.00	2.25
1DLC (584)	41.0	58.0	98.0	0.00	2.5
1JI6 (589)	38.0	58.0	99.0	0.00	2.4

These theoretical structural models of Cry1Id were ranked based on their normalized discrete optimized protein energy (DOPE) scores. The model with the lowest value of the normalized DOPE score is considered as the best model for energy minimization in Discovery Studio3.5 (Accelrys, Inc. San Diego, USA).

Energy minimization

The best model with lowest DOPE score was subjected to energy minimization by DS3.5 with the minimization protocol. The minimization protocol employs the steepest descent and conjugate gradient methods of minimization algorithms with a generalized Born implicit solvent model. In the present study the following parameters are considered for the structural minimization: distance-dependent dielectric constant=1, non-bonded radius of 14 Å with CHARMM force field [44], spherical electrostatic cut-off, and the steepest descent algorithm to remove close van der Waals contacts for a maximum steps of 5000 with 0.1 minimizing RMS gradient. Finally the potential energy, van der Waals energy and electrostatic energy for the minimized model of Cry1Id was determined using the calculate energy protocol in DS3.5.

Molecular dynamics simulations of Cry1Id

Molecular dynamics simulations were preformed to optimize the obtained homology model. The MD simulation was performed with GROMOS96 43A1 [45] force field in GROMACS4.5.4 [46] package running on a high performance CentOS6.2 cluster computer. The protonation states of all ionizable residues in the modeled protein were set to their normal states at pH 7.0. During the MD simulations, all atoms of the protein were surrounded by a octahedron water box of SPC3 water molecules that extended 0.9 nm from the protein and periodic boundary conditions were applied in all directions. The system was neutralized with four Na^+ counter ions replacing the water molecules. In the present study, the system was composed of 4498 atoms. Energy minimization was performed using steepest descent algorithm for 2000 steps. A

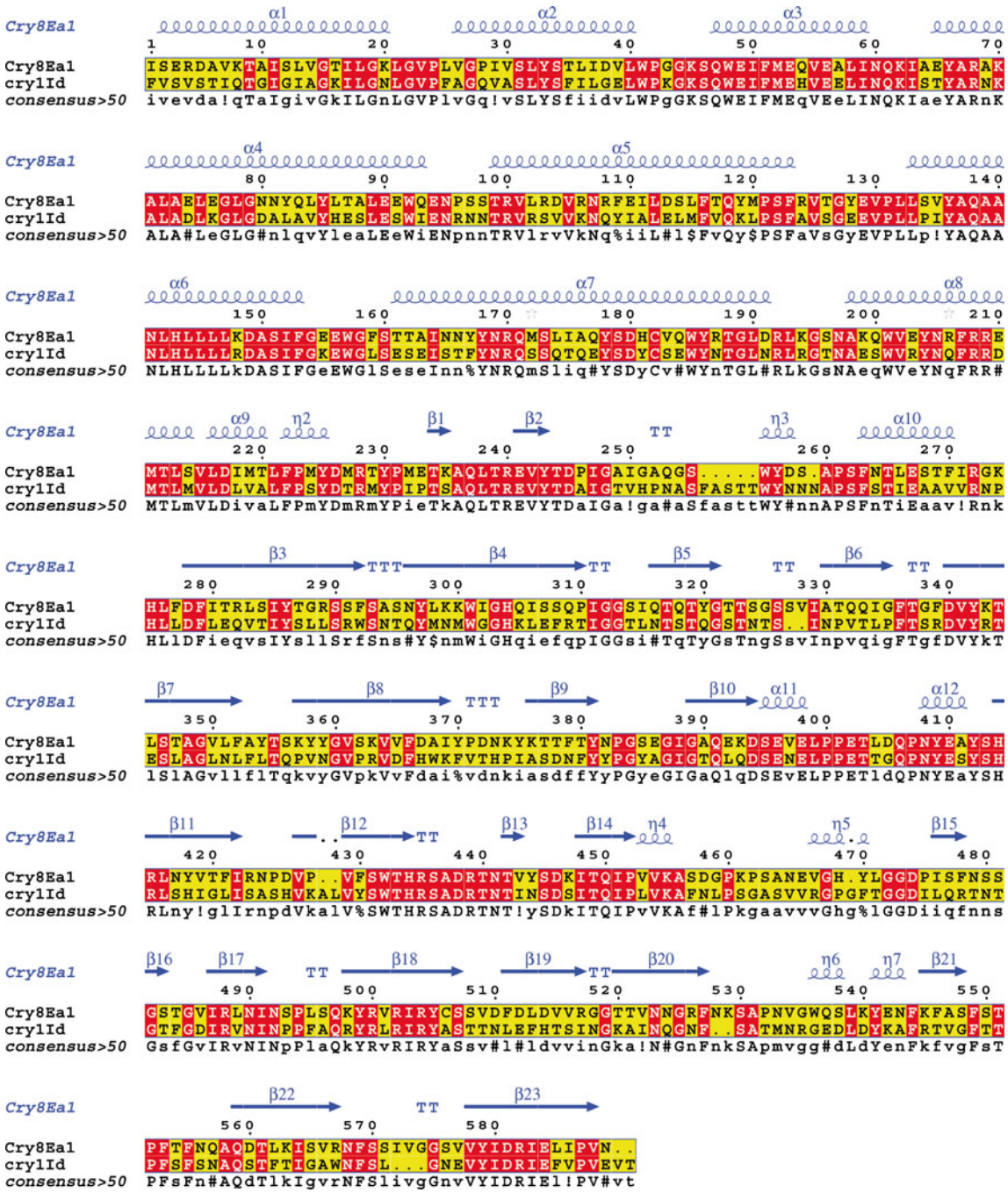


Fig. 1 The Pair-wise sequence alignment of the target Cry1Id and template Cry8Ea1 was constructed using ClustalX and ESPript. The secondary structural elements were identified from the Cry8Ea1 structure using ESPript. The α-helices, η-helices, β-sheets and strict β-turns are denoted α, η, β and TT respectively. The gray stars indicate side chains for which multiple conformations were modeled. Similar amino acids are highlighted in boxes, and completely conserved residues are indicated by white lettering on a red background

100 ps position restrained MD simulation was performed for the system followed by 10 ns MD simulations at constant pressure (1 atm) and temperature (300 K). The electrostatic interactions were calculated by the particle meshEwald (PME) algorithm [47] and all bonds were constrained using LINCS algorithm [48]. A cut-off value was set for long-range interactions including 0.9 nm for van der Waals and 1.4 nm for electrostatic interactions using the PME method. The snapshots were collected at every 1 ps and stored for further analyses. The system stability and differences in the trajectories, root mean square deviation (RMSD), root mean square fluctuations (RMSF) and the energies of the system was analyzed using tools available with GROMACS package.

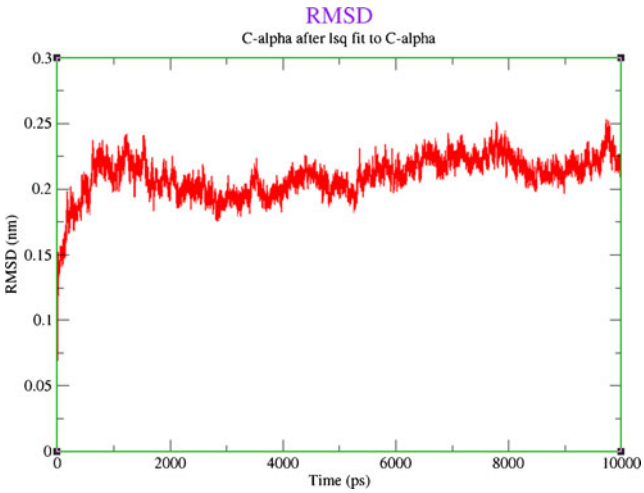


Fig. 2 The RMSD (Cα atoms) values with respect to simulation time for a 10 ns MD simulation on the Cry1Id model. The red line represents the value for Cry1Id

Model quality assessment

The refined model of Cry1Id was evaluated by a number of tools to test the internal consistency and reliability of the model. PROCHECK [49] analysis which quantifies the amino acid residues in the available zones of Ramachandran plot, was used to assess the stereo chemical quality of the model. ERRAT tool [50], which finds the overall quality factor of the protein, was used to check the statistics of non-bonded interactions between different atom types. The VERIFY-3D program [51] was used to determine the compatibility of the atomic model (3D) with its own amino acid sequence (1D). The average magnitude of the volume irregularities in the model was calculated using PROVE program [52]. PROVE program uses an algorithm which treats the atoms like hard spheres and calculates a statistical Z-score (*i.e.*, deviation) for the model from highly resolved (2.0 Å or better) and refined (R-factor of 0.2 or better). Standard bond lengths and bond angles of the model were determined using WHAT IF web server [53]. The estimated energy of the Cry1Id model was calculated by the ANOLEA server [54]. Furthermore, the stereochemical calculations were also performed using the MetaMQAP [55], ProQ [56], and ModFOLD version4.0 [57] servers. Also MolProbity web server (http://molprobity.biochem.duke.edu/) [58] was used in the model validation process which provides a detailed atomic contact analysis of any steric problems within the molecules as well as the dihedral-angle diagnostics. Subsequently the Protein structure analysis (ProSA-web) (https://prosa.services.came.sbg.ac.at/prosa.php) tool [59] was employed in the refinement and validation process to check the native protein folding energy of the model by comparing the energy of the model with the potential mean force derived from a large set of known protein structures. Structural superimposition of proposed 3-D model with its closest homologue Cry8Ea1 (PDB ID: 3 EB7) was performed in iPBA (http://www.dsimb.inserm.fr/dsimb_tools/ipba/) web server [60]. The iPBA web server presented the root mean square deviation (RMSD) between the Cα-atoms and all atoms of the homology model and template. To have a knowledge on the conservedness in the secondary structure of the refined model and the template Cry8Ea1, the pair-wise 3-D structural alignment was performed in the pair-wise 3-D alignment tool MATRAS (MArkovian TRAnsition of Structure evolution) (http://strcomp.protein.osaka-u.ac.jp/matras/) [61]. So as to ensure the accuracy in the assignment of secondary structure elements in the proposed model, the results of secondary structure elements assigned by STRIDE (http://webclu.bio.wzw.tum.de/stride/) [62] and DSSP (http://swift.cmbi.ru.nl/gv/dssp/) [63] was compared with the results of CONCORD web server.

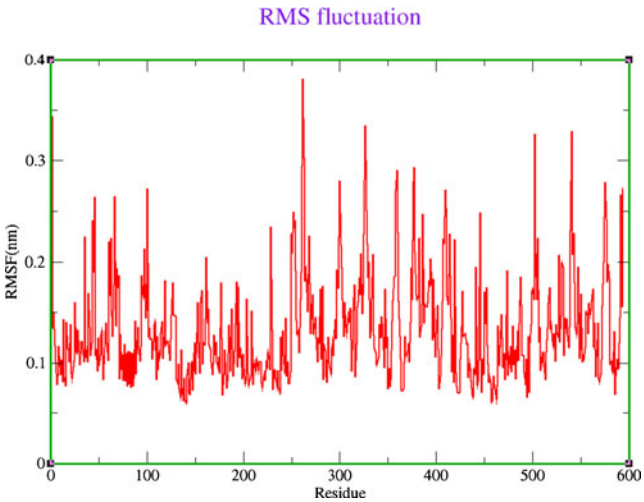


Fig. 3 The root mean square fluctuations (RMSF) values of Cry1Id (marked in red) during 10 ns simulation time, respectively

Results and discussion

Sequence analysis

The reviewed 81 kDa pesticidal crystal protein of Cry1Id (719 amino acids) from *Bacillus thuringiensis* belonging to delta endotoxin family was retrieved from UniProtKB database. SMART search of the core protein comprised of 592 amino acids (*i.e.*, Phe54-Thr646) revealed three putative domains *viz*, delta endotoxin N- terminal domain (Ile60-Met282), the middle (M) domain of delta endotoxin (Thr287-Asp497) and the C-terminal endotoxin domain (Ile507-Glu644). The result of SMART was affirmed from the prediction made by CDD and Pfam. SignalP predicted Cry1Id without any signal peptide cleavage sites.

Primary structural analysis of Cry1Id showed that this protein is acidic in nature (isoelectric point=5.74), which

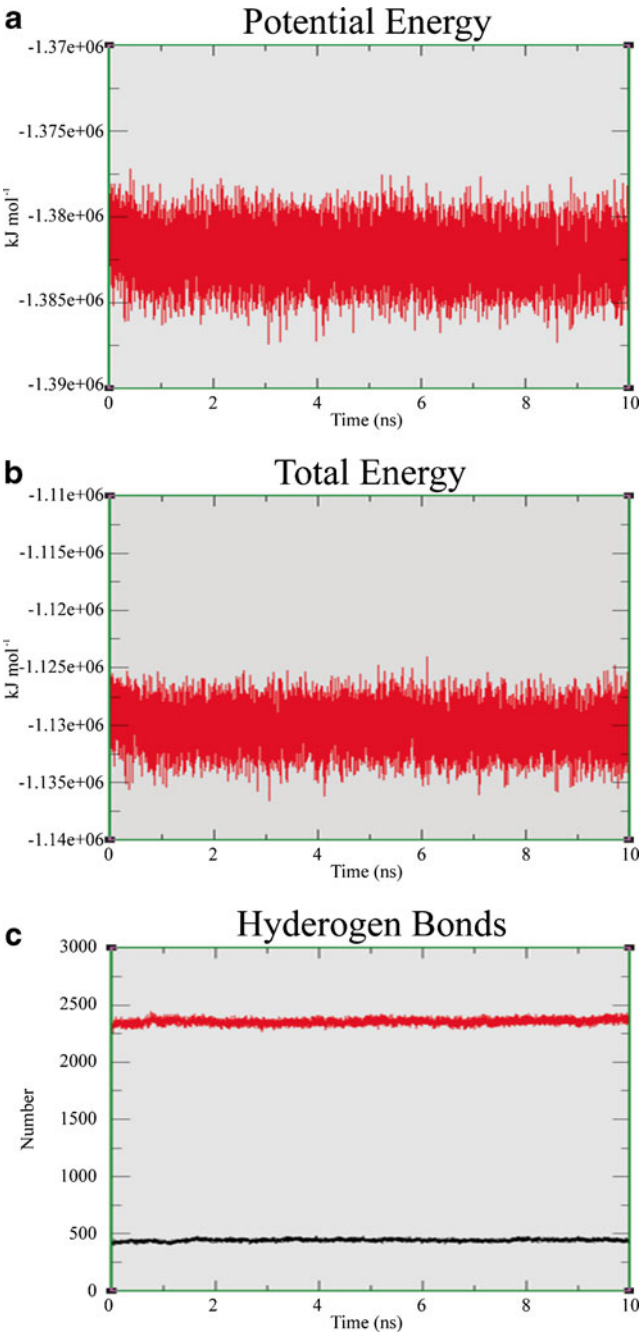


Fig. 4 The potential energy, total energy and the hydrogen bond count of the Cry1Id during 10 ns MD simulation. **a** The potential energy of the Cry1Id molecular system during 10 ns MD. **b** The total energy of the Cry1Id molecular system during 10 ns MD. **c** The hydrogen bond count (marked in black) of the modeled Cry1Id during the 10 ns equilibrium MD simulation. The number of hydrogen bonds formed between the donor and acceptor pairs within the distance of 0.35 nm

might be making it easier to dissolve in the midgut of insect pests. The aliphatic index (AI) which is defined as the relative volume of a protein occupied by aliphatic side chains is regarded as a positive factor for the increase of thermal stability of globular proteins [64]. The aliphatic index was very high (79.19) reflecting the stable nature of the protein for a wide range of temperature. It is well known that a protein whose instability index is smaller than 40 is predicted as stable, whereas a value above 40 predicts that the protein may be unstable [65]. The Cry1Id was predicted to be stable in nature as its instability index was reported to be 36.04 (<40). The grand average hydropathicity (GRAVY) value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids, divided by the total number of residues in the sequence. The GRAVY index of Cry1Id was found to be very low (−0.368), indicating the possibility of better interaction with water. Secondary structure prediction made by CONCORD server revealed random coils (48.67 %) dominated over the other secondary structure elements followed by helices (34.07 %) and strands (17.36 %).

Homology modeling of Cry1Id

Comparative modeling offers tremendous potential in the development of theoretical three-dimensional protein models and often a method of choice when a clear relationship of homology between the sequences of target protein and at least one known structure is found. This approach would give reasonable results based on the assumption that the tertiary structure of two proteins will be similar if their sequences are related [66]. A high level of sequence identity promises a more reliable alignment between the target sequence and the template structure. DELTABLAST search revealed four putative templates (PDB ID: 3 EB7, 1CIY, 1DLC and 1JI6), all are insecticidal delta-endotoxin from *Bacillus thuringiensis* with sequence identity of 49 %, 43 %,41 % and 38 % respectively with target protein Cry1Id as shown in Table 1.

The suitable template identification through various meta-servers also revealed the same templates as that of DELTA-BLAST search. The 3-D modeling of Cry1Id protein was done using the above mentioned templates using advance modeling techniques in Modeller. Furthermore, to ensure the suitability in the model generated through comparative modeling, cross-checking of the model was done with the best models obtained from I-TASSER and Phyre. Our model showed better secondary structure conservation than the I-TASSER and Phyre models. The number of secondary structure elements (α helices and β-sheets) within the domains in Cry1Id generated by I-TASSER and LOMETS were very few (data not shown) as compared to our proposed model and were occupied by turns. Thus, the comparative modeling through multi-template approach in Cry1Id suggested a reliable model for structural analysis.

The selected model was finally subjected to energy minimization using CHARMM force field in DS3.5. The potential energy, van der Waals energy and electrostatic energy and RMS gradient of Cry1Id were determined (data not shown) using the DS3.5. It was observed that there is a decrease in the force field energies of the model before and after refinement which confirmed the proposed model was minimized. To

Table 2 Comparison of Ramachandran plot statistics of model *Cry*1Id and its closest structural homologue *Cry*8Ea1

Ramachandran plot analysis parameters	Homology modeled Cry1Id		Crystallographic structure of template Cry8Ea1 (PDB ID-3 EB7) chain A	
	Number of Residues	Percentage (%)	Number of Residues	Percentage (%)
Residues in most favored regions	488	93.3	478	92.1
Residues in additionally allowed regions	34	6.5	40	7.7
Residues in generously allowed regions	1	0.2	1	0.2
Residues in disallowed regions	0	0.0	0	0.00
Number of non-glycine and non-proline	523	100.0	519	100.0
Number of end residues (excluding Glycine and Proline)	2	–	2	–
Number of Glycine residues	41	–	42	–
Number of proline residues	27	–	26	–
Total number of residues	593		589	

assess the conservedness among the secondary structure elements, the secondary structure of both Cry1Id and templates were compared. The comparison of secondary structure revealed that the N-terminal domain, middle-domain and C-terminal domain of Cry1Id shares maximum percentage of identity in secondary structure elements with the template Cry8Ea1. Even though Cry1Id sequence shares strong homology throughout the length of templates notably, domain I shares maximum percentage of sequence similarity with the corresponding domain of Cry8Ea1 as compared to the other domains (Fig. 1). The conservation of the secondary structure elements of Cry1Id with its closest homologue showed the reliability of our proposed model, predicted by Modeller based on the target-template alignment.

The closet structural homologue of Cry1Id protein was 3HB7 predicted by I-TASSER and Phyre. So a cross-check validation approach was employed to further validate the accuracy of the structure of the homolog and the method used to generate the 3D model of closest template. In this case, the proposed Cry1Id model was chosen as the template and the 3HB7 sequence was considered the target. Modeller program, assisted in the generation of 3D model of 3HB7. A cross comparison was performed to evaluate the accuracy of the modeled 3HB7 as compared to crystal structure of 3HB7 as presented in Supplementary Table 1. Furthermore, the RMSD values between the modeled 3HB7 generated by Modeller and the PDB coordinate of 3HB7 was calculated to be 0.531 Å for all atoms and 0.81 Å for backbone atoms by the PyMOL superimposition program. These data supported the reliability of our proposed model of Cry1Id.

Molecular dynamics of Cry1Id

To gain insight into the stability and MD properties of the structure of Cry1Id model, explicit solvent MD simulation was performed. The results of MD simulation suggested that monomeric structure of Cry1Id toxin is stable in aqueous solution during 10.0 ns. Compared to the starting coordinates, the RMSD of the Cα atoms increased in the first 2 ns and then reached a plateau in the subsequent simulation time (Fig. 2). The calculated average Cα RMSD value for the Cry1Id was reported to be 0.2 nm whereas the radius of gyration (Rg) value lies within ~2.52 to ~2.44 A° till the end of the simulation suggested that an accepted structure was obtained by the end of the simulation. Furthermore, to understand the structural flexibility of the Cry1Id model, the mean RMSF values (Fig. 3) were calculated and the flexible regions (peaks in the plot) of Cry1Id have been observed in domain I, II and III. It was observed that most of the loops connecting adjacent helices of domain II possess most of the flexible regions. Similarly the potential energy (Fig. 4a), total energy (Fig. 4b), and the hydrogen bonding property (Fig. 4c) confirmed the stable nature of the protein during MD simulations. It was observed that after 2 ns of molecular simulation overall hydrogen bonds of the protein remained stable ranging from ~400 to ~500. All of the above properties converged after 10 ns MD simulation highlighting that the model is stable in nature which is suitable for further studies. The final snapshots obtained at the end of the simulations were considered to represent the structure of the Cry1Id model.

Model evaluation and quality assessment

The quality of the final simulated Cry1Id model including geometric properties of the backbone confirmations, compatibility of residues interactions and overall qualities were assessed using three different structural evaluation programs; PROCHECK, ERRAT, Verify-3D through SAVES server. PROCHECK was first used to check the reliability of the backbone of torsion angles Φ, Ψ of the model, which quantifies the residues fall in the available zones of Ramachandran plot (Table 2).

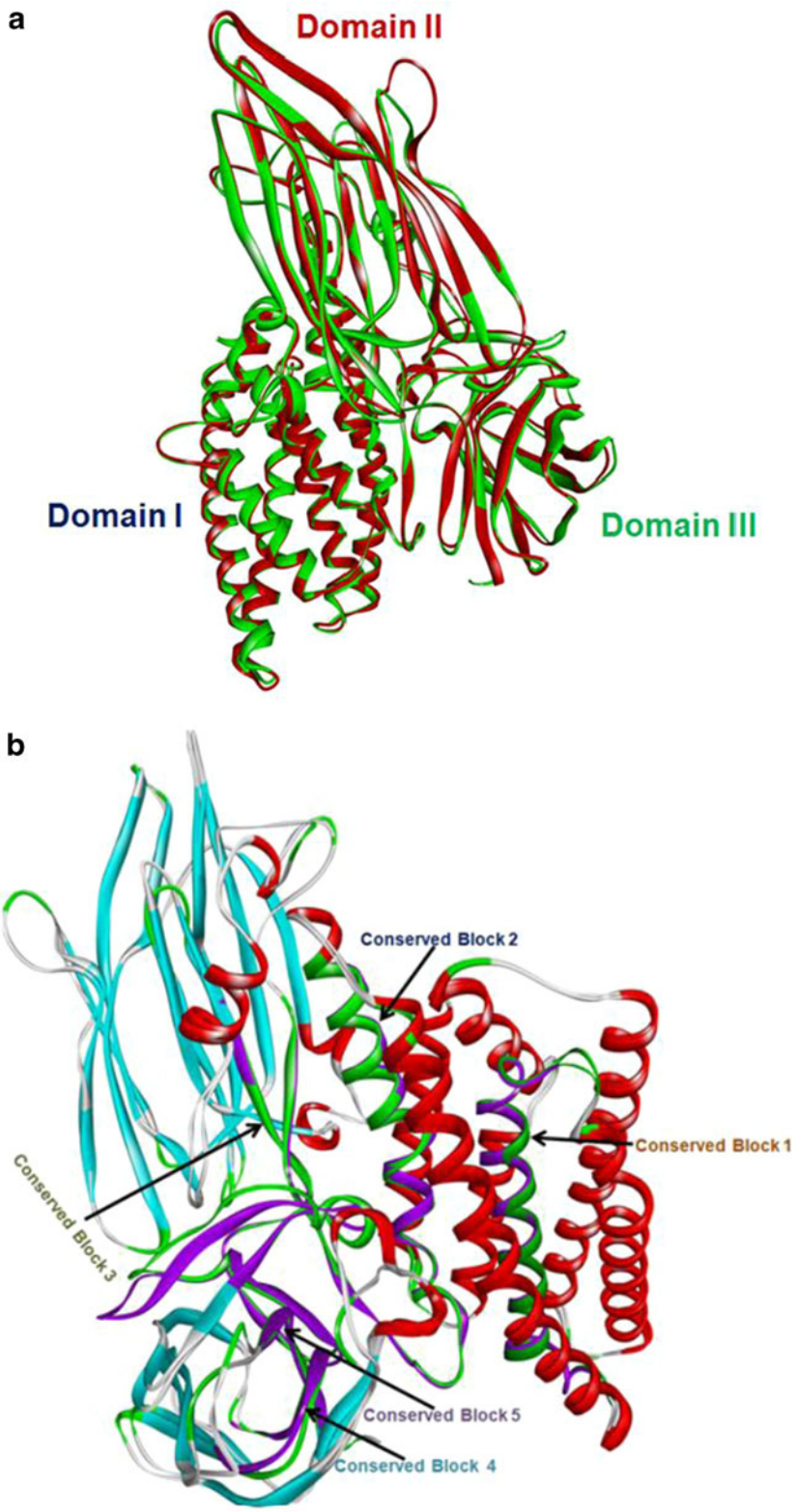


Fig. 5 Superimposition of the Cry1Id model (*Green*) and the template (Cry8Ea1, PDB ID: 3 EB7) (*Red*) by iPBA web server. The superposition of corresponding domains was marked **a** The conserved blocks (*violet*: model; *green*: template) in the superposed target-template were highlighted **b** The image was generated using DS3.5

Ramachandran plot analysis for the modeled protein of Cry1Id showed 93.5 % residues fell in the most favored regions, 6.5 % residues in additional allowed regions, 0.2 % residues generously allowed regions and no residue in the disallowed regions. As compared with the template Cry8Ea1, the built 3-D model had a similar Ramachandran plot which signifies the predicted model is reliable in terms of its backbone conformation as reported in Table 2. The high quality of the structure is further evident by the

Table 3 Model validation statistics of the homology modeled *Cry*1Id

Homology modeled protein	Overall G factor	Verify-3D	PROVE Z-score	Errat	ProSA (Z-score)	RMSD (A°)
*Cry*1Id	−0.07	95.62 %	0.584	74.658	−9.73	0.71

G-factor of −0.07 computed in PROCHECK. The quality of our model Cry1Id was further supported by the ERRAT score of 74.658 which indicates acceptable protein environment [50]. The Verfiy-3D results of Cry1Id model showed 95.62 % of the amino acids had an average 3D-1D score of >0.2 and 92.32 % of the residues showed positive scores (cut-off score was >0) indicating the reliability of the proposed model. PROVE program was used to measure the average magnitude of the volume irregularities in terms of the Z-score root mean square deviation of the model. The Z-score RMS values for the model and template was 0.584 and 1.199, respectively (a Z-score RMS value of ~1.0 indicates good resolution of structures). WHAT IF server analyzed the coarse packing quality, anomalous bond length, planarity, packing quality and the collision with symmetry axis, distribution of omega angles, proline puckering and anomalous bond angles of the model protein reflecting its acceptance of good quality.

The packing qualities of the Cry1Id model showed that most of the regions had negative ANOLEA score with very low energy and favorable energy environment. The proposed model had a QMEAN6 score of 0.451, and an average Z-score of −3.41. The ProQ analysis of the Cry1Id model revealed an LG score of 5.842 (>4 for extremely good model) and MaxSub score of 0.244 (>0.5 very good model). The quality assessment of the model in MetaMQAPII server showed a global distance test (GDT_TS) score of 82.799 with an RMSD value of 1.509 Å. A global quality score of 0.9069 and p value of 0.00032 was predicted by ModFOLD server.

Analysis of Cry1Id in MolProbity server showed 0 % of the residues had bad bonds (goal 0 %), 0.84 % of the residues had bad angles (goal < 0.1 %), and 0 % of the Cβ deviations were >0.25 Å (goal 0 %) which further confirmed the reliability of our model.

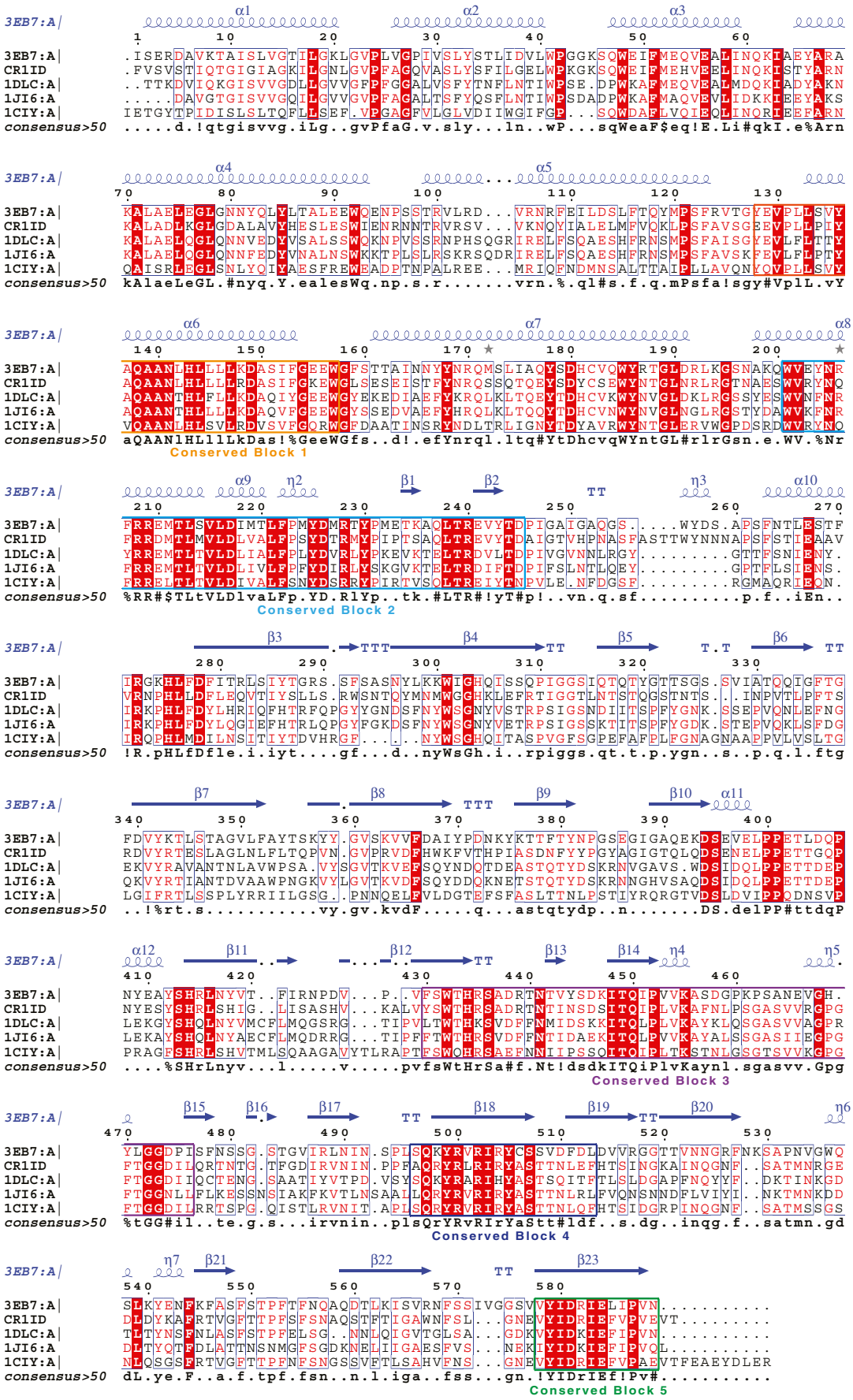
Energy profile of the proposed model and the Z-score value (a measure of model quality as it measures the total energy of the structures) was obtained using ProSA program which calculates the interaction energy per residue using a distance-based pair potential. The ProSA analysis of the model Cry1Id achieves a Z-score of −9.73 and that of template was −9.13, (where the negative PROSA energy reflects reliability of the model) reflecting the quality of the model. These results all together validated the quality of the Cry1Id model.

The quality of the model was also assessed by comparing the predicted structures to the experimentally determined structure by structural superimposition and atoms RMSD assessment. The superimposition of Cry1Id with respect to its closest homologue Cry8Ea1 (3 EB7) was executed by combinatorial extension of polypeptides. The RMS deviation of Cα trace between modeled structure and template was 0.71 A° (for 577 aligned residues), which indicates the generated model is reasonably good and quite similar to template (Fig. 5a and b). The model validation statistics of Cry1Id from different tools and web servers are reported in Table 3. The pairwise 3-D alignment of the model and the closest template (3HB7), predicted by MATRAS server revealed that the key secondary structure elements (within the functional domains) are strongly conserved (secondary structure elements identity=95.2 %) in the alignment where there exists a sequence similarity of 49.1 % (Supplementary Fig. 1). As compared to secondary structure elements assigned by CONCORD, Cry1Id shared the same statistics predicted through DSSP and STRIDE as shown in Table 4 signifies the accuracy in the assignment of secondary structure elements through homology modeling.

Furthermore, to support the accuracy of the proposed model the all-atom based superimposition RMSD value for each domain (I, II and III) between Cry1Id model and the templates were calculated by PyMOL program, and these values are shown in Supplementary Table 2. The domain wise all-atom superimposition RMSDs of the proposed model with the closest homologue 3HB7 was very low as compared to other templates. Moreover, the all-atom superposition of conserved secondary structure elements using PyMOL (helix-helix, sheet-sheet) for the functional domains revealed a very low deviation, indicating the acceptance of the Cry1Id model (Supplementary Table 3).

Table 4 Statistics of the predicted secondary structure elements of the model (*Cry*1Id) and the template (*Cry*8Ea1) from their 3-D structure using STRIDE and DSSP

The predicted model of Cry1Id						The crystal structure of the template Cry8Ea1					
DSSP			STRIDE			DSSP			STRIDE		
Helix (%)	Sheet (%)	Others (%)	Helix (%)	Sheet (%)	Others (%)	Helix (%)	Sheet (%)	Others (%)	Helix (%)	Sheet (%)	Others (%)
35.2	29.0	35.8	35.8	31.7	32.5	35.8	28.5	35.7	36.3	30.7	32.9


3EB7:A
CR1ID
1DLC:A
1JI6:A
1CIY:A
consensus>50
Conserved Block 1
Conserved Block 2
Conserved Block 3
Conserved Block 4
Conserved Block 5

◀ **Fig. 6** Structure based sequence alignment of *Cry*1Id and other *Cry* toxins of *Bacillus thuringiensis*. From top to bottom, the sequences are *Cry*8Ea1 (PDB ID: 3 EB7), *Cry*1Id, *Cry*1Ia (PDB ID: 1CIY), *Cry*3Aa (PDB ID: 1DLC) and *Cry*3bb1 (PDB ID: 1JI6). Highly conserved residues are highlighted in red, and similar residues are in yellow. The highly conserved boxes of *Cry* family are indicated by gray frames. The search model we used in molecular replacement is the crystal structure of *Cry*8Ea1 toxin (PDB ID: 3 EB7), which shares a primary sequence identity of 49.0 % with *Cry*1Id toxin. The conserved blocks were highlighted in colored square boxes. Image was prepared using ESPript2.2

Structural features of Cry1Id

A multiple sequence alignment of Cry1Id along with its closest structural homologues (crystal structure of known Cry proteins) (Fig. 6) showed a relatively high degree of sequence similarities (40–50 %). Comparison of structures among the members of Cry toxin family revealed that Cry1Id shares similar architecture with them and forming a wedge shape. The predicted structure of core toxin (Phe54-Thr646) is comprised of three putative domains (DI, DII and DIII) (Fig. 7). Domain I (Ile60-Met282) is comprised of eight helices whereas domain II (Thr287-Asp497) formed a prism shape and consists of three antiparallel β-sheets. But domain III (Ile506-Glu644) was comprised of antiparallel β-sheets forming a jelly roll topology. The core protein was comprised of five sheets, seven beta hairpins, nine beta bulges, 24 strands, 16 helices, 35 helix-helix interacs, 41 beta turns and three gamma turns (Fig. 8a) where the hydrogen bond forming residues within the β-turn stabilizes the overall folds of Cry1Id protein. The coordinates of the model has been submitted to Protein Model DataBase (PMDB), which is accessible at http://mi.caspur.it/PMDB/ using the PMDB ID: PM0079143.

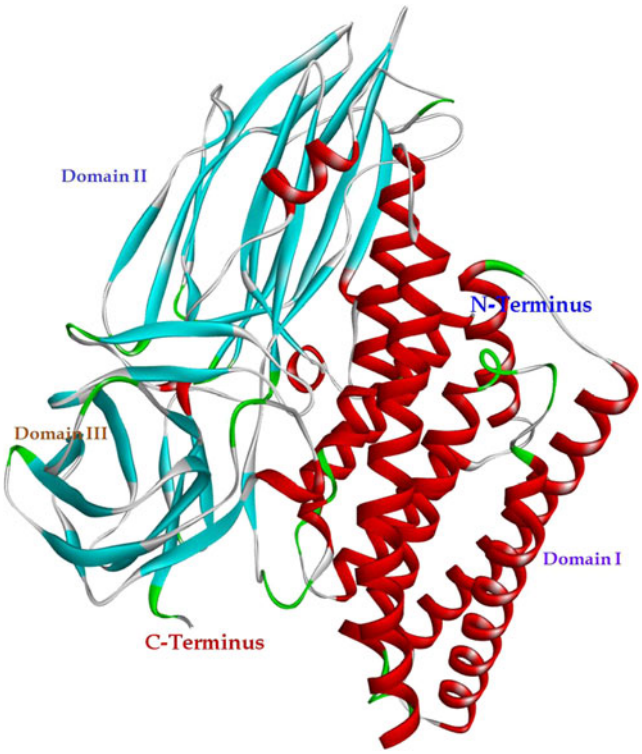


Fig. 7 Overall homology modeled Cry1Id toxin structure from *Bacillus thuringiensis* with secondary structure assignments. Homology modeled *Cry*1Id from *Bacillus thuringiensis*. Solid ribbon representation of domain I, II and III colored by their secondary structure elements

The result from I-TASSER and Phyre revealed 3HB7 as the closest structural homologue of Cry1Id protein. So as to understand the structure-function relationship of Cry1Id, the overall architecture of toxin was compared with its closest structural homologue Cry8Ea1 (PDB ID: 3HB7) (an insecticidal toxin toxic to underground pests, the larvae of *Holotrichia parallela*). The structural superposition of (overlapping of Cα atoms) of the domains of Cry1Id with that of corresponding domains of Cry8Ea revealed the spatial position and orientation of helices and sheets are highly conserved. Although both the proposed model and template overall share the same topology, the surface electrostatic potential distribution in Cry1Id (Fig. 8b) was quite different from that of Cry8Ea (data not shown).

Domain I

Domain I of Cry proteins, which is helical in nature, is thought to be directly involved in membrane penetration and pore formation after binding to the specific receptor on the surface of midgut. Cry1Id domain I is typically a helical domain comprised of eight helices of the wedge shape protein. Li and co-workers [13] have suggested that the helical hairpin α4-α5 acts as the initiator of the membrane related allosteric mechanism of penetration commonly known as umbrella model. In the absence of recognition specificity of the endotoxin toward any receptor, the lid on the helix bundle functions as a protector to the offensive loop α4-α5, avoiding unexpected hydrophobic binding. After a correct recognition, the lid comprising helix α2 and loop α3 must be removed from the top of the helix bundle, such that helical hairpin α4-α5 can be released normally. As compared to domain I of crystal structure Cry8Ea1, Cry1Id had the same conserved region, thereby it is reasonable that domain I has the same role of pore-formation. Guo and co-workers [19] reported that domain I of Cry8Ea1 possess kinked helices in domain I, similar kinked helices were observed in domain I of Cry1Id. Three helices α2, α4 and α7 are merely a regular α-helix in Cry1Id. In addition to the structural alignment of domain I with corresponding domain of reported Cry toxins, it was observed that similar to the highly conserved break of helix α2, the kink of helices α4 and α7 were also conserved. Domain I of Cry1Id superposes well onto domain I of the template protein with an RMSD of 0.276 A° (backbone atoms) (Fig. 9a).

The most important structural feature of Cry toxin family is the highly conserved five blocks of amino acid residues [2, 6]. From the multiple sequence alignment of Cry1Id with its structural homologs, it was observed that Cry1Id shares the same highly conserved amino acid residue blocks (Fig. 6). Our

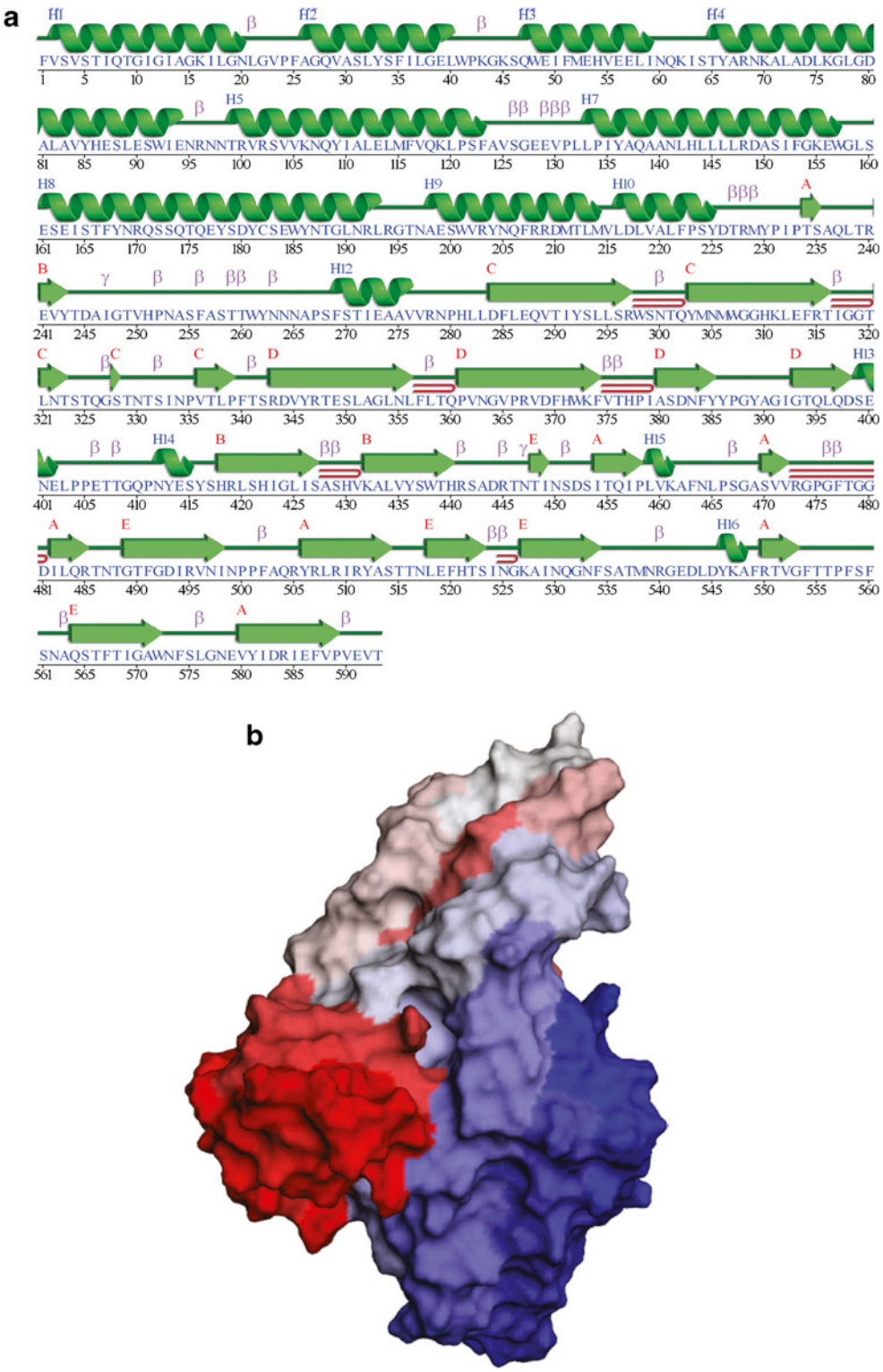


Fig. 8 The graphical representation (wiring diagram) of modeled Cry1Id with its secondary structure elements (a). Helices are labeled with (H1, H2…); Strands with their sheets are labeled with (A, B…); beta turns are labeled with β and gamma turns are labeled with γ. (b) The surface representation of Cry1Id

structure along with the other known Cry toxin structures revealed that conserved block 1 (Lavender in Fig. 6) covers helix α6, the central helix of domain I helix bundle, and nine residues in the adjacent loops at both ends. However the conserved block 2 (blue in Fig. 6) which forms the major part of domain I and domain II interface covers the C-terminal helix of domain I and the N-terminal strand of domain II, which are antiparallel to each other. It is evident from Fig. 6 that most of the residues in conserved block 3 (shown in purple) entirely located at the center of three domains, thus involved in the formation of the interfaces between any two domains. As compared to the other three blocks (1, 2 and 3), conserved blocks 4 (light blue) and 5 (aqua) (Fig. 6)

are very small in size. These two blocks form the core of domain III where they represent antiparallel strand β18 and β23 respectively and possess partial contact with domain II. It was also observed that except for conserved block 1, all of the conserved blocks are either entirely or partially involved in domain-domain interactions, suggesting that the central helix of domain I and the inter-domain structural communications are both essentially important to the function of every Cry toxins. The high homology of the inter-domain regions of Cry1Id with its closest structural homolog implies that the function of the Cry1Id is not only restricted to stabilization of the overall Cry toxin structure, but also to propagate the conformational changes properly to ensure the functional disassembly of Cry toxins.

Domain II

The structural superposition of domain II in both Cry1Id and template (Cry8Ea1) protein revealed that domain II of the former is structurally divergent from the later with an RMSD of 0.249 A° (backbone atoms; ~11 % sequence identity) (Fig. 9b). In addition as compared to other known Cry protein, domain II is diverse in terms of its structural features especially in the apical loops. The variable surface accessibility of these apical loops of Cry proteins thought to be the sole factor for receptor recognition. It has been reported that site-directed mutagenesis of the loop residues in related Cry toxins influence the binding affinity and toxicity of Cry proteins. A comparison of the apical loops of domain II (data not shown) in the known Cry proteins with the Cry1Id revealed that these loops are variable in length and in the composition of amino acids. As compared to known Cry toxins (Cry1Aa, Cry3Aa, Cry3bb1 and Cry8Ea1), Cry1Id is the most divergent member but interestingly the apical loop of Cry8Ea1 superpose well with the Cry1Id. But the loop is very small in the case of Cry1Id comprised of only four residues (Phe357-Leu358-Thr359-Gln360). We hypothesize that variation of amino acids and length of the apical loop (domain II) in Cry1Id leads it to target various midgut surface receptors or target different kinds of specificity-determinants on shared receptors with structurally similar apical loops. Further it was noticed that the presence of only one aromatic amino acid (Phe357) residue in the apical loop might be influencing the binding of domain II with multiple receptors which can explain the specificity of Cry1Id toward different receptors.

Domain III

Domain III which plays the key role in prevention of the toxin against action of gut protease thus protect from the cleavage of Cry proteins. Domain III of Cry1Id (Ile506-Glu644) which adopts a β-sandwich fold showed a jelly-roll-like topology comprised of two antiparallel β-sheets (6 beta strands) and a small α-helix. This domain lies on the top of domain II and

Fig. 9 Domain-wise superposition of Cry1Id with the template Cry8Ea1. **a** Superposition of domain I of Cry1ID (*Green*) with the domain I of Cry8Ea1 (Mauve). **b** Superposition of domain II of Cry1ID (*Green*) with the domain II of Cry8Ea1 (Mauve). **c** Superposition of domain III of Cry1ID (*Green*) with the domain III of Cry8Ea1 (Mauve). All the above images were generated using DS3.5

against the side of domain I. Each of the outer and inner sheets of domain III was comprised of three strands where the inter-sheet connections between the domain II and III are dominated by hydrogen bonding and hydrophobic interactions. In contrast to domain II, domain III of Cry1Id shared almost the same architecture as the template protein (with an RMSD of 0.689 A° on backbone atoms, ~19.0 % sequence identity) (Fig. 9c). From the multiple sequence alignment of Cry1Id with other Cry proteins (closest structural homologues) (Fig. 6) it was observed that domain III holds the last three conserved blocks (block3, block4 and block5). The structural superposition of domain III of Cry1Id and its template showed close structural similarity between them with few exceptions within the loop regions. For instance, the larger loop (Arg526-Asp546) of Cry1Id was missing in Cry8Ea1. Schwartz and co-workers [14] had shown that mutation in domain III of Cry1Aa affects the ion channel activity and membrane permeability. Furthermore, domain swapping experiments of domain III of Cry proteins suggested that, Cry protein with swapped domain possess better activity than the native Cry proteins. Tajne and co-workers [67] showed, when domain III of Cry1Ac was replaced by ASAL (*Allium sativum* lectin), the binding affinity of the new protein toward aminopeptidase N (APN) receptor of *Manduca sexta* increased more significantly than in the native fold. The MD simulations showed that the structure retains most of the secondary structure elements stable throughout; such a domain swapping experiment of Cry1Id involving replacement of domain III with suitable plant lectins is underway in our laboratory to build novel fusion proteins. The development of novel fusion protein ultimately will provide greater insights into the binding specificity toward multiple receptors of Cry protein. In addition to fusion proteins, novel hybrid Cry proteins can be constructed by interchanging domain III among different Cry proteins to enhance the insect specificity of Cry proteins. Both of the techniques eventually not only provide a sole opportunity for design of novel fusion protein but also in development of hybrid Cry proteins with improved insect specificity for better insect pest management in various crop plants.

Conclusions

Genetic engineering of Bt Cry proteins has resulted in the synthesis of various novel toxins with enhanced insecticidal activity and specificity toward different insect pests. Among different classes of Cry family proteins, the Cry1I class displays several unique characteristics. As of now more than 15 Cry1I proteins have been grouped into six subclasses ranging from Cry1Ia to Cry1If. In this work the theoretical 3-D structural model of a novel delta-endotoxin Cry1Id was constructed by comparative modeling using available crystal structures as templates and a 10 ns molecular dynamics (MD) simulation was carried out to characterize its structural and dynamical features, which was further validated by the SAVES, WHAT IF and MolProbity web servers. In spite of sharing a very low sequence identity with its closest structural homologs, Cry1Id shares a common three-dimensional structure comprised of three domains. It also shares the common five conserved blocks of amino acid residues, an important characteristics of the entire Cry toxin family. Domain II of Cry1Id superposes well with its closest structural homolog (Cry8Ea1). However, variation of amino acids and length renders variable surface accessibility of the apical loop2 in domain II of Cry1Id to be observed. These differences may be the sole factor for recognition of specificity toward different receptors of Cry proteins. Domain III which contains the three conserved blocks superposed well with the corresponding domain of Cry8Ea1. As the action of insecticidal toxin Cry1Id depends on the delicate equilibrium between the conformational stability and protein stability, MD simulation was performed. MD simulation has suggested that the proposed monomer is stable in aqueous solution. In addition, in silico studies are underway in our laboratory to design and build a fusion protein by replacing domain III of Cry1Id with plant lectins to evaluate the functional role of the fusion protein in terms of its toxicity and binding ability toward the various receptors. This is the first ever collective report on the structural characteristics of the novel insecticidal toxin Cry1Id which would help in the development of new potent fusion proteins with novel specificities against different insect pests for improved pest management of crop plants. However, further works involving site-directed mutagenesis along with MD simulation experiments may be carried to shed more light on the toxin oligomerization process.

Acknowledgments This work was financially supported by the grant of Bioinformatics Initiative Division, Department of Electronics and Informatics Technology (DEITY), Ministry of Communications and Information Technology, Government of India. We also gratefully acknowledge the financial support by Biotechnology Information System Network (BTISNET), Department of Biotechnology, Government of India.

References

1. Aronson AL, Beckman W, Dunn P (1986) *Bacillus thuringiensis* and related insect pathogens. Microbiol Rev 50:1–24
2. Schnepf E, Crickmore N, Van Rie J, Lereclus D (1998) *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol Mol Biol 62:775–806
3. Manimaran P, Ramkumar G, Mohan M, Mangrauthia SK, Padmakumari AP, Muthuraman P, Bentur JS, Viraktamath BC, Balachandran SM (2011) Bt rice evaluation and deployment strategies. GM Crops 2:135–137
4. Huang DF, Zhang J, Song FP, Lang ZH (2007) Microbial control and biotechnology research on *Bacillus thuringiensis* in China. J Invertebr Pathol 95:175–180
5. Pigott CR, Ellar DJ (2007) Role of receptors in *Bacillus thuringiensis* crystal toxin activity. Microbiol. Mol Biol Rev 71:255–281

6. Hofte H, Whiteley HR (1989) Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbiol Rev 53:242–255
7. Knowles BH, Ellar DJ (1987) Colloid-osmotic lysis is a general feature of the mechanism of action of *Bacillus thuringiensis* δ-endotoxins with different insect specificity. Biochim Biophys Acta 924:507–518
8. Choma CT, Surewicz WK, Carey PR, Pozsgay M, Raynor T, Kaplan H (1990) Unusual proteolysis of the protein and toxin from *Bacillus thuringiensis* structural implications. Eur J Biochem 189:523–527
9. Hofmann C, Vanderbruggen H, Hofte H, Van Mellaert H (1988) Specificity of *Bacillus thuringiensis* delta-endotoxins is correlated with the presence of high-affinity binding sites in the brush border membrane of target insect midgets. Proc Natl Acad Sci USA 85: 7844–7848
10. de Maagd RA, Bravo A, Berry C, Crickmore N, Schnepf HE (2003) Review Structure, diversity, and evolution of protein toxins from spore-forming entomopathogenic bacteria. Annu Rev Genet 37: 409–433
11. Shelton AM, Zhao JZ, Roush RT (2002) Economic, ecological, food safety, and social consequences of the deployment of bt transgenic plants. Annu Rev Entomol 47:845–881
12. Gazit E, La Rocca P, Sansom MS, Shai Y (1998) The structure and organization within the membrane of the helices composing the pore-forming domain of Bacillus thuringiensis delta-endotoxin are consistent with an "umbrella-like" structure of the pore. Proc Natl Acad Sci USA 95:12289–12294
13. Li JD, Carroll J, Ellar DJ (1991) Crystal structure of insecticidal delta-endotoxin from *Bacillus thuringiensis* at 2.5 A° resolution. Nature 353:815–821
14. Schwartz JL, Lu YJ, Sohnlein P, Brousseau R, Laprade R, Masson L, Adang MJ (1997) Ion channels formed in planar lipid bilayers by *Bacillus thuringiensis* toxins in the presence of *Manduca sexta* midgut receptors. FEBS Lett 412:270–276
15. Boonserm P, Mo M, Angsuthanasombat C, Lescar J (2006) Structure of the functional form of the mosquito larvicidal Cry4Aa toxin from *Bacillus thuringiensis* at a 2.8 angstrom resolution. J Bacteriol 188: 3391–3401
16. Gomez I, Arenas I, Benitez I, Miranda-Ríos J, Becerril B, Grande R, Almagro JC, Bravo A, Soberon M (2006) Specific epitopes of domains II and III of *Bacillus thuringiensis* Cry1Ab toxin involved in the sequential interaction with cadherin and aminopeptidase-N receptors in *Manduca sexta*. J Biol Chem 281:34032–34039
17. Bravo A (1997) Phylogenetic relationships of *Bacillus thuringiensis* delta-endotoxin family proteins and their functional domains. J Bacteriol 179:2793–801
18. de Maagd RA, Bravo A, Crickmore N (2001) How *Bacillus thuringiensis* has evolved specific toxins to colonize the insect world. Trends Genet 17:193–199
19. Guo S, Ye S, Liu Y, Wei L, Xue J, Wu H, Song F, Zhang J, Wu X, Huang D, Rao Z (2009) Crystal structure of *Bacillus thuringiensis* Cry8Ea1: An insecticidal toxin toxic to underground pests, the larvae of *Holotrichia parallela*. J Struct Biol 168:259–266
20. Crickmore N, Zeigler DR, Feitelson J, Schnepf E, Van Rie J, Lereclus D, Baum J, Dean DH (1998) Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins. Microbiol Mol Biol Rev 62:807–813
21. Grochulski P, Masson L, Borisova S, Pusztai-Carey M, Schwartz JL, Brousseau R, Cygler M (1995) *Bacillus thuringiensis* CryIA(a) insecticidal toxin: crystal structure and channel formation. J Mol Biol 254:447–464
22. Morse RJ, Yamamoto T, Stroud RM (2001) Structure of Cry2Aa suggests an unexpected receptor binding epitope. Structure 9:409–417
23. Galitsky N, Cody V, Wojtczak A, Ghosh D, Luft JR, Pangborn W, English L (2001) Structure of the insecticidal bacterial-endotoxin Cry3Bb1 of *Bacillus thuringiensis*. Acta Crystallographica Section D: Biological Crystallography 57:1101–1109
24. Boonserm P, Davis P, Ellar DJ, Li J (2005) Crystal structure of the mosquito-larvicidal toxin Cry4Ba and its biological implications. J Mol Biol 348:363–382
25. Choi SK, Shin BS, Kong EM, Rho HM, Park SH (2000) Cloning of a new *Bacillus thuringiensis* cry1I-type crystal protein gene. Curr Microbiol 41:65–69
26. Gleave AP, Williams R, Hedges RJ (1993) Screening by polymerase chain reaction of Bacillus thuringiensis serotypes for the presence of cryV-like insecticidal protein genes and characterization of a cryV gene cloned from *B. thuringiensis* subsp. kurstaki. Appl Environ Microbiol 59:1683–1687
27. Tabashnik BE, Finson N, Chilcutt CF, Cushing NL, Johnson MW (1993) Increasing efficiency of bioassays: evaluation of resistance to *Bacillus thuringiensis* in diamondback moth (Lepidoptera: Plutellidae). J Econ Entomol 86:635–644
28. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31:3784–3788
29. Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunesekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A (2010) The Pfam protein families database. Nucleic Acids Res 38:D211–222
30. Letunic I, Doerks T, Bork P (2009) SMART 6: recent updates and new developments. Nucleic Acids Res 37:D229–D232
31. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39:D225–D229
32. Wei Y, Thompson J, Floudas CA (2011) CONCORD: A consensus method for protein secondary structure prediction via Mixed Integer Linear Optimization. Proc R Soc A. doi:10.1098/rspa.2011.0514
33. Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M (1995) Evaluation of comparative protein modeling by MODELLER. Proteins 23:318–326
34. Boratyn GM, Schaffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL (2012) Domain enhanced lookup time accelerated BLAST. Biol Direct 7:12
35. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402
36. Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) 3D-Jury: a simple approach to improve protein structure predictions. Bioinformatics 19:1015–8
37. Wallner B, Larsson P, Elofsson A (2007) Pcons.net: protein structure prediction meta server. Nucleic Acids Res 35:W369–374
38. Kurowski MA, Bujnicki JM (2003) GeneSilico protein structure prediction meta-server. Nucleic Acids Res 31:3305–3307
39. Combet C, Jambon M, Deléage G, Geourjon C (2002) Geno3D: automatic comparative molecular modeling of protein. Bioinformatics 18:213–214
40. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5: 725–738
41. Kelley LA, Sternberg MJE (2009) Protein structure prediction on the web: a case study using the Phyre server. Nat Protoc 4:363–371
42. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948

43. Gouet P, Courcelle E, Stuart DI, Métoz F (1999) ESPript: multiple sequence alignments in PostScript. Bioinformatics 15:305–308
44. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J Comp Chem 4:187–217
45. Walter RPS, Philippe HH, Ilario GT, Alan EM, Salomon RB, Jens F, Andrew ET, Thomas H, Peter K, Wilfred FG (1999) The GROMOS biomolecular simulation program package. J Phys Chem 103:3596–3607
46. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible, and free. J Comput Chem 26:1701–1718
47. Darden T, Perera L, Li L, Pedersen L (1999) New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. Structure 7:R55–R60
48. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18:1463–1472
49. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK-a program to check the stereochemical quality of protein structures. J App Cryst 26:283–291
50. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of non-bonded atomic interactions. Protein Sci 2:1511–1519
51. Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356:83–85
52. Pontius J, Richelle J, Wodak SJ (1996) Deviations from standard atomic volumes as a quality measure for protein crystal structures. J Mol Biol 264:121–136
53. Hekkelman ML, Te Beek TA, Pettifer SR, Thorne D, Attwood TK, Vriend G (2010) WIWS: a protein structure bioinformatics Web service collection. Nucleic Acids Res 38:W719–W723
54. Melo F, Devos D, Depiereux E, Feytmans E (1997) ANOLEA: a www server to assess protein structures. Proc Int Conf Intell Syst Mol Biol 97:110–113
55. Pawlowski M, Gajda MJ, Matlak R, Bujnicki JM (2008) MetaMQAP: a meta-server for the quality assessment of protein models. BMC Bioinformatics 9:403
56. Wallner B, Elofsson A (2003) Can correct protein models be identified? Protein Sci 12:1073–1086
57. McGuffin LJ, Buenavista MT, Roche DB (2013) The ModFOLD4 server for the quality assessment of 3D protein models. Nucleic Acids Res 41:W368–W372
58. Chen VB 3rd, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66:12–21
59. Wiederstein S (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:W407–W410
60. Gelly JC, Joseph AP, Srinivasan N, de Brevern AG (2011) iPBA: a tool for protein structure comparison using sequence alignment strategies. Nucleic Acids Res 39:W18–W23
61. Takeshi K (2003) MATRAS: a program for protein 3D structure comparison. Nucleic Acids Res 31:3367–3369
62. Frishman D, Argos P (1995) Knowledge-based protein secondary structure assignment. Proteins 23:566–579
63. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637
64. Ikai AJ (1980) Thermostability and aliphatic index of globular proteins. J Biochem 88:1895–1898
65. Guruprasad K, Reddy BV, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng 4:155–161
66. Kroemer RT, Doughty SW, Robinson AJ, Richard WG (1996) Predictions of the three-dimensional structure of human interleukin-7 by homology modeling. Protein Eng 9:493–498
67. Tajne S, Sanam R, Gundla R, Gandhi NS, Mancera RL, Boddupally D, Vudem DR, Khareedu VR (2012) Molecular modeling of Bt Cry1Ac (DI-DII)-ASAL (*Allium sativum* lectin)-fusion protein and its interaction with aminopeptidase N (APN) receptor of *Manduca sexta*. J Mol Graph Model 33:61–76