Research Journal of Recent Sciences _____ ISSN 2277-2502
Vol. **2(2),** 40-49, February **(2013)** *Res. J. Recent Sci.*

Theoretical structure prediction of TcaA from *Photorhabdus luminescens* and aminopeptidase N receptor from *Helicoverpa armigera*

Maithri S.K., Ramesh K.V.* and Mutangana Dieudonné
Department of Biotechnology, Center for Postgraduate studies, Jain University, Jayanagar, Bangalore – 560011, INDIA

Available online at: www.isca.in
Received 12th October 2012, revised 22nd November 2012, accepted 22nd December 2012

Abstract

Due to resistance developed by various agricultural pests towards Bacillus thuringiensis (BT) toxins, there is a necessity of developing alternative biopesticide. The TcaA toxin produced by Photorhabdus luminescens is a high molecular weight insecticidal toxin having toxicity against wide range of agricultural pests. Phylogenetic tree constructed for TcaA revealed that this toxin does not have any ancestral relationship with BT toxins. Present study focused on the modeling the TcaA toxin from Photorhabdus luminescens and aminopeptidase N receptor from Helicoverpa armigera using various computational approaches. Structural validation using various tools such as ProSA and PROCHECK revealed that Φ and ψ angles of these theoretical models were present in the core and allowed region. The theoretical toxin structure was subsequently docked onto the homology modeled aminopeptidase N receptor. Outcome of the docking study showed that first domain of TcaA had highest docking energy when compared to remaining domains.

Keywords: TcaA, APN, docking, molecular modelling, domain.

Introduction

Various problems are associated with the application of synthetic insecticides. Many of these insecticides are non-biodegradable, undergo slow degradation and persist in the environment for a long time. Also excessive and prolonged uses of synthetic insecticides can lead to the development of insect resistance[1]. To overcome the environmental hazards caused by the application of synthetic insecticides, biopesticides offer a better alternative. For instance, *Bacillus thuringiensis* accounts for 90% of the bio insecticide market; it produces insecticidal toxins called delta endotoxins (BT toxin) which are proteinaceous, readily biodegradable and also have a short half-life inside the insect midgut[2]. In the recent past, many of the crop pests have developed resistance towards BT toxins and this includes: *Ostrinia nubilalis* (European corn borer), *Heliothis virescens* (Tobacco bud worm), *Pectonophora gossypiella* (Pink bollworm moth), *Culex quinquefasciatus* (Mosquito), *Aedes aegypti* (Yellow fever mosquito), *Trichloro plusiani* (Tiger moth), *Leptinotarsa decemlineata* (Colorado potato beetle), *Spodoptera exigua* (Beet armyworm), *Spodoptera littoralis* (Egyptian cotton leaf worm), and *Chryosomela scripta* (Cottonwood leaf beetle)[3-7]. Therefore, insecticidal toxins of *Photorhabdus luminescens* can be considered as a potential substitute to the BT toxin.

Photorhabdus luminescens is a gram-negative entomopathogenic enterobacterium that exists in mutualistic symbiosis with nematodes of the family *Heterorhabditidae* found in the gut of infective insect host *Heterorhabditis bacteriophora*[8]. After entering inside the insect host, the nematode releases the bacteria by regurgitation directly into the insect's hemocoel. In the hemocoel, these bacteria replicate rapidly and cause lethal sepsis by producing different high molecular weight toxins that kill the insect within 48–72 hours[9]. Among various insecticidal toxins produced by the *Photorhabdus luminescens* purified Tca was shown to disrupt the insect midgut epithelium in a manner similar to the δ-endotoxins from *Bacillus thuringiensis*[8].

There are several studies reporting that glycosyl-phosphatidyl-inositol (GPI) anchored aminopeptidase N (APN)[10] and cadherin-like protein[11] of *Manduca sexta* act as receptors for the BT toxins. Among these two receptors, the APN receptor belongs to the Zn-binding metalloprotease family. The C-terminal region of the APN binding site is rich in N-acetylgalactosamine (GalNAc) and acts as the binding site of the Cry1Ac toxin[10,12].

Although the toxin protein produced by *Photorhabdus* bacterium has been proved to be toxic against wide variety of insects[8], structural information of these proteins is yet to be resolved through X- ray / NMR experiments. Therefore in the present study we report 3D models for TcaA toxin (*Photorhabdus luminescens* w14) as well as APN receptor (*Helicoverpa armigera*). Further we have also examined the molecular interaction of TcaA toxin with the APN receptor through the docking studies.

Material and Methods

Phylogenetic analysis: Amino acid sequence of TcaA from *Photorhabdus luminescens* in FASTA format along with the insecticidal cry toxins from *Bacillus thuringiensis* such as 1DLC_A[13], 1JI6_A [14], 3EB7[15], 1CIY_A[16], 2C9K_A[17],

1W99_A[18] and 1I5P_A[19] were submitted to the CLUSTALW[20]. Output generated by CLUSTALW tool was saved as 'allign.phy'. Then the CONSENSE tree for 100 data sets was generated with the help of Phylip[21]. Molecular modeling of TcaA toxin from *Photorhabdus luminescens* and Aminopeptidase N from *Helicoverpa armigera.*

TcaA toxin sequence from *Photorhabdus luminescens* was submitted to ProDom[22] server for detecting the presence of any domains. Templates selection for TcaA and APN were based on the output generated by PSI-BLAST tool[23]. To predict the tertiary structure, primary amino residues of these two proteins along with the template suggested by PSI-BLAST tool were submitted to SWISS-MODEL sever[24] in the alignment mode. As an alternative to homology modeling, threading method was also attempted by accessing the I-TASSER server[25]. The initial 3D models of the toxins as well as APN were subjected to loop refinement based on the output generated by ERRAT[26], ProSA[27] and PROCHECK[28] server. Loop regions of the proteins showing high percentage of error were further refined through ModLoop server[29]. After re-validating the refined structures, their structural homologs were searched using the services of DaliLite server[30].

Docking studies: Tertiary structures of all domains of TcaA were taken for docking studies with APN receptors from *Helicoverpa armigera* using Hex software[31]. Docked conformations and interaction energies were recorded at the end of the docking exercise. During the dock operation, the total energies were calculated based on shape as well as electrostatics using a default grid spacing of 0.6 Å.

Results and Discussion

Phylogenetic analysis: Phylogenetic analysis of TcaA from *Photorhabdus luminescens* was done by including similar group of toxin sequences from other organisms along with Cry toxins of *Bacillus thuringiensis*. The consense tree generated for TcaA toxin branched into two distinct groups, wherein the first group consisted of only Cry toxins from *B. thuringiensis,* while *the* second group comprised of only TcaA toxins from different species. Further, based on the location of TcaA sequence of *P. luminescens* clade in the consensus tree, it is possible to deduce that, the sequence was evolutionary unrelated to the remaining Cry toxins of *Bacillus thuringiensis*; however, it showed close evolutionary relationship with TcaA toxin of *Burkholderia rhizoxinica* (figure-1).

Molecular modeling studies: The results obtained from ProDom server suggest that TcaA sequence from *Photorhabdus luminescens* has six domains (table-1). Although, search for template sequences from PDB using PSI Blast tool did not yield any sequence homologs for any of these 6 domains of TcaA, homologous sequences for the receptor sequence "aminopeptidase N" (APN) from *Helicoverpa armigera* could be retrieved from PDB. Among the various template sequences reported by the PSI Blast tool, human endoplasmic reticulum aminopeptidase which had the maximum sequence identity (39%) was selected for modeling APN from *H. armigera* (table-2).

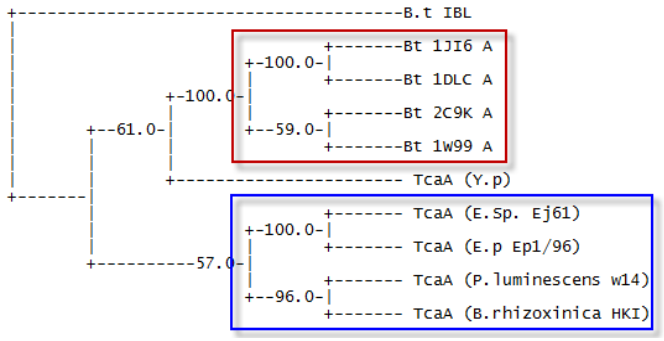


Figure-1
Consensus tree generated for TcaA from *Photorhabdus luminescens* using Promlk program of PHYLIP

Table-1
Detection of domains in TcaA sequence from *Photorhabdus luminescens* using the ProDom server

Position	ProDom domain	Score E value
1-117	#PD287783	615 5e-63
118-224	#PD156536	555 5e-56
225-595	#PD321543	1989 0
596-836	#PD185778	1296 6e-142
860-999	#PDA1W468	741 1e-77
1000-1095	#PD230859	405 1e-38

The 3D structure of APN from *H. armigera* generated by SWISS MODEL server displayed four domains. While the first (D-I) and third domain (D-III) arranged in sandwich form, consisted of β-sheets (D- I=16; D - III= 6), the fourth (D-IV) domain is made up of α-helices (D-IV=15). The second domain (D-II) had both α-helices (10) and β-sheets (5). Except for minor differences in D-I and D-II, same number of helices / sheets were present in the corresponding domains of the PDB template, 3MDJ_A.

Comparison of homology modeled APN structure of *H. armigera* with crystal structure of crystal structure of human aminopeptidase (3MDJ_A), shows that overall topology of all the four domains (D I, II, III and IV) of the predicted models appears to have more or less similar orientation to that of 3MDJ_A. The second domain (D- II) is having a good sequence homology with the corresponding domain of the crystal structure of human aminopeptidase (3MDJ_A), a catalytic domain determined by Nguyen *et al*. (2011). Therefore, it is possible to speculate similar activity for the theoretical structure of D-II of APN. Sequence analysis revealed that, except for the residue T in the motif "323GATEN327, remaining residues followed by the motif "359HExxHx18E382" of D-II was well conserved with human aminopeptidase 3MDJ_A (figure-2). Methionine located at 267th position of 3MDJ_A was replaced with threonine at the corresponding position of D-II. Studies carried out by Nguyen *et al*. on human aminopeptidase have clearly shown that while "265GAMEN269" motif plays roles in

the exopeptidase specificity of aminopeptidase through interaction with the N-terminal amino acid of the substrate, the motif region "301HEXXH324" functions as receptors for zinc ligands and are essential for the catalytic activity of the enzyme.

Table-2

Summary of (A) PDB templates generated by PSI blast results and (B) sequence alignment of APN sequence from *Helicoverpa armigera* at the end of 20th iteration

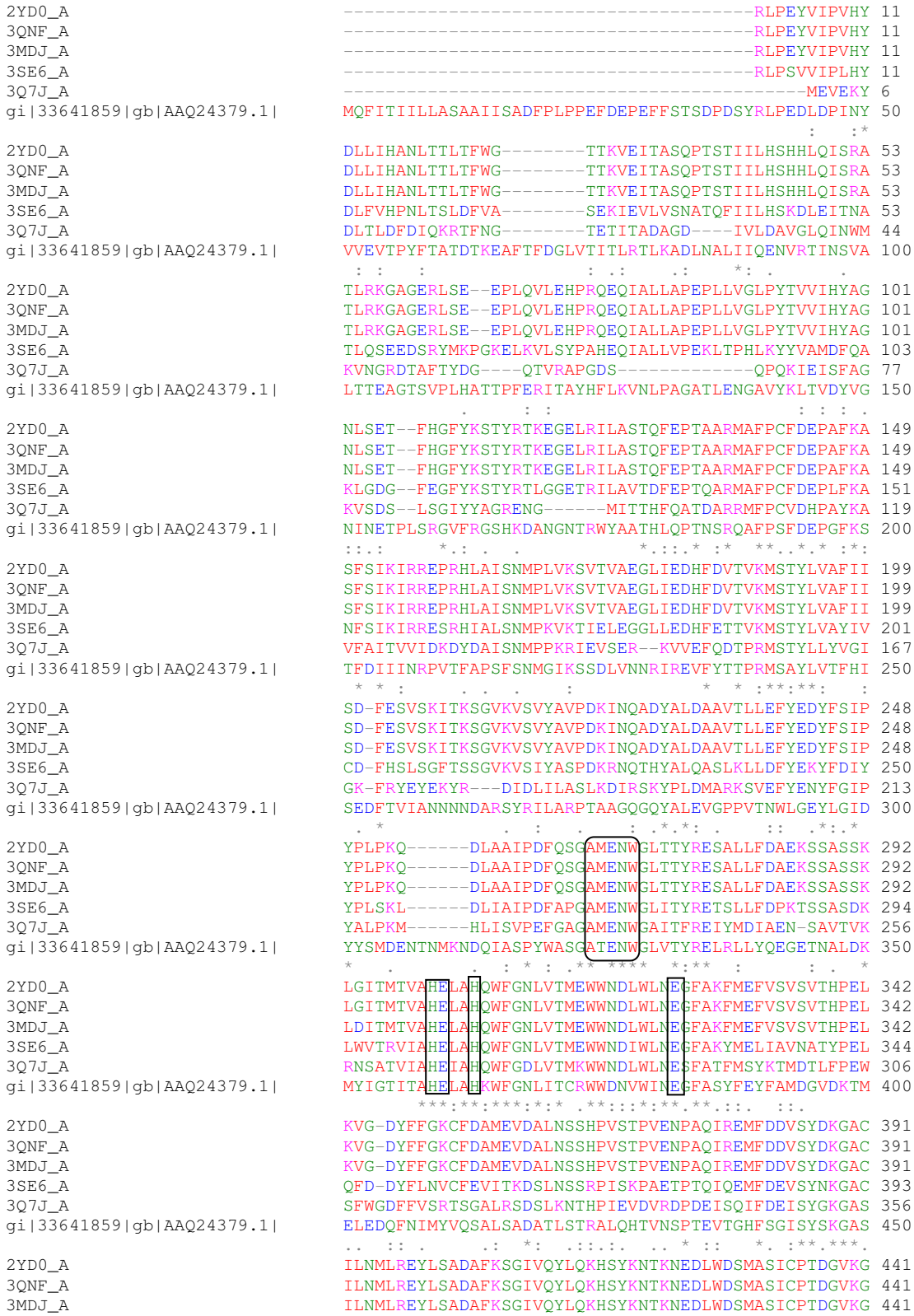
(A)

E-value BETTER than threshold Sequences producing significant alignments:		Score (Bits)	E Value
pdb|3MDJ|A	Chain A, Er Aminopeptidase, Erap1, Bound To The Zi...	1022	0.0
pdb|2YD0|A	Chain A, Crystal Structure Of The Soluble Domain O...	1004	0.0
pdb|3QNF|A	Chain A, Crystal Structure Of The Open State Of Hu...	1002	0.0
pdb|3SE6|A	Chain A, Crystal Structure Of The Human Endoplasmi...	960	0.0
pdb|3Q7J|A	Chain A, Engineered Thermoplasma Acidophilum F3 Fa...	720	0.0
pdb|1Z1W|A	Chain A, Crystal Structures Of The Tricorn Interac...	720	0.0

B)

```
Query  40   RLPEDLDPINYVVEVTPYFTATDTKEAFTFDGLVTITLRTLKADLNALIIQENVRTINSV  99
            RLPE + P++Y + +    T        TF G   + + T     + +I+  +   I+
Sbjct  17   RLPEYVIPVHYDLLIHANLTTL------TFWGTTKVEI-TASQPTSTIILHSHHLQISRA  69
Query  100  ALTTEAGTSVPLHATTPFERITAYHFLKVNLPAGATLENGAVYKLTVDYVGNINETPLSR  159
             L   AG  +      P + +      ++ L A   L  G  Y + + Y GN++ET
Sbjct  70   TLRKGAGERLS---EEPLQVLEHPRQEQIALLAPEPLLVGLPYTVVIHYAGNLSET--FH  124
Query  160  GVFRGSHKDANGNTRWYAATHLQPTNSRQAFPSFDEPGFKSTFDIIINRPVTFAPSFSNM  219
            G ++ +++   G  R  A+T  +PT +R AFP FDEP FK++F I I R        +
Sbjct  125  GFYKSTYRTKEGELRILASTQFEPTAARMAFPCFDEPAFKASFSIKIRREPRHLAISNMP  184

Query  220  GIKSSDLVNNRIREVFYTTPRMSAYLVTFHISEDFTVIANNNNDARSYRILARPTAAGQG  279
             +KS  +    I + F  T +MS YLV F IS+ F  ++          + A P    Q
Sbjct  185  LVKSVTVAEGLIEDHFDVTVKMSTYLVAFIISD-FESVSKITKSGVKVSVYAVPDKINQA  243

Query  280  QYALEVGPPVTNWLGEYLGIDYYSMDENTNMKNDQIASPYWASGATENWGLVTYRELRLL  339
             YAL+    +  +  +Y  I Y         K D  A P + SGA ENWGL TYRE  LL
Sbjct  244  DYALDAAVTLLEFYEDYFSIPY------PLPKQDLAAIPDFQSGAMENWGLTTYRESALL  297

Query  340  YQEGETNALDKMYIGTITAHELAHKWFGNLITCRWWDNVWINEGFASYFEYFAMDGVDKT  399
            +   +++A  K+ I    AHELAH+WFGNL+T  WW+++W+NEGFA + E+ ++
Sbjct  298  FDAEKSSASSKLDITMTVAHELAHQWFGNLVTMEWWNDLWLNEGFAKFMEFVSVSVTHPE  357

Query  400  MELEDQFNIMYVQSALSADATLSTRALQHTVNSPTEVTGHFSGISYSKGASLLLMLKHFL  459
            +++ D F       A+  DA  S+  +   V +P ++   F  +SY KGA +L ML+ +L
Sbjct  358  LKVGDYFF-GKCFDAMEVDALNSSHPVSTPVENPAQIREMFDDVSYDKGACILNMLREYL  416

Query  460  TENTFKKALNIFLEARKFEHAFPADLFSAFATAVQQDGVP-------------------S  500
            + + FK  +  +L+   +++    DL+ + A+    DGV
Sbjct  417  SADAFKSGIVQYLQKHSYKNTKNEDLWDSMASICPTDGVKGMDGFCSRSQHSSSSSHWHQ  476

Query  501  NTFDIASFMKYWVEEPGYPVLEVSVNSAAGRIELSQKRFLVSATATP-TDQVWPLPLTYT  559
               D+ + M  W  + G+P++ ++V      + + Q+ ++  +   P T  +W +PLT+
Sbjct  477  ERVDVKTMMNTWTLQRGFPLITITVRGR--NVHMKQEHYMKGSDGAPDTGYLWHVPLTFI  534

Query  560  TESNPDWQNLLPSKVMTAKTDFIERNVGTNEWVIFNVQQKGIYRVNYDTRNWELLAAALS  619
T  +   +++   ++  KTD +       EW+ FNV   G Y V+Y+   W+ L   L
Sbjct  535  TSKS----DMVHRFLLKTKTDVLILPEEV-EWIKFNVGMNGYYIVHYEDDGWDSLTGLLK  589
Query  620  RDHTAIHHLNRAQIVDDVFALMRSGQITYRLGFKVLDFLKKDTSYYSWYPAITGFNWLRN  679
              HTA+   +RA ++++ F L+  G+++      +  +LK +T      P   G N L
Sbjct  590  GTHTAVSSNDRASLINNAFQLVSIGKLSIEKALDLSLYLKHETE---IMPVFQGLNELIP  646

Query  680  RF-LHLPTTLAAFDEILYGFLDAVITDL-GYDVVANE-PLTRTLNRFFTLSFACNIGHKG  736
             + L     +   +     FL  ++ DL       +E  ++  + R   L  AC   ++
Sbjct  647  MYKLMEKRDMNEVETQFKAFLIRLLRDLIDKQTWTDEGSVSERMLRSELLLLACVHNYQP  706

Query  737  CVDNAVQKFVALKDNSV--AVNPNLRRHVFCEGLRAGGLDEWQYLYNRRQASNNQGDEVA  794
CV  A   F   K+++   ++  ++   VF  G  A   + W +LY++ Q S +  ++
Sbjct  707  CVQRAEGYFRKWKESNGNLSLPVDVTLAVFAVG--AQSTEGWDFLYSKYQFSLSSTEKSQ  764

Query  795  MLRSLGCTSNTAAGQAYLKMILDDDVVKAQDRVNAFSFFYMGHRDNAKAGLQFLKDNVDA  854
+  +L  T N    Q  L      D +K Q+           +        QFL+ N +
Sbjct  765  IEFALCRTQNKEKLQWLLDESFKGDKIKTQE-FPQILTLIGRNPVGYPLAWQFLRKNWNK  823
Query  855  IRKAVVLPAWFNN--VLTTTAGYLDEAGLRDME---EWLLANQNAVPEFAVGISAITSAR  909
            + +   L +      V+ TT  +     L +++     L  N + +      I  I
Sbjct  824  LVQKFELGSSSIAHMVMGTTNQFSTRTRLEEVKGFFSSLKENGSQLRCVQQTIETI---E  880
Query  910  NNMQWGSDNAATI---IAAANDEDPPEDGGSGEE  940
             N+ W   N   I   + +   E  PE   +G E
Sbjct  881  ENIGWMDKNFDKIRVWLQSEKLEHDPEADATGLE  914
```

Figure-2
Multiple sequence alignment of Aminopeptidase N from *H. armigera* with selected PDB homologs using ClustalW.



```
3SE6_A                          ILNMLKDFLGEEKFQKGIIQYLKKFSYRNAKNDDLWSSLSNSCLESDFTS 443
3Q7J_A                          ILRMIEDYAGYEEFRKGISKYLNDHKFGNAEGSDLWTAIEDVS------- 399
gi|33641859|gb|AAQ24379.1|      LLLMLKHFLTENTFKKALNIFLEARKFEHAFPADLFSAFATAVQQDGVPS 500
                                :* *:..:   : *:..:  :*:  .: ::   **: ::
2YD0_A                          MDGFCSRSQHSSSSSHWHQEGVDVKTMMNTWTLQKGFPLITITVRGRN-- 489
3QNF_A                          MDGFCSRSQHSSSSSHWHQEGVDVKTMMNTWTLQKGFPLITITVRGRN-- 489
3MDJ_A                          MDGFCSRSQHSSSSSHWHQERVDVKTMMNTWTLQRGFPLITITVRGRN-- 489
3SE6_A                          GGVCHSDPKMTSNMLAFLGENAEVKEMMTTWTLQKGIPLLVVKQDGCS-- 491
3Q7J_A                          --------------------GKPVKRVMEYWIKNPGYPVIKLKRNGRK-- 427
gi|33641859|gb|AAQ24379.1|      N-------------------TFDIASFMKYWVEEPGYPVLEVSVNSAAGR 531
                                                    :  .*  *  : * *:: :.  .
2YD0_A                          VHMKQEHYMKGSD------GAPDTGYLWHVPLTFITSKS----DMVHRFL 529
3QNF_A                          VHMKQEHYMKGSD------GAPDTGYLWHVPLTFITSKS----DMVHRFL 529
3MDJ_A                          VHMKQEHYMKGSD------GAPDTGYLWHVPLTFITSKS----DMVHRFL 529
3SE6_A                          LRLQQERFLQGVFQEDPEWRALQERYLWHIPLTYSTSSS----NVIHRHI 537
3Q7J_A                          ITMYQTRFLLNGE----------EEGRWPVPVNIKKK------DGVERIL 461
gi|33641859|gb|AAQ24379.1|      IELSQKRFLVSAT-------ATPTDQVWPLPLTYTTESNPDWQNLLPSKV 574
                                : : * ::: .             * :*:.  ..      : :   :
2YD0_A                          LKTKTDVLIL-PEEVEWIKFNVGMNGYYIVHYEDDGWDSLTGLLKGTHTA 578
3QNF_A                          LKTKTDVLIL-PEEVEWIKFNVGMNGYYIVHYEDDGWDSLTGLLKGTHTA 578
3MDJ_A                          LKTKTDVLIL-PEEVEWIKFNVGMNGYYIVHYEDDGWDSLTGLLKGTHTA 578
3SE6_A                          LKSKTDTLDL-PEKTSWVKFNVDSNGYYIVHYEGHGWDQLITQLNQNHTL 586
3Q7J_A                          LEDEAS-----IEADGLIKINADSAGFYRVLYDDATFSDVMGHYR----D 502
gi|33641859|gb|AAQ24379.1|      MTAKTDFIERNVGTNEWVIFNVQQKGIYRVNYDTRNWELLAAALSRDHTA 624
                                :  ::.           : :*.   * * * *:   :. :
2YD0_A                          VSSNDRASLINNAFQLVSIGKLSIEKALDLSLYLKHETEIMPVFQGLNEL 628
3QNF_A                          VSSNDRASLINNAFQLVSIGKLSIEKALDLSLYLKHETEIMPVFQGLNEL 628
3MDJ_A                          VSSNDRASLINNAFQLVSIGKLSIEKALDLSLYLKHETEIMPVFQGLNEL 628
3SE6_A                          LRPKDRVGLIHDVFQLVGAGRLTLDKALDMTYYLQHETSSPALLEGLSYL 636
3Q7J_A                          LSPLDRIGLVDDLFAFLLSGHIDPETYRQRIRNFFDDEDHNVITAIVGQM 552
gi|33641859|gb|AAQ24379.1|      IHHLNRAQIVDDVFALMRSGQITYRLGFKVLDFLKKDTSYYSWYPAITGF 674
                                :   :*  ::.: * ::  *::      .    : .: .       :  :
2YD0_A                          IPMYKLMEKRDMNEVETQFKAFLIRLLRDLIDKQ--TWTDEGSVSERMLR 676
3QNF_A                          IPMYKLMEKRDMNEVETQFKAFLIRLLRDLIDKQ--TWTDEGSVSERMLR 676
3MDJ_A                          IPMYKLMEKRDMNEVETQFKAFLIRLLRDLIDKQ--TWTDEGSVSERMLR 676
3SE6_A                          ESFYHMMDRRNISDISENLKRYLLQYFKPVIDRQ--SWSDKGSVWDRMLR 684
3Q7J_A                          EYLRMLT--HAFDDDARAFCRSRMQFLTGKQDEN--LKIALGRVSR---- 594
gi|33641859|gb|AAQ24379.1|      N--WLRNRFLHLPTTLAAFDEILYGFLDAVITDLGYDVVANEPLTRTLNR 722

                                           :      :       :                :
2YD0_A                          SQLLLLACVHNYQPCVQRAEGYFRKWKESNGNLSLPVDVTLAVFAVGAQS 726
3QNF_A                          SQLLLLACVHNYQPCVQRAEGYFRKWKESNGNLSLPVDVTLAVFAVGAQS 726
3MDJ_A                          SELLLLACVHNYQPCVQRAEGYFRKWKESNGNLSLPVDVTLAVFAVGAQS 726
3SE6_A                          SALLKLACDLNHAPCIQKAAELFSQWMESSGKLNIPTDVLKIVYSVGAQT 734
3Q7J_A                          ------LYVMVDESYAEEMSKLFKDFDSAEP------EMRSSIATAYALV 632
gi|33641859|gb|AAQ24379.1|      FFTLSFACNIGHKGCVDNAVQKFVALKDNSVAVNPNLRRHVFCEGLRAGG 772

                                              :.    *    . .                 *
2YD0_A                          TEGWDFLYSKYQFSLSSTEKSQIEFALCRTQNKEKLQWLLDESFKGDKIK 776
3QNF_A                          TEGWDFLYSKYQFSLSSTEKSQIEFALCRTQNKEKLQWLLDESFKGDKIK 776
3MDJ_A                          TEGWDFLYSKYQFSLSSTEKSQIEFALCRTQNKEKLQWLLDESFKGDKIK 776
3SE6_A                          TAGWNYLLEQYELSMSSAEQNKILYALSTSKHQEKLLKLIELGMEGKVIK 784
3Q7J_A                          TGDLKGLLEKFRSVDRDEDRVRIISAFGKLKSNTDLSTVYGMVEKTEIKK 682
gi|33641859|gb|AAQ24379.1|      LDEWQYLYNRRQASNNQGDEVAMLRSLGCTSNTAAGQAYLKMILDDDVVK 822

                                   . * .: .    . :.  :  ::   .              . .  *
2YD0_A                          TQ-EFPQILTLIGRNPVGYPLAWQFLRKNWNKLVQKFELGSSSIAHMVMG 825
3QNF_A                          TQ-EFPQILTLIGRNPVGYPLAWQFLRKNWNKLVQKFELGSSSIAHMVMG 825
3MDJ_A                          TQ-EFPQILTLIGRNPVGYPLAWQFLRKNWNKLVQKFELGSSSIAHMVMG 825
3SE6_A                          TQ-NLAALLHAIARRPKGQQLAWDFVRENWTHLLKKFDLGSYDIRMIISG 833
3Q7J_A                          QD-MISFFSSALETLP-----GREFIFANLDRIIRLVI------------ 714
gi|33641859|gb|AAQ24379.1|      AQDRVNAFSFFYMGHRDNAKAGLQFLKDNVDAIRKAVVLP--AWFNNVLT 870

                                 :  .  :             . :*:  *   : : .
2YD0_A                          TTNQFSTRTRLEEVKGFFSSLKENGSQLRCVQQTIETIEENIGWMDKNFD 875
3QNF_A                          TTNQFSTRTRLEEVKGFFSSLKENGSQLRCVQQTIETIEENIGWMDKNFD 875
3MDJ_A                          TTNQFSTRTRLEEVKGFFSSLKENGSQLRCVQQTIETIEENIGWMDKNFD 875
3SE6_A                          TTAHFSSKDKLQEVKLFFESLEAQGSHLDIFQTVLETITKNIKWLEKNLP 883
3Q7J_A                          ---------------------------------------------------
gi|33641859|gb|AAQ24379.1|      TTAGYLDEAGLRDMEEWLLANQNAVPEFAVGISAITSARNNMQWGSDNAA 920
```

A large cavity was observed between D-II and D-IV of APN model which had resemblance to the template 3MDJ_A**.** Based on this, it is hypothesized that this cavity could provide an easy access to the catalytic site for substrates and might also represent binding site for peptide based substrates[32]. The D-III is the smallest among all the domains and act as a connecting domain between D-II and D-IV (Figure-3A). Superimposition of Cα backbone of predicted model onto 3MDJ_A shows very low RMSD value of 1.1Å.

I-TASSER software was successful in predicting the 3D structures for all the six domains of TcaA from *Photorhabdus luminescens* (Figure-3B). Among these six domains D-I, II, V and VI are purely made up of α-helices (D-I=6, D-II=4, D-V=9 and D-VI =5). Helices of D-II are arranged as antiparallel helix-loop-helix motif in a manner similar to that of solution NMR structure of metalloprotein from *Escherichia coli* (PDB ID: 2HZ8) determined by Calhoun *et a*l (2008)[33]. However, both D-III and D-IV had α-helices (D-III=11 and D-IV= 15) as well β-sheets (D III = 5 and D IV= 2). While the folding pattern of D-III is like a horseshoe comparable to that of the crystal structure of ribonuclease inhibitor (PDB ID 2BEX)[34], D-IV architecture resembled to a half-doughnut which is identical to the crystal structure of mitochondrial transcription termination factor 3 from human (PDB ID: 3M66)[35]. Superimposition of Cα backbone of all these structures onto the top PDB template used by I-TASSER showed a very low RMSD (table-3).

Table-3
RMSD of all the six domains of TcaA from *Photorhabdus luminescens* generated by the Dali server

Domains	Templates used by I-TASSER	RMSD generated by DALI server
Domain I	3V42_A	2.2
Domain II	2HZ8_A	1.5
Domain III	2BEX_ A	2.6
Domain IV	3M66_A	1.9
Domain V	2J0O_A	3.1
Domain VI	3KPH_B	3.3

Structure validation: Validation of the theoretical models of TcaA and APN by the structural assessment tools such as ProSA, ERRAT and PROCHECK showed that there was an improvement in the quality of the predicted structure upon loop refinement. Quality factor of the loop refined APN model calculated by ERRAT reached 57.65 from 48.941. The Z score value computed by the ProSA tool for APN model was -8.89. This suggests that the models are lying within the Z score values of native conformational structures. Ramachandran plot generated by PROCHECK indicates that most of the residues from APN model have φ and ψ angles in the core and allowed regions. This indicates that most of the main-chain conformations of models are consistent with their side-chain conformations. Summary of the Ramachandran plot revealed that, 85.5% of the residues of APN structure were distributed in the core region, followed by 12.8% in the allowed, 1.0% in the general and remaining 0.9% in the disallowed. The overall G-score calculated for APN model was found to be -0.08 which was above the threshold value indicating the predicted model was satisfactory.

Improvement in the quality of the structure upon loop refinement was also seen in all the six TcaA domains of *Photorhabdus*. The Φ and ψ angles of these theoretical models were present in the core and allowed region. Ramachandran plot statistics indicates that the predicted models were of good quality (table-4).

Table-4
PROCHECK summary for all the seven domains of TcaA from *Photorhabdus luminescens*

Domains	Residues Present in				G-factor
	Core region	Allowed region	General region	Disallowed region	
I	84.8%	11.4%	1.9%	1.9%	-0.05
II	90.8%	9.2%	0.0%	0.0%	0.01
III	68.8%	25.0%	3.1%	3.1%	-0.49
IV	84.8%	10.1%	3.2%	1.8%	-0.06
V	82.8%	12.5%	3.1%	1.6%	-0.18
VI	87.4%	9.2%	1.1%	2.3%	0.15

Docking studies: HEX was successful in docking the homology modeled aminopeptidase receptor of *Helicoverpa armigera* on to all the six domains of TcaA toxin model of *Photorhabdus luminescens*. Based on the docking energies for all the six domains of TcaA presented in table 5, the D-I model got docked onto the APN receptor model with the highest dock energy of -792.9 KJ / mol. Visualization of the docked output using DeepView package reveal that a total of 121 residues of both D-I domain and APN model are involved in the interaction (D-I = 59 and APN = 62 residues) (figure-4). Among these 121 residues, maximum residues are non polar in nature (table-6) (figure-5). The docked complex was stabilized by the formation of strong inter-molecular hydrogen bonds.

Table-5
Summary of the docking energy of APN from *Helicoverpa armigera* onto the different domains of TcaA from *Photorhabdus luminescens*

Domains	Binding energy with APN from *Helicoverpa armigera*
I	-792.9 kj/mol
II	-719.0 kj/mol
III	-304.1 kj/mol
IV	-274.9 kj/mol
V	-588.5 kj/mol
VI	-656.8 kj/mol

Conclusion

Based on the *in-silico* studies conducted on the interaction of TcaA toxin from *Photorhabdus luminescens* with the aminopeptidase N receptor of *Helicoverpa armigera* we can conclude that TcaA toxin can be considered as a potential biopesticide for controlling the pest population affecting the agricultural crops.

Table-6
Summary of residues interacting between DI of TcaA and homology modeled APN receptor from *Helicoverpa armigera*

Acidic	Basic	Polar	Non Polar
DI of TcaA			
asp30,97,106,108,111, glu63,77,101	lys8,66,81,85, arg36,61,70,75,115	gln6,26,41,60,73,76,89, asn7,64,91, ser10,33,67,90,117, tyr25	val14,28,40,94, met5,113, ile9,34, phe11,109, ala37,71,98,116, gly24,105,107 leu65,68,74,86,89,95,112, trp72, pro78
APN receptor (*Helicoverpa armigera*)			
asp780, glu109,390,841	lys812, arg118,289,752,781, his121,394,	gln367,779, asn108,806,783, ser113,120,277,396,399,786, thr29, 110,389,362, tyr274,307,400	val54,748,782,814,815 met364,790, ile363,398,810 phe395,787,802,820 ala276,372,778,813,818 gly26,119,278,393,397 leu112,291,751,816 trp130,275,631,842 pro111

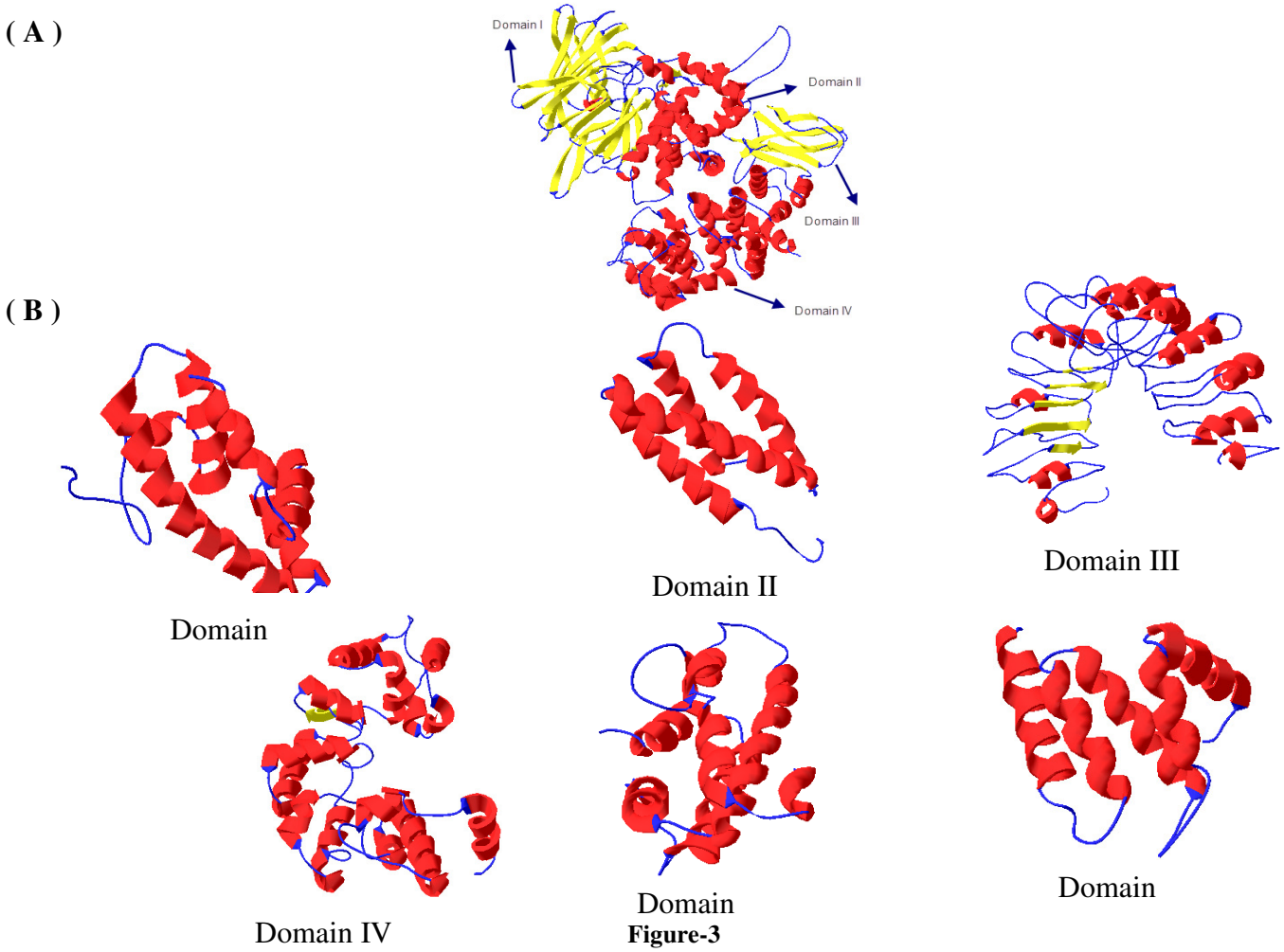


Figure-3
3D structure of (A) APN receptor from *Helicoverpa armigera* (B) TcaA toxin from *Photorhabdus luminescens*

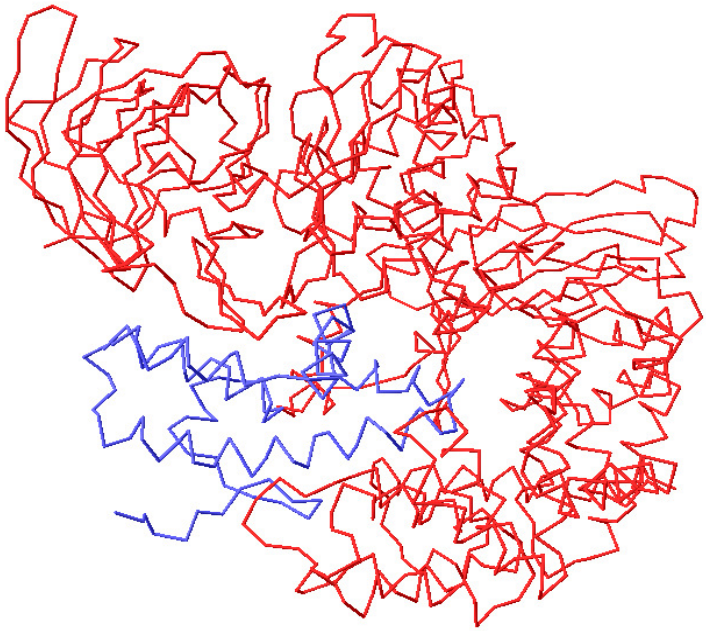


Figure-4
Docking of D-I of TcaA onto homology modeled aminopeptidase N from *Helicoverpa armigera* using HEX software (Total energy = –792.9 KJ / mol)

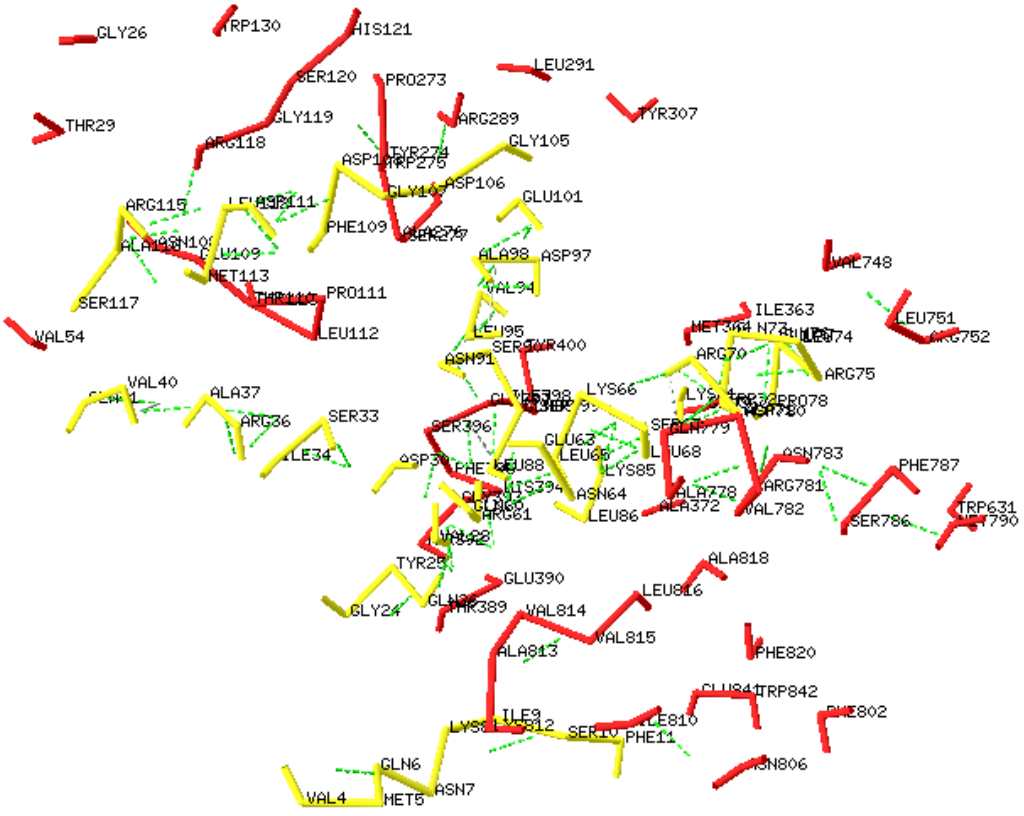


Figure-5
Amino acid residues of TcaA of *Photorhabdus luminescens* (Yellow) interacting with homology modeled aminopeptidase N receptor from *Helicoverpa armigera* (Red) within 5Å distance.

References

1. Roush R., Can we slow adaptation by pests to insect transgenic crops? In: Parsley GJ (ed) Biotechnology and integrated pest management. CAB International, Willinford. **(1996)**

2. Poinar G. O, Thomas G. M. and Hess R., Characteristics of the specific bacterium associated with Heterorhabditis bacteriophora (Heterorhabditidae: Rhabditida), *Nematologica.,* **23**, 97–102 **(1977)**

3. Wirth M. C., Georghiou G. P., and Federici, B. A., Cyt A enables CryIV endotoxins of BT to overcome high levels of CryIV resistance in the mosquito, Culex quinquefasciatus, *Proc. Natl. Acad. Sci,* **94,** 10536-10540. **(1997)**

4. Gould F. Anderson A. Jones A. Sumerford D. Heckel D. Lopez, J., et al. Initial frequency of alleles for resistance to Bacillus thuringiensis toxins in field populations of Heliothis virescens. *Proc. Natl. Acad. Sci,* **94**, 3519-3523**(1997)**

5. Liu Y. B. Tabashnik B. Dennehy T. Patin A. and Bartlet A., Development time and resistance to Bt crops, *Nature,* **400**, 519 **(1999)**

6. Frutos R. Rang C. and Royer M., Managing Insect Resistance to Plants Producing Bacillus thuringiensis Toxins, *Crit. Rev. Biotechnol,* **19(3)**, 227-276 **(1999)**

7. Tabashnik B., Delaying Insect Adaptation to Transgenic Plants: Seed Mixtures and Refugia Reconsidered, *Proc. R. Soc. Lond., 255***(1342),** 7-12 **(1994)**

8. Blackburn M. Golubeva E. Bowen D. and Ffrench-Constant R., A novel insecticidal toxin from photorhabdus luminescens, toxin complex a (Tca), and its histopathological effects on the midgut of manduca sexta, *Applied and environmental microbiology,* **64(8)**, 3036-3041**(1998)**

9. Han R. and Ehlers R. U., Effect of Photorhabdus luminescens phase variants on the in vivo and in vitro development and reproduction of the entomopathogenic nematodes Heterorhabditis bacteriophora and Steinernema carpocapsae, *FEMS microbiology ecology,* **35(3)**, 239-247 **(2001)**

10. Knight P. J. Crickmore N. Ellar D. J., The receptor for Bacillus thuringiensis CrylA(c) delta-endotoxin in the brush border membrane of the lepidopteran Manduca sexta is aminopeptidase N, *Mol. Microbiol* **11,** 429–436 **(1994)**

11. Francis B. R. and Bulla Jr L. A., Further characterization of BT-R1, the cadherin- like receptor for Cry1Ab toxin in tobacco hornworm (Manduca sexta) midgets, *Insect Biochem. Mol. Biol*, **27,** 541–550 **(1997)**

12. Knight P. J. Carroll J. and Ellar D. J., Analysis of glycan structures on the 120 kDa aminopeptidase N of Manduca sexta and their interactions with Bacillus thuringiensis Cry1Ac toxin, *Insect Biochem. Mol. Biol* **34,** 101–112 **(2004)**

13. Li J. D. Carroll J. and Ellar D. J. Crystal structure of insecticidal delta-endotoxin from Bacillus thuringiensis at 2.5 A resolution, *Nature*. 353, 815-821 **(1991)**

14. Galitsky N. Cody V. Wojtczak A. Ghosh D. Luft J. R. Pangborn W. and English L., Structure of the insecticidal bacterial delta-endotoxin Cry3Bb1 of Bacillus thuringiensis, *Acta crystallogr D Biol Crystallogr* **57,** 1101-1109 **(2001)**

15. Guo S. Ye S. Liu Y. Wei L. Xue J. Wu H. Song F. Zhang J. Wu X. Huang D. Rao Z., Crystal structure of Bacillus thuringiensis Cry8Ea1: An insecticidal toxin toxic to underground pests, the larvae of *Holotrichia parallela*, *J struct Biol* , **168,** 259-266 **(2009)**

16. Grochulski P. Masson L. Borisova S. Pusztai-Carey M. Schwartz J. L. Brousseau R. Cygler M., Bacillus thuringiensis CryIA(a) insecticidal toxin: crystal structure and channel formation, *J Mol Biol.* **254,** 447-464 **(1995)**

17. Boonserm P. Mo M. Angsuthanasombat C. and Lescar J., 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 **(2006)**

18. Boonserm P. Davis P. Ellar D. J. and Li J., Crystal structure of the mosquito-larvicidal toxin Cry4Ba and its biological implications, *J Mol Biol,* **348,** 363-382 **(2005)**

19. Morse R. J. Yamamoto T. and Stroud R. M., Structure of Cry2Aa suggests an unexpected receptor binding epitope, Structure 9, 409-417 (2001)

20. Thompson J. D. Higgins D. G. and Gibson T. J., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res* , **22,** 4673-4680 **(1994)**

21. Felsenstein J. *Inferring phylogeny*, Sinauer Associates, Sunderland, MA **(2003)**

22. Servant F.C. Bru E. Courcelle J. Gouzy D. Peyruc D. and Kahn D., ProDom: Automated clustering of homologous domains, *Briefings in Bioinformatic,* **3,** 246-251**(2002)**

23. Altschul S. F. Madden T. L. Schäffer1 A. A. Zhang J. Zhang Z. Miller W. and Lipman D. J., Gapped BLAST Nucleic Acids Res and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res*, **17,** 3389-3402 **(1997)**

24. Schwede T. Kopp J. Guex N. and Peitsch M. C., SWISS-MODEL: an automated protein homology-modeling serve,. *Nucl. Acids Res*, **31,** 3381-3385 **(2003)**

25. Zhang Y., I-TASSER server for protein 3D structure prediction, *BMC bioinformatics*, **9,** 40-47 **(2008)**

26. Colovos C. and Yeates T. O., Verification of protein structures: Patterns of nonbonded atomic interactions, *Protein Science,* **2,**1511-1519 **(1993)**

27. Wiederstein M. and Sippl M. J., ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Res* , **35**, W407-W410 **(2007)**

28. Laskowski R. A. MacArthur M. W. Moss D. and Thornton J. M., PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Cryst,* **26,** 283-291**(1993)**

29. Fiser A. Do R. K. and Sali A. Modeling of loops in protein structures. *Protein Science* **9,** 1753-1773**(2000)**

30. Holm L. and Rosenstrom P. Dali server: conservation mapping in 3D, *Nucl Acids Res* **38,** 545-549 **(2010)**

31. Ritchie D. W., Evaluation of protein docking predictions using Hex 3.1 in CAPRI rounds 1 and 2. Proteins, *Structure, Function, and Bioinformatics,* **52**, 98–106 **(2003)**

32. Nguyen T. T. Chang S. Evnouchidou I. York I. A. et al., Structural Basis For Antigenic Peptide Precursor Processing by the Endoplasmic Reticulum Aminopeptidase ERAP1, *Nat Struct Mol Biol.* **18**, 604-613 **(2011)**

33. Calhoun J. R. Liu W. Spiegel K. et al., Solution NMR Structure of a Designed Metalloprotein and Complementary Molecular Dynamics Refinement, *Structure* **16**, 210–215 **(2008)**

34. Iyer S. Holloway D. E. Kumar K. Shapiro R. and Acharya K. R., Molecular Recognition of Human Eosinophil-derived Neurotoxin (RNase 2) by Placental Ribonuclease Inhibitor, *J. Mol. Biol.* **347,** 637–655 **(2005)**

35. Spahr H. Samuelsson T, Hällberg B. M. and Gustafsson C. M., Structure of mitochondrial transcription termination factor 3 reveals a novel nucleic acid-binding domain, *Biochem. Biophys. Res. Commun.* **397,** 386–390 **(2010)**