Letters

Homology Modeling of Mosquitocidal Cry30Ca2 of *Bacillus thuringiensis* and Its Molecular Docking with N-acetylgalactosamine*

ZHAO Xin Min[1,2], ZHOU Pan Deng[1] and XIA Li Qui[2,#]

1. Department of Chemistry and Environmental Engineering, Hunan City University, Yiyang 413000, Hunan, China; 2. Key Laboratory of Microbial Molecular Biology of Hunan Province, College of Life Science, Hunan Normal University, Changsha 410081, Hunan, China

Abstract

Objective To investigate the theoretical model of the three-dimensional structure of mosquitocidal Cry30Ca2 and its molecular docking with N-acetylgalactosamine.

Methods The theoretical model of Cry30Ca2 was predicted by homology modeling on the structure of the Cry4Ba. Docking studies were performed to investigate the interaction of Cry30Ca2 with N-acetylgalactosamine on the putative receptor.

Results Cry30Ca2 toxin is a rather compact molecule composed of three distinct domains and has approximate overall dimensions of 95 by 75 by 60Å. Domain I is a helix bundle, Domain II consists of three antiparallel β-sheets, Domain III is composed of two β-sheets that adopt a β-sandwich fold. Residue 321Ile in loop1, residues 342Gln 343Thr and 345Gln in loop2, residue 393Tyr in loop3 of Cry30Ca2 are responsible for the interactions with GalNAc via 7 hydrogen bonds, 6 of them were related to the oxygen atoms of hydroxyls of the ligand, and one to the nitrogen of the ligand.

Conclusion The 3D structure of Cry30Ca2 resembles the previously reported Cry toxin structures but shows still some distinctions. Several residues in the loops of the apex of domain II are responsible for the interactions with N-acetylgalactosamine.

Key words: Cry30Ca2; *Bacillus thuringiensis*; Homology modeling; Mosquitoes; Molecular docking

Biomed Environ Sci, 2012; 25(5):590-596 *doi: 10.3967/0895-3988.2012.05.014* *ISSN:0895-3988*

www.besjournal.com(full text) *CN: 11-2816/Q*

INTRODUCTION

Mosquito-borne diseases, including malaria, filariasis, dengue and the viral encephalitides, remain the most important diseases of humans, with an estimated two billion people worldwide living in areas where these diseases are endemic. Thus, there is an urgent need for novel agents and new strategies to control these diseases. Entomopathogenic bacteria, namely *Bacillus thuringiensis* (Bt), have been known from the early 1900's, but the control of dipteran species has been envisaged only since the discovery of Bt serovar *israelensis* in 1977[1].

The virulence of Bt is largely attributable to large crystalline inclusions produced during sporulation. These inclusions are composed of pore-forming proteins known as crystal (Cry) toxins. The mode of action of the Cry toxins is still a matter of investigation; generally, following ingestion by insects, they are activated by gut proteases, and bind to specific receptors on midgut epithelial cells[2].

*This research was supported by grants from Hunan Provincial Natural Science Foundation of China (No.12JJ3021), the National Natural Science Foundation of China (No.30670052, 30570050) and 863 Program of China (2006AA02Z187).

#Correspondence should be addressed to XIA Li Qui. Tel: 86-731-88872905. Fax: 86-731-88872298. E-mail: xialq@hunnu.edu.cn

Biographical note of the first author: ZHAO Xin Min, male, born in 1962, Ph. D, majoring in microbial molecular biology.

Received: July 10, 2011; Accepted: March 15, 2012

The binding to receptors induces a conformational change in the toxin necessary for membrane insertion, forming ion selective channels by oligomerization of toxin monomers, and the insect dies by colloid osmotic lysis[3].

From then on, several toxins from several isolates of Bt were found to have some activity on mosquitoes[4]. Cry30Ca2 is a Cry toxin produced by Bt serovar *jegathesan* during its stationary phase. This δ-endotoxin is active against mosquitoes and has great potential for mosquitoes control. Crystal structures of several Cry toxins including the Diptera-specific Cry4Ba[5] and Cry4Aa[6] toxins have been elucidated. The three dimensional structures of these toxins are remarkably similar in spite of different insect specificities. They are composed of three structurally conserved domains.

The seven α-helices that form the N-terminal domain I is involved in ion channel formation[7]. Domain II consists of three antiparallel β-sheets with exposed loop regions, which vary significantly in length and amino acid sequence in different toxins. Previous studies have showed that surface exposed charged and hydrophobic residues in domains II loop regions of several Cry toxins are involved in receptor binding (both irreversible and reversible)[8]. Domain III is a β-sandwich, and implicated in both correct folding of the whole active toxin and receptor recognition[9].

Early report showed that the binding of the insecticidal Bt Cry1Ac toxin to the putative receptor aminopeptidase N(APN) is specifically inhibited by N-acetylgalactosamine(GalNAc), suggesting that this toxin recognize GalNAc on the receptor[10].

Recently, mosquitocidal Cry11Ba can bind to APNs from *An. quadrimaculatus* and *An. Gambiae*[11-12]. APN is supposed to interact with domain II, loop 2, and loop 3[13]. More recently, mosquitocidal Cry4Ba toxin-binding proteins APNs have been identified in gut brush border membranes of the *Aedes aegypti* mosquito larvae[14]. All those prompted us to investigate the implications that the residues in other areas on the surface of domain II may interact with GalNAc on the receptor aminopeptidase .

In this study , we report a model for the structure of the Cry30Ca2 toxin based on a hypotheses of structural similarity with Cry4Ba toxin, and A more complete understanding of the 3D structure of mosquitocidal 30Ca2 and its molecular docking with the putative receptor will be important in addressing the question of how Cry toxins target mosquitoes. Such insights will lead to a better understanding of the basis of specificity and the practical application of improved toxins in mosquito control.

MATERIALS AND METHODS

An alignment of the amino acid sequences of Cry4Ba (PDB entries 1W99) and Cry30Ca2 was produced with Clustalw program (http://www.ebi.ac.uk/c-lustalw/#) at first, and then corrected manually with the structural alignment tool of the program Swiss-PdbViewer[15] Until a satisfactory placement of conserved blocks and amino acid identities was obtained. Cry30Ca2 contains four of the five protein motifs conserved among the main family of Cry proteins. This alignment project file was submitted to Swiss-Model in the expasy server (http://www.expasy.ch/spdbv/) and a preliminary model for Cry30Ca2 was retrieved. The server pipeline builds the model purely based on this alignment. During the modelling process, implemented as rigid fragment assembly in the SWISS-MODEL pipeline, the modelling engine might introduce minor heuristic modifications to the placement of insertions and deletions.

The model was geometry optimized with calculation of the Hyperchem program[16] and recalculated by molecular dynamics at the same time. To perform the molecular mechanics geometry optimization and molecular dynamics, the following parameters were chosen. Dielectric: Distance dependent, Scale factor: 1, both Electrostatic and van der Waals 1-4 scale factors: 0.5, Cutoffs: None, Minimization algorithm: RMS gradient: 0.1, Heat time: 0.1 picoseconds. Staring temperature: 100 K, Simulation temperature: 300 K, Temperature step: 30 K, and the default values for the other variables.

Sequence identities were calculated with Molsoft program[17]. Figures, electrostatic potentials calculations were generated with Pymol program[18]. The PDB file of GalNAc was built with Molsoft program. Docking studies were performed using Molegro software[19]. This docking program is based on a new hybrid search algorithm, called guided differential evolution. The guided differential evolution algorithm combines the differential evolution optimization technique with a cavity prediction algorithm which is dynamically used during the docking.

In homology modeling, it is very important that appropriate steps are built into the process to assess

the quality of the model. For this, we used Ramachandran plot, VERIFY-3D[20] and ERRAT[21] to assess the quality of the deduced model.

RESULTS AND DISCUSSION

Overall Architecture

The sequence identity of Cry30Ca2 with Cry4Ba is 38%. It is possible and reasonable to build the theoretical model by manual alignment. The final model comprises 582 amino acid residues (Figure 1). The Ramachandran plot indicated that most (95%) of residues have ϕ and ψ angles in the core and allowed regions (Figure 2).

VERIFY-3D uses energetic and empirical methods to produce averaged data points for each residue to evaluate the quality of protein structures. By using this scoring function, if more than 80% of the residue has a score of >0.2 then the protein structure, is considered high quality[20]. For our Cry30Ca2 model, 85.52% of the residues have a score of >0.2 (Figure 3).

ERRAT is a so-called 'overall quality factor' for non-bonded atomic interactions, and higher scores mean higher quality. Practically, the normally accepted range is >50 for a high-quality model[21]. In the current case, the ERRAT score for Cry30Ca2 model was 53.240, which again proved the good quality of the predicted structure.

Cry30Ca2 toxin is a rather compact molecule composed of three distinct domains and has approximate overall dimensions of 95 by 75 by 60Å. It resembles the previously reported Cry4Ba toxin structure but shows some distinctions (Figure 4). The root mean square deviation (RMSD) of aligned Ca atoms of Cry30Ca2 compared to Cry4Ba is 1.10. The surface electrostatic potential distribution of Cry30Ca2 and Cry1Aa is different (Figure 5).

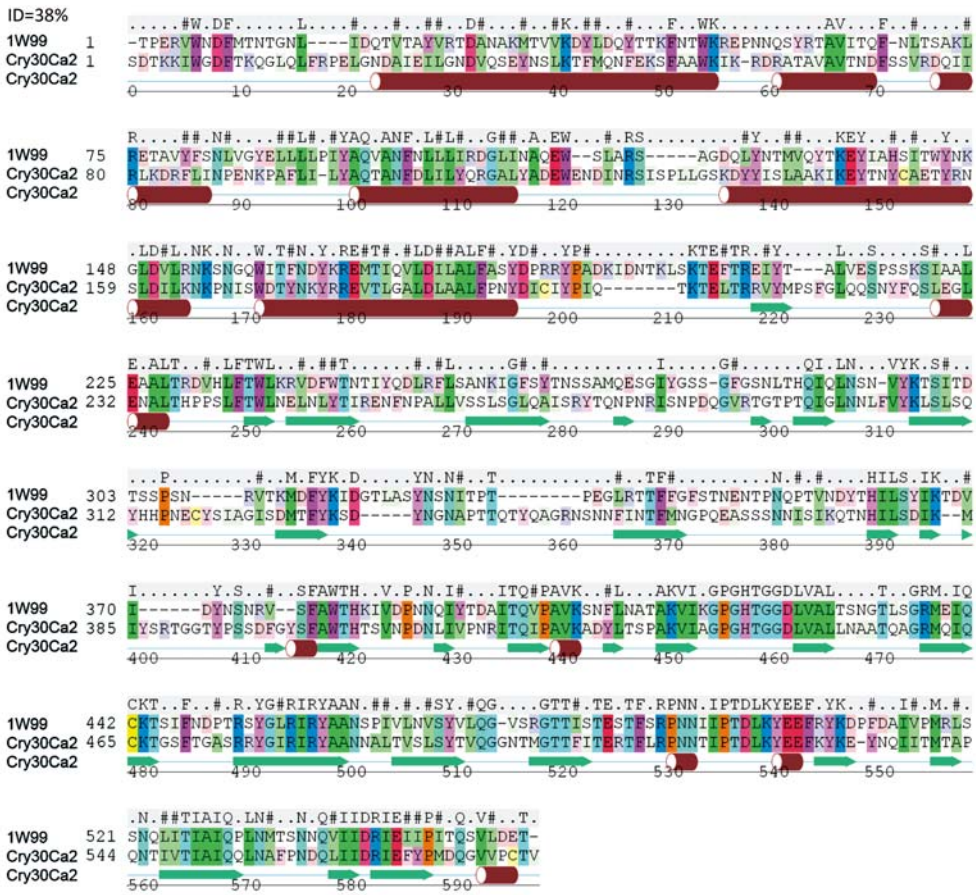

Figure 1. Alignment of the amino acid sequences between Cry30Ca2 with Cry4Ba (1W99). Secondary structure elements of Cry30Ca2 are indicated under the sequence. α-Helices are indicated by cylinders, and β-sheets by arrows. ID: sequence identity.

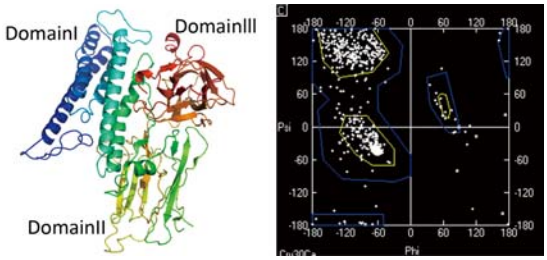

Figure 2. Ribbon representations of Cry30Ca2 (right) and Ramachandran plot of ψ and ϕ of Cry30Ca2 model (left). The 3D structure of the deduced model shows a topology composed of three domains.

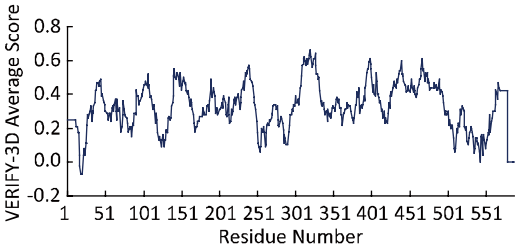

Figure 3. VERIFY-3D score profiles calculated for Cry30Ca2 model. 85.52% of the residues of Cry30Ca2 have a score of >0.2. The percentage of residues with a score >0.2 should be more than 80% for a reliable model.

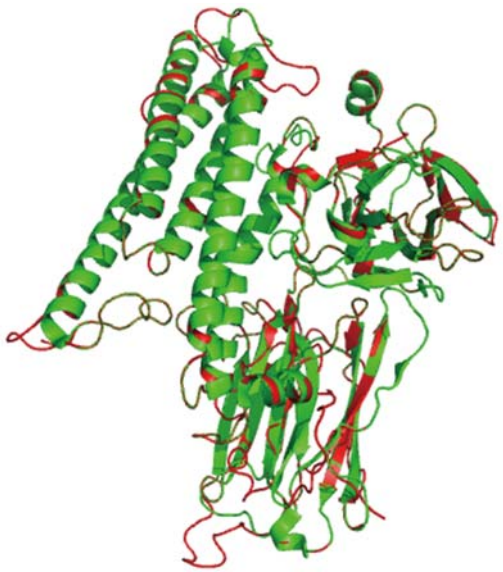

Figure 4. Superimposition of the overall ribbon structures of Cry30Ca2 (red) and Cry4Ba (green). The N-terminal domain I of the two toxins superimpose best. Domain II is the most divergent domain.

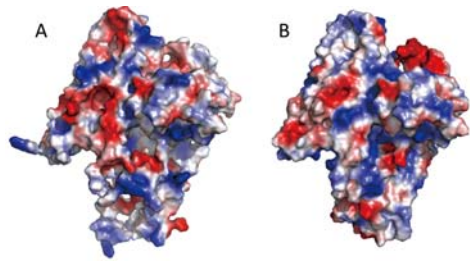

Figure 5. Surface representations of the electrostatic potential of Cry30Ca2(A) and Cry4Ba (B). The distribution of electrostatic potential of the two toxins is different. Positive electrostatic potentials are blue, and negative electrostatic potentials are red.

Domain I

Domain I contains residues 1-203. The most hydrophobic helix, α5, is located centrally and surrounded by the four remaining helices (Figure 6). These helices correspond to helices α3-α7 in the previously reported Cry toxin structures[22]. This suggests that the N-terminal fragment have undergone proteolysis during crystallization of those Cry toxins. The five helices in domain I are all longer than 31 Å and amphipathic in character. Helices α3, α4, α6, and α7 are arranged in an arc around helix α5, in a counter-clockwise order when viewed from the top of Figure 2. The N-terminal part of α3 is exposed on the molecular surface. In Cry4Ba, site-directed mutagenesis studies showed that a single proline substitution in the middle of α4 abolished toxicity, whereas introducing proline into the middle of α3 had no effect[23]. This indicates α4 is important for pore formation and α3 is probably not part of the pore structure. When domain I of Cry30Ca2 and Cry4Ba are superimposed, all the helices superimposed very well (Figure 4). That means the two domains share a common pore forming action.

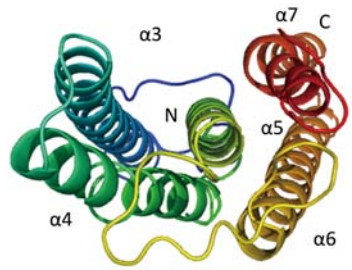

Figure 6. Cartoon representation of domain I of Cry30Ca2. Domain I is a helix bundle with α5 located centrally and surrounded by the remaining helices.

Domain II

Receptor binding domain II, made of residues 204-406, consists of three antiparallel β-sheets packed through formation of a central hydrophobic core as in Cry4Ba. Comparison of the structures of the known Cry toxins pointed out their structural diversities and traced the most variable part of the Cry toxin family to domain II, especially its apical loops[24]. The apical loop of Cry30Ca2 is much longer than that of Cry4Ba. Their surface accessibility added to their variability, favors a receptor-recognition role for these loops. Site-directed mutagenesis of the loop residues in related toxins was reported to affect binding affinity and toxicity[8]. The structural differences of the apical loops of the two toxins may indicate they are toxic to very species of mosquitoes.

Domain III

The C-terminal domain, domain III, extending from residues 407 to 582, is composed of two antiparallel β-sheets that adopt a β-sandwich fold and show a jelly-roll-like topology. It stacks on top of domain II and against the side of domain I. The outer sheet composed of three strands is exposed to the solvent. The inner sheet, containing four strands faces the other two domains. Domain II and III are associated via the intersheet connection through hydrogen bonds and hydrophobic interactions. As three of the conserved blocks are included in domain III[1], superposition of domain III of Cry30Ca2 and domain III of Cry4Ba revealed close structural similarity. The function of the C-terminal domain III remains largely unknown. Mutations in domain III of Cry1Aa toxin had an effect on both ion channel activity and membrane permeability[9]. Domain III could play a role in protecting the toxin against extended proteolytic cleavage by gut proteases[25]. Swapping domain III between toxins suggested the changes in specific activity[26]. Recently, our reports showed that the loop linking β18-β19 in domain III of the Cry1Ac toxin played an essential role in its insecticidal activity[27-28].

Molecular Docking

Five binding sites in Cry30Ca2 were predicted by Molegro program. One was located in the apex of domain II that was identified as the receptor binding cavity (Figure 7). Early reports showed that GalNAc was a part of APN recognized uniquely by Cry1Ac[29], and a loop in domain III is part of the GalNAc binding site[30]. Recent reports with mosqutocidal Cry toxins support the role of domain II in binding receptor in mosquto midgut. Cry11Ba binding to APNs from *An. quadrimaculatus* and *An. Gambiae*[11-12]. More recently, A novel Bt subsp. *israelensis* (Bti) Cry4Ba toxin-binding proteins aminopeptidase has been identified in gut brush border membranes of the *Aedes aegypti* mosquito larvae[14]. All those reports make us to hypothesize that GalNAc moiety of APN is involved in the interactions with domain II of the toxin. The pocket in domain II is the reasonable binding site for molecular docking.

Residue 321Ile in loop1, residues 342Gln 343Thr and 345Gln in loop2, residue 393Tyr in loop3 of Cry30Ca2 are responsible for the interactions with GalNAc via 7 hydrogen bonds, 6 of them were related to the oxygen atoms of hydroxyls of the ligand, and one to the nitrogen of the ligand (Figure 8).

A cluster of aromatic amino acids form a potential binding site with dimensions that could accommodate a short oligosaccharide was found in mosquitocidal Cry 4Aa. These five aromatic amino acids form a potential binding site with dimensions that could accommodate a short oligosaccharide[6]. The architecture suggests that domain II probably bind to the carbohydrate moiety of a receptor of the target insect membrane. This notion is further reinforced by the finding that *Caenorhabditis elegans* resistance to nematicidal Cry5Ba toxicity is linked to the loss of a gene encoding a galactosyltransferase[31]. There is only one aromatic amino acid (393) in the apical loops of domain II of Cry30Ca2, but there exists the similar architecture that form a potential binding site with dimensions that could accommodate a carbohydrate group. Here, our study supports that GalNAc on APN interacts with residues in the three loops of domain II of Cry30Ca2. We can hypothesize that mutation in this section may improve the toxicity to mosquitoes and ever change the specificity.

In conclusion, based on structural homology modeling, the findings from this study show that despite the low amino acid sequence identity among Cry30Ca2 and Cry4Ba, the two toxins share a common three-dimensional structure. The more structural similarity between domain I of the two toxins suggests that they have the same role in pore-forming activity, and the more divergence in domain II, especially their apical loops, indicates that they may recognize different receptor or bind the same receptor in a different way. Molecular docking

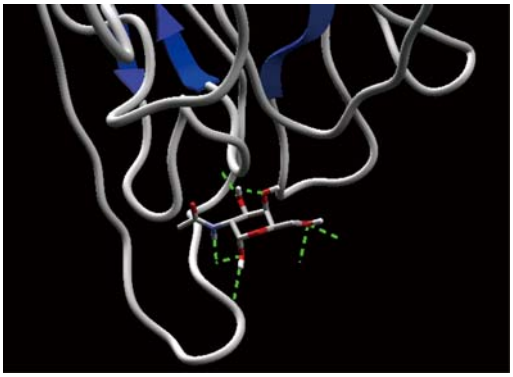

Figure 7. Diagram of the hydrogen-bond (green dotted lines) interactions of GalNAc with domain II of Cry30Ca2.

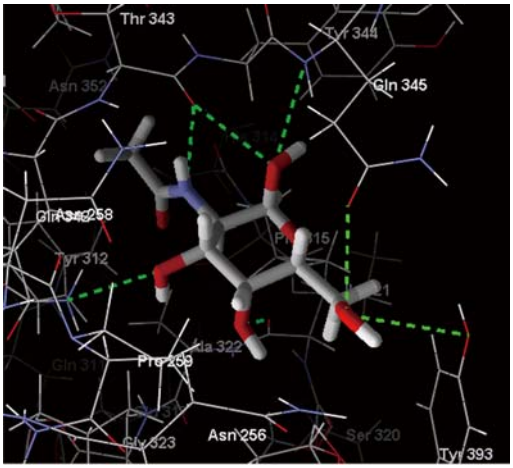

Figure 8. Visualization of the binding mode of GalNAc with the residues in the apex of domain II of Cyt2Ca1. The toxin is presented in wire style, GalNAc in stick, and hydrogen bonds in green dotted lines.

shows that GalNAc moiety of APN is involved in the interactions with domain II of Cry30Ca2. The deduced model and the binding simulation may help design mutagenesis experiments aimed to improving of the toxicity, and shed light on the mechanism of action of mosquitocidal toxins.

ACKNOWLEDGEMENT

We thank Dr. ZHANG You Min (Gene Bridge GmbH, Dresden, Germany) for a critical reading of the manuscript.

REFERENCES

1. Goldberg LJ, Margalit J. A bacterial spore demonstrating rapid larvicidal activity against *Anopheles sergentii*, *Uranotaenia unguiculata*, *Culex univitattus, Aedes aegypti,* and *Culex pipiens.* Mosq. News, 1997; 37, 355-8.
2. Hofte H, Whiteley H. Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbiol Rev, 1998; 53, 242-55.
3. Roh JY, Choi JY, Li MS. *Bacillus thuringiensis* as a specific, safe, and effective tool for insect pest control. J Microbiol Biotechnol, 2007; 17, 547-59.
4. Balaraman K. Occurrence and diversity of mosquitocidal strains of *Bacillus thuringiensis*. J Vect Borne Dis, 2005; 42, 81-6.
5. Boonserm P, Davis D, Ellar J, et al. Crystal structure of the mosquito-larvicidal toxin Cry4Ba and its biological implications. J Mol Biol, 2005; 348, 363-82.
6. Boonserm P, Min M, Angsuthanasombat C, et al. Structure of the functional form of the mosquito larvicidal Cry4Aa toxin from *Bacillus thuringiensis* at a 2.8-angstorm resolution. J Bacteriol, 2006; 188, 3391-401.
7. Li J, Carroll J, Ellar DJ. Crystal structure of insecticidal δ-endotoxin from *Bacillus thuringiensis* at 2.5 Å resolution. Nature, 1991; 353, 815-21.
8. Rajamohan F, Hussain SR, Cotrill JA, et al. Mutations at domain II, loop 3, of *Bacillus thuringiensis* CryIAa and CryIAb delta-endo-toxins suggest loop 3 is involved in initial binding to lepidopteran midguts. J Biol Chem, 1996; 271, 25220-6.
9. Schwartz JL, Potvin L, Chen XJ, et al. Single-site mutations in the conserved alternating-arginine region affect ionic channels formed by CryIAa, a *Bacillus thuringiensis* toxin. Appl Environ Microbiol, 1997; 63, 3978-84.
10. Burton SL, Ellar DJ, Li J, et al. N-acetylgalactosamine on the putative insect receptor aminopeptidase N is recognised by a site on the domain III lectin-like fold of a *Bacillus thuringiensis* insecticidal toxin. J Mol Biol, 1999; 287, 1011-22.
11. Abdullah MA, Valaitis AP, Dean DH. Identification of a *Bacillus thuringiensis* Cry11Ba toxin-binding aminopeptidase from the mosquito *Anopheles quadrimaculatus*. BMC Biochem, 2006; 7, 16-22.
12. Zhang R, Hua G, Andacht TM, et al. A 106-kDa aminopeptidase is a putative receptor for *Bacillus thuringiensis* Cry11Ba toxin in the mosquito *Anopheles gambiae*. Biochemistry, 2008; 47, 11263-72.
13. Jenkins, JL, Lee, MK, Valaitis, AP, et al. Bivalent Sequential Binding Model of a *Bacillus thuringiensis* Toxin to Gypsy Moth Aminopeptidase N Receptor. J Biol Chem, 2000; 275, 14423-31.
14. Bayyareddy K, Andacht TM, Abdullah MA, et al. Proteomic identification of *Bacillus thuringiensis* subsp. *israelensis* toxin Cry4Ba binding proteins in midgut membranes from *Aedes* (*Stegomyia*) *aegypti Linnaeus* (*Diptera, Culicidae*) larvae. Insect Biochemistry and Molecular Biology, 2009; 39, 279-86.
15. Arnold K, Bordoli L, Kopp J, et al. The SWISS-MODEL workspace: a web-based environment. Bioinformatics, 2006; 2, 195-201.
16. Hypercube. HyperChem Release 7.0 for windows Reference Mannual. Hypercube Inc Publication, 2002.
17. Abagyan RA, Totrov MM, Kuznetsov DN. ICM - a new method for protein modeling and design. Applications to docking and structure prediction from the distorted native conformation. J Comp Chem, 1994; 15, 488-506.
18. DeLano WL. "The PyMOL Molecular Graphics System." DeLano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org
19. Thomsen R, Christensen MH. MolDock: A new technique for high-accuracy molecular docking. J Med Chem, 2006; 49, 3315-21.
20. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci, 1993; 2, 1511-9.
21. Zhang KY, Eisenberg D. The three-dimensional profile method using residue preference as a continuous function of residue

environment. Protein Sci, 1994; 3, 687-95.
22. Morse RJ, Yamamoto T, Stroud RM. Structure of Cry2Aa suggests an unexpected receptor binding epitope. Structure, 2001; 9, 409-17.
23. Uawithya P, Tuntitippawan T, Katzenmeier G, et al. Effects on larvicidal activity of single proline substitutions in alpha3 or alpha4 of the *Bacillus thuringiensis* Cry4B toxin. Biochem Mol Biol Int, 1998; 44, 825-32.
24. Zhao XM, Xia LQ, Wang FX, et al. Comparison and Analysis of Cry1Aa, Cry2Aa, Cry3Aa and Cry4Aa of *Bacillus thuringiensis* Toxins with Computer. Acta Chimi Sin, 2008; 66, 108-11.
25. Masson L, Tabashnik BE, Mazza A, et al. Mutagenic analysis of a conserved region of domain III in the Cry1Ac toxin of *Bacillus thuringiensis*. Appl Environ Microbiol, 2002; 68, 194-200.
26. de Maagd RA, Bravo A, Crickmore N. How *Bacillus thuringiensis* has evolved specific toxins to colonize the insect world. Trends Genet, 2001; 17, 193-9.
27. Xia LQ, Wang FX, Ding XZ, et al. The role of β18-β19 loop structure in insecticidal activity of Cry1Ac toxin from *Bacillus thuringiensis*. Chinese Science Bulletin, 2008; 53, 3178-84.
28. Wang FX, Xia LQ, Ding XZ, et al. N546 in β18-β19 loop is important for binding and toxicity of the *Bacillus thuringiensis* Cry1Ac toxin. Journal of Invertebrate Pathology, 2009; 101, 119-23.
29. Knowles BH, Ellar DJ. Characterization and partial purification of a plasma membrane receptor for *Bacillus thuringiensis* var. *kurstaki* lepidopteran-specific delta-endotoxin. J Cell Sci, 1986; 83, 89-101.
30. Pootanakit K, Angsuthanasombat C, Panyim S. Identification of Two Isoforms of Aminopeptidase N in *Aedes aegypti* Larval Midgut. J Biochem Mol Biol, 2003; 36, 508-13.
31. Griffitts JS, Haslam SM, Yang T, et al. Glycolipids as receptors for *Bacillus thuringiensis* crystal toxin. Science, 2005; 307, 922-5.