The complement of family M1 aminopeptidases of *Haemonchus contortus* – biotechnological implications

Namitha Mohandas^a, Neil D. Young^a, Abdul Jabbar^a, Pasi K. Korhonen^a, Anson V. Koehler^a, Ross S. Hall^a, Min Hu^c, Andreas Hofmann^{a,b,*}, Robin B. Gasser^{a,c,*}

^a *The University of Melbourne, Faculty of Veterinary and Agricultural Sciences, Parkville, Victoria, Australia*

^b *Structural Chemistry Program, Eskitis Institute for Drug Discovery, Griffith University, Brisbane, Queensland 4111, Australia*

^c *State Key Laboratory of Agricultural Microbiology, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan 430070, Hubei, China*

* Corresponding author. robinbg@unimelb.edu.au (R. B. Gasser) or a.hofmann@griffith.edu.au (A. Hofmann)

ABSTRACT

Although substantial research has been focused on the ‘hidden antigen’ H11 of *Haemonchus contortus* as a vaccine against haemonchosis in small ruminants, little is known about this and related aminopeptidases. In the present article, we reviewed genomic and transcriptomic data sets to define, for the first time, the complement of aminopeptidases (designated *Hc*-AP-1 to *Hc*-AP-13) of the family M1 with homologs in *C. elegans*, characterised by zinc-binding (HEXXH) and exo-peptidase (GAMEN) motifs. The three previously published H11 isoforms (accession nos. X94187, FJ481146 and AJ249941) had most sequence similarity to *Hc*-AP-2 and *Hc*-AP-8, whereas unpublished isoforms (accession nos. AJ249942 and AJ311316) were both most similar to *Hc*-AP-3. The aminopeptidases characterised here had homologs in *C. elegans*. *Hc*-AP-1 to *Hc*-AP-8 were most similar in amino acid sequence (28-41%) to *C. elegans* T07F10.1; *Hc*-AP-9 and *Hc*-AP-10 to *C. elegans* PAM-1 (isoform b) (53-54% similar); *Hc*-AP-11 and *Hc*-AP-12 to *C. elegans* AC3.5 and Y67D8C.9 (26% and 50% similar, respectively); and *Hc*-AP-13 to *C. elegans* C42C1.11 and ZC416.6 (50-58% similar). Comparative analysis suggested that *Hc*-AP-1 to *Hc*-AP-8 play roles in digestion, metabolite excretion, neuropeptide processing and/or osmotic regulation, with *Hc*-AP-4 and *Hc*-AP-7 having male-specific functional roles. The analysis also indicated that *Hc*-AP-9 and *Hc*-AP-10 might be involved in the degradation of cyclin (B3) and required to complete meiosis. *Hc*-AP-11 represents a leucyl/cystinyl aminopeptidase, predicted to have metallopeptidase and zinc ion binding activity, whereas *Hc*-AP-12 likely encodes an aminopeptidase Q homolog also with these activities and a possible role in gonad function. Finally, *Hc*-AP-13 is predicted to encode an aminopeptidase AP-1 homolog of *C. elegans* with hydrolase activity, suggested to operate, possibly synergistically with a PEPT-1 ortholog, as an oligopeptide transporter in the gut for protein uptake and normal development and/or reproduction of the worm. Appraisal of structure-based amino acid sequence alignments revealed that all conceptually translated *Hc*-AP proteins, with the exception of *Hc*-AP-12, adopt a topology similar to those observed with the two subgroups of mammalian M1 aminopeptidases which possess either three (I, II and IV) or four (I-IV) domains. In contrast, *Hc*-AP-12 lacks the N-terminal domain (I), but possesses a substantially expanded domain III. Although further work needs to be done to assess amino acid sequence conservation of the different aminopeptidases among individual worms within and among *H. contortus* populations, we hope that these insights will support future localisation, structural and functional studies of these molecules in *H. contortus* as well as facilitate future assessments of a recombinant subunit or cocktail vaccine against haemonchosis.

Keywords:

Haemonchus contortus

H11

Hidden antigen

Aminopeptidases (family M1)

Immunogens

Transcription profiles

Structure models

1. Introduction

The best-characterised hidden antigen of *Haemonchus contortus*, called H11, is a 110 kDa gut membrane glycoprotein (Smith et al., 1993) belonging to the M1 aminopeptidase family. Native H11, isolated from adult *H. contortus*, had been evaluated in numerous vaccine experiments in sheep of different breeds, including very young lambs, and shown to induce high levels of antibody-mediated protection, reflected in (on average) a >90% decrease in faecal egg counts (FECs) and > 75% reduction in worm burdens (reviewed by Newton and Munn, 1999; Newton and Meeusen, 2003; Knox, 2011). Molecular characterisation showed that H11 is a microsomal aminopeptidase (Graham et al., 1993; Smith et al., 1997), believed to be involved in the degradation of small peptides from dietary proteins (Munn and Munn, 2002). At the time, it was proposed that protection was achieved via antibody-induced disruption of nutrient absorption by the worm (cf. Newton and Meeusen, 2003).

Originally, three isoforms of H11 (designated H11-1, -2 and -3; GenBank accession nos. AJ249941, AJ249942 and AJ311316, respectively), sharing ~ 61-70% amino acid sequence identity, were reported (Graham et al., 1993; Hederer et al. 2001, direct submission to GenBank). Subsequently, two other isoforms (H11-4 and -5) were sequenced and deposited in the GenBank database (accession nos. FJ481146 and X94187; Smith et al., 1997; Zhou et al., 2010). Although vaccination success was achieved using purified, native H11, and cDNA clones were patented (Smith et al., 1993; Graham et al., 1993), to date, vaccination of sheep with recombinant forms of H11 has not achieved the level of protection required for a practical or commercial vaccine (cf. Knox et al., 2003; Newton and Meeusen, 2003; Knox, 2011; Roberts et al., 2013).

At the time, enzymatically active H11-S forms of H11-1 and -2 had been expressed in insect cells using recombinant baculovirus (cf. Newton and Meeusen, 2003). Although the immunisation of sheep with these recombinant H11-S molecules induced high serum antibody levels, which cross-reacted with native H11 and inhibited the aminopeptidase activities of native and recombinant H11 *in vitro*, sheep were not protected against *H. contortus* challenge infection (cf. Newton and Meeusen, 2003). On the other hand, immunisation of sheep with the active site domain of H11-1, expressed in inactive form in *Escherichia coli* as inclusion bodies, induced partial protection, reflected in a 30-40% reduction in faecal egg counts (FECs). Levels of serum antibodies, which cross-reacted with native H11 in sheep immunised with *E. coli*-expressed active site domains, appeared to relate to decreased FECs, but varied significantly among sheep (Newton et al., unpublished; cf. Newton and Meeusen, 2003). However, these antibodies were not able to inhibit the aminopeptidase activity of H11, which rejected the hypothesis that the mechanism of protection induced by H11 relates to enzymatic inhibition (cf. Newton and Meeusen, 2003).

Questions remained as to why recombinant molecules had failed to achieve protection against challenge infection. Given that partial protection could be achieved using inactive *E. coli*-expressed H11-1 active site domain, the possibility that the aminopeptidase-active site domain is not involved in induction of protective immunity seems unlikely. Since prokaryotically expressed H11 was obtained from inclusion bodies, the possibility of reduced protective immunity due to incorrectly folded structural epitopes cannot be excluded (Newton and Meeusen, 2003). However, there is also a possibility that the unique glycosylation of the baculovirus-expressed H11-S molecules had a significant effect on the nature and/or extent of immune responses in sheep following immunisation. In this regard, the N-linked glycosylation of native H11 from *H. contortus* is distinct from that produced by insect cells (Haslam et al., 1996), and insect cell glycosylation might have masked protective protein epitopes, or the carbohydrate moiety may be part of the protective epitope(s) of native H11. Similar explanations appear to apply to recombinant H11 expressed in *C. elegans* (see Roberts et al., 2013). Moreover, to us, it also seems plausible that as yet uncharacterised

aminopeptidases or isoforms thereof might provide functional redundancy in the face of antibody inhibition of the activity of H11-1 and -2 (cf. Newton and Meeusen, 2003), and their incorporation as antigens in a multivalent vaccine might be needed to achieve high levels of protection in sheep.

The availability of draft genomes for *H. contortus* (see Laing et al., 2013; Schwarz et al., 2013) and advanced informatics (Cantacessi et al., 2012, 2015; Gasser et al., 2016) now enable deep explorations of particular molecules or molecular groups, including aminopeptidases. In recent work, we predicted a considerable number of aminopeptidase-encoding genes in *H. contortus* (see Schwarz et al., 2013), which provided a unique opportunity to characterise H11 members and other aminopeptidases of the family M1. Therefore, in the present article, we curate the full-length open reading frames (ORFs) and define the complement of aminopeptidases of the family M1, establish their phylogenetic relationships, explore the transcription profiles of their genes in all key developmental stages of *H. contortus* and model selected proteins *in silico*. We hope that these insights can guide molecular explorations of these peptidases and evaluations of recombinant forms of immunogenic aminopeptidases as a cocktail vaccine against haemonchosis in the future.

2. Methodology

We employed published draft genome as well as transcriptomes of all key developmental stages (egg, L1-L4 and adult) and both sexes (L4 and adult) of *H. contortus* (NCBI BioProject accession no. PRJNA205202; WormBase (www.wormbase.org) (Schwarz et al., 2013). This draft genome is ~ 320 Mb in size and has been predicted to encode 23,610 proteins (Schwarz et al., 2013).

2.1. Identification of genomic scaffolds containing genes encoding M1 aminopeptidases

From the complete, assembled transcriptome representing all eight stages or sexes of *H. contortus* (see Schwarz et al., 2013), we identified and isolated 1,066 assembled transcripts based on their homology matches (*E*-value cut off: 10^{-5}) to all five known *h11* genes (designated *h11-1* to *h11-5*; GenBank accession numbers AJ249941, AJ249942, AJ311316, FJ481146 and X94187, respectively; Graham et al., 1993; Smith et al., 1997; Zhou et al., 2010). Then, we identified genomic scaffolds containing regions of homology to known H11 genes (Smith et al., 1997; Roberts et al., 2013) by mapping (*E*-value cut-off: 10^{-5}) these assembled transcripts ($n = 1,066$) using BLAT (Kent, 2002). We also used *h11* genes from the *H. contortus* draft genome, predicted previously using MAKER2 (Holt and Yandell, 2011; cf. Schwarz et al. 2013). Open reading frames (ORFs) of individual assembled transcripts were inferred using the program GeneMark-ES (Lomsadze et al., 2005; Ter-Hovhannisyan et al., 2008). Using the Integrative Genomics Viewer (IGV) (Robinson et al., 2011; Thorvaldsdóttir et al., 2012), we then visually integrated all of these data to obtain a consensus sequence for individual coding regions.

2.2. Identification of protein domains, families and subfamilies

Identifying aminopeptidase (M1 family) genes encoded in the draft genome allowed us to then define the complete set of full-length transcripts. ORFs were verified, and corresponding coding region was inferred from these full-length transcripts using ORF-finder (<http://www.ncbi.nlm.nih.gov/projects/gorf/>). Each predicted protein was characterised by its primary amino acid sequence and structural and/or functional domains, inferred using all databases (i.e., PROSITE, HAMAP, Pfam, PRINTS, ProDom, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, CATH-Gene3D and PANTHER) within InterProScan v.5.7.48 (Zdobnov et al., 2001; Jones et al., 2014). Individual predicted proteins were classified according to family

and/or subfamily using information in the PANTHER database v.9.0 (Thomas et al., 2003; Mi et al., 2013). Then, using the program HMMER 3.1 (Finn et al., 2011), we generated hidden Markov model (HMM) profiles for each of these four protein groups and searched the full complement of 23,610 predicted proteins of *H. contortus* (see Schwarz et al., 2013) for H11 and related M1 aminopeptidase homologs using *hmmsearch* (<http://www.ebi.ac.uk/Tools/hmmer/search/hmmsearch>). Following a comparison of the retrieved sequences with those inferred from transcripts and genomic exons encoding *h11* and related genes, we were able to infer the **entire** complement of full-length transcripts and protein sequences.

2.3. Differential transcription

Each set of quality-filtered, paired-end RNA-seq reads for each individual developmental stage or sex of *H. contortus* was mapped to the final complement of full-length *h11* and related genes using Burrows-Wheeler Aligner (BWA) software (Li and Durbin, 2010). For each stage, the numbers of reads that mapped to individual transcripts were established using the SAMtools algorithm (Li et al., 2009). The resultant read counts per transcript per developmental stage were used as input data for DESeq2 and EdgeR (Robinson et al., 2010; Love et al., 2014). Differential transcription was calculated by pairwise comparison of all free-living (egg, L1, L2 and L3) and parasitic stages (L4 and adult). Genes were recorded as differentially transcribed, using edgeR-calculated common and gene-wise dispersion factors, if the $\log_2$ fold change ($\log_2$ FC) between free-living and parasitic stages compared with the normalised read count data was ≥ 2 , with a false discovery rate (FDR) of ≤ 0.05 . A heat map (representing mapped reads) was produced using the *heatmap.2* function in the *gplots* package in R (<http://www.R-project.org/>).

2.4. Phylogenetic analyses

Amino acid sequences inferred from individual full-length transcripts were aligned using MUSCLE v.3.8 (Edgar, 2004), ensuring an accurate alignment of homologous characters. Aligned sequence data were then subjected to phylogenetic analysis using Bayesian inference (BI) in the program MrBayes v.3.2.2 (Ronquist and Huelsenbeck, 2003); the optimal model of sequence evolution was assessed using a mixed amino acid substitution model employing four chains and 200,000 generations, sampling every 100th generation; the first 25% of the sampled generations were removed from the analysis as burn-in.

2.5. Prediction of protein structure

Preliminary annotation and fold appraisal were guided by results from automated model predictions performed using the program I-TASSER (Yang et al., 2015). For a more detailed analysis, structure-based amino acid sequence alignments were prepared for predicted M1 aminopeptidases of *H. contortus* using programs PSIPRED (Bryson et al., 2005) and SBAL (Wang et al., 2012; Wang and Hofmann, 2015). Based on the phylogenetic analysis and the structure-based amino acid sequence alignment (Supplementary Fig. 1), three different protein topologies were identified, and the structure of a representative protein for each case was generated by comparative modelling using a template structure identified employing the program pGenThreader (Bryson et al., 2005). As a representative for the four-domain proteins, *Hc*-AP-2 was chosen and modelled using human aminopeptidase A (PDB 4kx7; pGenThreader p-value: $2e-26$). The predicted three-domain protein *Hc*-AP-13 was modelled based on leukotriene A4 hydrolase (PDB 4gaa; pGenThreader p-value: $2e-25$). *Hc*-AP-12 was identified as the most distant from the other 12 aminopeptidases predicted. This predicted

protein appears to be a three-domain protein, but due to uncertainty as to the structural fold in the area of domain III, an indicative illustration was generated. For comparative modelling, the software MODELLER (Sali and Blundell, 1993) and the structure-based amino acid sequence alignment (Supplementary Fig. 1) was employed; 20 independent models were computed without restraints, and the model with the lowest objective function was selected and visually inspected.

3. Defining the complement of full-length transcripts

A total of 1,066 assembled transcripts from the transcriptome of *H. contortus* (see Schwarz et al., 2013) had homology (*E*-value cut off: 10^{-5}) to the five known (full-length) *h11* gene sequences (designated here as *h11-1* to *h11-5*; GenBank accession numbers AJ249941, AJ249942, AJ311316, FJ481146 and X94187; Graham et al., 1993; Smith et al., 1997; Zhou et al., 2010). These transcripts mapped to 33 different *H. contortus* genomic scaffolds containing exons identified based on *ab initio* prediction from the genome (Fig. 1; cf. Schwarz et al., 2013).

In total, 13 full-length transcripts (ORFs of 1791-3324 nt) were identified (designated *Hc-ap-1* to *Hc-ap-13*). Significant differences in transcription (FDR of < 0.05) were recorded among some of the 13 genes encoding M1 aminopeptidases (Fig. 2A), of which *Hc-ap-1* to *Hc-ap-8* were highly transcribed in parasitic stages of *H. contortus*, whereas *Hc-ap-9* to *Hc-ap-13* were relatively highly transcribed in all life cycle stages, with the exception of *Hc-ap-11* which had limited transcription in the egg (Fig. 2B). Genes *Hc-ap-4* and *Hc-ap-7* were transcribed at significantly higher levels in males than in females in both haematophagous (L4 and adult) stages (Fig. 2B).

4. Classification of predicted proteins

We then inferred protein sequences (*Hc-AP-1* to *Hc-AP-13*; 779 to 1002 amino acids in length) from respective transcripts (Table 1). All 13 proteins belonged to the family of M1 zinc metalloproteases (PTHR11533) (Fig. 1) and shared 11% to 100% sequence identity upon pairwise comparison and 13-99% with the previously reported genes/gene products (*H11-1* to *H11-5*; cf. Table 2). In spite of substantial variation among some of the M1 aminopeptidase protein sequences predicted, the zinc-binding (HEXXH) and the exo-peptidase (GAMEN) motifs, characteristic of all microsomal N-aminopeptidases (cf. Smith et al., 1997; Roberts et al., 2013) were both conserved among all amino acid sequences studied here.

We then proceeded to classify the predicted aminopeptidases based on InterProScan domains and protein signatures. Predicted proteins *Hc-AP-1* to *Hc-AP-8* and *Hc-AP-11* belonged to subfamily SF147 (slamdance; isoform b); *Hc-AP-9* and *Hc-AP-10* to SF163 (i.e. puromycin-sensitive aminopeptidase-related protein); *Hc-AP-12* to SF21 (protein Y67D8C.9); and, *Hc-AP-13* to SF150 (protein ZC416.6) (Figure 1; Table 1). For individual amino acid sequences, we inferred transmembrane, 5-element protein fingerprint aladiptase (PR00756), neutral protease 2 (1.10.390.10), peptidase M1 (PF01433), metalloprotease catalytic (SSF55486), leukotriene A4 hydrolase N-terminal (SSF63737) and/or ERAP1-like C-terminal (PF11838) domains (Fig. 1; Table 1).

5. Phylogenetic relationships

Subsequently, we explored the phylogenetic relationships of the 13 aminopeptidase sequences of *H. contortus* as well as sequences *H11-1* to *H11-5* derived from previous studies (Graham et al., 1993; Smith et al., 1997; Zhou et al., 2010). Phylogenetic analysis showed strongly supported nodes (pp = 0.92-1.00), according to the metalloprotease subfamilies (SFs) to which sequences belonged. Specifically, *Hc-AP-9* and *Hc-AP-10* in subfamily SF163

formed a separate cluster (pp = 1.00) to the exclusion of those in subfamilies SF21 and SF150 (pp = 0.92-0.99). In addition, nine amino acid sequences in SF147 formed a cluster to the exclusion of *Hc*-AP-11 (pp=1.00) (Fig. 3A). Then, we compared the sequences (designated here as H11-1 to H11-5) inferred from publicly available (full-length) *h11* transcripts (*h11-1* to *h11-5*; see Fig. 3B). *Hc*-AP-8 clustered together with H11-4 (pp = 1.00), and *Hc*-AP-1 and *Hc*-AP-2 clustered closely with H11-2 and H11-5 (pp = 1.00) to the exclusion of other predicted M1 aminopeptidase sequences (0.95-1.00) (Fig. 3B). The phylogenetic analysis also revealed that M1 aminopeptidases (Fig. 3) encoded by genes were transcribed at high levels in parasitic stages grouped separately from those transcribed in the free-living stages of *H. contortus* (Fig. 3). In addition, the sequences *Hc*-AP-4 and *Hc*-AP-7, both encoded by genes exhibiting significantly higher transcription in males than in female L4 and adult *H. contortus* (see Fig. 2B), grouped together (pp = 1.00; Fig. 3), to exclusion of others.

6. Comparative analyses with *C. elegans*, and functions predicted for aminopeptidase orthologs

All 13 predicted aminopeptidases (family M1) of *H. contortus* had homologs amongst the 12 known aminopeptidases (same family) of *C. elegans* (see Table 2; Fig. 4), and all had conserved zinc-binding and exopeptidase motifs (cf. Iturrioz et al., 2001; Laustsen et al., 2001; Althoff et al., 2014). Upon pairwise comparison, the sequence identities ranged from 26% to 58% between the *H. contortus* and *C. elegans* homologs (Table 2).

Hc-AP-1 to *Hc*-AP-8 had most sequence similarity (28-41%) to *C. elegans* T07F10.1, and *Hc*-AP-9 and *Hc*-AP-10 were most similar (53-54%) to *C. elegans* PAM-1 (two homologs), whereas *Hc*-AP-11 and *Hc*-AP-12 had most similarity to *C. elegans* AC3.5 and Y67D8C.9 (26% and 50%), respectively; and *Hc*-AP-13 to *C. elegans* C42C1.11 and ZC416.6 (50-58%). A phylogenetic analysis of amino acid sequence data for predicted aminopeptidases (family M1) and respective homologs from *C. elegans* showed strongly supported clades (pp: usually 0.99-1.00), according to the subfamilies to which they belonged. *Hc*-AP-1 to *Hc*-AP-8, which shared most similarity/homology to T07F10.1 (cf. Geldhof et al., 2007) formed a separate clade to the exclusion of *Hc*-AP-11, which clustered with AC3.5 isoform b (pp = 1.00). Predicted proteins *Hc*-AP-9 and *Hc*-AP-10 clustered with *C. elegans* PAM-1 isoform b (cf. Lyczak et al., 2006) and *Hc*-AP-12 with *C. elegans* Y67D8C.9 isoform b with strong nodal support (pp = 1.00). Predicted protein *Hc*-AP-13 formed a separate cluster with the two *C. elegans* aminopeptidases ZC416.6 isoform a (Benner et al., 2011) and C42C1.11 isoform a (Baset et al., 1998), with high support (pp = 1.00). The *C. elegans* homologs were also subjected InterProScan analysis to identify their functional domains and protein signatures. The InterProScan results for individual predicted aminopeptidases of *H. contortus* matched those of their corresponding *C. elegans* homologs (see Table 1), and the PANTHER subfamilies to which they belonged.

Based on information available for *C. elegans* (WormBase), we propose that the T07F10.1 isoform a homologs *Hc*-AP-1 to *Hc*-AP-8 play key roles in digestion, metabolite excretion, neuropeptide processing and/or osmotic regulation principally in parasitic stages of *H. contortus*, with *Hc*-AP-4 and *Hc*-AP-7 proposed to have male-specific functions (cf. Fig. 2B). In *C. elegans*, promoter-reporter constructs of T07F10.1 have shown expression in the intestine, excretory cell, pharynx and nervous system. In transgenic *C. elegans*, the expression of *H. contortus* *Hc*-AP-2 (which has 99% amino acid sequence similarity to *Hc*-AP-8) was localised to the intestine, and also to amphid cells, similar to findings in the parasite (cf. Zhou et al., 2010).

Hc-AP-9 and *Hc*-AP-10 are most similar in sequence (53-54%) to *C. elegans* PAM-1 isoform b (i.e. puromycin-sensitive aminopeptidase), which is involved in the degradation of cyclin (B3) and needed for the completion of meiosis, consistent with evidence for mammals

(Constam et al., 1995). PAM-1 is recognised as an essential molecule (Brooks et al., 2003; Lyczak et al., 2006), with embryonic lethality upon gene knockdown; mutant embryos fail to exit meiosis following fertilization (McCloskey and Kempfues, 2012). In *C. elegans*, PAM-1 is expressed in the intestine and nerve cells of the head throughout development (Brooks et al., 2003) and has also been localised to spermatids in the spermatheca as well as developing oocytes/embryos (see Fortin et al., 2010), indicating both paternal and maternal roles for this molecule. Interestingly, the *pam-1* gene has known interactions with *par-2*, *par-3*, *pgl-1* and *pie-1*, and is predicted to interact with *fust-1*, *his-71*, *his-72*, *mdt-10* and *prp-6*, many of which have key roles in development, reproduction and/or embryonic polarity (e.g., Huang et al., 2002; Brooks et al., 2003; Lyczak et al., 2006; Fortin et al., 2010).

Hc-AP-11 is a homolog of the *C. elegans* gene designated AC3.5 isoform b, which encodes a human leucyl/cystinyl aminopeptidase ortholog. Based on its domains, this protein is predicted to have metallopeptidase activity as well as zinc ion-binding activity; this gene does not appear to have interactions with other genes in *C. elegans*.

Hc-AP-12 is a homolog of *C. elegans* Y67D8C.9 isoform b, which encodes an ortholog of human aminopeptidase Q with likely metallopeptidase activity and zinc ion-binding activity based on its protein domains. In *C. elegans*, gene knockdown experiments using RNAi (Althoff et al., 2014) have shown that this protein affects the initiation of nucleolar disassembly during oogenesis, resulting in a significant decrease in brood size and oocyte maturation, indicating a role for this molecule in gonad function.

Hc-AP-13 shows 58% sequence similarity to aminopeptidase AP-1 encoded by the *C. elegans* orphan gene C42C1.11 isoform a, which encodes a human leukotriene A4 (LTA4) hydrolase ortholog (Baset et al., 1998). Although its function in *C. elegans* is currently unknown, this aminopeptidase is believed to have evolved to gain hydrolase activity (Baset et al., 1998). In addition, *Hc-AP-13* shows slightly less (50%) sequence similarity/identity to the protein predicted from encoded by ZC416.6 isoform a, and relates to bifunctional leukotriene A4 hydrolases (Benner et al., 2011). In *C. elegans*, knockdown by double-stranded RNA interference (RNAi) results in decreased intestinal dipeptide transport, and increased free fatty acid uptake and fat storage (Benner et al., 2011); this phenotype resembles loss-of-function of the intestinal peptide transporter PEPT-1 (Spanier et al., 2009; Benner et al., 2011; Geillinger et al., 2014). Therefore, it is proposed that *H. contortus* *Hc-AP-13* and its *C. elegans* homologs regulate PEPT-1 transporter activity via the regulation of intracellular amino acids levels and, ultimately, of the PEPT-1 amount on membranes of intestinal cells. PEPT-1 itself is a high capacity/low affinity oligopeptide transporter (Benner et al., 2011), which localises to the apical membrane and is required for the uptake of intact peptides from the intestine, and is therefore essential for normal development, growth and reproduction. In addition, the *pept-1* gene has been shown to interact with both the *let-363/TOR* and *daf-2/insulin*-signalling pathways to regulate lifespan, and to influence stress response via the *daf-2* pathway (Nehrke, 2003; McElwee et al., 2004; Meissner et al., 2004; Hertweck and Baumeister, 2005; Walker et al., 2005). Together with transcription data (Fig. 2), this information suggests that *Hc-AP-13* operates, possibly synergistically with a PEPT-1 ortholog in *H. contortus* (cf. Schwarz et al., 2013), as an oligopeptide transporter in the gut of *H. contortus* to ensure adequate protein uptake (from various sources, including microbes in free-living stages, and blood in parasitic stages) for proper development, growth and/or reproduction of the worm.

7. Structural appraisal of selected M1 aminopeptidases

Family M1 aminopeptidases occur throughout all kingdoms (archaea, bacteria and eukaryotes), but only the 12 mammalian members (Table 3) have been characterised in detail. Structurally, the mammalian M1 aminopeptidases can be classified into two groups that differ in the number of their domains (Table 3). Currently, three-dimensional structures are available for eight M1 aminopeptidases from different organisms: Ap-N (Wong et al., 2012), LTA₄H

(Thunnissen et al., 2001), ERAP1 (Nguyen et al., 2011) and ERAP2 (Birtley et al., 2012) from humans as well as Ap-N from pigs (Chen et al., 2012), *Plasmodium falciparum* (see McGowan et al., 2009), *E. coli* (see Addlagatta et al., 2006) and *Neisseria meningitidis* (see Nocek et al., 2008). Structure-based alignment of amino acid sequences representing the M1 aminopeptidases predicted in this study (Supplementary Fig. 1) shows that all proteins, except *Hc*-AP-12 and *Hc*-AP-13, adopt the same topology as the Ap-N subgroup of the mammalian M1 aminopeptidases with four structural domains (illustrated by *Hc*-AP-2 in Fig. 5, left panel). The topology of *Hc*-AP-13 shows more similarity to the LTA₄H subgroup, which is characterised by three domains (Fig. 5, central panel). In contrast, the folding of *Hc*-AP-13 is less clear due to the absence of domain I and a long insertion within domain III, with the predicted secondary structure deviating from the typical pattern of that domain (Fig. 5, right panel). It is likely that the three-dimensional structure of *Hc*-AP-12 will possess a different type of domain III than those known for the Ap-N subgroup from mammals (Chen et al., 2012; Wong et al., 2012). Notably, an arginine-aspartate motif located in the first domain of aminopeptidase A has been proposed to be crucial for the orientation of the substrate (Rozenfeld et al., 2002) and is strictly conserved among all predicted aminopeptidases, except *Hc*-AP-12 (Supplementary Fig. 1) that lacks domain I. The important functional residues required for metal ion coordination, proteolysis and stabilisation of the transition state are fully conserved among all predicted proteins, with the exception of *Hc*-AP-11. The latter protein possesses a lysine in position 519 rather than the expected histidine (metal coordination), and asparagine at position 601, instead of tyrosine (transition state stabilisation). These variations indicate that the proteolytic ability of *Hc*-AP-11, as a functional aminopeptidase, might be impaired. Based on observations made in an early study (Smith et al., 1997), H11 is generally regarded as a type II integral membrane protein with a single transmembrane helix, a short N-terminal region in the cytoplasm and remaining domains located extracellularly. Notably, all previously reported H11 sequences (i.e. H11-1 to H11-5) adopt this topology. Within the complement of family M1 aminopeptidases predicted here, we identified seven molecules (*Hc*-AP-2, -3, -6, 9, -10, -12 and -13) that lack the transmembrane domain and are predicted to be soluble (see Table 1; Supplementary Fig. 1).

8. Concluding remarks

Although an anti-*H. contortus* vaccine (Barbervax; based on the molecular complex H-gal-GP; cf. Knox, 2011) was released in Australia in 2014 to support parasite treatment programs, its efficacy in routine, large scale applications in the field is yet unknown. Consequently, work toward effective recombinant vaccines needs to continue, given the substantial problems associated with drug resistance in *H. contortus* and related nematodes (reviewed by Wolstenholme et al., 2004; Kaplan and Vidyashankar, 2012; Wolstenholme and Martin, 2014).

Extending previous studies of H11 (reviewed by Newton and Meeusen, 2003) and employing our genomic and transcriptomic resources (Schwarz et al., 2013), we curated and defined the complement of aminopeptidases of the family M1 of *H. contortus*. Being enzymes that hydrolyse peptides, aminopeptidases are often of crucial importance for the uptake of nutrition by an organism. Capitalising on this pivotal role, these enzymes have been popular targets for development of therapeutics against cancer and infectious diseases (Hitzerd et al., 2014; Stratikos, 2014; Atkinson et al., 2015; Jengarn et al., 2015). In this context, *H. contortus* H11 has been a major focus of many efforts aimed at developing a vaccine against haemonchosis. However, in contrast to native H11, no satisfactory level of protection was achieved by immunisation of sheep with recombinant forms of H11 produced in *E. coli*, insect cells or *C. elegans* (see Newton and Meeusen, 2003; Roberts et al., 2013). Although this problem might relate to biotechnological issues, such as the absence of the native

carbohydrate conjugates and/or incorrectly folded epitopes, the findings of this article might suggest a more complex situation.

Interestingly, the present findings showed that none of the aminopeptidases curated here was identical to any previously identified H11 isoforms (H11-1 to H11-5; Table 2; Fig. 3B), indicating sequence variation in H11 isoforms among distinct *H. contortus* populations from distinct geographic provenances. In our opinion, prior to further work on H11 and related aminopeptidases (family M1), it would be important and prudent to establish sequence conservation (or otherwise) for individual molecules within and among individual worms, and within and among different populations of *H. contortus* from around the globe. Given the amount of work and expense required to develop recombinant vaccines, this would seem to be a worthwhile endeavour.

Although it is not possible to exclude the basic biotechnological issues surrounding glycosylation and protein-folding that might account for the lack of success of recombinant H11 vaccine candidates assessed thus far, there are two other aspects that might have implications in the search of an aminopeptidase-based vaccine. First, we predicted 12 of the 13 aminopeptidases (family M1) from *H. contortus* to be catalytically active, which leaves ample opportunity for functional redundancy in the collective protein target. Second, it is possible that some effective immunogens are amongst the complement of M1 aminopeptidases predicted here, but have not yet been studied. Although we appreciate that native H11 (hidden antigen complex) isolated from membrane-bound components of *H. contortus* gut was an effective vaccine, it is tempting to hypothesize, based on the present findings, that related aminopeptidases with secretory signals, such as *Hc*-AP-4 and *Hc*-AP-5 (Table 1; Fig. 1), might also be antigenic or immunogenic. This aspect might be worthy of investigation, following the localisation of individual aminopeptidases (*Hc*-AP-1 to *Hc*-AP-13) in different stages, sexes and tissues and their characterisation in secretions from *H. contortus*. In this context, a cocktail of distinct, conserved epitopes, sourced from several of the aminopeptidases characterised here, might have some promise for designing a recombinant vaccine in the future. This approach could have merit, given recent success achieved for *Teladorsagia circumcincta* with a cocktail of recombinant proteins (Nisbet et al., 2013).

Acknowledgements

Research funding from the Australian Research Council (ARC) and National Health and Medical Research Council of Australia (NHMRC) is gratefully acknowledged (R.B.G.). This project was also supported by a Victorian Life Sciences Computation Initiative (VLSCI) (grant number VR0007) on its Peak Computing Facility at the University of Melbourne, an initiative of the Victorian Government (R.B.G.). N.D.Y. holds an NHMRC Early Career Research Fellowship. P.K.K. is the recipient of a scholarship (STRAPA) from the University of Melbourne.

References

- Addlagatta, A., Gay, L., Matthews, B.W., 2006. Structure of aminopeptidase N from *Escherichia coli* suggests a compartmentalized, gated active site. *Proc. Natl. Acad. Sci. USA* 103,13339-13344.
- Althoff, M.J., Flick, K., Trzepacz, C., 2014. Collaboration within the M1 aminopeptidase family promotes reproductive success in *Caenorhabditis elegans*. *Dev. Genes Evol.* 224, 137-146.
- Atkinson, S.C., Armistead, J.S., Mathias, D.K., Sandeu, M.M., Tao, D., Borhani-Dizaji, N., Tarimo, B.B., Morlais, I., Dinglasan, R.R., Borg, N.A., 2015. The Anopheles-midgut APN1 structure reveals a new malaria transmission-blocking vaccine epitope. *Nat. Struct. Mol. Biol.* 22, 532-539.
- Baset, H. A., Ford-Hutchinson, A. W., O'Neill, G. P., 1998. Molecular cloning and functional expression of a *Caenorhabditis elegans* aminopeptidase structurally related to mammalian leukotriene A4 hydrolases. *J. Biol. Chem.* 273, 27978-27987.
- Benner, J., Daniel, H., Spanier, B., 2011. A glutathione peroxidase, intracellular peptidases and the TOR complexes regulate peptide transporter PEPT-1 in *C. elegans*. *PLoS One* 6, e25624.
- Birtley, J.R., Saridakis, E., Stratikos, E., Mavridis, I.M., 2012. The crystal structure of human endoplasmic reticulum aminopeptidase 2 reveals the atomic basis for distinct roles in antigen processing. *Biochemistry* 51, 286-295.
- Brooks, D.R., Hooper, N.M., Isaac, R.E., 2003. The *Caenorhabditis elegans* orthologue of mammalian puromycin-sensitive aminopeptidase has roles in embryogenesis and reproduction. *J. Biol. Chem.* 278, 42795-42801.
- Bryson, K., McGuffin, L.J., Marsden, R.L., Ward, J.J., Sodhi, J.S., Jones, D.T., 2005. Protein structure prediction servers at University College London. *Nucl. Acids Res.* 33, W36-38.
- Cantacessi, C., Campbell, B.E., Gasser, R.B., 2012. Key strongylid nematodes of animals – Impact of next-generation transcriptomics on systems biology and biotechnology. *Biotechnol. Adv.* 30, 469-88.
- Chen, L., Lin, Y.L., Peng, G., Li, F., 2012. Structural basis for multifunctional roles of mammalian aminopeptidase N. *Proc. Natl. Acad. Sci. USA.* 109, 17966-17971.
- Constam, D.B., Tobler, A.R., Rensing-Ehl, A., Kemler, I., Hersh, L.B., Fontana, A., 1995. Puromycin-sensitive aminopeptidase. Sequence analysis, expression, and functional characterization. *J. Biol. Chem.* 270, 26931-9.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792-1797.
- Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research* 39, W29-W37.
- Fortin, S.M., Marshall, S.L., Jaeger, E.C., Greene, P.E., Brady, L.K., Isaac, R.E., Schrandt, J.C., Brooks, D.R., Lyczak, R., 2010. The PAM-1 aminopeptidase regulates centrosome positioning to ensure anterior-posterior axis specification in one-cell *C. elegans* embryos. *Dev. Biol.* 344, 992-1000.
- Gasser, R.B., Schwarz, E.M., Korhonen, P.K., Young, N.D., 2016. Understanding *Haemonchus contortus* better through genomics and transcriptomics. *Adv. Parasitol.* R0_submitted 1 August 2016. To be published in 2016.
- Geillinger, K.E., Kuhlmann, K., Eisenacher, M., Giesbertz, P., Meyer, H. E., Daniel, H., Spanier, B., 2014. Intestinal amino acid availability via PEPT-1 affects TORC1/2 signaling and the unfolded protein response. *J. Proteome Res.* 13, 3685-3692.

- Geldhof, P., Clark, D., Molloy, C., Knox, D.P., 2007. Assessment of *Caenorhabditis elegans* as a model in *Haemonchus contortus* vaccine research. *Mol. Biochem. Parasitol.* 152, 220-223.
- Graham, M., Smith, T.S., Munn, E.A., Knox, D.P., Oliver, J.J., Newton SE., 1993. Recombinant DNA molecules encoding amino-peptidase enzymes and their use in the preparation of vaccines against helminth infections. Patent no. WO 93/23542.
- Haslam, S.M., Coles, G.C., Munn, E.A., Smith, T.S., Smith, H.F., Morris, H.R., Dell, A., 1996. *Haemonchus contortus* glycoproteins contain N-linked oligosaccharides with novel highly fucosylated core structures. *J. Biol. Chem.* 271, 30561–30570.
- Hertweck, M., Baumeister, R., 2005. Automated assays to study longevity in *C. elegans*. *Mech. Ageing Dev.* 126, 139-145.
- Hitzerd, S.M., Verbrugge, S.E., Ossenkoppele, G., Jansen, G., Peters, G.J., 2014. Positioning of aminopeptidase inhibitors in next generation cancer therapy. *Amino Acids*, 46, 793-808.
- Holt, C., Yandell, M., 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12, 491.
- Huang, N.N., Mootz, D.E., Walhout, A.J., Vidal, M., Hunter, C.P., 2002. MEX-3 interacting proteins link cell polarity to asymmetric gene expression in *Caenorhabditis elegans*. *Development* 129, 747-759.
- Iturrioz, X., Rozenfeld, R., Michaud, A., Corvol, P., Llorens-Cortes, C., 2001. Study of asparagine 353 in aminopeptidase A: characterization of a novel motif (GXMEN) implicated in exopeptidase specificity of monozinc aminopeptidases. *Biochemistry* 40, 14440-14448.
- Jengarn, J., Wongthida, P., Wanasen, N., Frantz, P.N., Wanitchang, A., Jongkaewwattana. A., 2015. Genetic manipulation of porcine epidemic diarrhea virus (PEDV) recovered from a full-length infectious cDNA clone. *J. Gen. Virol.*, In press. doi: 10.1099/vir.0.000184.
- Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236-1240.
- Kaplan, R.M., Vidyashankar, A.N., 2012. An inconvenient truth: global worming and anthelmintic resistance. *Parasitology* 186, 70-78.
- Kent, W.J., 2002. BLAT - The BLAST-like alignment tool. *Genome Res.* 12, 656-664.
- Knox, D.P., 2011. Proteases in blood-feeding nematodes and their potential as vaccine candidates. *Adv. Exp. Med. Biol.* 712, 155-176.
- Knox, D.P., Redmond, D.L., Newlands, G.F., Skuce, P.J., Pettit, D., Smith, W.D., 2003. The nature and prospects for gut membrane proteins as vaccine candidates for *Haemonchus contortus* and other ruminant trichostrongyloids. *Int. J. Parasitol.* 33, 1129-1137.
- Laing, R., Kikuchi, T., Martinelli, A., Tsai, I.J., Beech, R.N., Redman, E., Holroyd, N., Bartley, D.J., Beasley, H., Britton, C., Curran, D., Devaney, E., Gilabert, A., Hunt, M., Jackson, F., Johnston, S.L., Kryukov, I., Li, K., Morrison, A.A., Reid, A.J., Sargison, N., Saunders, G.I., Wasmuth, J.D., Wolstenholme, A., Berriman, M., Gilleard, J.S., Cotton, J.A., 2013. The genome and transcriptome of *Haemonchus contortus*, a key model parasite for drug and vaccine discovery. *Genome Biol.* 14, R88.
- Laustsen, P.G., Vang, S., Kristensen, T., 2001. Mutational analysis of the active site of human insulin-regulated aminopeptidase. *Eur. J. Biochem.* 268, 98-104.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079.
- Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589-595.

- Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., Borodovsky, M., 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* 33, 6494-6506.
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi.org/10.1186/s13059-014-0550-8.
- Lyczak, R., Zweier, L., Group, T., Murrow, M.A., Snyder, C., Kulovitz, L., Beatty, A., Smith, K., Bowerman, B., 2006. The puromycin-sensitive aminopeptidase PAM-1 is required for meiotic exit and anteroposterior polarity in the one-cell *Caenorhabditis elegans* embryo. *Development* 133, 4281-4292.
- McCloskey, R.J., Kempfues, K.J., 2012. Deubiquitylation machinery is required for embryonic polarity in *Caenorhabditis elegans*. *PLoS Genet.* 8, e1003092.
- McElwee, J.J., Schuster, E., Blanc, E., Thomas, J.H., Gems, D., 2004. Shared transcriptional signature in *Caenorhabditis elegans* dauer larvae and long-lived daf-2 mutants implicates detoxification system in longevity assurance. *J. Biol. Chem.* 279, 44533-44543.
- McGowan, S., Porter, C.J., Lowther, J., Stack, C.M., Golding, S.J., Skinner-Adams, T.S., Trenholme, K.R., Teuscher, F., Donnelly, S.M., Grembecka, J., Mucha, A., Kafarski, P., Degori, R., Buckle, A.M., Gardiner, D.L., Whisstock, J.C., Dalton, J.P., 2009. Structural basis for the inhibition of the essential *Plasmodium falciparum* M1 neutral aminopeptidase. *Proc. Natl. Acad. Sci. USA* 106, 2537-2542.
- Meissner, B., Boll, M., Daniel, H., Baumeister, R., 2004. Deletion of the intestinal peptide transporter affects insulin and TOR signaling in *Caenorhabditis elegans*. *J. Biol. Chem.* 279, 36739-45.
- Mi, H., Muruganujan, A., Casagrande, J.T., Thomas, P.D., 2013. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* 8, 1551-1566.
- Munn, E.A., Munn, P.D., 2002. In *The Biology of Nematodes*, Lee, D.L. (ed.). London: Taylor and Francis; pp. 211-232.
- Nehrke, K., 2003. A reduction in intestinal cell pH_i due to loss of the *Caenorhabditis elegans* Na⁺/H⁺ exchanger NHX-2 increases life span. *J. Biol. Chem.* 278, 44657-66.
- Newton, S.E., Munn, E.A., 1999. The development of vaccines against gastrointestinal nematode parasites, particularly *Haemonchus contortus*. *Parasitol. Today* 15, 116-22.
- Newton, S.E., Meeusen, E.N., 2003. Progress and new technologies for developing vaccines against gastrointestinal nematode parasites of sheep. *Parasite Immunol.* 25, 283-96.
- Nguyen, T.T., Chang, S.C., Evnouchidou, I., York, I.A., Zikos, C., Rock, K.L., Goldberg, A.L., Stratikos, E., Stern, L.J., 2011. Structural basis for antigenic peptide precursor processing by the endoplasmic reticulum aminopeptidase ERAP1. *Nat. Struct. Mol. Biol.* 18, 604-613.
- Nisbet, A.J., McNeilly, T.N., Wildblood, L.A., Morrison, A.A., Bartley, D.J., Bartley, Y., Longhi, C., McKendrick, I.J., Palarea-Albaladejo, J., Matthews, J.B., 2013. Successful immunization against a parasitic nematode by vaccination with recombinant proteins. *Vaccine* 31, 4017-4023.
- Nocek, B., Mulligan, R., Bargassa, M., Collart, F., Joachimiak, A., 2008. Crystal structure of aminopeptidase N from human pathogen *Neisseria meningitidis*. *Proteins* 70, 273-279.
- Roberts, B., Antonopoulos, A., Haslam, S.M., Dicker, A.J., McNeilly, T.N., Johnston, S.L., Dell, A., Knox, D.P., Britton, C., 2013. Novel expression of *Haemonchus contortus* vaccine candidate aminopeptidase H11 using the free-living nematode *Caenorhabditis elegans*. *Vet. Res.* 44, 111.
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.

- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative Genomics Viewer. *Nature Biotechnol.* 29, 24–26.
- Ronquist, F., Huelsenbeck, J., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.
- Rozenfeld, R., Iturrioz, X., Maigret, B., Llorens-Cortes, C., 2002. Contribution of molecular modelling and site-directed mutagenesis to the identification of two structural residues, Arg-220 and Asp-227 in aminopeptidase A. *J. Biol. Chem.* 277, 29242–29252.
- Sali, A., Blundell, T., 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815.
- Schwarz, E.M., Korhonen, P.K., Campbell, B.E., Young, N.D., Jex, A.R., Jabbar, A., Hall, R.S., Mondal, A., Howe, A.C., Pell, J., Hofmann, A., Boag, P.R., Zhu, X.Q., Gregory, T.R., Loukas, A., Williams, B.A., Antoshechkin, I., Brown, C.T., Sternberg, P.W., Gasser, R.B., 2013. The genome and developmental transcriptome of the strongylid nematode *Haemonchus contortus*. *Genome Biol.* 14, R89.
- Smith, T.S., Munn, E.A., Graham, M., Tavernor, A.S., Greenwood, C.A., 1993. Purification and evaluation of the integral membrane protein H11 as a protective antigen against *Haemonchus contortus*. *Int. J. Parasitol.* 23, 271–280.
- Smith, T.S., Graham, M., Munn, E.A., Newton, S.E., Knox, D.P., Coadwell, W.J., McMichael-Phillips, D., Smith, H., Smith, W.D., Oliver, J.J., 1997. Cloning and characterization of a microsomal aminopeptidase from the intestine of the nematode *Haemonchus contortus*. *Biochim. Biophys. Acta* 1338, 295–306.
- Spanier, B., Lasch, K., Marsch, S., Benner, J., Liao, W., Hu, H., Kienberger, H., Eisenreich, W., Daniel, H., 2009. How the intestinal peptide transporter PEPT-1 contributes to an obesity phenotype in *Caenorhabditis elegans*. *PLoS One* 4, e6279.
- Stratikos, E., 2014. Regulating adaptive immune responses using small molecule modulators of aminopeptidases that process antigenic peptides. *Curr. Opin. Chem. Biol.* 23, 1–7.
- Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O., Borodovsky, M., 2008. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. *Genome Res.* 18, 1979–1990.
- Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., Narechania, A., 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13, 2129–2141.
- Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2012. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics* 14, 178–192.
- Thunnissen, M.M., Nordlund, P., Haeggström, J.Z., 2001. Crystal structure of human leukotriene A(4) hydrolase, a bifunctional enzyme in inflammation. *Nat. Struct. Biol.* 8, 131–135.
- Vazeux, G., Iturrioz, X., Corvol, P., Llorens-Cortes, C., 1997. A tyrosine residue essential for catalytic activity in aminopeptidase A. *Biochem. J.* 327, 883–889.
- Walker, G., Houthoofd, K., Vanfleteren, J.R., Gems, D., 2005. Dietary restriction in *C. elegans*: from rate-of-living effects to nutrient sensing pathways. *Mech. Ageing Dev.* 126, 929–937.
- Wang, C.K., Broder, U., Weeratunga, S., Gasser, R.B., Loukas, A., Hofmann, A., 2012. SBAL: a practical tool to generate and edit structure-based amino acid sequence alignments. *Bioinformatics* 28, 1026–1027.
- Wang, C.K., Hofmann, A., 2015. An integrated Java tool for generating amino acid sequence alignments with mapped secondary structure elements. *Biotechnology* 5, 87–92.
- Wolstenholme, A.J., Fairweather, I., Prichard, R., von Samson-Himmelstjerna, G., Sangster, N.C., 2004. Drug resistance in veterinary helminths. *Trends Parasitol.* 20, 469–476.
- Wolstenholme, A.J., Martin, R.J., 2014. Anthelmintics – from discovery to resistance. *Int. J. Parasitol. Drugs Drug Resist.* 4, 218–219.

- Wong, A.H., Zhou, D., Rini, J.M., 2012. The X-ray crystal structure of human aminopeptidase N reveals a novel dimer and the basis for peptide processing. *J. Biol. Chem.* 287, 36804-36813.
- Yang, J., Yan, R., Roy, R., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods* 12, 7-8.
- Zdobnov, E.M., Apweiler, R., 2001. InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847-848.
- Zhou, Q.J., Zhang, H.L., Jiang, X.L., Du, A.F., 2010. The gene structure and promoter region of the vaccine target aminopeptidase H11 from the blood-sucking nematode parasite of ruminants, *Haemonchus contortus*. *Funct. Integr. Genomics* 10, 589-601.

TABLES AND FIGURES

Table 1

Full length transcripts representing the different *Hc-ap* genes and the genomic scaffolds to which they map, including the lengths of inferred amino acid sequences and InterProScan matches (including domains and motifs) for individual predicted proteins (this study) as well as for published H11 isoforms (Smith et al., 1997; Zhou et al., 2010) and *Caenorhabditis elegans* homologs (Baset et al., 1998; Lyczak et al., 2006; Geldhof et al., 2007; Benner et al., 2011).

[-] not present; SF = subfamily

Table 2

Pairwise comparisons of sequence differences (%) among 13 predicted *Hc*-AP (family M1) protein sequences of *H. contortus*, published H11 isoforms and other aminopeptidases from *Caenorhabditis elegans*.

Supplementary Table 1

The 12 mammalian aminopeptidases (family M1) can be classified into two subgroups based on the number of domains.

Aminopeptidase	Abbreviation	Domains
Aminopeptidase A	Ap-A	Four (I, II, III & IV)
Aminopeptidase N	Ap-N	
Aminopeptidase Q	Ap-Q	
Endoplasmic reticulum aminopeptidase 1	ERAP1	
Endoplasmic reticulum aminopeptidase 2	ERAP2	
Insulin-responsive aminopeptidase	IRAP	
Puromycin-sensitive aminopeptidase	PSA	
Thyrotropin-releasing hormone degrading enzyme	TRH-DE	
Aminopeptidase B	Ap-B	Three (I, II & IV)
Aminopeptidase O	Ap-O	
Leukotriene A ₄ hydrolase	LTA ₄ H	
Arginyl aminopeptidase-like 1	RNPEP-L1	

Fig. 1. A schematic representation of members of the aminopeptidase family M1 of *Haemonchus contortus* inferred from full-length transcripts and their functional domains and motifs as predicted by InterProScan (cf. Table 1). Signal peptide (SP); coiled coil (CC); cytoplasmic domain (CD); transmembrane domain (TM); PTHR11533 represents the family of M1 zinc metalloproteases; bars filled with grey or black/white patterns represent five distinct subfamily (SF) members.

Fig. 2. Fig. 2A: Differential gene transcription where the log fold change of each gene is plotted against the total number of counts recorded for that gene. Differentially expressed genes (FDR < 0.05) are marked in red. The predicted *Hc-ap* genes are marked in blue. DeSeq2 is an R programming language package that analyzes differential gene expression based on negative binomial distribution. Fig. 2B: Heat-map displaying transcription profiles for genes *Hc-ap-1* to *Hc-ap-13*; each of the genes *Hc-ap-1* to *Hc-ap-8* (see panel A) are differentially transcribed between parasitic and free living stages of *H. contortus*.

Fig. 3. Relationships among predicted amino acid sequences of 13 aminopeptidases (family M1) of *Haemonchus contortus* (Haecon5) (Fig. 3A) as well as published H11 sequences (Fig. 3B).

Fig. 4. Phylogenetic analysis of all 13 aminopeptidase (family M1) sequences predicted from *Haemonchus contortus* and their respective homologs from *Caenorhabditis elegans*.

Fig. 5. Homology models for *Hc*-AP proteins. Left: Proteins *Hc*-AP-1 to *Hc*-AP-11 are likely to exhibit the same topology as the Ap-N subgroup of the mammalian M1 aminopeptidases with four structural domains. The model shown is a cartoon of *Hc*-AP-2 with the four domains indicated by different colours (I: blue, II: green, III: orange and IV: red). Centre: *Hc*-AP-13 is more similar to the three-domain aminopeptidases. The cartoon model is coloured to indicate the different domains (I: blue, II: green and IV: red). Right: *Hc*-AP-12 is likely to possess three structural domains, but lacks the first domain shared by all other aminopeptidases. Insertion of a sequence of ~50 amino acids in domain III (represented by a question mark), which is predicted to be largely unstructured (see Supplementary Fig. 1), renders the three-dimensional structure of this protein ambiguous.

Supplementary Fig. 1. Structure-based amino acid sequence alignment of the H11 proteins. Amino acid sequences of the 13 *Hc*-AP proteins identified in this study (*Hc*-AP-1 to *Hc*-AP-13), *Hc*-H11 sequences from public databases (H11-1: accession no. AJ249941, H11-2: AJ249942, H11-3: AJ311316, H11-4: FJ481146 and H11-5: X94187; cf. Graham et al., 1993; Smith et al., 1997; Zhou et al., 2010) and the sequence of human aminopeptidase N (PDB 4fyt) were aligned, as described in section 2 (Materials and methods). Alpha-helical (green) and beta-strand (red) secondary structure elements (as observed in the experimental structure for human aminopeptidase N, predicted for all other sequences) are mapped; cysteine residues (yellow) and the transmembrane helices are indicated (light blue background). Annotation: 'm' indicates metal-coordinating residues, 'a' the catalytically active glutamate residue, 's' the tyrosine residue stabilising the transition state (Vazeux et al., 1997), and 'o' the arginine-aspartate motif implicated in substrate orientation. The image was prepared using the program SBAL (Wang et al., 2012).