MINIREVIEW AND HYPOTHESIS: HOMOLOGY MODELLING OF *SPODOPTERA LITURA* (LEPIDOPTERA: NOCTUIDAE) AMINOPEPTIDASE N RECEPTOR

By

Silvio Alejandro López Pazos[1] & Jairo Alonso Cerón Salamanca[1*]

Resumen

López Pazos, S. A. & J. A. Cerón Salamanca: Minireview and hypothesis: homology modelling of *Spodoptera litura* (Lepidoptera: Noctuidae) aminopeptidase N receptor . Rev. Acad. Colomb. Cienc. **32**(123): 139-144, 2008. ISSN 0370-3908.

La aminopeptidasa N (APN) es un receptor para las proteínas Cry1 de *Bacillus thuringiensis*. Una APN de 108-kDa ha sido caracterizada in *Spodoptera litura*, aquí se determinó su estructura tridimensional; esta tiene 4 dominios estructurales. El dominio I es la región que reconoce a las toxinas Cry1, un loop de esta sección puede ser muy importante en este papel. Probablemente, el dominio II tiene funciones en la interacción proteínas Cry1-APN. El dominio III tiene una topología de sándwich y el dominio IV es una superhélice. La APN tiene estructuras conservadas, con ligeras mutaciones como resultado a la coevolución con las proteínas Cry.

Palabras clave: aminopeptidasa N, *Bacillus thuringiensis*, Modelo 3D, *Spodoptera litura.*

Abreviaturas: APN, aminopeptidasa N; *Bt, Bacillus thuringiensis*, 3D, tridimensional

Abstract

Pest insect aminopeptidase N (APN) is a receptor for *Bacillus thuringiensis* Cry1 proteins. A 108-kDa APN has been characterised in *Spodoptera litura*, here their APN three-dimensional structure was determined; it has 4 structural domains. Domain I is the region that recognizes Cry1 toxins, a loop of this section might be very important in this role. Probably, domain II has functions in Cry1 protein-APN interaction. Domain III has a sandwich topology and domain IV is a superhelix. The APN have conserved structures, with slightly mutations just like result of coevolution with Cry proteins.

Key words: aminopeptidase N, *Bacillus thuringiensis*, 3D model, *Spodoptera litura.*

Abbreviations: APN, aminopeptidase N; *Bt, Bacillus thuringiensis*, 3D, three-dimensional

1 Instituto de Biotecnología, Universidad Nacional de Colombia, A.A. 14-490, Santafé de Bogotá D.C., Colombia. Fax: 571-3165412.

* Correo electrónico: jacerons@unal.edu.co

Introduction

Aminopeptidase N (APN) is found as a soluble cytoplasmic enzyme and a membrane-bound ectoenzyme; the ectoenzyme form is attached to epithelial cells of intestinal brush borders and respiratory tracts. In intestinal epithelial cells, APN is important for the final hydrolysis step of ingested proteins (**Luo *et al*.,** 1999). The different APN isoforms in an insect have differential specificities to the N-terminal residues of the protein substrates; their differential expression may play an important role in adaptation of an insect to various dietary proteins from different host-plants (**Wang *et al*.,** 2005).

APN isoforms, together with the cadherin-like protein and membrane bound alkaline phosphatase of the midgut, are known to serve as receptors for insecticidal *Bacillus thuringiensis* (*Bt*) endotoxins (**Wang *et al*.,** 2005) APN is an *N*-acetyl-D-galactosamine (GalNAc)-bearing glycoprotein that is linked to the cell membrane by a glycosylphosphatidylinositol (GPI) anchor, with a molecular weight of 120-kDa approximately (**Chen *et al*.,** 2005). More than 60 different APNs from different Lepidoptera have been sequenced and registered in databases showing the high diversity in isoforms; APNs have a similarity level ranging from 26% to 65% (**Herrero *et al*.,** 2005, **Nakanishi *et al*.,** 1999).

The lepidopteran moth *Spodoptera litura* (Fabricius) (Lepidoptera: Noctuidae) is a polyphagous pest affecting various economically important crops. The insect is susceptible to *Bt* Cry1C toxin. A 2,856-bp gene, encoding a 108-kDa APN, has been isolated from *S. litura* and expressed in Sf21 insect cells. The recombinant APN is glycosylated, is attached to the membrane by a GPI anchor, retains aminopeptidase activity and interacted with the Cry1 toxin (**Agrawal *et al*.,** 2002).

Comparative or homology modelling can provide a useful three-dimensional (3D) model for a protein based on its aligning to one or more proteins of known structure. The prediction process consists of fold assignment, target–template alignment, model-building and evaluation of models. The number of protein sequences that can be modelled accurately is increasing steadily because of the growth in the number and variety of experimentally determined structures and because of improvements in the modelling software (**Marti-Renom,** 2003). Homology modelling is the most reliable method to predict the 3D structure of a protein with accuracy comparable with a low-resolution, experimentally determined structure. Even models with errors can be useful because some aspects of function can be predicted from coarse structural features (**Marti-Renom,** 2003). In this work we show a 3D model of the *S. litura* APN, determined by comparative modelling.

Materials and methods

The *S. litura* APN sequence was obtained from the National Center for Biotechnology Information (NCBI) protein database (http://www.ncbi.nih.gov, accession number: AAK69605). The experimental structures used for construction of the model were tricorn interacting factor F3 of *Thermoplasma acidophilum* (PDB: 1z1w) and leukotrien A4 hydrolase D375n mutant of human (PDB: 1gw6). Functional domains were identified from the NCBI conserved domain database (CDD) (http://www. ncbi.nlm.nih.gov/betweenz/query.fcgi?db=cdd). The structural alignments was generated using DeepView-Swiss-PdbViewer software (**Guex y Peitsch,** 1997) (http://www.expasy.org/spdbv/) and submitted to SWISS-MODEL (http://swissmodel.expasy.org/cgi-bin/sm-submit-request.cgi); a preliminary model for *S. litura* APN was retrieved. The energy minimisation was realized using GROMOS 96 (**Gunsteren *et al*.,** 1996) (http://iqc.ethz.ch/gromos) and the manual correction with the help of the COLORADO3D server (**Sasin y Bujnicki,** 2004) (http://asia.genesilico.pl/colorado3d) until an acceptable model was obtained. The model was validated with WHATIF (**Vriend,** 1990) (http://swift.cmbi.ru.nl/whatif/).

Results

Based on structural alignment a theoretical model for *S. litura* APN was obtained, and corresponds to residues 58-871 of the primary structure, omitting the N-terminal cleavable signal peptide (residues 1-20) and the C-terminal GPI modification site (residues 930-952) (**Agrawal *et al*.,** 2002). Analysis of the primary sequence indicated the presence of the peptidase M1 (residues 46-445) and peptidase N (residues 177-854) functional domains. The APN showed the presence of the HEXXH zinc-binding motif (residues 355-359) characteristic of the zinc peptidases; the third zinc-binding ligand is also conserved in the sequence motif NEXFA (residues 377-381); the amino acid sequence shared the GAMEN motif (residues 319-323) which characterizes them as gluzincin aminopeptidases and involved in their aminopeptidase activity (**Herrero *et al*.,** 2005; **Kyrieleis *et al*.,** 2005). Regions of amino acid sequences of the 11 APN from several insects were compared with the *Bombix mori* APN Cry1Aa toxin-binding region. These 64 residues are believed to be important for Cry1 toxin binding; the region includes highly conserved and variable structures. Cry1 toxins is thought to bind to highly conserved structure of APN or APN family protein

that consists of these conserved amino acid residues; every insect species is believed to produce several APNs containing of theses conserved structures (**Nakanishi *et al.***, 1999). The Cry1 protein-binding region is composed of residues Leu129 to Pro193 in *S. litura* APN.

The molecule is composed of four domains, which together form a hook like-structure (Figure 1). The overall dimensions of the model are 91 Å x 55 Å x 65 Å. The left sector is formed by the N-terminal domain I (residues Asn58-Ile266) and the domain II (residue Ser267-Gly506); the right section is built by the domain III (the smallest) (residues Asn507-Leu581) and the C-terminal domain IV (residues Ser582-Ala871). The zinc binding motif, the third zinc-binding ligand and GAMEN motif are part of the domain II. N-glycosylation sites are in the domain I (Asn104-Asn106), domain II (residues Asn377-Ser379), domain III (Asn574-Thr576) and domain IV (Asn782-Ser784). The four Cys residues, which are highly conserved among APN molecules of higher eukaryotes, are located in the domain IV (residues 728, 735, 763 and 799) (**Agrawal *et al.***, 2002). Cry1 toxin-binding region is localized in the domain I (Figure 2).

Domain I consists of three â-sheets and a short α-helix. The first β-sheet contains five mixed β-strands conformation: S1 (Leu59-Thr70), S2 (Ile78-Met89), S3 (Asn109-Gln111) and S4 (Thr143-Asn153) strands are in antiparallel form, and the S7 strand (Ile201-Ile203) runs parallel to S1. The second â-sheet is composed of S5 (Arg161-Tyr163) and S6 (Glu167-Thr169) antiparallel strands. The last β-sheet of the domain I is built by S8 (Glu222-Arg229) and S9 (Thr231-Ile237) antiparallel strands. Domain I contains an H1 α-helix (Phe187-Cys189) located between strands S6 and S7. The interaction site of *Bt* Cry1 protein with the APN began at the loop that join the strands S3 and S4, continue in strand S4 and the loop joining it to S5, followed by the strand S6, finalizing with the loop binding it to the H1 helix and then continuing until reaching the Pro193 residue. This domain forms a large, saddle-like structure based on strands S1, S2, S3, S4 and S7, and is solvent-exposed (Figure 2).

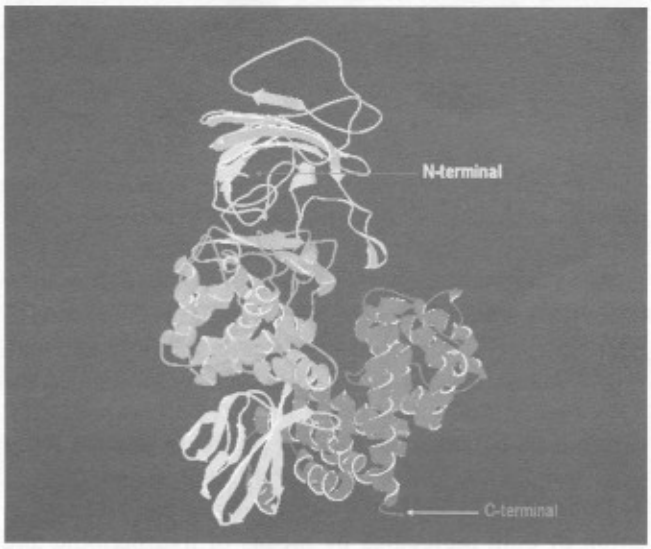

Figure 1. *S. litura* APN 3D structure. The N-terminal and C-terminal extremes are indicated with arrows. Domain I is in orange, domain II in green, domain III in yellow and domain IV in red.

Domain II is the catalytic region for this class of proteins. It presents a thermolysin-like disposition and consists of two β-sheets and ten α-helices. The first β-sheet has two β-strands: S10 (Ser267-Ala271) and S11 (Asn308-Ile311), these are parallel. The following β-sheet is antiparallel and formed by β-strands S12 (Ala320-Glu322) and S13 (Tyr329-Glu331). The α-helices from this domain begin with H2 (Ala276-Tyr277) and H3 (Gln288-His298) that are between S10 and S11 of the secondary structure. Also, this domain is composed of the H4 (Val344-Trp361), H5

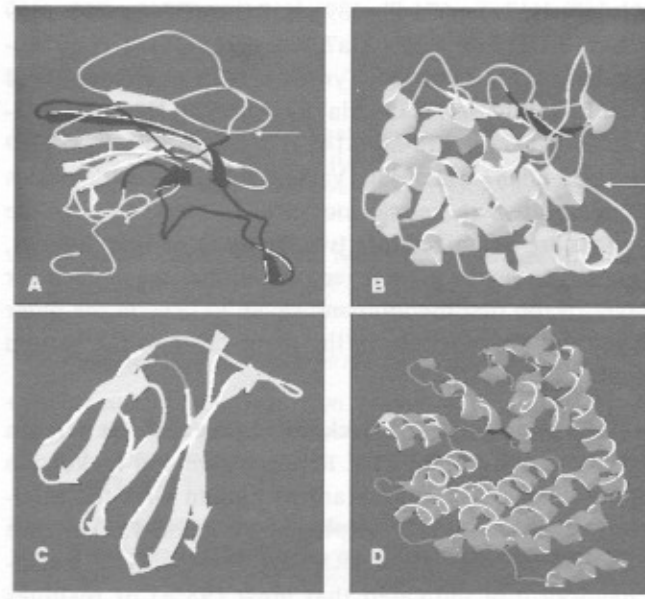

Figure 2. Structural domains of ***S. litura*** APN and their functional characteristics. A. Domain I, in blue Cry1 toxins binding site, the arrow indicates (in green) the N-glycosylation site. B. The zinc-binding motif (HEXXH) in domain II is shown in red, the third zinc-binding ligand (NEXFA) in yellow and the GAMEN motif in blue; the arrow shows the large loop (orange) covering the front of the active site. C. Blue represents residues in the N-glycosylation site of domain III. D. Two superhelix modules can be observed in the C-terminal domain, blue represents the N-glycosylation site and green indicates conserved Cys residues.

(Trp375-Ala392), H6 (Leu398-Asp414), H7 (Ile433-Gln435), H8 (Thr439-Phe455), H9 (Arg459-Asp471), H10 (Pro479-Lys488) and H11 (His494-Ile503) helices. The conserved zinc-binding motif (HEXXH) is part of helix H4; the third zinc-binding ligand (NEXFA) is located in helix H5, and the GAMEN motif contains the β-strand S12. Helices H3 to H11 are interlocked in two planes forming a compact, densely packed α-helical structure. A large loop (Ala415-Thr432) connects the helices H6 and H7; this loop is highly flexible and covers the front of the active site (Figure 2).

Domain III is formed of seven β-strands: S14 (Asn507-Asn515), S15 (Gly517-Asn526), S16 (Arg538-Glu544), S17 (Thr547-Pro553), S18 (Thr557-His561), S19 (Ser566-Lys568) and S20 (Thr576-Leu580). They are organized in a β-sandwich. This domain forms the base of the hook-like APN structure (Figure 2).

The C-terminal domain (IV) is an α-helical domain formed by 19 α-helices that are organized into a superhelix. The helices are: H12 (Ser582-Val589), H13 (Ser596-Ser612), H14 (Ser617-Arg628), H15 (Asn633-Thr647), H16 (Phe655-Ala668), H17 (Trp674-Phe689), H18 (Leu693-Asn702), H19 (Ser713-Ile723), H20 (Ala727-Cys735), H21 (Leu740-Lys747), H22 (Leu757-Tyr762), H23 (Leu766), H24 (Leu777-Asn782), H25 (Ala787-Ala798), H26 (Asp802-Leu806), H27 (Ser813-Met814), H28 (Lys818-Phe828), H29 (Leu832-Glu837) and H30 (Val842-Val867). This domain is divided into two helical modules. In both modules the helices are organized into two layers of parallel helices, four helices in the outer layer and four helices in the inner layer of the first module, and seven helices in the outer layer and three helices in the inner layer of the second module (Figure 2).

The superimposed backbone traces of the tricorn interacting factor F3 of *T. acidophilum* and leukotrien A4 hydrolase D375n mutant of human with the theoretical model of APN displayed 0.57 and 1.70 Å root mean square deviation (RMSD) for Cα. The Ramachandran plot indicated that most (98%) of residues have φ and ψ angles in the core and allowed regions (Figure 3). Most bond angles, bond lengths and torsion angles were in the range of values expected for a naturally folded protein. The structure is more closely related to tricorn interacting factor F3 of *T. acidophilum* (**Kyrieleis *et al.*,** 2005). The theoretical model of *S. litura* APN model obtained here has been deposited in the Model Protein Database (http://www.caspur.it/PMDB/) (ID: PM0074654).

Discussion

The evidence supplied by experiments such as expression of an APN in *Drosophila* made larvae sensitive to Cry1Ac toxin, and the silencing of midgut APN of *S. litura* determined by interference RNA technology (resulted in reduced sensitivity to the Cry1C toxin) suggested that APN plays an important role in Cry toxin susceptibility (**Atsumi *et al.*,** 2005). It has been proposed that the receptor binding sites on Cry1 proteins must have two basic characteristics: a highly conserved structure, because the Cry1 proteins have similar primary sequences and 3D structures and can recognise similar APNs found in the midguts of several lepidopteran insects, and a nonconserved structure, because the Cry1 toxins also exhibit highly specific insecticidal activity and can distinguish host species in the lepidopteran range (**Atsumi *et al.*,** 2005).

The 3D model proposed here will probably be conserved amongst APNs from different types of Lepidoptera on which *Bt* has activity, especially the Cry1 protein binding region, bearing in mind that phylogenetically distant Cry toxin domains II and III (implicated in receptor recognition) have conserved residues (**Shinkawa *et al.*,** 1999).

It was proposed that this Cry1-APN interaction have two steps: carbohydrate recognition and irreversible protein-protein interaction (**Chen *et al.*,** 2005; **Jenkins *et al.*,** 2000). Sixty-four residues form the Cry1-binding epitope of *S. litura* APN, located in domain I. The epitope is formed by various β-strands and an α-helix, as well as the large loops providing the flexibility for binding to Cry1 protein; also, characteristics such as the position in the molecule (in the most distal part of the membrane anchor segment), the orientation towards the solvent and the location in the bottom of the saddle-like structure (conformed by N-terminal domain β-sheets) have supplied their activity and providing stability necessary to the binding. This region has a dimension of 62 Å x 32 Å x 21 Å, implying a considerable surface, probably it allow their interaction with the toxin. A N-glycosylation site (residue ASN103) was found close to this area, possibly allowing initial interaction and then an irreversible binding on the APN recognition site, maybe this might involve initial recognition between the binding pocket in domain III of Cry1 protein and GalNAc on APN (**Atsumi *et al.*,** 2005**; Jenkins *et al.*,** 1999), and subsequently the irreversible union of both molecules by means of hydrophobic and electrostatic interactions and probably salt bridges. Most residues in this region are charged, possibly being important in recognition of the toxin (Figure 4). The

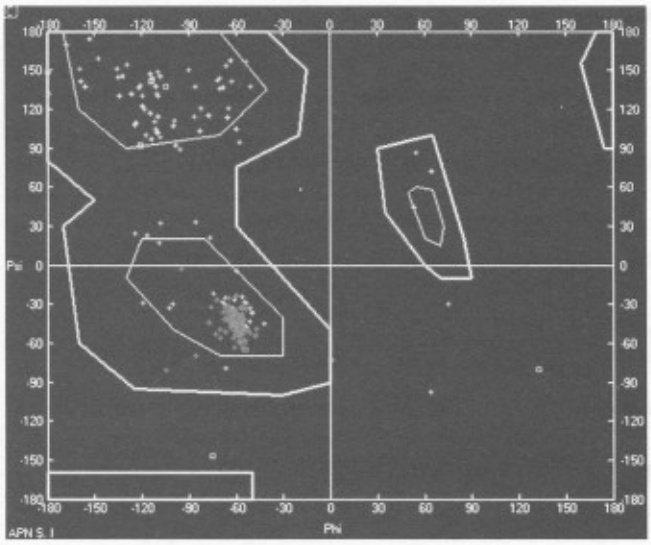

Figure 3. ***S. litura*** APN Ramachandran plot. Most residues were found in the allowed regions. Yellow residues correspond to β-sheets and red residues to α-helices. The squares/boxes indicate the glycine residues which might be located in non-allowed regions.

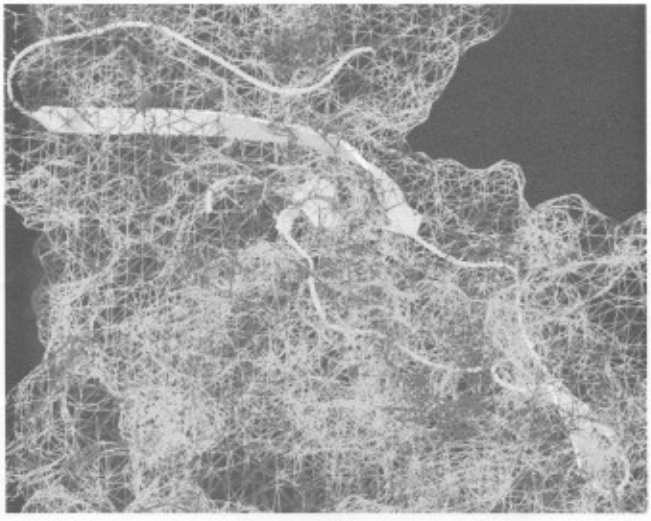

Figure 4. The surface electrostatic potential of the domain I of APN binding site of *Bt* toxins. Yellow indicates the Cry1 binding site, the net represents the electrostatic potential (red: negative, blue: positive).

presence of two binding sites has been suggested in *Manduca sexta* APN, meaning that it is possible that the catalytic domain plays an important role in receptor-toxin interaction (**Masson *et al.***, 2002). The role of domain II in binding to different substrates has been demonstrated in several zinc-metallopeptidases, including APN from mammals (specifically due to the presence of arginin residues in the active site) (**Cristofoletti y Terra,** 2000).

It is probable that the structure of different APNs keep a conserved conformation; however, the recognition site of *Bt* toxin could be a region susceptible to change. This might be a result of the insect and microorganism coevolution, the conserved Cry1 toxin-binding structure of insect APNs has changed slightly, divergent *Bt* Cry toxins might be the result of adaptation to these slight changes in the insect receptor. APNs have conserved structures such as signal peptides, zinc-binding motifs, catalytic domains, N-glycosylation sites and conserved residues. Cry1 proteins can easily develop toxic activity in a broad spectrum of pest insects by recognizing the conserved structures (**Herrero *et al.***, 2005, **Nakanishi *et al.***, 1999; **Agrawal *et al.***, 2002; **Kyrieleis *et al.***, 2005). Maybe, this phenomenon might be profitable for produce novelty Cry proteins with a greater activity, or new specificities, for this reason APN must be structurally characterised in different pest insects, and the interactions with Cry toxins must be analysed; the results will be valuable for directed mutagenesis or domain exchange experiments in *Bt* toxins.

References

Agrawal N., Malhotra P., and Bhatnagar R.K. 2002. Interaction of gene-cloned and insect cell-expressed aminopeptidase N of *Spodoptera litura* with insecticidal crystal Cry1 protein C. Applied and Environmental Microbiology. 68: 4583-4592.

Atsumi S., Mizuno E., Hara H., Nakanishi K., Kitami M., Miura N., Tabunoki H., Watanabe A., Sato R. 2005. Location of the *Bombyx mori* aminopeptidase N type 1 binding site on *Bacillus thuringiensis* Cry1Aa toxin. Applied and Environmental Microbiology. 71: 3966-3977.

Cristofoletti P.T., Terra W.R. 2000. The role of amino acid residues in the active site of a midgut microvillar aminopeptidase from the beetle *Tenebrio molitor*. Biochimica et Biophysica Acta. 1479: 185-195.

Chen J., Brown M.R., Hua G., Adang M.J. 2005. Comparison of the localization of *Bacillus thuringiensis* Cry1A δ-endotoxins and their binding proteins in larval midgut of tobacco hornworm, *Manduca sexta*. Cell and Tissue Research. 321: 123-129.

Guex, N. and Peitsch, M. C. 1997. Swiss-model and the Swiss-PdbViewer: An environment for comparative protein modelling. Electrophoresis. **18**: 2714-2723.

Gunsteren W.F. van, et al. 1996. In Biomolecular simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag ETHZ.

Herrero S., Gechev T., Bakker P.L., Moar W.J. de Maagd R.A. 2005. *Bacillus thuringiensis* Cry1Ca-resistant *Spodoptera exigua* lacks expression of one of four aminopeptidase N genes. BMC Genomics. 6: 96.

Jenkins J.L., Lee M. K., Sangadala S., Adang M.J., Dean D.H. 1999. Binding of *Bacillus thuringiensis* Cry1Ac toxin to *Manduca sexta* aminopeptidase-N receptor is not directly related to toxicity. FEBS Letters. 462: 373-376.

Jenkins J. L., Lee M. K., Valaitis A. P., Curtiss A., Dean D. H. 2000. Bivalent Sequential binding Model of a *Bacillus thuringiensis* Toxin to Gypsy Moth Aminopeptidase N Receptor. The Journal of Biological Chemistry. 275: 14423-14431.

Kyrieleis O.J.P., Goettig P., Kiefersauer R., Huber R. Brandstetter H. 2005. Crystal structures of the tricorn interacting factor F3 from *Thermoplasma acidophilum*, a zinc aminopeptidase in three different conformations. Journal of Molecular Biology. 349: 787-800.

Luo K., McLachlin J.R., Brown M.R., Adang M.J. 1999. Expression of a glycosylphosphatidylinositol-linked *Manduca sexta* aminopeptidase N in insect cells. Protein Expression and Purification. 17: 113-122.

Marti-Renom M. A. 2003. Protein structure modelling for structural genomics. Business briefing. Future Drug Discovery, 59-63.

Masson L., Mazza A., Sangadala S., Adang M.J., Brousseau R. 2002. Polydispersity of *Bacillus thuringiensis* Cry1 toxins in solution and its effect on receptor binding kinetics. Biochimica et Biophysica Acta. 1594: 266-275.

Nakanishi, K., Yaoi K., Shimada N., Kadotani T., Sato T. 1999. *Bacillus thuringiensis* insecticidal Cry1Aa toxin binds to a highly conserved region of aminopeptidase N in the host insect leading to its evolutionary success. Biochimica et Biophysica Acta. 1432: 57-63.

Sasin J. M., Bujnicki J. M. 2004. COLORADO3D, a web server for the visual analysis of protein structures, Nucleic Acids Research. **32** (Web Server issue): W586-W589.

Shinkawa A., Yaoi K., Kadotani T., Imamura M., Koizumi N., Iwahana H., Sato R. 1999. Binding of phylogenetically distant *Bacillus thuringiensis* Cry toxins to a *Bombyx mori* aminopeptidase N suggests importance of Cry toxin's conserved structure in receptor binding. Current Microbiology. 39: 14-20.

Vriend G. 1990. WHAT IF: a molecular modeling and drug design program. Journal of Molecular Graphics. 8: 52-56.

Wang P., Zhang X., Zhang J. 2005. Molecular characterization of four midgut aminopeptidase N isozymes from the cabbage looper, *Trichoplusia ni*. Insect Biochemistry and Molecular Biology. 35: 611-620.

Recibido: 13 de febrero de 2007

Aceptado para su publicación: 9 de junio de 2008