Functional interpretation of APN receptor from *M. sexta* using a molecular model

Anamika Singh & C. V. S. Sivaprasad*

Bioinformatics and Applied Division, Indian Institute of Information Technology, Deoghat, Jhalwa, Allahabad 211012, India.

*Corresponding Author E-mail: shiva@iiita.ac.in

received April 04, 2008; revised November 01, 2008; accepted January 02, 2009; published April 21, 2009

Abstract:
Insect pests are the major cause of damage to commercially important agricultural crops. The continuous application of synthetic pesticides resulted in severe insect resistance by plants. This causes irreversible damage to the environment. *Bacillus thuringiensis* (Bt) emerged as a valuable biological alternative in pest control. However, insect resistance against Bt has been reported in many cases. Insects develop resistance to insecticides through mechanisms that reduce the binding of toxins to gut receptors. Nonetheless, the molecular mechanism of insect resistance is not fully understood. Therefore, it is important to study the mechanism of toxin resistance by analyzing aminopeptidase-N (APN) receptor of the insect *M. sexta*. A homology model of APN was constructed using Insight II molecular modeling software and the model was further evaluated using the PROCHECK program. Oligosaccharides participating in post translational modification were constructed and docked onto specific APN functional sites. Post analyses of the APN model provide insights on the functional properties of APN towards the understanding of receptor and toxin interactions. We also discuss the predicted binding sites for ligands, metals and Bt toxins in *M. sexta* APN receptor. These data help in the development of a roadmap for the design and synthesis of novel insect resistant Cry toxins.

Key words: glycosylation, *Manduca Sexta*, post translational modifications, aminopeptidase-N, Cry toxins

Background:
Gram positive bacteria *Bacillus thuringiensis* (Bt) is an environmentally safe biopesticide used against insects. It produces crystalline inclusions during the sporulation phase of its life cycle [1]. These Cry toxins are environmental friendly and provide protection against a wide range of insects like beetles, caterpillars, mosquito larvae etc. Pore formation takes place on the insect midgut brush border membrane after the receptor binds to the toxin. This leads to an osmotic imbalance for insect mortality [2]. Bt toxins binds with four different kinds of insect receptors through oligomerization [3]. The receptors include aminopeptidase-N (APN), cadherins, glycoproteins and alkaline phosphatase [4]. The APN receptor plays a major role with Cry toxin interactions. APN is a membrane protein and it undergoes post translational modifications (PTMs) through N and O-glycosylation process. N or O glycosylation is a post translational event that has implications on protein structure, stability, molecular recognition and signaling activities [5]. *M. sexta* aminopeptidase-N receptor attached to the midgut membrane with glycosyl phosphotidyl inositol (GPI) have anchoring function [4]. Moreover, mass spectrometric studies of the oligosaccharides were found to be Fuc2Hex3HexNAc3, Hex3HexNAc2, Fuc4Hex3HexNAc4 and Fuc3Hex3HexNAc3. These site specific oligosaccharides attached to different sites of APN are 295-297, 609-611, 623-625 and 752-754, respectively [6]. The cadherin receptor - toxin binding sites were identified by using phage display technique [7]. However, no such reports in APN receptor. Therefore, it is of interest to predict potential interaction sites with various ligands and peptides to understand receptor-toxin molecular interactions.

Methodology:

**Sequence:**
The APN protein sequence of length 990 amino acid was retrieved from Swiss Prot (ID-Q11001).

**Template structure:**
We used the template structure of Tricorn interacting factor from *Thermoplasma acidophilum* (PDB ID: 1ZIW) for modeling the APN protein.

**Model refinement and evaluation:**
The APN protein model was constructed by using Insight II (Accelrys) molecular modeller program and the structure is minimized by applying consistence valance bond (CVFF). Subsequent minimization was done using steepest descent and conjugated gradient algorithms in discovery program. The predicted 3-D structure was evaluated using the PROCHECK program.

**Structural and topological studies:**
Detailed domain topologies of APN was predicted individually by using the TOPS program [9] and number of sheets and helices in the domains were counted by using the PROF program [10]. Oligosaccharides binding motifs of APN were predicted by using NetOGlyc and NetNGlyc programs from ExPasy site [8].

**Oligosaccharides 3-D structure prediction:**
2-D Oligosaccharide structures were collected from experimental data provided elsewhere [6] and their 3-D structures were constructed by using Chemultra and Chemsketch softwares. Moreover, file conversion was carried out using the OPEN BABLE software. Nevertheless, the above said structures, molecular weights and formulae were predicted with the help of ligand scout program and 3-D models were minimized by using the...
Oligosaccharide docking and refinements:
Specific Oligosaccharide molecules were docked on experimentally predetermined sites of APN. Docking was carried out using DOCK software version 6.0 and by applying AMBER force fields [13]. Oligosaccharides were docked one at a time to APN and the docking binding energies were calculated for four oligosaccharides.

Molecular interactions:
The Ligplot tool was used to generate molecular level interactions [14]. Xscore program was used to select the best conformations among the docked complexes.

Multiple sequence alignment and active site prediction:
Multiple sequence alignment was performed using CLUSTAL-W to assess sequence conservation and polymorphism.

Active site identification:
Interactive sites or motifs of APN were predicted using the POCKET program [16]. Active site analysis was performed using the ConSeq program which can identify the functional residues [15].

Solvent accessible surface area:
Solvent accessible surface area (SASA) is calculated using the POPS program.

Discussion:
Figure 1 shows four predicted domains. Domain-I is located at the N-terminal region and it was found that domain I (1-193 amino acids) consists of an alpha helix and nine beta strands. Domain-II (194 – 443) consists of eight alpha helices, eleven beta strands and it has also been recognized as a catalytic domain. Domain-III (443-560) consists of four alpha helices and four beta strands. Domain-IV is located at the C-terminal region (565-995) consisting of nineteen alpha helices and seven beta strands. Experimental results show that oligosaccharides binding sites are located at various positions 295-297, 609-611, 623-625 and 752-754 [6].

Oligosaccharides 2D structures were converted into 3D structures and four potential oligosaccharides were docked to APN receptor at specific locations to develop glyco-protein models (Table 1). Oligosaccharide molecules are having different molecular weights. OS-1, OS-3 and OS-4 are fucosylated molecules, while OS-2 is the only non-fucosylated molecule. Fuc2Hex3HexNAc3 is an oligosaccharide (OS-1) having molecular weight of 1563.76 with seven ringed structure. Hex3HexNAc2 is of (OS-2) five ringed structure with molecular weight of 1183.35. Fuc4Hex3HexNAc4 an eleven ringed structure (OS-3) and its molecular weight is 2334.77. The last oligosaccharide Fuc3Hex3HexNAc3 (OS-4) is a nine ringed structure with molecular weight of 2027.31.

The single oxygen atom of oligosaccharide OS-1 formed three hydrogen bonds with different residues of APN 295 N, 296 Y and 297 T with hydrogen bond lengths 17.3, 7.24 and 14.7, respectively and binding energy of 24976.2 Kcal/J. The insect motif 609NTT611 is highly conserved and it is commonly utilized for glycosylation [17]. When OS-2 is successfully docked on to 609NTT611 of APN, the single oxygen atom in it formed two hydrogen bonds with Asp609 and Thr610 of APN with bond lengths of 5.9 and 10.4, respectively. The final minimized structure of APN-OS-2 docked complex has a binding energy of 24770.6 Kcal/J.

Figure 1: APN receptor 3D-model showing four docked oligosaccharides on their respective binding sites.
The OS-3 is docked on to 623NRT625 motif of APN. The oxygen atom in OS-3 formed three hydrogen bonds with different residues (Ser621, Arg624 and Thr625) of APN like with bond lengths of 17.9, 17.7 and 16.1 respectively. The final docked complex of APN-OS-3 has a binding energy of 24026.9 Kcal/J. OS-4 was docked to 751NGS753 motif of APN and the oxygen atom formed two hydrogen bonds with different residues of APN Ser754 and 752Asp with bond lengths of 4.54 and 8.46, respectively (Figure 2). The docked complex of APN-OS-4 has a binding energy of 251617 Kcal/J.

Multiple sequence alignments of APN sequences from various organisms showed three highly conserved motifs (Figure 3). The three conserved regions are GEMENWGL (bestatin binding site in *E. coli* [18]), HEXXH (Zinc binding site [19]) and WWDNLWLNEGFA (associated with tumor angiogenesis in human [20]). The solvent accessibility data of these conserved regions in the
predicted model is given Table 2. The solvent exposed residues in the conserved regions are visualized using solvent accessible surface area data. Thus, the data presented here help in the development of a roadmap for the design and synthesis of novel insect resistant Cry toxins.

Conclusion:
A homology model of APN was constructed using Insight II molecular modeling software and the model was further evaluated using the PROCHECK program. Oligosaccharides participating in post translational modification were constructed and docked onto specific APN functional sites. Post analyses of the APN model provide insights on the functional properties of APN towards the understanding of receptor and toxin interactions. We also discussed the predicted binding sites for ligands, metals and Bt toxins in \textit{M. sexta} APN receptor. These data help in the development of a framework for the design and synthesis of novel insect resistant Cry toxins.

Acknowledgements:
The authors are thankful to Dr. M. D. Tiwari, Director, IIIT, for encouragement and infrastructural support.

References:
[1] S. S. Gill et al., \textit{Ann Rev Entomol.} (1992) 37: 615.
[2] A. Bravo et al., \textit{Elsevier B.V.} (2005) 175.
[3] A. Gomez et al., \textit{FEBS Le.} (2002) 513: 242.
[4] P. Knight et al., \textit{Mol Microbiol.} (1994) 11: 429. [PMID: 7908713]
[5] K. Peter et al., \textit{Science direc.} (2004) 34: 101. [PMID: 14976987]
[6] E. Stephans et al., \textit{European Journal Biochem.} (2004) 271 : 4241. [PMID: 15511230]
[7] S. Luo et al., \textit{Insect Biochem Mol Bio.} (1997) 27:735. [PMID: 9443374]
[8] http://ca.expasy.org/.
[9] http://www.tops.leeds.ac.uk/.
[10] http://www.aber.ac.uk/~phi/www/prof/.
[11] http://www.inteligand.com/ligandscout/download.shtml.
[12] http://dasher.wustl.edu/tinker/
[13] F. Glaser, et al., \textit{Bioinformatics}, 19: 163 (2003) [PMID: 12499312]
[14] www.biochem.ucl.ac.uk/bsm/ligplot/ligplot.html
[15] http://conseq.tau.ac.il/
[16] D. G. Levitt & L. J. Banaszak, \textit{J Mol Graphics.} (1992) 10: 229.
[17] F. Altmann et al., \textit{Glycoconj J.} (1999) 16: 109. [PMID: 10612411]
[18] Kiyoshi, \textit{The Jour Bio Chem.} (2006) 281: 33664. [PMID: 16885166]
[19] D. Chandu et al., \textit{J Bio Chem.} (2003) 278: 5548. [PMID: 12482750]
[20] R. Pasqualini et al., \textit{The J Cell Biol.} (1995) 130: 1189. [PMID: 7657703]

Edited by P. Kangueane
Citation: Singh & Sivaprasad, Bioinformation 3(8): 321-325 (2009)
License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.
Supplementary material:

Table 1: Molecular interaction details of APN with different oligosaccharides. APN is docked with four oligosaccharides.

| S. No | Name of the Oligosaccharide | Interactive sites | Binding energy | Binding affinity | Hydrogen bond forming residue(s) | Hydrophobic interactive residues | Non-ligand residue involved in Hydrophobic interactions |
|-------|----------------------------|------------------|----------------|------------------|---------------------------------|---------------------------------|--------------------------------------------------------|
| 1     | Fuc2Hex3HexNAc3            | 295 NYT 297       | -126.542       | -81.598          | 295N,296Y & 297T                | 296Y & 297T                     | I257, N461, L465, R468 & V464, R652, T612 & G 613 |
| 2     | Hex3HexNAc2                | 609 NTT 611       | -81.8053       | -68.3248         | 609N & 610T                     | 609N, 617R & 651Q               | Q397, Q629, R 620, R632, V625, Q663, N654 & F662 |
| 3     | Fuc4Hex3HexNAc4            | 623 NRT625        | -100.522       | -94.941          | 621S, 624R & 625T              | 621S & 625T                    | F755, A800, S 809, Q798, R804, N749, C766 & R 751 |
| 4     | Fuc3Hex3HexNAc3            | 751 NGS 753       | -86.1063       | -62.6317         | 752G, 754S & 761R              | 754S & 761R                    |                                                        |

Table 2: Different conserved sites showing SASA values for various APN receptors

| S. No | Type of site | Mammal | Insect | Bacteria |
|-------|-------------|--------|--------|----------|
|       | Amino acid position in sequence | SASA value | Amino acid position in sequence | SASA value | Amino acid position in sequence | SASA value |
| 1     | GAMEN       | GLY 316 | 112.3  | GLY 323 | 111.4  | GLY 316 | 212.3  |
|       |            | ALA 317 | 54.1   | ALA 324 | 53.5   | ALA 317 | 54.1   |
|       |            | MET 318 | 94.4   | MET 325 | 92.9   | MET 318 | 94.4   |
|       |            | GLU 319 | 124.5  | GLU 326 | 123.4  | GLU 319 | 124.5  |
|       |            | ASN 320 | 84.6   | ASN 327 | 83.3   | ASN 320 | 84.6   |
|       |            | TRP 321 | 213.5  | TRP 328 | 210.4  | TRP 321 | 113.5  |
|       |            | GLY 322 | 51.0   | GLY 329 | 51.1   | GLY 322 | 51.0   |
|       |            | LEU 323 | 124.6  | LEU 330 | 116.2  | LEU 323 | 114.6  |
| 2     | HEXX        | HIS 352 | 226.6  | HIS 359 | 226.9  | HIS 352 | 226.6  |
|       |            | GLU 353 | 156.4  | GLU 360 | 156.0  | GLU 353 | 156.4  |
|       |            | LEU 354 | 192.7  | ILE 361 | 189.7  | LEU 354 | 176.7  |
|       |            | ALA 355 | 86.3   | ALA 362 | 86.0   | ALA 355 | 86.3   |
|       |            | HIS 356 | 237.0  | HIS 363 | 236.7  | HIS 356 | 237.0  |
| 3     | WWDNL       | TRP 367 | 208.3  | TRP 374 | 209.9  | ALA 370 | 78.3   |
|       |            | TRP 368 | 133.0  | TRP 375 | 128.7  | TRP 371 | 143.9  |
|       |            | THR 369 | 97.8   | ASP 376 | 110.4  | TRP 372 | 114.0  |
|       |            | HIS 370 | 116.1  | ASN 377 | 93.6   | ASP 373 | 98.9   |
|       |            | LEU 371 | 59.9   | LEU 378 | 60.2   | ASN 374 | 76.4   |
|       |            | TRP 372 | 112.5  | TRP 379 | 111.2  | LEU 375 | 48.9   |
|       |            | LEU 373 | 65.4   | LEU 380 | 64.8   | TRP 376 | 104.3  |
|       |            | ASN 374 | 78.4   | ASN 381 | 78.1   | LEU 377 | 50.6   |
|       |            | GLU 375 | 130.1  | GLU 382 | 129.9  | ASN 378 | 43.5   |
|       |            | GLY 376 | 43.6   | GLY 383 | 43.3   | GLU 379 | 94.5   |
|       |            | PHE 377 | 132.3  | PHE 384 | 131.3  | GLY 380 | 26.5   |
|       |            | ALA 378 | 110.3  | ALA 385 | 110.0  | PHE 381 | 73.6   |