Published in final edited form as:

Insect Biochem Mol Biol. 2008 July ; 38(7): 685–696. doi:10.1016/j.ibmb.2008.03.010.

Diversity of aminopeptidases, derived from four lepidopteran gene duplications, and polycalins expressed in the midgut of *Helicoverpa armigera*: Identification of proteins binding the δ -endotoxin, Cry1Ac of *Bacillus thuringiensis*

Constanza Angelucci^a, Gregory A. Barrett-Wilt^b, Donald F. Hunt^{c,d}, Raymond J. Akhurst^a, Peter D. East^a, Karl H.J. Gordon^a, and Peter M. Campbell^{a,*}

^aCSIRO Entomology, PO Box 1700, Canberra, ACT 2601, Australia

^bUniversity of Wisconsin Biotechnology Center, 425 Henry Mall, Madison, WI 53706, USA

^cDepartment of Chemistry, University of Virginia, Charlottesville, VA 22904, USA

^dDepartment of Pathology, University of Virginia, Charlottesville, VA 22904, USA

Abstract

Helicoverpa armigera midgut proteins that bind the *Bacillus thuringiensis* (Bt) δ -endotoxin Cry1Ac were purified by affinity chromatography. SDS-PAGE showed that several proteins were eluted with *N*-acetylgalactosamine and no further proteins were detected after elution with urea. Tandem mass spectral data for tryptic peptides initially indicated that the proteins resembled aminopeptidases (APNs) from other lepidopterans and cDNA sequences for seven APNs were isolated from *H. armigera* through a combination of cloning with primers derived from predicted peptide sequences and established EST libraries. Phylogenetic analysis showed lepidopteran APN genes in nine clades of which five were part of a lepidopteran-specific radiation. The Cry1Ac-binding proteins were then identified with four of the seven *HaAPN* genes. Three of those four APNs are likely orthologs of APNs characterised as Cry1Ac-binding proteins in other lepidopterans. The fourth Cry1Ac-binding APN has orthologs not previously identified as Cry1Ac-binding partners. The *HaAPN* genes were expressed predominantly in the midgut through larval development. Each showed consistent expression along the length of the midgut but five of the genes were expressed at levels about two orders of magnitude greater than the remaining two. The remaining mass spectral data identified sequences encoding polycalin proteins with multiple lipocalin-like domains. A polycalin has only been previously reported in another lepidopteran, *Bombyx mori*, but polycalins in both species are now linked with binding of Bt Cry toxins. This is the first report of hybrid, lipocalin-like domains in shorter polycalin sequences that are not present in the longest sequence. We propose that these hybrid domains are generated by alternative splicing of the mRNA.

Keywords

Bacillus thuringiensis; Cry1Ac; *Helicoverpa armigera*; *Bombyx mori*; Binding site; Aminopeptidase N; Polycalin; Alternative splicing

1. Introduction

The polyphagous, lepidopteran, agricultural pest *Helicoverpa armigera* is often controlled with the Cry1Ac toxic protein from *Bacillus thuringiensis* (*Bt*) in sprays and transgenic plants, especially the cotton crop. The propensity of field populations of *H. armigera* to develop resistance to chemical insecticides (Forrester et al., 1993), its capacity for resistance to Cry1A toxin (Akhurst et al., 2003), and the already evident field-developed resistance of *Plutella xylostella* (Tabashnik et al., 1990), has highlighted the need for resistance management when controlling field populations of *H. armigera*. Such precautions require detecting resistant insects in field populations, which can be handled efficiently by using resistance markers. In developing these markers an understanding of how resistance to the *Bt* toxins arises is paramount. In the majority of Cry-resistant insect species investigated, resistance has been linked to an alteration in the binding of *Bt* toxins to midgut epithelia, often correlated with reduced expression of a protein ligand (reviewed by Ferré and Van Rie, 2002). In *H. armigera*, Cry1Ac resistance was linked to an alteration in Cry1Ac binding to a cadherin-like protein (Akhurst et al., 2003; Xu and Wu, 2008). However, two isoenzymes of aminopeptidase (APN) from *H. armigera* interacted with Cry1Ac after expression in *Trichoplusia* cells. Both were membrane-associated, catalytically active, and glycosylated (Rajagopal et al., 2003). Furthermore, reducing the expression of one of those APNs reduced the cytotoxicity of Cry1Ac toward the cells demonstrating the potential for resistance to develop by down-regulation or modification of an APN (Sivakumar et al., 2007). However, no instance of *Bt* resistance has been linked with a loss of binding to an APN.

The mode of action of *Bt* Cry toxins was recently reviewed by Pigott and Ellar (2007). Multiple steps lead to cell lysis through the formation of pores in the brush border membrane of the midgut of susceptible insects. Prior to pore formation Cry toxins bind to receptor molecules, for which many different proteins have been identified. For example, transgenic expression of an APN from the lepidopteran, *Manduca sexta*, in the gut of *Drosophila melanogaster* caused the latter to become susceptible to Cry1Ac (Gill and Ellar, 2002). Binding partners for *Bt* Cry toxins in Lepidoptera have included the GPI-anchored molecules, aminopeptidase N (APN) and alkaline phosphatase (ALP), and cadherin-like molecules. Other binding partners for Cry toxins include a glyco-conjugate receptor, V-ATP synthase subunit and actin (Krishnamoorthy et al., 2007; Valaitis et al., 2001). Further binding investigations for both the cadherin-like and GPI-anchored proteins have led to proposals that both protein groups interact with Cry toxins to form pores. Models were suggested where binding commences with the cadherin, and after some toxin modification, binding to GPI-anchored receptors follows, with *N*-acetylgalactosamine (GalNAc) moieties regarded as essential for the binding of Cry1Ac to GPI-anchored proteins and in the formation of pores (Bravo et al., 2007; Gómez et al., 2007; Jurat-Fuentes and Adang, 2006). GalNAc inhibited 68% of binding to posterior midgut brush border membrane vesicles (BBMVs) from *H. armigera* but had only slight effect on the binding of Cry1Ac to anterior midgut BBMVs (Rodrigo-Simón et al., 2008). Nonetheless, GalNAc completely inhibited Cry1Ac-induced pore formation in BBMVs from both anterior and posterior regions of the midgut of *H. armigera* (Rodrigo-Simón et al., 2008). These results are consistent with multiple binding sites for Cry1Ac, some GalNAc mediated and some not, but add to the evidence that GalNAc-mediated binding plays a key role in pore formation.

Previously Cry1Ac affinity purification was used to isolate a set of proteins of about 100–160 kDa from the brush border membrane of *H. armigera* (Liao et al., 2005). Of these, one showed an N-terminal sequence most similar at the time to an APN from the lepidopteran, *M. sexta*, but the other proteins could not be identified clearly. It was notable that these binding proteins were eluted with GalNAc and that there was little material eluted subsequently with urea, a general denaturant. Here we have repeated the study using mass spectral analysis of peptides from affinity purified Cry1Ac-binding proteins to compare with an extensive set of midgut-

expressed genes from *H. armigera*. Full-length cDNA sequences from five Cry1Ac-binding proteins are reported herein.

2. Methods

2.1. Preparation of the Cry1Ac affinity matrix

Cry1Ac crystals were purified from *B. thuringiensis* var. *kurstaki* HD-73 by sucrose density gradients as described by Liao et al. (2002). Crystals were solubilised in 50 mM CAPS, 1 mM DTT, pH 10 and incubated for 1 h at 37°C. Solubilised protoxin was converted to the active form (~60 kDa) by digestion with trypsin (0.25 mg/ml) for 1 h at 37°C. The reaction was stopped with the addition of 1 mM PMSF, an irreversible inhibitor of serine proteinases. The residual trypsin (22 kDa) and small peptides produced during activation were removed by centrifugal ultrafiltration (Vivaspin 15 MWCO 30,000, Australasian Medical and Scientific Limited) at 2000g for 15–20 min. The buffer was replaced through ultrafiltration with coupling buffer (100 mM NaHCO₃, 0.5 M NaCl, pH 10). Activated Cry1Ac was coupled with cyanogen bromide (CNBr)-activated sepharose 4B matrix following the manufacturer's instructions (Amersham Biosciences). The Cry1Ac-coupled matrix was washed with Buffer III (50 mM Na₂CO₃, pH 8.5, 150 mM NaCl, 5 mM EGTA, 1.7 mM CHAPS) including proteinase inhibitors, PMSF (1 mM) and E-64 (10 µM), and disulphide reductant thioglycerol (9.5 mM). The matrix was stored in Buffer III containing 0.02% thiomersal.

2.2. Purification of Cry1Ac-binding proteins

The ANGR strain of *H. armigera* (Liao et al., 2002) was reared on the artificial diet of Teakle and Jensen (1985), with the exclusion of pollen and sunflower oil. Midgut tissue was dissected from fifth instars, washed in ice-cold MET buffer (0.3 M mannitol, 5 mM ethyleneglycol-tetraacetic acid (EGTA, Sigma), 17 mM Tris/HCl, pH 7.5), snap-frozen in liquid nitrogen, and stored at -80°C. Brush border membrane vesicles (BBMV) were prepared from the stored tissue by the method of Wolfersberger et al. (1987), and subsequently solubilised with CHAPS following the protocol of Liao et al. (2005), except that two of the protease inhibitors, pepstatin and 1,10-phenanthroline, were not included. The affinity protocol was essentially as described by Liao et al. (2005). The Cry1Ac-coupled matrix was washed with solution containing 50 mM Na₂CO₃, 150 mM NaCl, 5 mM EGTA, 1.7 mM CHAPS, 9.5 mM thioglycerol and the proteinase inhibitors, PMSF (1 mM) and E-64 (10 µM), pH 8.5. For storage 0.02% thiomersal was included. The affinity resin was incubated for about 20 h at 4°C with 11 mg solubilised BBMV in 3 ml, with rotation at 4 rpm. The matrix was washed again with washing solution above. The column was eluted with the washing solution sequentially containing 2 M KCl, 0.3 M GalNAc (two elutions which were pooled), 4 M urea, or 6 M urea.

The fractions were assayed for protein using the Bio-Rad Protein Assay Kit (Bradford, 1976). The APN and alkaline phosphatase assays were performed following the method of Liao et al. (2005). Alkaline phosphatase activity can be used as a marker for insect midgut membranes (Eguchi et al., 1990). Its absence from the affinity purified material was taken as an indication of purity (although its presence, had this occurred, would have been consistent with a report of alkaline phosphatase acting as a Cry1Ac receptor in *Heliothis virescens* (Krishnamoorthy et al., 2007)).

2.3. Identification of protein bands

Fractions from the affinity purification were separated by 10% SDS-PAGE and stained with Coomassie blue R250. Most of the width of each band was excised for in-gel digestion with trypsin and mass spectral analysis of the peptides as described previously (Martin et al., 2000) except that a Finnigan LCQ ion trap mass spectrometer was used. *De novo* sequence interpretations for at least six amino acid residues and searching of public domain sequence

databases were also as previously described (Martin et al., 2000). At the time of this analysis in 1999 there were only about 50 protein sequences available from *H. armigera* but it was obvious that at least some of the tryptic peptides resembled sequences from APNs of other lepidopterans. These peptide sequences helped to isolate full-length cDNA sequences of aminopeptidases from *H. armigera*, as described below.

The remaining SDS-PAGE gel was stored dry between cellophane at room temperature for eight years before we excised the remainder of each stained protein band in 2007. These were also processed by in-gel digestion and LC/MS analysis, this time with an Agilent XCT ion trap mass spectrometer as described elsewhere (Campbell et al., 2001, 2008). In the intervening period, a more extensive collection of EST and full-length cDNA sequences from the midgut of *H. armigera* was developed (below) against which we could compare both the earlier peptide sequence interpretations and the later ion trap MS peptide fragmentation data.

2.4. Predictions of post-translational modifications

Post-translational modifications were predicted using on-line bioinformatic software, SignalP 3.0, NetOGlyc, NetNGlyc and DGPI available through the ExPASy website (<http://kr.expasy.org/>) (Kronegg and Buloz, 1999; Bendtsen et al., 2004).

2.5. Cloning and sequencing of HaAPN cDNAs

The mRNA sequences of the *HaAPN* 1, 2, 3 and 4 genes were determined using mRNA and cDNA prepared for 3'- and 5'-Rapid Amplification of cDNA Ends (RACE) as previously described (Campbell et al., 2001). For 3'-RACE primers were designed from the peptide sequence reported in Liao et al. (2005), highly conserved regions of insect APNs and certain peptides revealed by *de novo* sequence interpretations of the mass spectral data described above, particularly where they resembled partial gene sequences available for *H. armigera* or near relatives. For 5'-RACE specific primers were designed from sequences obtained through 3'-RACE.

PCR products were gel purified using the Nucleotrap Nucleic Acid Purification Kit and ligated into the pGEM-T Easy vector (Promega) and transformed into DH5 *Escherichia coli*. PCR products were sequenced with BigDye Terminator (ABI PRISM) and then later a CEQ2000 (Beckman Coulter) was used for sequencing using Dye Terminator Cycle.

Further *HaAPN* cDNAs were cloned from the cDNA libraries described below. Clones containing aminopeptidase sequences were further sequenced with a CEQ™ 2000 (Dye Terminator Cycle Sequencing with Quick Start Kit) from Beckman Coulter. Sequencing reactions contained, in 10 µl total volume, 100–300 ng plasmid DNA, 2 µl Quick Start Master Mix, 2 µl Better Buffer (Microzone Ltd., West Sussex, UK), and 10 pmol reverse primer. All sequence traces were read using Phred (Ewing et al., 1998) to yield sequence and quality data in FASTA format. The quality scores were automatically evaluated using Perl scripts and all sequences with a minimum Phred20 score >100 collated. Sequences were then assembled using phrap and confirmed by hand.

All sequences were confirmed by repeated sequencing of full-length clones (or nearly full-length in the case of *HaAPN7*) from the libraries described below.

2.6. cDNA Libraries

Three cDNA libraries were generated from midgut tissue of *H. armigera* larvae. The first two cDNA libraries were constructed from mRNA prepared from midguts of second and early third instars; guts were dissected from both moulting and intermoult larvae. Random sequencing of cDNAs from library 1 yielded 1214 expressed sequence tags (ESTs), whereas library 2 was

generated by a combination of size- and subtraction-selection, yielding 3669 ESTs. The third library was derived from fifth-instar larvae; after size- and subtraction-selection of cDNA, it yielded 5229 ESTs. The total of 10,112 clones were then clustered with Stackpack (Christoffels et al., 2001) to yield 3917 unique sequences.

Of these, clones corresponding to 450 distinct sequences were selected for full-length sequencing on the basis of likely importance for midgut function following automated gene ontology annotation. To this 'in house' database of sequences we added all sequences from *H. armigera* and its near relatives, *Helicoverpa zea* and *H. virescens* that were available in public databases for a total of 4873 sequences. Sequences from the 'in-house' library have been deposited with Genbank if they were the better matches with peptide mass spectral data reported here than sequences already in the public domain.

2.7. Expression analysis of HaAPNs

The expression of four of the *HaAPN* genes was assessed through larval development from frozen, triplicate samples of midgut and carcass tissues from all instars, except first instars, which were used whole. cDNA was prepared as above. The Robust RT-PCR kit (FINNZYMES) was used for amplifying the *HaAPN* and *actin* cDNAs using primers specific to each of the genes.

The expression of seven *HaAPN* genes was assessed by quantitative real-time PCR (qRT-PCR). RNA samples (300 µg) were prepared from anterior, middle and posterior midgut sections from duplicate batches of ten fourth instars using TriReagent (Molecular Research Center) according to the manufacturer's instructions. Genomic DNA was removed with 1 unit of DNaseI per 1 µg of mRNA, which was then deactivated with EDTA and heating at 65 °C for 10 min according to the manufacturer's instructions (Invitrogen). RNA (500 ng) was then used directly in a reverse transcription reaction using an Invitrogen Superscript III RT Kit (a component of the SYBR GreenER 2-step qRT-PCR kit) following the manufacturer's instructions.

Primers for qRT-PCR were designed with Primer Express software to amplify a product of less than 150 bp. qRT-PCR was carried out using the SYBR GreenER 2-step qRT-PCR kit (Invitrogen) following the manufacturer's instructions and the following conditions on an ABI Prism 7000 Sequence Detection System: 1 cycle at 50 °C for 2 min; 1 cycle at 95 °C for 2 min; 40 cycles at 95 °C for 15 s and 60 °C for 1 min. Melting/dissociation analysis was then performed for 1 cycle at 60 °C. A pair of technical replicates was performed for each of the biological duplicates. A housekeeping gene, GAPDH, was used as an internal control to normalise RNA levels between the samples. GFP was used as negative control.

2.8. Genome searches and phylogenetic analysis

Various publicly available databases were searched for aminopeptidase sequences. Silkworm-specific databases searched included the two partially assembled genomes available for *B. mori*, the SilkDB from the Beijing Genomics Institute in China (<http://silkworm.genomics.org.cn/index.jsp>, searched on 30/6/05) and KAIKObase from the National Institute of Agrobiological Sciences in Japan (<http://sgp.dna.affrc.go.jp/>, release date 10/5/04), plus the EST libraries at SilkBase (<http://papilio.ab.a.u-tokyo.ac.jp/silkbase/index.html>, release date 10/11/03). Other insect specific databases searched included the proteins predicted from the assembled genomes of *Apis mellifera*, *Tribolium castaneum*, and *Drosophila melanogaster*. The non-redundant protein databases at NCBI (<http://www.ncbi.nlm.nih.gov/>) were also searched. Searches were conducted using BLAST, TBLASTN and BLASTP as appropriate, as well as using perl scripts to identify proteins containing selected Interpro motifs. Multiple sequences were aligned using

the default parameters of ClustalW. From these alignments, phylogenetic analysis was conducted using several methods, including the neighbour-joining method with 1000 bootstrap trials in Clustal X (Thompson et al., 1997).

3. Results

After washing the Cry1Ac-affinity column with a low salt buffer, a further 24 µg of non-specifically bound protein was eluted with 2 M KCl (Fig. 1). The column was then eluted twice with GalNAc yielding 69 and 11 µg of protein, respectively. A total of 9 µg of protein was eluted with 4 M urea, but no further protein was detected with elution by 6 M urea. The GalNAc and 4 M urea eluates were concentrated by lyophilisation and were analysed by SDS-PAGE, and the resultant gel image closely resembled results presented previously (Liao et al., 2005). No protein was detected in the 2 M KCl and 6 M urea eluates. Six bands were observed in the GalNAc eluate with estimated molecular weights of 180, 130, 121, 116, 109, and 94 kDa. A doublet band of ~65 kDa in this sample was confirmed as Cry1Ac by Western blot and peptide mass spectrometry (not shown). The Cry1Ac was not observed as a doublet prior to coupling (not shown) suggesting that Cry1Ac had undergone slight digestion/modification during or after coupling to the affinity matrix.

The specific activity of APN in the GalNAc fraction increased 6.9-fold from the homogenised midgut and 2-fold from the BBMVs, whereas alkaline phosphatase activity was at the limit of detection in the GalNAc fraction (Table 1). This indicates that APN activity but not alkaline phosphatase co-purified with Cry1Ac-binding proteins.

Mass spectral data for tryptic peptides from the six GalNAc-eluted bands were first screened to remove from consideration those with matches to obvious contaminants such as human keratin. From the remaining data some peptides matched sequences of aminopeptidases previously identified as Cry1Ac-binding proteins from either *H. virescens* (Banks et al., 2001; Luo et al., 1997) or *M. sexta* (Knight et al., 1994). All the high-quality mass spectral data were used to generate *de novo* sequence interpretations of six residues or more. Thus each of the six bands yielded between nine and 43 peptide sequences. When all the then-available (1999) lepidopteran APN sequences were aligned it was clear, firstly, that they segregated into distinct phylogenetic clusters, and secondly, that most of the peptide sequences were similar to APN sequences from other lepidopterans.

The peptide sequences were then used to aid in the cloning and sequencing of four full-length APN cDNAs from *H. armigera*, *HaAPN1* (AY038607), *HaAPN2* (AY038608), *HaAPN3* (AF535166) and *HaAPN4* (AF535165). All four sequences predict motifs present in other lepidopteran APNs (Pigott and Ellar, 2007): secretion signal peptides, a GPI cleavage site and a hydrophobic tail, indicating that they would be retained on the cell membrane, the zinc-binding/gluzincin motif, HEX₂HX₁₈E, required by APN proteins for enzymatic function, and the gluzincin APN motif GAMENWG, also believed to form part of the active site. *HaAPN4* differed from the other three by the lack a polythreonine (mucin) region upstream of the GPI prepeptide anchor.

Three further APN sequences were identified from our cDNA libraries, *HaAPN5* (EU325551), *HaAPN6* (EU328183) and *HaAPN7* (EU328182). While *HaAPN5* and *HaAPN6* have the HEX₂HX₁₈E and GAMENWG motifs, *HaAPN7* does not, suggesting that it is not an active APN (Fig. 2). All three contain a putative secretion signal peptide and are likely to have a GPI anchor.

Reinspection of the *de novo* peptide sequence data more recently (2007) showed that all but 14 of the 174 peptide sequences matched with HaAPNs 1, 2, 3 or 4, but no peptides matched HaAPNs 5, 6 or 7. An alignment of the *H. armigera* APN sequences, with the matched peptides

underlined, shows that each peptide unambiguously identifies only one of the APN sequences (Fig. 2). All peptides were exact matches with cDNA sequences as shown (albeit leucine/isoleucine is ambiguous in the mass spectral data). In addition there was one peptide from each of APNs 1, 3 and 4 that showed one or two generally conservative amino acid differences (not shown).

Finding some previously unknown APNs in the *H. armigera* set led us to ask whether these corresponded to a complete set of genes likely to be present in the genomes of Lepidoptera. The only lepidopteran for which extensive genome sequence data is available is *B. mori* (Mita et al., 2004; Xia et al., 2004). Four APNs from *B. mori* have been reported previously (Entrez gene IDs: *BmAPN1*, 692563; *BmAPN2*, 692370; *BmAPN3*, 100127057; *BmAPN4*, 100127099), in contrast to the seven from *H. armigera* reported here. Since the genome of *B. mori* is not currently available in a completely assembled form, we searched genomic contigs and EST sequences. *BmAPN1* and *BmAPN4* were found together on contig AADK01000258 but the other two known *BmAPNs* were each found on different contigs. We then identified three further *B. mori* sequences orthologous with the sequences from *H. armigera*. A sequence most closely similar to *HaAPN2* was identified on approximately 11,000 nucleotides at the 3' end of contig AADK01003974 and it was supported by several ESTs (BY920762, CK495040, CK495048, CK494950, BY929236, BY915618). Although the sequence is incomplete due to the incomplete nature of the EST and cDNA support, it clearly corresponds to a new APN termed *BmAPN5*. A further APN, termed *BmAPN6*, was identified on contig AADK01039313. This sequence was incomplete (by comparison to its closest relative in the *H. armigera* set, *HaAPN6*) but the remaining portion was located on the 5' portion of contig AADK01003974. An ortholog of *HaAPN7*, *BmAPN7*, is encoded on contigs AADK01003439 (Bmb020723) and AADK01002744 (Bmb021898). Like *HaAPN7*, *BmAPN7* appears to lack crucial catalytic residues. Finally, an APN sequence not closely similar to others, Bmb025220, comprises the following peptides predicted from the genome (in this order): Bmb025220, Bmb042445 and Bmb037589 (Xia et al., 2004). It was confirmed by the following ESTs in Genbank: BB983452 BB991210 BB985776, BP117449, BY931685, BY929985, BY939134, BY923137.

Nine distinct phylogenetic clusters of lepidopteran APN genes were identified (Fig. 3). Of these, five formed a lepidopteran-specific radiation comprising the clades termed Classes I, II, III, IV and VI. Classes I, III, IV and V correspond with four of the lepidopteran clades reported previously (Nakanishi et al., 2002; Herrero et al., 2005; Wang et al., 2005; Pigott and Ellar, 2007). Classes II, VI and VII are noteworthy because they each contain a single APN from *H. armigera* and *B. mori* with only one other sequence from *P. xylostella* in class II.

For Bmb025220, no orthologue has yet been identified from *H. armigera* or any other lepidopteran insect, but its closest related proteins were leucyl aminopeptidases, for example, CG7340 from *Drosophila* and PEL1_HUMAN. Interestingly this new protein is more similar to these canonical leucyl aminopeptidases than is the earlier described *BmAPN-L* (Koike et al., 2003).

The genes encoding the four Cry1Ac-binding APNs were the most highly expressed, at similar levels to each other and across all instars, predominantly in the midgut tissue. Only slight expression was detected in the remaining tissues, consistent with the enzymes' probable function in the digestion of dietary protein (Fig. 4). *HaAPN6* was expressed at slightly lower levels than *HaAPN1-4* but *HaAPN5* and *HaAPN7* were expressed at about two orders of magnitude lower level. All seven genes showed only slight variation of their expression levels along the length of the midgut (Fig. 4).

All but two of the 14 *de novo* peptide sequences that were not matched with *HaAPNs* 1, 2, 3, or 4 were matched with a set of three closely related, full-length 'polycalin' cDNA sequences

from our libraries (Fig. 5). ‘Polycalin’ is a term coined by Mauchamp et al. (2006) from a single example in *B. mori* to describe proteins with multiple, lipocalin-like domains. Three full-length cDNAs from *H. armigera* predicted polycalins of different lengths: 570 (EU325567), 747 (EU325566) and 927 (EU325561) amino acids. The N- and C-terminal regions of the predicted proteins are nearly identical (>98%, Fig. 5) including, respectively, at their extreme ends, a predicted signal peptide and GPI anchor attachment and cleavage sites. No polythreonine, mucin-like sequence is present but there are several potential sites for N- and O-linked glycosylation (Fig. 5). The tryptic peptide matches are all from the regions of shared, nearly identical sequence so we can not claim greater confidence for identification with any one of these cDNA sequences. However, the longest sequence predicts a protein with five lipocalin domains, which is closest to the observed size (~100 kDa). The highly similar regions across the three polycalin cDNA sequences suggest that they derive from a single gene with the three forms generated by alternative splicing (Fig. 5). One of the putative polymorphic differences within the shared C-terminal region was observed by a pair of tryptic peptides that differed by only one amino acid (Fig. 5).

In addition to the *de novo* peptide sequences we later recovered tryptic peptides from the GalNAc-eluted bands for analysis with another LC/ion trap MS instrument. These later data identified an overlapping set of peptides and extended the sequence coverage by two or three peptides for each of HaAPN3, HaAPN4 and polycalin but did not identify any gene products that had not been identified already. Taking all the data together, four HaAPNs were each identified by between 19 and 25 distinct peptides giving 24–37% sequence coverage. Polycalin sequences were matched by ten distinct peptides giving between 12 and 19% sequence coverage.

A notable feature of the data is that peptides matching particular gene sequences were not in any case restricted to only one of the bands. However, consideration of the numbers of peptides and their ion current intensities (not shown) allowed us to identify the major components in each band (Fig. 1). The 180 kDa band contained similar amounts of HaAPN1 and HaAPN2 but none of the other proteins were detected. The 130 kDa band contained predominantly HaAPN3. The 121, 116 and 109 kDa bands were predominantly HaAPN4 with traces of the other APNs. The 109 band also had traces of polycalin while the 94 kDa band was predominantly polycalin with traces of all the APNs. It is not unusual to find cross-contamination from adjacent bands, but in this case it is also likely that the APNs are distributed across the bands due to heterogeneous post-translational modifications such as glycosylation (Oltean et al., 1999). All the APN sequences predict proteins of about 110 kDa so the observed gel mobilities are consistent with variable additions of glycan. In another study (Campbell et al., 2008) we identified HaAPNs from SDS-PAGE bands after a chemical deglycosylation treatment (Edge, 2003). The APNs migrated in the same order (HaAPN1 and HaAPN2 comigrated, HaAPN3 migrated a little further and HaAPN4 was the furthest) but all migrated a little further than a 97 kDa marker protein.

4. Discussion

Cry1Ac-binding proteins were affinity purified from the midgut brush border membrane of *H. armigera* and identified as the products of four distinct APN genes and a polycalin gene. Previously two APN genes, *HaAPN1* and *HaAPN3* (*HaAPN3* here corresponds with *HaAPN2* of Rajagopal et al., 2003) from *H. armigera* had been expressed in *Trichoplusia* cells and shown to bind Cry1Ac by ligand blot (Rajagopal et al., 2003). Cells expressing HaAPN1 showed aberrant morphology when exposed to Cry1Ac that was less severe when the expression of HaAPN1 had been reduced by RNAi (Sivakumar et al., 2007). A previous study employing the purification technique employed here had identified only one of the proteins, HaAPN1, by its N-terminal sequence (Liao et al., 2005). The proteins were identified by

comparing the mass spectra of tryptic digest peptides with a database of genes expressed in the midgut of *H. armigera*. The polycalin and four APNs reported here were also recovered from the peritrophic matrix (PM) of *H. armigera* (Campbell et al., 2008) consistent with observations that the PM of *H. armigera* is a major site of *Bt* toxin binding (Rodrigo-Simón et al., 2006).

Previous phylogenetic analyses identified four or five sets of orthologous lepidopteran APN genes (Herrero et al., 2005; Wang et al., 2005; Pigott and Ellar, 2007) but additional cDNA and genomic sequence data from *Bombyx mori* and *H. armigera* have extended that to seven sets (Fig. 3). The products of four of the seven APN genes from *H. armigera*, HaAPNs 1–4, were purified by Cry1Ac affinity in this study. HaAPNs 1, 3 and 4 each have orthologues in other lepidopteran species that have been identified as Cry1Ac-binding proteins. Examples of these purified APNs include *M. sexta* X89081 (Knight et al., 1995) and *H. virescens* AF173552 (Oltean et al., 1999) orthologous with HaAPN1 in Class I; *H. virescens* U35096 (Gill et al., 1995) orthologous with HaAPN3 in Class III, and *H. virescens* AF378666 (Banks et al., 2001) orthologous with HaAPN4 in Class IV. On the other hand, neither of the known orthologues of HaAPN2, *P. xylostella* (AJ222699) or *B. mori* (Bm01003974b) (Class II), are established Cry1Ac-binding proteins. Nakanishi et al. (2002) reported equivocal binding of the *P. xylostella* APN to Cry1Aa and Cry1Ab (on a ligand blot but not BBMV). Therefore, HaAPN2 represents a putative new class of APNs that may bind Cry toxin.

While five Cry1Ac-binding proteins were isolated from *H. armigera* in this study, *in situ* studies predicted only two binding sites (Akhurst et al., 2003; Liao et al., 1996). These *in situ* studies were conducted with BBMVs, which, without the membrane solubilisation step, might better reflect *in vivo* binding whereas the treatment of BBMVs with CHAPS as part of the affinity purification process could have exposed binding epitopes that are concealed *in vivo*. Multiple Cry1Ac-binding proteins have also been isolated in purification studies with other lepidopterans for which fewer binding sites had been predicted by *in situ* binding studies, for example, *H. virescens* (Gill et al., 1995; Jurat-Fuentes and Adang, 2001; MacIntosh et al., 1991), *M. sexta* (Garczynski et al., 1991; Herrero et al., 2001; Keeton and Bulla, 1997; Knight et al., 1994; Sangadala et al., 1994), and *Epiphyas postvittana* (Simpson and Newcomb, 2000; Simpson et al., 1997). The greater number of binding proteins might be reconciled with fewer predicted binding sites if some of the proteins, or parts of the proteins, are equivalent as ligands. As discussed below, one possibility is that the mucin-like regions of some APNs might be one type of site while a GalNAc-modified GPI anchor might be another type of site.

Finding the four *H. armigera* Cry1Ac-binding APNs distributed through multiple bands is most likely due to variable glycosylation although other post-translational modifications or simple degradation cannot be discounted. Certainly glycosylation of other lepidopteran APNs is demonstrated (Knight et al., 2004; Sangadala et al., 2001), toxin-binding forms of an *H. virescens* APN differed by about 40 kDa (Oltean et al., 1999), and we observed distinctly reduced apparent molecular weights for the four *H. armigera* APNs when they were exposed to a deglycosylation treatment that would remove both N- and O-linked glycan (Edge, 2003; Campbell et al., 2008).

Glycosylation is generally regarded as an important determinant of Cry1Ac-binding (Pigott and Ellar, 2007). Therefore it is likely the Cry1Ac-binding HaAPNs all share the minimum amount of necessary GalNAc-containing glycan for interaction with activated Cry1Ac toxin since the toxin is known to bind to GalNAc residues and the proteins were eluted with GalNAc. HaAPNs 1, 2 and 3 have 39, 32 and 14 predicted O-linked sites, respectively, in mucin-like polythreonine sequences close to their GPI membrane anchors at the C-terminus (Fig. 2), a combination that is well supported as important for Cry1Ac binding (Oltean et al., 1999; Knight et al., 2004). Consistent with the probable role of the mucin-like sequences is the observation

that HaAPN2 is the only protein from the Class II clade with a predicted mucin-like region and it is the only one that has been isolated as a Cry1Ac-binding protein. The three HaAPNs that were not recovered in this experiment, HaAPNs 5, 6 and 7, are likely to be GPI anchored but lack the mucin-like sequence (Fig. 2).

In contrast with HaAPNs 1, 2, and 3, HaAPN4 has no predicted O-linked sites. Current understanding of glycosylation suggests that GalNAc moieties are generally O-linked and rarely N-linked in eukaryotes (Knight et al., 2004; Stephens et al., 2004). Thus, while it is reasonable to assume that HaAPNs 1, 2 and 3 were bound through their O-linked glycan, it is not obvious how HaAPN4 was isolated through GalNAc elution without any predicted O-linked glycan. However, this does tend to discount the mucin-like sequence from being uniquely required for Cry1Ac binding. Consequently we have considered four alternative mechanisms by which HaAPN4 might have bound to the affinity column. Firstly, HaAPN4 might bind Cry1Ac directly at a site that overlaps where Cry1Ac would bind GalNAc, and consequently binding is competitively displaced by GalNAc. Secondly, HaAPN4 might be bound indirectly through an association with one or more of the other Cry1Ac-binding proteins. Thirdly, HaAPN4 has the greatest number of N-linked sites (8) among the HaAPNs and perhaps these are unconventionally glycosylated to include GalNAc. Lectin-binding studies appear to show the presence of terminal GalNAc on the N-linked glycan of alkaline phosphatase and other BBMV proteins in *H. virescens* (Wu et al., 1997; Jurat-Fuentes and Adang, 2004, 2007). Also, while APN1 (AMPM, Q11001) from *M. sexta*, contains unusual, highly fucosylated, N-linked glycans (probably without GalNAc), GalNAc can occur on insect N-glycans since it has been found in these structures on honeybee PLA₂ (Stephens et al., 2004). On the other hand, contrary to our findings, the HaAPN4 orthologue, *H. virescens* AF378666 bound Cry1Ac independent of glycosylation (Banks et al., 2001). Finally, it is possible that GPI anchors are modified with GalNAc residues (Orlean and Menon, 2007) and it may be through this structure that the HaAPN4 protein binds Cry1Ac.

Current theories of the mode of action of the Cry1Ac suggest that the toxin binds to GPI-anchored proteins prior to insertion into the membrane but it is unclear whether the anchor has any role in that binding (Bravo et al., 2007; Gómez et al., 2007; Pigott and Ellar, 2007). In at least one case the GPI-anchor is not required for binding since an APN from *M. sexta* was still able to bind Cry1Ac after cleavage of its anchor (Lu and Adang, 1996), however, that particular APN (from class I) has the C-terminal, mucin-like sequence with demonstrated GalNAc content (Knight et al., 2004). All the proteins isolated in this study have predicted GPI anchor sites but only some have the mucin-like domain. These data might be reconciled by the suggestion that the two classes of binding sites identified kinetically (Liao et al., 1996; Akhurst et al., 2003) might correspond with GalNAc contained in mucin-like domains and GalNAc-modified GPI-anchors rather than two particular proteins. If this is correct then either of these features might be sufficient to enable an APN to bind Cry1Ac under the conditions reported here.

H. armigera mRNA (HaAPNs 1–4) was present in midguts at similar levels throughout larval development (Fig. 4A). mRNA corresponding to HaAPNs 1–4 and 6 were present at similar levels from anterior to posterior midgut, whereas HaAPNs 5 and 7 were present at significantly lower levels (Fig. 4B). While we observed little difference in HaAPN mRNA levels along the midgut, Rodrigo-Simón et al. (2008) found APN activity was about 3-fold greater in BBMVs prepared from the posterior portion of the midgut of *H. armigera* than the anterior portion. Our results are also not completely congruous with a near relative, *Helicoverpa punctigera*, in which *hpapn1*, *hpapn2* and *hpapn3* (homologues, respectively, of HaAPN1, HaAPN4 and HaAPN3) showed faint expression for the latter two genes in neonates, while *hpapn1* was expressed more abundantly (Emmerling et al., 2001). Through the second and third instars expression for *hpapn1* increased while remaining low for *hpapn2* and 3. The authors suggested

that the differing expression levels between the three *hpapn* genes was an adaptation to a particular diet type. Two of the APNs from *H. armigera* showed some difference of specificity with artificial substrates (Rajagopal et al., 2003) and presumably heterogeneity among APNs would be an advantage for digesting peptides of diverse sequences as one expects to derive from most diets. However, we are unaware of any study that shows the expression of APNs responding to dietary differences in the manner of serine proteinases whose expression levels respond to inhibitor-containing diets (Duncan et al., 2006).

The identification of the *B. mori* APN genes on genome sequence contigs also provided evidence for their juxtaposition on the genome. Two APN genes (or fragments thereof) were identified on each of contigs AADK01000258 (*BmAPN1* and *BmAPN4*) and AADK01003974 (*BmAPN5* and partial *BmAPN6*). These pairs of adjacent genes are consistent with the phylogeny of the genes, and supports clustering of the APN genes in the *B. mori* genome. In view of the phylogenetic analysis (Fig. 3), including the evidence that the class II and VI clades are common to *H. armigera* and *B. mori*, this gene duplication is likely to be present in all lepidoptera. It is therefore likely that all the APN genes are clustered as a result of multiple gene duplications during evolution of the lepidopteran ancestor. The phylogenetic analysis suggests that the earliest duplication within the lepidopteran-specific radiation yielded the ancestor of Classes I and III and the ancestor of Classes II, IV and VI. The duplications between Classes I and III, and II and VI would be more recent. The existence of such a gene cluster common to phylogenetically distant lepidoptera has been confirmed by analysis of genomic DNA fragments cloned in BACs for *H. armigera* and *Spodoptera frugiperda* (Fournier, Feyereisen et al., unpublished) and in a more complete *B. mori* genome assembly (Xia, Mita et al., unpublished).

No cadherin-like protein was observed in this study but Liao et al. (2005) did note high-molecular-weight material (>200 kDa) from BBMV that bound to the toxin on a ligand blot. Xu and Wu (2008) also show a very similar banding pattern with biotinylated Cry1Ac on a ligand blot of BBMVs from *H. armigera*. In the latter case the high-molecular-weight band is absent in a Cry1Ac resistant strain (GYBT) in which the cadherin-like *Ha_BtR* gene is disrupted. Lower-molecular-weight bands in the region of the HaAPNs and polycalin were also reduced in the resistant strain (Xu and Wu, 2008).

It is not clear why Cry1Ac-binding proteins were only recovered with GalNAc elution and not with the urea elution (Fig. 1). Elution with urea might have been expected to yield the known cadherin-like protein at least (Ihara et al., 1998). Such proteins might not have been observed if they were much less abundant than the observed proteins, more susceptible to proteolysis (in spite of our inclusion of proteinase inhibitors), or remained tightly bound in spite of urea. It is notable that GalNAc had little effect on the binding of Cry1Ac to BBMVs prepared from the anterior portion of the midgut of *H. armigera* suggesting a significant population of non-GalNAc-mediated binding sites, but GalNAc-mediated binding dominated for BBMVs from the posterior of the midgut (Rodrigo-Simón et al., 2008).

Only once before has a protein with more than two repeated lipocalin domains been described. That protein from *B. mori* had 15 lipocalin domains and the term polycalin was coined. This polycalin was characterised as a binding protein for chlorophyllid-A, the prosthetic group from chlorophyll (Mauchamp et al., 2006). The polycalin sequence of *B. mori* was subsequently matched with two peptides from a previously unidentified, very high-molecular-weight protein that bound Cry1Aa, Cry1Ab and Cry1Ac in heterologous competition binding studies (Hossain et al., 2004; Hossain et al., 2005; Pigott and Ellar, 2007).

Similarly, Malik et al. (2001) found proteins from BBMVs of *H. armigera* that showed Cry1Ac binding on ligand blots using classical chromatography. Their preparation was enriched for

aminopeptidase activity and major protein bands were observed at about 70 and 120 kDa. The 120 kDa band yielded an N-terminal sequence VIQTGQCDQSIADVVTNFNLSA which they attributed to a novel aminopeptidase. However, this sequence is nearly identical with residues 21–41 of the polycalin sequences reported here and consistent with the predicted signal peptide (Fig. 5). Thus polycalins appear to be Cry1Ac-binding proteins in both of the species in which they have been reported to occur.

Differential lectin binding of *B. mori* polycalin suggests that it may have both N- and O-linked GalNAc, including terminal GalNAc, yet binding by Cry1Ac was not inhibited by GalNAc (Hossain et al., 2004), a result that contrasts with purification method reported here for the protein from *H. armigera*. The *H. armigera* polycalin has two Asn-Xaa-Ser/Thr sequons that may carry N-linked glycan, five predicted O-linked sites and a predicted GPI anchor site but a mucin-like region is not predicted. However, like the polycalin from *B. mori*, a mucin-like domain cannot be required for binding by the *H. armigera* polycalin and alternative binding mechanisms may apply such as were discussed for HaAPN4.

The peptide data for the polycalin obtained in this study did not distinguish between three cDNA sequences. These sequences appear to be derived from a single gene by alternative splicing and they clearly encode proteins with multiple, non-identical lipocalin-like domains (Fig. 5). The consensus of the lipocalin-like sequences reported here shows 41–55% identity with each of the 15 lipocalin-like domains of the polycalin from the gut of the lepidopteran *B. mori* (Mauchamp et al., 2006). In *B. mori* only one cDNA species was observed and it included all 15 of the lipocalin-like sequences from the genomic gene sequence. In *H. armigera* the observed mobility on SDS-PAGE of the affinity purified polycalin was consistent with the predicted size of the longest, pentacalin sequence, and there were no peptides that were unique matches with either of the shorter cDNA sequences. However, polycalin-derived peptides were recovered from SDS-PAGE of an extract of the peritrophic matrix across a very wide size range (retained in the first millimetre of the gel and gel slices in the range 40–150 kDa, Campbell et al., 2008).

Alignment of the predicted translations of the three *H. armigera* polycalin cDNA sequences suggests that the most N-terminal and C-terminal sequences are constant but the central regions of sequences vary (Fig. 5). A simple form of alternative splicing might have involved the variable inclusion or omission of sequences for complete lipocalin domains from an array of lipocalin sequences in the genome. However, here it appears that alternative splicing yields hybrid lipocalin domains. For example, the first lipocalin domain of the tricalin is not the same as any of the domains of the other polycalins. Instead it is identical towards the N-terminus with the first lipocalin domains of the other polycalins but its C-terminal sequence is identical with the C-terminal portion of the third domain of the pentacalin. In other words, this first domain of the tricalin appears as a hybrid between the first and third domain of the pentacalin.

The genomic sequence of the polycalin gene is not available for *H. armigera* but this proposed alternative splicing model is consistent with the genomic structure in *B. mori* (Mauchamp et al., 2006). The mRNA sequence of the polycalin of *B. mori* spans several contigs in the partially assembled bombyx genome, but there is no evidence for the existence of more than a single copy of this gene (Mita et al., 2004; Xia et al., 2004). Mauchamp et al. (2006) did not observe any differential splicing in *B. mori* but recognised the potential for this to occur and called for further investigation. In *B. mori* each of the polycalin's fifteen lipocalin domains is encoded by three exons of consistent lengths. It is plausible then that alternative splicing could generate complete, novel lipocalin domains by joining the first exon of one lipocalin sequence with the second exon of another lipocalin sequence from further downstream, or similarly, a second exon could be joined to a third exon from further downstream. Several additional cDNA sequences (for example, EE399911) suggest that the putative, single, homologous gene in *H.*

armigera contains more lipocalin-like sequences than are shown in Fig. 5 but they were not included in the analysis because they were clearly not full-length and unconfirmed by repeated sequencing.

Each of the predicted lipocalin domains from the three full-length polycalin sequences show two pairs of conserved cysteine residues and three structurally conserved regions that are typical of lipocalins (Kayser, 2005; Ganfornina et al., 2006). These lipocalin domains are most similar to a radiation of lipocalins in the Lepidoptera that include the well-characterised bilin-binding proteins (Ganfornina et al., 2006). The canonical role of lipocalins is to bind small lipophilic molecules in the barrel-shaped interior of the protein, often with high specificity. A function of the protein in *B. mori* appears to be binding of chlorophyllid A which is a tetrapyrrole-like bilin (Mauchamp et al., 2006). This might have functions in defence against viruses or oxidative damage (Campbell et al., 2008). However, the proposed alternative splicing might enhance the diversity of binding specificities among the lipocalin domains. The proteins might then protect the midgut from diverse toxins (Campbell et al., 2008).

In summary, the *B. thuringiensis* Cry1Ac toxin bound five proteins from the brush border membrane of the midgut of *H. armigera* in a non-denaturing affinity column and these were eluted with GalNAc. Three of the proteins (HaAPN1, 3 and 4) are from clades of lepidopteran APNs with previously identified Cry1Ac-binding proteins (Fig. 3, Pigott and Ellar, 2007). A fourth Cry1Ac-binding APN from *H. armigera* (HaAPN2) is from a clade sparsely populated with known sequences but at least one other of these may be a Cry-binding protein (Nakanishi et al., 2002).

H. armigera polycalin is most similar to the only other known polycalin, from *B. mori*, and both were identified as Cry toxin-binding proteins before their cDNA sequences were obtained and characterised (Malik et al., 2001; Pigott and Ellar 2007; this paper). Thus class II APNs and the polycalins most likely represent two new classes of *Bt* toxin-binding proteins. It is noteworthy that the four APNs and polycalin reported here were also recovered from the PM of *H. armigera* (Campbell et al., 2008) consistent with observations that the PM is a major site Cry1Ac toxin binding in midguts of *H. armigera* (Rodrigo-Simón et al., 2006). Heterologous expression of the five proteins, or parts thereof, may assist in an investigation of the roles of GalNAc moieties, mucin regions, GPI anchors and any other regions in the binding of toxin. The functional relevance of these five proteins for Cry1Ac toxicity in *H. armigera* could be assessed with techniques such as RNA interference or subtraction cloning with a Cry1Ac-resistant strain of *H. armigera*, while bearing in mind the possibility that the proteins act in concert.

Acknowledgments

We thank Drs. Chunyan Liao, and Peter Hughes for assistance. This work was supported by a Cotton Research and Development Corporation postgraduate scholarship (to CA) the National Institutes of Health Grant, GM 37537 (to D.F.H.). We also thank Anh Cao, Sri Sriskantha, Michelle Williams, Heather Domasch, Mira Dumancic and Eva Zinkovsky for the creation of our database of *H. armigera* sequences, and Anh Cao also for performing the Q-RT-PCR. Early development of the sequence collection was funded by Syngenta and later through the Grain Protection Genes joint agreement between CSIRO and the Grains Research and Development Corporation.

References

- Akhurst RJ, James W, Bird LJ, Beard C. Resistance to the Cry1Ac delta-endotoxin of *Bacillus thuringiensis* in the cotton bollworm, *Helicoverpa armigera* (Lepidoptera: Noctuidae). *J Econ Entomol* 2003;96:1290–1299. [PubMed: 14503603]
- Banks DJ, Jurat-Fuentes JL, Dean DH, Adang MJ. *Bacillus thuringiensis* Cry1Ac and Cry1Fa delta-endotoxin binding to a novel 110 kDa aminopeptidase in *Heliothis virescens* is not *N*-acetylgalactosamine mediated. *Insect Biochem Mol Biol* 2001;31:909–918. [PubMed: 11439250]

- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004;340:783–795. [PubMed: 15223320]
- Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal Biochem* 1976;72:248–254. [PubMed: 942051]
- Bravo A, Gill SS, Soberón M. Mode of action of *Bacillus thuringiensis* Cry and Cyt toxins and their potential for insect control. *Toxicon* 2007;49:423–435. [PubMed: 17198720]
- Campbell PM, Harcourt RL, Crone EJ, Claudianos C, Hammock BD, Russell RJ, Oakeshott JG. Identification of a juvenile hormone esterase gene by matching its peptide mass fingerprint with a sequence from the *Drosophila* genome project. *Insect Biochem Mol Biol* 2001;31:513–520. [PubMed: 11267890]
- Campbell PM, East PD, Gordon KHJ. Proteomic analysis of peritrophic matrix from the gut of the caterpillar, *Helicoverpa armigera*. *Insect Biochem Mol Biol*. 2008 accepted subject to revision.
- Christoffels A, van Gelder A, Greyling G, Miller R, Hide T, Hide W. STACK: sequence tag alignment and consensus knowledgebase. *Nucleic Acids Res* 2001;29:234–238. [PubMed: 11125101]
- Duncan A-M, Ren H, Bound F, Tully J, Chandler DS, Sandeman RM. Assessment of novel inhibitors of *Helicoverpa* aminopeptidases as anti-insect agents. *Pest Manage Sci* 2006;62:1098–1108.
- Edge ASB. Deglycosylation of glycoproteins with trifluoromethanesulphonic acid: elucidation of molecular structure and function. *Biochem J* 2003;376:339–350. [PubMed: 12974674]
- Eguchi M, Azuma M, Yamamoto H, Takeda S. Genetically defined membrane-bound and soluble alkaline phosphatases of the silkworm: their discrete localisation and properties. *Prog Clin Biol Res* 1990;344:267–287. [PubMed: 2392471]
- Emmerling M, Chandler D, Sandeman M. Molecular cloning of three cDNAs encoding aminopeptidases from the midgut of *Helicoverpa punctigera*, the Australian native budworm. *Insect Biochem Mol Biol* 2001;31:899–907. [PubMed: 11439249]
- Ewing B, Hillier L, Wendl M, Green P. Base-calling of automated sequencer traces using *Phred*. I. Accuracy assessment. *Genome Res* 1998;8:175–185. [PubMed: 9521921]
- Ferré J, Van Rie J. Biochemistry and genetics of insect resistance to *Bacillus thuringiensis*. *Annu Rev Entomol* 2002;47:501–533. [PubMed: 11729083]
- Forrester NW, Cahill M, Bird LJ, Layland JK. Management of pyrethroid and endosulfan resistance in *Helicoverpa armigera* (Lepidoptera: Noctuidae) in Australia. *Bull Entomol Res Suppl Ser Suppl* 1993;1:1–132.
- Ganforina, MD.; Kayser, H.; Sanchez, D. Åkerström, B.; Borregaard, N.; Flower, DR.; Salier, J-P., editors. Lipocalins in Arthropoda: Diversification and Functional Explorations. 2006. Lipocalins <<http://landesbioscience.com>>
- Garczynski SF, Crim JW, Adang MJ. Identification of putative insect brush border membrane-binding molecules specific to *Bacillus thuringiensis* δ -endotoxin by protein blot analysis. *Appl Environ Microbiol* 1991;57:2816–2820. [PubMed: 1746942]
- Gill M, Ellar D. Transgenic *Drosophila* reveals a functional *in vivo* receptor for the *Bacillus thuringiensis* toxin Cry1Ac1. *Insect Mol Biol* 2002;11:619–625. [PubMed: 12421420]
- Gill SS, Cowles EA, Francis V. Identification, isolation, and cloning of a *Bacillus thuringiensis* Cry1Ac toxin-binding protein from the midgut of the lepidopteran insect *Heliothis virescens*. *J Biol Chem* 1995;270:27277–27782. [PubMed: 7592988]
- Gómez I, Pardo-López L, Muñoz-Garay C, Fernandez LE, Pérez C, Sánchez J, Soberón M, Bravo A. Role of receptor interaction in the mode of action of insecticidal Cry and Cyt toxins produced by *Bacillus thuringiensis*. *Peptides* 2007;28:169–173. [PubMed: 17145116]
- Herrero S, Gonzalez-Cabrera J, Tabashnik BE, Ferré J. Shared binding sites in Lepidoptera for *Bacillus thuringiensis* Cry1Ja and Cry1A toxins. *Appl Environ Microbiol* 2001;67:5729–5734. [PubMed: 11722929]
- Herrero S, Gechev T, Bakker PL, Moar WJ, de Maagd R. *Bacillus thuringiensis* Cry1Ca-resistant *Spodoptera exigua* lacks expression of one of four Aminopeptidase N genes. *BMC Genomics* 2005;6:96–105. [PubMed: 15978131]
- Hossain DM, Shitomi Y, Moriyama K, Higuchi M, Hayakawa T, Mitsui T, Sato R, Hori H. Characterization of a novel plasma membrane protein, expressed in the midgut epithelia of *Bombyx mori*, that binds to Cry1A toxins. *Appl Environ Microbiol* 2004;70:4604–4612. [PubMed: 15294792]

- Ihara H, Uemura T, Masuhara M, Ikawa S, Sugimoto K, Wadano A, Himeno M. Purification and partial amino acid sequences of the binding protein from *Bombyx mori* for CryIAa δ -endotoxin of *Bacillus thuringiensis*. *Comp Biochem Physiol* 1998;B120:197–204.
- Jurat-Fuentes JL, Adang MJ. Importance of CryI δ -endotoxin domain II loops for binding specificity in *Heliothis virescens* (Lepidoptera). *Appl Environ Microbiol* 2001;67:323–329. [PubMed: 11133462]
- Jurat-Fuentes JL, Adang MJ. Characterization of a CryI Ac receptor alkaline phosphatase in susceptible and resistant *Heliothis virescens* larvae. *Eur J Biochem* 2004;271:3127–3135. [PubMed: 15265032]
- Jurat-Fuentes JL, Adang MJ. Cry toxin mode of action in susceptible and resistant *Heliothis virescens* larvae. *J Invertebr Pathol* 2006;92:166–171. [PubMed: 16797583]
- Jurat-Fuentes JL, Adang MJ. A proteomic approach to study CryI Ac binding proteins and their alterations in resistant *Heliothis virescens* larvae. *J Invertebr Pathol* 2007;95:187–191. [PubMed: 17467006]
- Kayser, H. Lipocalins and structurally related ligand-binding proteins. In: Gilbert, LI.; Iatrou, K.; Gill, SS., editors. *Comprehensive Molecular Insect Science*. Vol. 4. Elsevier Science BV; Oxford: 2005. p. 267-306.
- Keeton TP, Bulla LAJ. Ligand specificity and affinity of BT-R1, the *Bacillus thuringiensis* CryIA toxin receptor from *Manduca sexta*, expressed in mammalian and insect cell cultures. *Appl Environ Microbiol* 1997;63:3419–3425. [PubMed: 9292994]
- Knight PJK, Crickmore N, Ellar DJ. The receptor for *Bacillus thuringiensis* CryIA(c) δ -endotoxin in the brush border membrane of the lepidopteran *Manduca sexta* is aminopeptidase. *N Mol Microbiol* 1994;11:429–436.
- Knight PJK, Knowles BH, Ellar DJ. Molecular cloning of an insect aminopeptidase N that serves as a receptor for *Bacillus thuringiensis* CryI Ac toxin. *J Biol Chem* 1995;270:17765–17770. [PubMed: 7629076]
- Knight PJK, Carroll J, Ellar DJ. Analysis of glycan structures on the 120 kDa aminopeptidase N of *Manduca sexta* and their interactions with *Bacillus thuringiensis* CryI Ac toxin. *Insect Biochem Mol Biol* 2004;34:101–112. [PubMed: 14976987]
- Koike Y, Mita K, Suzuki MG, Maeda S, Abe H, Osoegawa K, deJong PJ, Shimada T. Genomic sequence of a 320-kb segment of the Z chromosome of *Bombyx mori* containing a kettin ortholog. *Mol Genet Genomics* 2003;269:137–149. [PubMed: 12715162]
- Krishnamoorthy M, Jurat-Fuentes JL, McNall RJ, Andacht T, Adang MJ. Identification of novel CryI Ac binding proteins in midgut membranes from *Heliothis virescens* using proteomic analyses. *Insect Biochem Mol Biol* 2007;37:189–201. [PubMed: 17296494]
- Kronegg, J.; Buloz, D. Detection/prediction of GPI cleavage site (GPI-anchor) in a protein (DGPI). 1999. Retrieved 2/7/07 from <<http://129.194.185.165/dgpi/>>
- Liao, C.; Trowell, SC.; Akhurst, RJ. Binding of *Bacillus thuringiensis* insecticidal crystal proteins to the cotton bollworm, *Helicoverpa armigera*. *Proceedings of the Second Pacific Rim Conference on Biotechnology of Bacillus thuringiensis and its Impact to the Environment*; Chiang Mai, Thailand. 1996. p. 180-192.
- Liao C, Heckel DG, Akhurst R. Toxicity of *Bacillus thuringiensis* insecticidal proteins for *Helicoverpa armigera* and *Helicoverpa punctigera* (Lepidoptera: Noctuidae), major pests of cotton. *J Invertebr Pathol* 2002;80:55–63. [PubMed: 12234543]
- Liao C, Trowell SC, Akhurst RJ. Purification and characterisation of CryI Ac toxin binding proteins from the brush border membrane of *Helicoverpa armigera* midgut. *Curr Microbiol* 2005;51:367–371. [PubMed: 16252132]
- Lu Y-J, Adang MJ. Conversion of *Bacillus thuringiensis* CryI Ac-binding aminopeptidase to a soluble form by endogenous phosphatidylinositol phospholipase C. *Insect Biochem Mol Biol* 1996;26:33–40.
- Luo K, Sangadala S, Masson L, Mazza A, Brousseau R, Adang MJ. The *Heliothis virescens* 170 kDa aminopeptidase functions as “Receptor A” by mediating specific *Bacillus thuringiensis* CryIA δ -endotoxin binding and pore formation. *Insect Biochem Mol Biol* 1997;27:735–743. [PubMed: 9443374]
- MacIntosh SC, Stone TB, Jokerst RS, Fuchs RL. Binding of *Bacillus thuringiensis* proteins to a laboratory-selected line of *Heliothis virescens*. *Proc Natl Acad Sci USA* 1991;88:8930–8933. [PubMed: 1924353]

- Malik K, Mahmood T, Riazuddin S. The Receptor for *Bacillus thuringiensis* Cry 1Ac delta-endotoxin in the brush border membrane of the lepidopteran *Helicoverpa armigera* is aminopeptidase. *N J Biol Sci* 2001;1:782–784.
- Martin SE, Shabanowitz J, Hunt DF, Marto JA. Subfemtomole MS and MS/MS peptide sequence analysis using nano-HPLC micro-ESI Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem* 2000;72:4266–4274. [PubMed: 11008759]
- Mauchamp B, Royer C, Garel A, Jalabert A, Da Rocha M, Grenier A-M, Labas V, Vinh J, Mita K, Kadono K, Chavancy G. Polycalin (chlorophyllid A binding protein): a novel, very large fluorescent lipocalin from the midgut of the domestic silkworm *Bombyx mori* L. *Insect Biochem Mol Biol* 2006;36:623–633. [PubMed: 16876705]
- Mita K, Kasahara M, Sasaki S, et al. The genome sequence of silkworm, *Bombyx mori*. *DNA Res* 2004;11:27–35. [PubMed: 15141943]
- Nakanishi K, Yaoi K, Nagino Y, Hara H, Kitami M, Atsumi S, Miura N, Sato R. Aminopeptidase N isoforms from the midgut of *Bombyx mori* and *Plutella xylostella*—their classification and the factors that determine their binding specificity to *Bacillus thuringiensis* Cry1A toxin. *FEBS Lett* 2002;519:215–220. [PubMed: 12023048]
- Oltean DI, Pullikuth AK, Lee H-K, Gill SS. Partial purification and characterisation of *Bacillus thuringiensis* Cry1A toxin receptor A from *Heliothis virescens* and cloning of the corresponding cDNA. *Appl Environ Microbiol* 1999;65:4760–4766. [PubMed: 10543783]
- Orlean P, Menon AK. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycopospholipids. *J Lipid Res* 2007;48:993–1011. [PubMed: 17361015]
- Pigott CR, Ellar DJ. Role of receptors in *Bacillus thuringiensis* crystal toxin activity. *Microbiol. Mol Biol Rev* 2007;71:255–281.
- Rajagopal R, Agrawal N, Selvapandiyan A, Sivakumar S, Ahmad S, Bhatnagar RK. Recombinantly expressed isoenzymic aminopeptidases from *Helicoverpa armigera* (American cotton bollworm) midgut display differential interaction with closely related *Bacillus thuringiensis* insecticidal proteins. *Biochem J* 2003;370:971–978. [PubMed: 12441000]
- Rodrigo-Simón A, de Maagd RA, Avilla C, Bakker PL, Molthoff J, González-Zamora JE, Ferré J. Lack of Detrimental effects of *Bacillus thuringiensis* Cry toxins on the insect predator *Chrysoperla carnea*: a toxicological, histopathological, and biochemical analysis. *Appl Environ Microbiol* 2006;72:1595–1603. [PubMed: 16461715]
- Rodrigo-Simón A, Caccia S, Ferré J. *Bacillus thuringiensis* Cry1Ac toxin-binding and pore-forming activity in brush border membrane vesicles prepared from anterior and posterior midgut regions of lepidopteran larvae. *Appl Environ Microbiol* 2008;74:1710–1716. [PubMed: 18223107]
- Sangadala S, Walters FS, English LH, Adang MJ. A mixture of *Manduca sexta* aminopeptidase and phosphatase enhances *Bacillus thuringiensis* insecticidal Cry1Ac toxin binding and $^{86}\text{Rb}^+ - \text{K}^+$ efflux *in vitro*. *J Biol Chem* 1994;269:10088–10092. [PubMed: 8144508]
- Sangadala S, Azadi P, Carlson R, Adang MJ. Carbohydrate analyses of *Manduca sexta* aminopeptidase N, co-purifying neutral lipids and their functional interactions with *Bacillus thuringiensis* Cry1Ac toxin. *Insect Biochem Mol Biol* 2001;32:97–107. [PubMed: 11719073]
- Simpson RM, Newcomb RD. Binding of *Bacillus thuringiensis* delta-endotoxins Cry1Ac and Cry1Ba to a 120-kDa aminopeptidase-N of *Epiphyas postvittana* purified from both brush border membrane vesicles and baculo-virus-infected Sf9 cells. *Insect Biochem Mol Biol* 2000;30:1069–1078. [PubMed: 10989294]
- Simpson RM, Burgess EPJ, Markwick NP. *Bacillus thuringiensis* δ -endotoxin binding sites in two Lepidoptera, *Wiseana* spp. and *Epiphyas postvittana*. *J Invertebr Pathol* 1997;70:136–142. [PubMed: 9281402]
- Sivakumar S, Rajagopal R, Venkatesh GR, Srivastava A, Bhatnagar RK. Knockdown of aminopeptidase-N from *Helicoverpa armigera* larvae and in transfected Sf21 cells by RNA interference reveals its functional interaction with *Bacillus thuringiensis* insecticidal protein Cry1Ac. *J Biol Chem* 2007;282:7312–7319. [PubMed: 17213205]
- Stephens E, Sugars J, Maslen SL, Williams DH, Packman LC, Ellar DJ. The N-linked oligosaccharides of aminopeptidase N from *Manduca sexta*. *Eur J Biochem* 2004;271:4241–4258. [PubMed: 15511230]

- Tabashnik BE, Cushing NL, Finson N, Johnson MW. Field development of resistance to *Bacillus thuringiensis* in Diamondback moth (Lepidoptera: Plutellidae). *J Econ Entomol* 1990;83:1671–1676.
- Teakle, RE.; Jensen, JM. *Handbook of Insect Rearing*. Elsevier; Amsterdam: 1985.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;25:4876–4882. [PubMed: 9396791]
- Valaitis AP, Jenkins JL, Lee MK, Dean DH, Garner KJ. Isolation and partial characterization of gypsy moth BTR-270, an anionic brush border membrane glycoconjugate that binds *Bacillus thuringiensis* Cry1A toxins with high affinity. *Arch Insect Biochem Physiol* 2001;46:186–200. [PubMed: 11304752]
- Wang P, Zhang X, Zhang J. Molecular characterization of four midgut aminopeptidase N isozymes from the cabbage looper, *Trichoplusia ni*. *Insect Biochem Mol Biol* 2005;35:611–620. [PubMed: 15857766]
- Wolfersberger MG, Luethy P, Maurer P, Parenti P, Sacchi VF, Giordana B, Hanozet GM. Preparation and partial characterisation of amino acid transporting brush border membrane vesicles from the larval midgut of the cabbage butterfly (*Pieris brassicae*). *Comp Biochem Physiol* 1987;86A:301–308.
- Wu AM, Song SC, Sugii S, Herp A. Differential binding properties of Gal/GalNAc specific lectins available for characterization of glycoreceptors. *Indian J Biochem Biophys* 1997;34:61–71. [PubMed: 9343930]
- Xia Q, Zhou Z, Lu C, et al. A draft sequence for the genome of the domesticated silkworm (*Bombyx mori*). *Science* 2004;306:1937–1940. [PubMed: 15591204]
- Xu X, Wu Y. Disruption of Ha_BtR alters binding of *Bacillus thuringiensis* δ -endotoxin Cry1Ac to midgut BBMVs of *Helicoverpa armigera*. *J Invertebr Pathol* 2008;97:27–32. [PubMed: 17681529]

Fig. 1.

Affinity purification of Cry1Ac binding proteins from Brush Border Membrane Vesicles of *H. armigera*. Proteins were separated by 10% SDS-PAGE and stained with Coomassie blue R250. From left to right are: molecular weight standards with their sizes indicated; BBMV before and after solubilisation; material eluted with the first column wash; material eluted from the column with 2 M KCl; material eluted with GalNAc; material eluted with 4 and 6 M urea; an aliquot of the activated Cry1Ac used to make the affinity column. Labels indicate molecular weights estimated from the gel using TotalLab (nonlinear dynamics) software and the major components that were identified from the bands as discussed in the text.

Aminopeptidase N sequences from *H. armigera*. Peptides matched with MS/MS data are underlined. Predicted signal sequences and GPI anchor cleavage sites are indicated in bold at the N-termini and near the C-termini, respectively. Elsewhere bold indicates potential sites for N-linked glycosylation (Asn-Xaa-Ser/Thr sequons). Potential sites for O-linked glycosylation on threonine and serine residues are boxed. The zinc-binding motif (HEX₂HX₁₈E) and the gluzincin APN motif (GAMENWG) are boxed with shading. HaAPNs 1 to 7 are translations of cDNA sequences AY038607, AY038608, AF535166, AF535165, EU325551, EU328183 and EU328182, respectively, except that the cDNA sequence encoding HaAPN7 was incomplete. That sequence was completed with 39 residues at the N-terminus from an unpublished BAC sequence (Fournier, Fevereisen et al., unpublished).

Fig. 3.

Unrooted phylogram showing the relationships among members of the aminopeptidase protein family. The unrooted neighbour-joining tree was based on an alignment of full-length lepidopteran aminopeptidase sequences using a set of representative non-lepidopteran proteins as an outgroup. All bifurcations with support in > 500 out of 1000 bootstraps are indicated. The scale indicates the number of changes/site. Species abbreviations are: Bm, *Bombyx mori*; Hv, *Heliothis virescens*; Se, *Spodoptera exigua*; Ld, *Lymantria dispar*; Ms, *Manduca sexta*; Px, *Plutella xylostella*; Tn, *Trichoplusia ni*; Ha, *Helicoverpa armigera*; Hp, *Helicoverpa punctigera* (Lepidoptera); Ae, *Aedes aegypti*; CG numbers, *Drosophila melanogaster* (Diptera); Hs, *Homo sapiens*. Accession numbers used to generate the phylogram are followed in some cases by equivalent numbers cited on the phylogram of Pigott and Ellar (2007). Class I: BmAPN1, O76803, AAC33301; HaAPN1, Q962B4, AAK85538; HpAPN1, Q9NHZ9, AAF37558; HvAPN120, AAF08254; LdAPN4, Q95WL0, AAL26895; MsAMPN (= MsAPN1), Q11001, CAA61452; PxAPN, O16851, AAB70755; SeAPN1, AAP44964; TnAPN1, AAX39863. Class II: BmAPN5, see text; HaAPN2, Q962B3, AAK85539; PxAPN2, Q46156, CAA10950. Class III: BmAPN3, Q8T7M0, AAL83943; HaAPN3, Q8MU78, AAN04900; HpAPN3, Q9NHZ7, AAF37560; HvAMPN, Q11000; LdAPN2, Q9XYV2, AAD31183; MsAPN4, Q8T3I0, AAM13691; PiAPN, O17484, ACC36148; PxAPN3f, Q9U8B0, AAF01259; SeAPN3, AAP44966; TnAPN3, AAX39865. Class IV: BmAPN4, O77046, BAA33715; HaAPN4, Q8MU79, AAM44056; HpAPN2, Q9NHZ8, AAF37559; HvAPN110, Q963G0, AAK58066; LdAPN3, Q95WL1, AAL26894; MsAPN3, Q8T5Q4, AAM18718; SeAPN4, AAP44967; TnAPN4, AAX39866. Class V: BmAPN2, O76151, BAA32140; HaAPN5, AAW72993; LdAPN2x, Q9XYV3, AAD31184; MsAMPN (= MsAPN2), P91885, CAA66466; PxAMPN, P91887, CAA66467; SeAPN2, AAP44965; TnAPN2, AAX39864. Class VI: BmAPN6, see text; HaAPN6, EU328183. Class VII: BmAPN7, see text; HaAPN7, EU328182. Other sequences: Bmb025220, see text; CG7340, AAF32329; HsPEL1, Q8NDH3; BmAPNL, Q8T105; CG32351, Q9VSM7; HsAPN-L, AAH65564; Ae Q95PB9, AAK55416; CG11951, Q9VAM2; CG14516, Q8SYT3; CG5839,

Q8MRN5; CG8775, Q8INH6; HsAMPN, P55786. Boxes indicate APNs reported to interact with *Bt* Cry toxins (Pigott and Ellar, 2007; this paper).

Fig. 4.

Spatial and temporal expression of aminopeptidase genes in *H. armigera*. (A) The gel panels show PCR products corresponding to the four APN genes (and *actin*) identified in whole first instars (1st) and midgut (MG) and carcass (C) tissues from second, third, fourth and fifth instars. 'MW' indicates DNA molecular weight markers (kilobase). Negative controls ('Neg') lack reverse transcriptase or RNA. (B) Quantitative RT-PCR showing expression levels of APN genes relative to a housekeeping gene in anterior, middle and posterior thirds of the midgut from fourth instars. Error bars show standard errors from four reactions. Apparent expression levels from negative control reactions did not exceed 6×10^{-5} .

Fig. 5. Predicted gene and primary protein structures for *H. armigera* polycalins. Exon boundaries are proposed by comparison with the gene structure of the *B. mori* polycalin. Each cDNA sequence (EU325567, EU325566, EU325561) predicts a protein with three, four or five lipocalin-like domains, respectively, a signal peptide, and a GPI anchor cleavage site. Tryptic peptides matched to the sequences by MS/MS fragmentation are underlined. Note for example that the N-terminal portion of the first lipocalin domain of the tricalin sequence is identical with the first domain of the other sequences but the C-terminal portion is identical with the third domain of the pentacalin sequence. Single amino acid residues in bold indicate differences that we suggest are due to allelic variation in otherwise identical regions of the predicted proteins. Elsewhere bold indicates potential sites for N-linked glycosylation (Asn-Xaa-Ser/Thr sequons). Boxed threonine residues indicate potential sites for O-linked glycosylation.

Table 1

Aminopeptidase N- and alkaline phosphatase-specific activities through the Cry1Ac affinity purification process

	Aminopeptidase N		Alkaline phosphatase	
	Specific activity ($\mu\text{mol/min/mg}$)	Enrichment factor	Specific activity ($\mu\text{mol/min/ mg}$)	Enrichment factor
Homogenised midgut	0.75	1	0.03	1
BBMV	2.69	3.6	0.13	4.3
Solubilised BBMV	3.53	4.7	0.17	5.7
GalNAc eluate	5.15	6.9	0.01	0.3