Tropical Biomedicine 34(4): 1000–1013 (2017)

Cloning and characterization of recombinant Hc-CPL-1 from gastrointestinal nematode *Haemonchus contortus* isolate from goat (*Capra hircus*) population in Penang, Malaysia

Abdul Latif, M.N.F.[1,2], Ismail, M.N.[2], Abdul Rahman, A.W.[1] and Yahaya, Z.S.[1*]
[1]School of Biological Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia
[2]Analytical Biochemistry Research Centre, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia
[*]Corresponding author e-mail: zary@usm.my
Received 16 January 2017; received in revised form 30 June 2017; accepted 1 July 2017

Abstract. Cathepsin L (CPL) cysteine protease is a proteolytic enzyme that involves in many biological processes in a wide range of organisms. In free-living nematode *Caenorhabditis elegans*, CPL plays important roles in embryogenesis and development processes. The CPL protein is also believed to have a role in degradation of blood meal in the gut of parasitic nematodes. Considering this enzyme might play the same functions in parasitic nematodes, CPL became a potential candidate for vaccination against *Haemonchus contortus*, a gastrointestinal nematode of small ruminants. *H. contortus* has been shown to have variations in term of morphology and genetic materials that correlated with different hosts and geographical areas. These variations could hinder the development of effective vaccines. Thus, the present study was conducted to clone and characterize recombinant Hc-CPL-1 from *H. contortus* isolated from a goat population in Penang, Malaysia. Reverse transcription polymerase chain reaction (RT-PCR) was used to amplify target complimentary DNA (cDNA) from total RNA and protein expression using *Escherichia coli* expression system was performed from constructed cDNA clone library. The identity of each protein band was confirmed by liquid chromatography-mass spectrometry (LC-MS) analysis followed by *De novo* sequencing and database matching. The protein structure and its evolutionary relationship were also studied using several bioinformatics approaches. Basic local alignment search tool (BLAST) analysis of the strain retrieved from clone library showed 99% sequence similarity to the *Haemonchus* cathepsin L cysteine protease and a 47 kDA protein was successfully expressed. Bioinformatic analysis indicated that this protease has a close relationship with Dv-CPL-1, Sv-CPL-1 and Ce-CPL-1. These data might provide an insight on manipulating this enzyme for future novel vaccine development.

INTRODUCTION

High prevalence of anthelmintic drug resistance among parasitic nematodes of small ruminants has been reported worldwide (Geerts & Gryseels, 2000). This includes resistance by parasitic nematode of sheep and goat, *Haemonchus contortus*, toward various anthelmintic drugs such as benzimidazoles, ivermectin and levamisole in South Africa, Paraguay, Uruguay, Brazil and Argentina (Jackson & Coop, 2000). Resistance has also been reported in all farms in Malaysia as well as temperate regions such as Sweden and Northern Europe (Waller & Chandrawathani, 2005). *H. contortus*, also known as the 'barber's pole' worm, is a bloodsucker that has caused significant losses worldwide to livestock due to its morbidity and mortality effects (Wahab & Suhaila, 2007; Chandrawathani *et al.*, 2009). The need for alternative treatment against this parasite has become urgent as the drug-resistance in this parasite has spread rapidly worldwide.

Since the last few decades, vaccination applications using parasite molecules have been studied to combat *H. contortus* infections (Smith & Munn, 1990). Among them, cysteine proteases became one of the major targets that gave promising results in the protective studies (Redmond & Knox, 2006; Muleke *et al.*, 2007; Murray *et al.*, 2007) with overall percentage of reduction in faecal egg count (FEC) and worm burdens up to 77% and 47%, respectively (Knox *et al.*, 1999). The cysteine protease or cathepsin is one of the important excretory/secretory (ES) proteins which fall into the papain protein family together with other major protein groups like calpains, caspases and bleomycin hydrolases (Greenbaum *et al.*, 2000). The common and well known members in this family includes cathepsin B, C, H, K, L, N, S and Z in which all of them contained an essential cysteine residue inside their active sites to carry their functions (Bogyo *et al.*, 2000; Hashmi *et al.*, 2002). Although there is a similarity in the residue of their active sites, cysteine proteases would differ in term of substrate specificity, pH stability and also in tissue distribution (Hashmi *et al.*, 2002).

Cathepsin L (CPL) cysteine protease in nematodes was speculated to be involved in the development, feeding and moulting processes where such roles were highly conserved between some related species of nematodes (Britton & Murray, 2002). Britton & Murray (2004) showed that CPL is strongly related with the embryogenesis and larval development of the nematode through the correct yolk processing. In addition, CPL also involved in the degradation of blood meal in the gut of parasitic nematodes (Brady *et al.*, 1999) and the development of stage 3 larvae (L3) into stage 4 (L4) through the moulting process depending on the level of this protease in the cuticle regions (Lustigman *et al.*, 1996).

According to Knox *et al.* (2001), the differences between geographical strains of nematodes would hinder development of worldwide effective vaccines against *H. contortus* infections. As shown by Wahab *et al.* (2007), the genetic material of *H. contortus* isolated in Malaysia itself slightly differs between hosts (goat and sheep). Moreover, morphological variation of *H. contortus* also have been observed between Malaysian isolate with those reported in Yemen (Gharamah *et al.*, 2012), Australia and North America (Wahab & Suhaila, 2007). Thus, in this study we aimed to clone and express the CPL protein obtained from Malaysian isolate of *H. contortus* among goat population as a comparison with those isolated from other countries. Mass spectrometric and bioinformatic approaches were also employed to characterize this protease.

MATERIALS AND METHODS

Haemonchus contortus L3 culture

Haemochus contortus L3 were obtained from goat faecal culture using modified Bearmann technique. The fecal samples were collected from one traditional husbandry located at Bayan Lepas in Penang, Malaysia. After 7 days of incubation with rapid monitoring of the moisture condition, L3 larvae were obtained by decantation after left to migrate for 3 hours.

The larvae were cleaned with sterile 0.2% sodium hypochlorite/phosphate buffer saline (PBS) solution. Then, followed by three washes with PBS containing Penicillin (250 U/ml), Streptomycin (50 µg/ml) and Amphotericin B (1.25 µg/ml) in order to concentrate the larvae and minimize contamination particles.

Total RNA Extraction

Water containing L3 larvae was centrifuged for 5 min at 5000 rpm to separate the larvae from the supernatant without breaking up their cuticle. Total RNA was extracted using prescribed protocol of RNAqueos Micro Scale Isolation Kit (Ambion, USA). The larvae cuticles were disrupted in 10 volume of guanidinium thiocyanate solution per mass of the tissue (10 mg equivalent to 100 µl). Total RNA was eluted twice with 20 µl preheated elution solution and immediately treated with DNase to remove trace amount of contaminating genomic DNA. The reaction solution was incubated at 37°C for 20 min before DNase was deactivated using DNase

Inactivation Reagent at room temperature for 2 min. The DNase Inactivation Reagent was pelleted by centrifugation and RNA was quantified by spectrophotometer at 260 nm absorbance.

RT-PCR and cloning of *Hc-CPL-1* cDNA

Total RNA (20 ng) was reverse transcribed using RT-PCR to produce cDNA of *H. contortus* cathepsin L (Hc-CPL-1) following the protocol of Superscript III One-Step RT-PCR System (Invitrogen, USA). A forward (5'-CAAACTAGTATGCTACGTCTGTTGT CGTTGGCGC-3') and reverse (5'-AGCGTCG ACCCTAGTGGTGGTGGTGGTGGTGGTGG TGGTGGTGAACGAGCGGGTAGCTGGCC-3') primers as suggested by Murray *et al.* (2007) were used to amplify the target gene. The RT-PCR was started with 1 cycle of cDNA synthesis at 50°C for 20 min, then pre-denaturation at 94°C for 2 min and followed by 30 cycles of 94°C for 15 s, 53°C for 30 s and 68°C for 60 s. Final extension was set at 68°C for 5 min using thermal cycler machine (BIO-RAD, USA). The PCR product was analyzed by 1.0% (w/v) agarose gel electrophoresis in 1× TBE (Tris-Boric Acid-EDTA) buffer (89 mM Tris, 89 mM Boric Acid, 2 mM Ethylene Diamine Tetraacetic Acid, EDTA at pH 8.3) and cDNA fragment was purified following the protocol of QIAquick Gel Extraction Kit (Qiagen, Germany).

The purified *Hc-CPL-1* cDNA (50 ng) was ligated directly into pGEM-T Easy Vector (Promega, USA) overnight at 4°C and transformed into JM109 *E. coli* competent cells. The transformants were grown at 37°C overnight on lysogeny broth (LB) agar plate supplemented with ampicillin (100 µg/ml), isopropyl β-D-1-thiogalactopyranoside, IPTG (0.5mM) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, X-Gal (80 µg/ml).

DNA sequencing

Plasmids were extracted from the cells using the prescribed protocol of the Wizard Plus SV Minipreps DNA Purification System (Promega, USA). The purified plasmids DNA were subjected to restriction enzyme, *Eco*RI (Promega, USA) digestion to verify the inserted DNA in the plasmids was indeed the *Hc-CPL-1* cDNA by gel electrophoresis. The positive recombinant plasmids were sequenced to confirm the correct gene sequence using universal primers of M13 forward and reverse before aligned with those in the GenBank using BLAST programme.

Construction of recombinant clones

To construct recombinant clones for expression, a new insert gene of *Hc-cpl-1* was generated using PCR with modified pair of primers from the previous RT-PCR. Previously digested gene fragments was amplified using primers Hc-CPL-1-FR1 (5'-CACC**GAATTC**ATGCTACGTCTGTTGTCG TTGG-3'; bold letters indicates restriction site for *Eco*RI used as clamp to anneal with specific site in the plasmid vector; underline indicates start codon) and Hc-CPL-1-RC1 (5'-GGC**GAGCTC**CTAGTGGTGGTGGTGGTGG TGGTGGTGGTGGTGAACGA-3'; bold letters indicates restriction site for *Sac*I; underline indicates complement stop codon).

The purified PCR product was simultaneously digested with two restriction enzymes, *Eco*RI and *Sac*I (NEB, USA) to produce specific nucleotides overhang on both end sites of each DNA fragments. After 5 hours incubation at 37°C, the digested DNA fragments were ligated into *Eco*RI-*Sac*I digested pET30a (+) expression vector and transformed into BLR (DE3) *E. coli* competent cells to create *Hc-CPL-1* recombinant clone. The cells were spread onto LB plates containing kanamycin (50 µg/ml) and left at 37°C overnight. Colony PCR was performed to verify the positive clones and sequencing was done to confirm correct orientation of *Hc-CPL-1* gene in the recombinant construct.

Expression and purification of recombinant Hc-CPL-1 protein

An amount of 5 ml overnight culture of pET30a (+)/Hc-CPL-1 clone was inoculated into 125 ml of pre-warmed LB broth containing 50 µg/ml kanamycin. The culture was grown at 37°C until an optical density (OD) at 600 nm wavelength reached 0.6 values. Then, the cultured cells were induced

with IPTG to a final concentration of 1 mM at 37°C for 4 hours. The cells were harvested by centrifugation and the pellet was stored at -20°C overnight prior purification.

Purification of recombinant Hc-CPL-1 protein was followed the prescribed protocol of Ni-NTA Spin Kit (Qiagen, Germany) under denaturing conditions. Briefly, the cells pellet was resuspended in buffer contained 7 M urea, 0.1 M Sodium dihydrogen phosphate (NaH_2PO_4), 0.01 M Tris-HCl (pH 8.0) and 3 Units/ml culture volume of Benzonase Nuclease (25 U/µl). Then, the cells lysate was sonicated and centrifuged before the supernatant was passed through a pre-equilibrated Ni-NTA spin column. The spin column was washed twice with buffer of 8 M urea, 0.1 M NaH_2PO_4 and 0.01 M Tris-HCl (pH 6.3). Lastly, the bound 6× His-tagged recombinant protein was eluted three times with one elution of buffer contained 8 M urea, 0.1 M NaH_2PO_4 and 0.01 M Tris-HCl (pH 5.9) followed by two elution of the same buffer (pH 4.5) in order to separate 6× His-tagged monomers from multimers. All eluted proteins were subjected to sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) using Mini-PROTEAN Tetra System (BIO-RAD, USA) to analyze the proteins. Each protein samples were heated at 95°C for 5 min and electrophoresis was conducted at a constant voltage of 150 V for 1 hour. Then, the 10% resolving gel was stained with 0.1% Coomassie Blue staining solution to screen the separated proteins.

Mass spectrometry analysis

Protein bands were excised and in-gel digested overnight with 12.5 ng/µl trypsin (Promega, USA) in 50 mM ammonium bicarbonate (Sigma-Aldrich, Germany). Resulting peptides were extracted from the gel by formic acid extraction (5% formic acid, 50% acetonitrile) and concentrated using PerfectPure C-18 Tip (Eppendorf, Germany). Then, the peptides were reconstituted in 0.1% formic acid before analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using LTQ Orbitrap Velos Pro (Thermo Scientific, USA) coupled with Easy-nLC II (Thermo Scientific, USA) as described previously by Kwan *et al.* (2016). Briefly, 5 µl of peptide mixtures from each sample were trapped in the pre-column (EASY-column C18-A1, column dimension 2 cm × 100 µm inner diameter; Thermo Scientific, USA) at a flow rate of 0.3 µl/min and transferred to the analytical column (EASY-column C18-A2, column dimension 10 cm × 75 µm inner diameter; Thermo Scientific, USA) using a gradient of acetonitrile/0.1% formic acid (5% – 100% in 100 min). Fragmentation of peptides was performed in data-dependent mode and mass spectra were acquired in full-scan mode. Two fragmentation techniques were employed to obtain maximum peptide coverage, namely Collision Induced Dissociation (CID) and High Energy Collision Dissociation (HCD). Peptide *De novo* sequencing and database matching were performed using PEAKS Studio Version 7.0 (Bioinformatics Solution, USA) for protein identification.

Bioinformatics and phylogenetic studies

The nucleotide sequence of *Hc-CPL-1* gene was compared with those in the non-redundant databases (GenBank) using the BLAST programme provided at National Center for Biotechnology Information (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to verify the identity of gene sequenced. Then, the verified nucleotide sequence was converted into an amino acid sequence using Translate programme (http://web.expasy.org/translate/) provided by Expert Protein Analysis System (ExPASy). The potential signal peptide was predicted by SignalP 4.1 Server (http://www.cbs.dtu.dk/services/SignalP/) meanwhile GlobPlot 2 Version 2.3 (http://globplot.embl.de/) was employed to analyze its tendency towards globularity and disorder motifs. The CPL protein sequence from other nematode species such as *Brugia malayi* (accession number AF331035.1), *Caenorhabditis elegans* (accession number NM_074798.5), *Dictyocaulus viviparous* (accession number AY057110.1), *Litomosoides sigmodontis* (accession number AF331037.1), *Onchocerca volvulus* (accession number AF331036.1), *Ancylostoma caninum* (accession number AF320084.1), *Strongylus vulgaris* (accession number EU000411.1) and

Toxocara canis (accession number U53172.1) were retrieved from the Expressed Sequence Tags database, dbEST (https://www.ncbi.nlm.nih.gov/genbank/dbest/) at NCBI and multiple sequence alignment with Hc-CPL-1 protein were performed using Clustal Omega programme (http://www.ebi.ac.uk/Tools/msa/clustalo/) from European Bioinformatics Institute (EBI). The phosphorylation and active site of each CPL proteins were predicted using ScanProsite programme (http://prosite.expasy.org/scanprosite/). Lastly, phylogenetic tree was constructed using MultAlin programme (http://multalin.toulouse.inra.fr/multalin/).

RESULTS

Hc-CPL-1 cDNA construct and protein expression

Figure 1 shows the electrophoresis result of *Hc-CPL-1* gene obtained from one step RT-PCR. The target fragment of 1252 bp was produced from good quality (OD ratio 1.8-2.0) RNA extracts. There was no genomic DNA contamination in the reaction as no band was produced in both minus-RT control and minus-template PCR. The fragment was sequenced and results from BLAST indicated 95-99% similarity with *H. contortus* cathepsin L cysteine protease in the GenBank. To produce recombinant protein, cDNA of *Hc-CPL-1* gene was ligated into pET30a (+) expression vector and transformed in BLR (DE3) *E. coli* competent cells. The recombinant protein (rHc-CPL-1) was purified using affinity chromatography under denaturing condition and analyzed by SDS-PAGE (Figure 2). Coomassie blue staining of the purified rHc-CPL-1 showed a visible signal corresponded to the expected 47 kDa amino acid mass of Hc-CPL-1. However, there were also a number of unknown protein bands that were non-specifically purified with Hc-CPL-1 (Figure 2; lane 3, 4 & 5). The bands were excised and digested by trypsin to perform protein identification using mass spectrometry analysis.

Protein identification

Liquid chromatography-mass spectrometry (LC-MS/MS) analysis revealed that the excised protein band was indeed Hc-CPL-1 with the highest PEAKS peptide score (-10lgP). As shown in the Table 1, the fragmentation techniques CID and HCD were able to identify Hc-CPL-1 up to 64 and 46 percent of coverage, respectively. In CID technique, the number of peptides found for Hc-CPL-1 was 45 where 24 peptides were unique for Hc-CPL-1. Whereas for HCD technique, 21 total peptides was found with 14 unique for Hc-CPL-1. Comparing to the most similar protein found which were Ac-CPL-1 and Sv-CPL-1, there were low number of peptides found as well as no unique peptides were identified for both proteins. Peptide *De novo* sequencing and matching to NCBI database (Figure 3) indicated that there were no other post-translational modifications detected except for carbamidomethylation at position 192 and oxidation at position 206 amino acid resulted from trypsin digestion procedure.

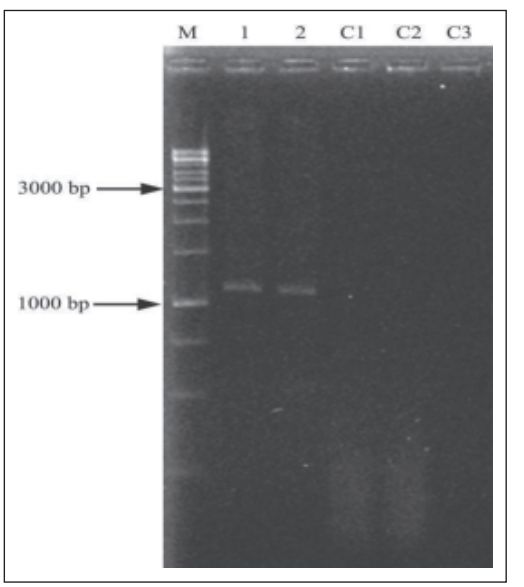


Figure 1. The cDNA synthesis of *Hc-CPL-1* gene from reverse transcription reaction using one step RT-PCR. Lane M: 1 kb DNA Ladder (Fermentas, Canada). Lane 1 & 2: The cDNA synthesized from RNA 1 & 2. Lane C1 & C2: Minus-RT controls of respective RNA. Lane C3: Minus-Template control.

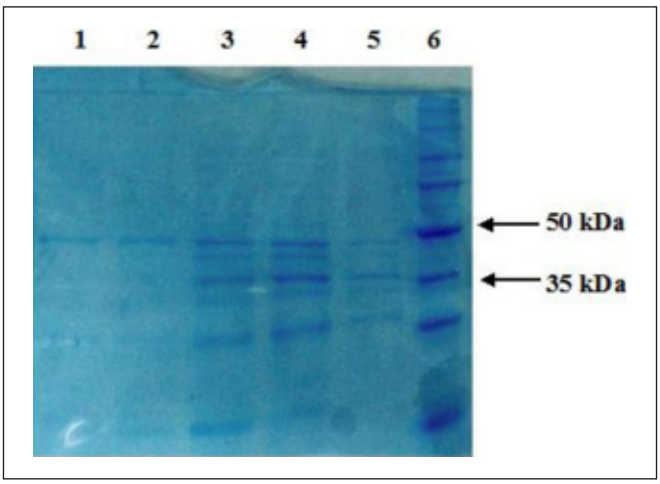


Figure 2. Recombinant Hc-CPL-1 purified under denaturing conditions from BLR (DE3) cells and eluted from a Nickel column. The elution fractions were determined by the flow of eluted protein from the column. The first 200 µl of eluted protein was the first elution fraction (E1) and this 200 µl amount continued until the fifth elution fraction (E5). Lane 1-5: Elution fraction 1-5. Lane 6: SDS-PAGE Low Molecular Weight Protein Marker (10-225 kDa).

Table 1. Protein identification obtained from PEAKS Studio Version 7.0 using fragmentation techniques, CID and HCD

Fragmentation techniques	Protein identity	-10lgP	Coverage (%)	No. of peptides	No. of unique peptides
Collision Induced Dissociation (CID)	*Haemonchus contortus* cathepsin L cysteine protease	299.44	64	45	24
	Ancylostoma caninum cathepsin L-like protease	213.60	40	16	0
	Strongylus vulgaris cathepsin L-like protease	210.71	24	15	1
High Collision Dissociation (HCD)	*Haemonchus contortus* cathepsin L cysteine protease	178.05	46	21	14
	Ancylostoma caninum cathepsin L-like protease	115.56	29	7	0
	Strongylus vulgaris cathepsin L-like protease	122.15	16	7	0

Bioinformatics and phylogenetic analysis

Bioinformatic study indicates that *Hc-CPL-1* encoded a 354 amino acids protein with a predicted mass of 39 974 Da and a calculated isoelectric point, pI of 5.93. Hc-CPL-1 comprises a signal peptide, a pro-domain and a mature domain. The putative signal peptide sequence has a predicted cleavage site between amino acid residue 16 and 17 (Alanine-Serine). Based on GlobPlot 2 programme prediction analysis (Figure 4), the Hc-CPL-1 protein was shown to have

Figure 3. Peptide *De novo* sequencing and database matching performed by PEAKS Studio Version 7.0 for Hc-CPL-1 protein identification.

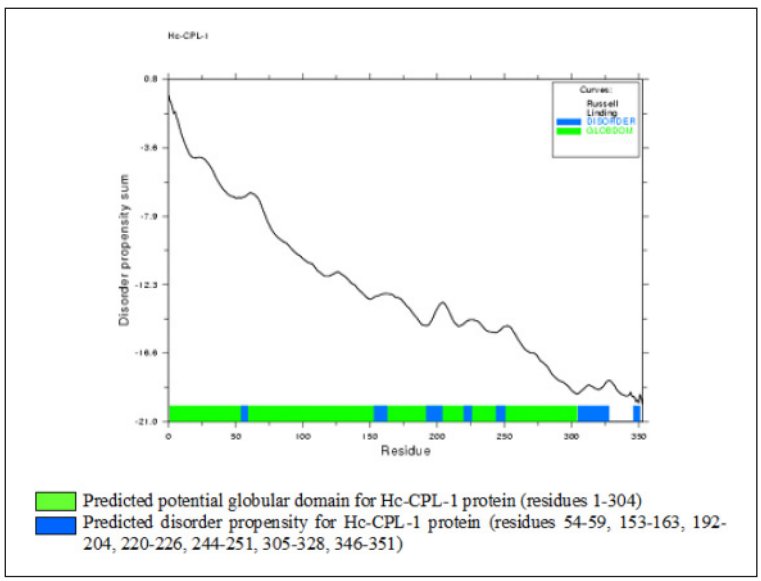


Figure 4. Prediction of Hc-CPL-1 globular domain and disorder using GlobPlot 2 programme.

a potential putative globular domain (GLOBDOM) between residues 1 to 304. However, there were 5 regions (residues 54-59, 153-163, 192-204, 220-226 and 244-251) along the residues (1-304) that were predicted to be disorder regions for Hc-CPL-1 protein.

The predicted amino acid sequence of Hc-CPL-1 was compared with the eight CPL sequences from other nematode species

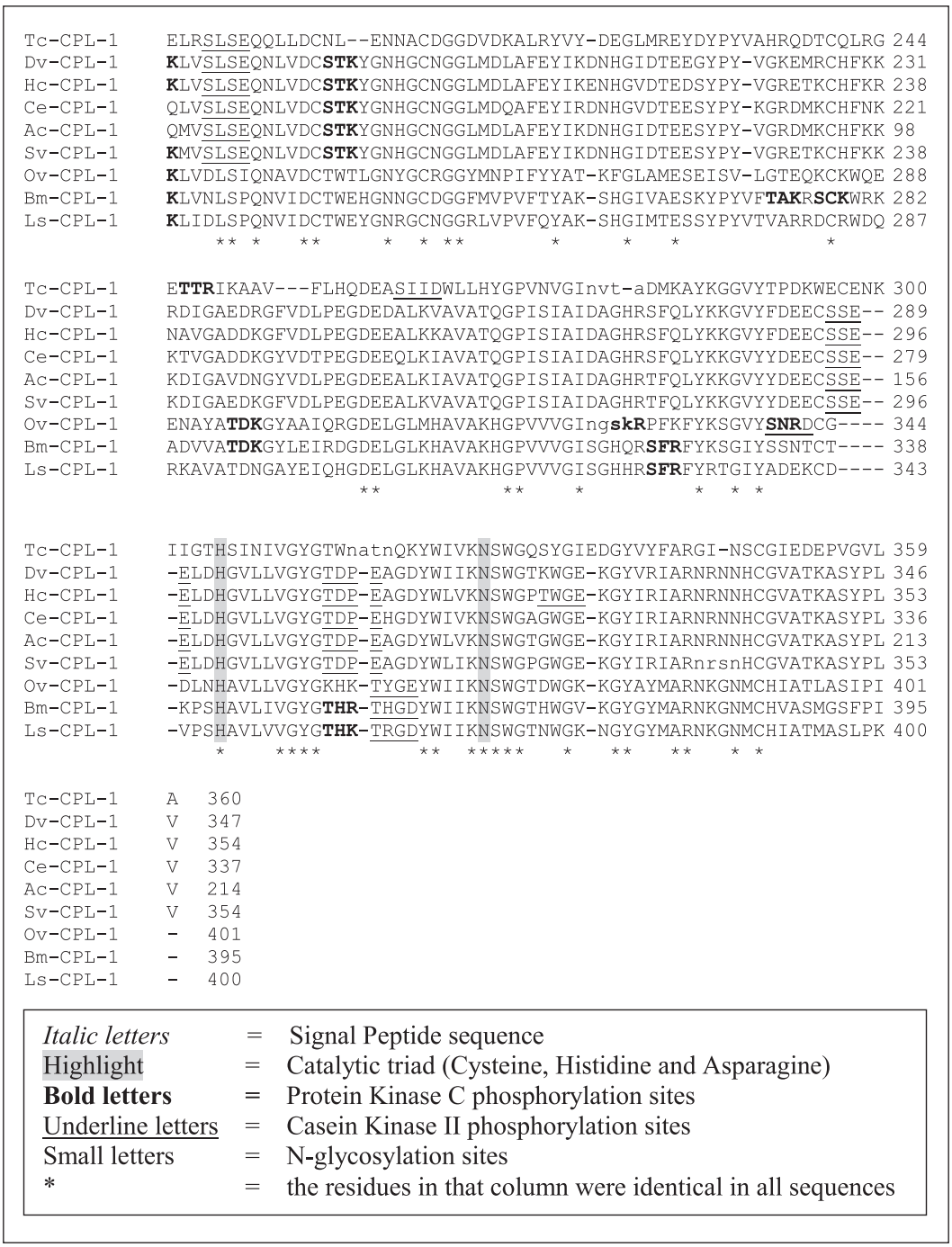
obtained from the database (Supplementary Figure 1). The mature domain of Hc-CPL-1 consisted of catalytic triad (cysteine, histidine and asparagine) residues located at position 161, 300 and 321 amino acids respectively as the characteristic of a cysteine protease (Supplementary Figure 1; highlight). Furthermore, Hc-CPL-1 was predicted to have three Protein Kinase C (PKC) phosphorylation sites, four Casein Kinase II (CK2) phosphorylation sites and one N-glycosylation site along the protein sequence (Supplementary Figure 1; bold, underline & small letters respectively). Comparison with other eight CPL sequences showed that there were 53 residues which are identical in all the sequences. Moreover, the Hc-CPL-1 had the greatest homology to CPLs from *S. vulgaris* (86.72%), *D. viviparous* (79.66%) and *C. elegans* (75.14%). This was in agreement with phylogenetic analysis where Hc-CPL-1 clustered with those 3 nematode species (Figure 5).

```
Tc-CPL-1    MPFGEAGSIMVLSLIAIQIAIILYVVAKnqs--------vKFEKEY--------------- 38
Dv-CPL-1    ----------LLLFLC------D--LAS-------------TK----------------- 12
Hc-CPL-1    -----MLRLLSLALLC------AVVLAS-------------ID----------------- 19
Ce-CPL-1    -----MNRFILLALVA------AVV---------------------------------- 14
Ac-CPL-1    ------------------------------------------------------------ 0
Sv-CPL-1    -----MFRLLSLVLLC------ASVFAS-------------ID----------------- 19
Ov-CPL-1    -------MLRIIVLLIVFAFLVDFTVTLNAQVQQLREVLGTFDQDYKRGnmtrLTTDFKK 53
Bm-CPL-1    --MWKIAQLAVLAFLA--------AFATAGEIEQLKEVLGKFDKDYKQGnmtrLASDFRS 50
Ls-CPL-1    -----MLRTAVFVFVAQFALIADSVVALSGEVLQLQQVLTELSNDYNQHnftkFTDDYQA 55

Tc-CPL-1    ----------------------------DLTRELRLLDRFEEFIRKYDKVYDSNEEFAER 70
Dv-CPL-1    ---IGTPRKHG----FY-----SEKQKSLRQRIDEAFGKWDEYKIKYDKHYDPEEE-NDY 59
Hc-CPL-1    ---GFRRHDHG----VR-----VHRQKSLRQKIDEAFNKWDDYKETFGKSYEPDEE-NDY 66
Ce-CPL-1    ----------------------AVNSAKLSRQIESAIEKWDDYKEDFDKEYSESEE-QTY 51
Ac-CPL-1    ------------------------------------------------------------ 0
Sv-CPL-1    ---SGSRRDHT----IR-----LHRVKSLRQKIDEAFKLWDDYKEAFGKSYNKDEE-NDY 66
Ov-CPL-1    AVKKYGDGKESQKSTVLQSFLKKMEDNGELRAMEKLETEWNDYVMALGKHYDSnes-nLR 112
Bm-CPL-1    ALKEYGDGQQG-ESTVVQEFLKKMESNGEQRAMEKLETEWKDYVTALGKHYDQKET-NFR 108
Ls-CPL-1    ALREYGEGDNG-EETVVQAFLKKLKS-GKTKAVKRLVREWGDYKVALGKHYneteN-NLR 112

Tc-CPL-1    FRIYVNNMLEAQKLNQRNRDYGT--IYGENEFADWNVNEFREILLPKDFFKNLRKKSTFI 128
Dv-CPL-1    MEAFVKNMIHIEEHNHEHRLGRKTFEMGLNNIADLPFSEYRKLNGYR--HRRL-----FG 112
Hc-CPL-1    MEAFVKNVIHIEEHNKEHRLGRKTFEMGLNEIADLPFSQYRKLNGYR--MRRQ-----FG 119
Ce-CPL-1    MEAFVKNMIHIENHNRDHRLGRKTFEMGLNHIADLPFSQYRKLNGYR----RL-----FG 102
Ac-CPL-1    ------------------------------------------------------------ 0
Sv-CPL-1    MEAFVKNVIHIDEHNQEHRLGRKTFEMGLNSIADLPFSQYRKLNGYR--HRRN-----FG 119
Ov-CPL-1    MAIFESNELMTEATNRKYEQGLISYTNGLNHLADLTDKKFKMMNGLR--FPnethLRTRR 170
Bm-CPL-1    MAIFESNELMTEKINKKYEQGLVSYTTALNDLADLTDEEFMVMNGLR--LPnqty---LR 163
Ls-CPL-1    MAIFESNEILTEKLNEQYKKGLISYATQLNDLADLTDEEFLTMNGLQ--AMnetdS--SR 168

Tc-CPL-1    DSFID-PPETVLARREEIPDHFDWRPYNVVTPVKSQFKCGSCWAFATVGTVESAYALGTG 187
Dv-CPL-1    DSMRKngtkFLVPFNVKVPDSVDWREHNLVTPVKNQGMCGSCWAFSATGALEGQHFRATG 172
Hc-CPL-1    DSLQSngtkFLVPFNVQIPESVDWREEGLVTPVKNQGMCGSCWAFSSTGALEGQHARATG 179
Ce-CPL-1    DSRIKnsssFLAPFNVQVPDEVDWRDTHLVTDVKNQGMCGSCWAFSATGALEGQHARKLG 162
Ac-CPL-1    ---------------------VDWRDKGLVTEVKNQGMCGSCWAFSATGALEGQHARASG 39
Sv-CPL-1    DSMQSngtkWLAPFNVEIPDSVDWRDKGLVTDVKNQGMCGSCWAFSATGALEGQHARASG 179
Ov-CPL-1    QTRHTVGQKYTYDPNEKLPVSVDWRKKGMVTPVKNQGVCGSCYAFAAIGALEAYNKKKTG 230
Bm-CPL-1    GKRQSFDTFYKYDQRERLPSSVDWRKKGVLTPVRNQGECGSCYAFGALAALEAYHKKRTG 223
Ls-CPL-1    RTRQASNRFYQYDRNEKLPAYVDWRKKGYVTPIRNQGQCGSCYAFGALAALEAYHKKKTG 228
                                ***     *    *  **** **      *       *
```



Supplementary Figure 1. Multiple sequence alignment of CPL protein sequences from parasitic and non-parasitic nematodes.

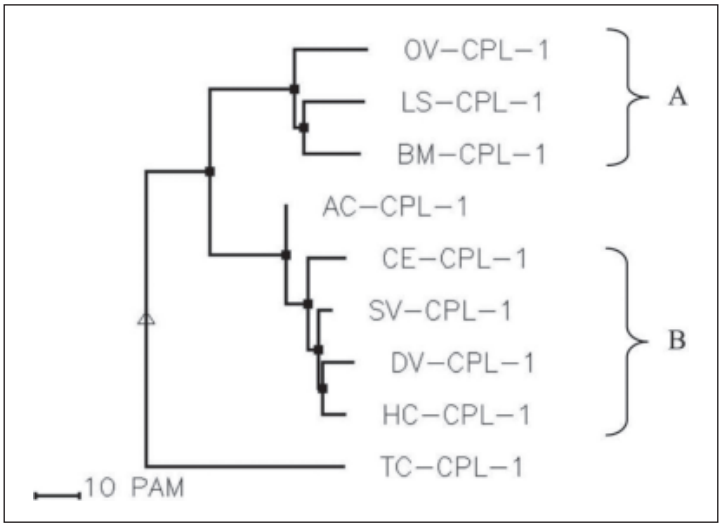


Figure 5. The plotted phylogenetic tree of Hc-CPL-1 sequence with other CPL proteins using MultAlin programme. The triangle in the constructed tree indicated that where was the tree rooted. The scale bar 10 PAM represented those 10 accepted point mutations. **A**: Branch A. **B**: Branch B.

DISCUSSION

One of the reliable prospects for the treatment of gastrointestinal nematode infection is vaccination application. The application utilizes important antigens from parasite to stimulate the host immune responses. In this study, the recombinant Hc-CPL-1 from Malaysian *H. contortus* isolate was expressed in *E. coli* system and purified antigen was studied to provide a better understanding about this parasite molecule. The fact that *H. contortus* is the predominant parasite of small ruminant in Malaysia as well as in the temperate regions (Waller & Chandrawathani, 2005) with a few geographical strains reported (Wahab & Suhaila, 2007; Gharamah *et al.*, 2012) have made this study as important comparison data before effective vaccine can be developed.

Although conventional Trizol method was known to be the best extraction method for RNA (Xiang *et al.*, 2001), method used in this study was seen to provide better quality and higher yields. This was due to a series of failure to isolate total RNA of *H. contortus* L3 larvae using Trizol method. Total RNA were isolated using commercialized kit (RNAqueos from Ambion, USA) that utilized chaotropic agent, guanidinium thiocyanate to disrupt membrane cells and destroyed the endogenous RNases. However, RT-PCR reaction have produced a fainted cDNA band from each RNA samples which have similar molecular weight to *Hc-CPL-1* target gene as stated by Geldhof *et al.* (2006). The possible reason for the poor yield could be the lower efficiency of cDNA synthesis during reverse transcription process before amplification in PCR reaction. In this case, low concentration of mRNA as the starting template might contribute to the difficulty in the cDNA synthesis thus affected the total yield after PCR reaction.

The expression of recombinant protein was successfully achieved from induced BLR (DE3) expression cells. A 47 kDa protein band was observed as predicted by PeptideMass programme for the size of recombinant Hc-CPL-1 protein. However, the protein was expressed in low intensity with some background proteins contamination. Purification of low level His-tagged recombinant protein in *E. coli* system using metal affinity chromatography usually result in coeluted with some contaminant proteins

from the host cells (Robichon *et al*., 2011). In this case, the purification protocol was optimized where the expressed protein was eluted in two different pH of elution buffer. The first two elution fractions were eluted using buffer with pH 5.9, slightly higher than normal buffer (pH 4.5). According to Robichon *et al*. (2011), the histidine side chain would lose its affinity toward the metal ions when the buffer pH less than 6, however this included those contaminant proteins that carried the non-consecutive histidine residues. As the result, two different pH of elution buffers were able to separate 6×His-tagged recombinant protein from those contaminants even though the recombinant protein was eluted first before the untagged proteins.

One of the possible explanations for the low intensity of the expressed protein with background contamination was the instability of the target protein. The target protein could be degraded and caused the weak bound to the Ni-NTA resin (Ni-NTA Spin Kit, Qiagen, Germany). In such cases, the recombinant protein would be lost easily each time the buffer moved through the silica-based filter due to the lost association between chromatography resins with the irreversible adsorption of non-specific region in the target protein (Christensen *et al*., 2007) and thus led to the lower yield in the final elution step. Meanwhile, nontagged proteins such as cellular protein with the presence of clustered histidine residues or metal binding motifs from the host cells (Robichon *et al*., 2011) might be increasingly bound non-specifically to the unoccupied sites of the filter resulted from the lost of tagged protein on the resin thus eluted together as the background contamination (Christensen *et al*., 2007).

Mass spectrometry is a more accurate method to identify the protein than polyacrylamide electrophoresis. In the analysis, two fragmentation techniques namely CID and HCD were employed to obtain maximum peptide coverage for Hc-CPL-1. Based on these techniques, we were able to confirm that expressed protein was Hc-CPL-1. Peptide *De novo* sequencing and NCBI database matching was performed by PEAKS Studio Version 7.0 where PEAKS peptide score (-10lgP) was calculated based on the maximum number of all peptide-spectrum match (PSMs) reported by PEAKS database, PEAKS post-translational modification (PTM) and SPIDER. However, there was no PTM found from the Hc-CPL-1 sequence as the protein was expressed in a prokaryote system. Comparison between the peptides found from the analysis with predicted peptides in PeptideMass (http://web.expasy.org/peptide_mass/) shown that 11 out of 26 total peptides (42.31%) were matched with the mass range between 700 Da to 2000 Da.

In the bioinformatic analysis, the multiple sequence alignment was constructed to determine the identical amino acid residues in all the protein sequences which may contribute to the functional and structural importance since the protein with biological functions will have homologues sequences in many organisms (Livingstone & Barton, 1993). Result from the constructed alignment using nine CPL protein sequences showed that 354 amino acid residues encoded by Hc-CPL-1 comprised of 53 identical residues with other nematodes. Among these identical residues, glycine was the most common residue (22.64%). The fact that glycine was the most stable amino acids due to its small side chain thus made it favourable to be abundance in a certain types of the protein.

In addition, the putative signal peptide sequence of Hc-CPL-1 was found to have a predicted cleavage site between residue 16 and 17 (Alanine-Serine). This signal peptide region was an indicator to all secreted proteins in which its plays a role in the translocation process. Alignment of the nine CPL protein sequences also showed the presence of catalytic triad cysteine, histidine and asparagine residues in all the sequences as these were the characteristic of cysteine proteases family (Ultaigh *et al*., 2009). Furthermore, the Hc-CPL-1 amino acid sequence was found to have the greatest homology with Sv-CPL-1, Dv-CPL-1 and Ce-CPL-1 amino acid sequences. This result could suggest that these cysteine proteases

have more similarity in term of protein structure and functions. Study by Britton & Murray (2002), demonstrated that transgenic expression of *Hc-CPL-1* gene from knockout CPL-1 *C. elegans* was able to rescue the embryonic lethality.

The Hc-CPL-1 was predicted to have three PKC, four CK2 phosphorylation sites and one N-glycosylation site along the protein sequence. Phosphorylation is an important process in post-translational modification for regulation and signal transduction within the cells that respond to their substrates through the presence of a motif near the phosphorylation site (Ritz *et al.*, 2009). In this case, conserved amino acid residues near to this site might contribute to the operational process and lead to a specific biochemical function of a certain protein family (Chen *et al.*, 2011). Thus, it is suggested to focus more on those phosphorylation sites that have conserved residues as well as those which conserved in most of the CPL protein sequences. On the other hand, glycosylation was important for the correct protein folding and provide stability to the structure of the protein. One N-glycosylation site predicted for Hc-CPL-1 was enough to be functional since there were no differences in function between multiple glycosylation sites as reported by Park & Zhang (2011).

Globular domain in a protein structure could demonstrate their functions where this domain was the characteristic of a catalytic enzyme, messenger hormone and also transporter of other molecule through the membrane. Usually globular protein was able to be soluble in aqueous solvents. The Hc-CPL-1 was predicted to have a potential putative globular domain from residue 1 until 304. There were also 5 short regions that were predicted to be disordered regions in Hc-CPL-1 which overlapped with the predicted globular domain. Interestingly, most of the predicted disordered regions were corresponded with the loop structures as predicted by PredictProtein programme (data not shown) and two of those regions have the predicted phosphorylation sites. These short linear disordered regions were suggested to be associated with some important function in Hc-CPL-1. This was in agreement with Linding *et al.* (2003) that disordered regions often contain short linear peptides which can have functional sites and always been found corresponded to the flexible regions in the protein such as loop or hinge structure.

In conclusion, Hc-CPL-1 from Malaysian isolate was successfully cloned and expressed in *E. coli* expression system with low background contamination. The Hc-CPL-1 was found to have no post-translational modifications (PTM) in our mass spectrometric analysis since the protein was expressed in a prokaryote system. In this study, the highly sensitive mass spectrometer used was able to detect and confirm that the isolate was Hc-CPL-1 with high degree of certainty even though in the presence of low amount of protein content. Furthermore, bioinformatics study on the isolate indicated that this protein consisted of loop-rich structures where have been known to be exposed to their surroundings. Besides, the protein also contained sequence that encodes signal peptide and several disordered regions thus gave an idea that this protein might be secreted to the environment. However, bioinformatic analysis could only give preliminary data on the structure and function based on the prediction through the other known proteins structures thus more detailed analysis need to be done using other approaches. This would provide broad range opportunities in exploring the functions of parasite proteins that can contribute in developing effective antiparasitic therapies.

Acknowledgements. The authors are grateful to Mr. Ahmed Abdul Malek Al-Rofaai from School of Biological Sciences, Universiti Sains Malaysia, Penang for providing us with *H. contortus* samples and people at Naran Village husbandry, Penang for giving us with goat faecal samples. This research was funded by Malaysian Research University (RU) grant (No. 305/PBiologi/613509) and Fundamental Research Grant Scheme (203/PDOPING/6711389).

REFERENCES

Bogyo, M., Verhelst, S., Bellingard-Dubouchaud, V., Toba, S. & Greenbaum, D. (2000). Selective targeting of lysosomal cysteine proteases with radiolabeled electrophilic substrate analogs. *Chemistry & Biology* **7**: 27-38.

Brady, C.P., Dowd, A.J., Brindley, P.J., Ryan, T., Day, S.R. & Dalton, J.P. (1999). Recombinant expression and localization of *Schistosoma mansoni* cathepsin L1 support its role in the degradation of host haemoglobin. *Infection and Immunity* **67**(1): 368-374.

Britton, C. & Murray, L. (2002). A cathepsin L protease essential for *Caenorhabditis elegans* embryogenesis is functionally conserved in parasitic nematodes. *Molecular and Biochemical Parasitology* **122**: 21-33.

Britton, C. & Murray, L. (2004). Cathepsin L protease (CPL-1) is essential for yolk processing during embryogenesis in *Caenorhabditis elegans*. *Journal of Cell Science* **177**: 5133-5143.

Chandrawathani, P., Nurulaini, R., Adnan, M., Premalaatha, B., Khadijah, S., Jamnah, O. & Zaini, C.M. (2009). A survey of parasitic infection on small ruminant farms in Kinta and Hilir Perak districts, Perak, Malaysia. *Tropical Biomedicine* **26**(1): 11-15.

Chen, Y.C., Aguan, K., Yang, C.W., Wang, Y.T., Pal, N.R. & Chung, I.F. (2011). Discovery of protein phosphorylation motifs through exploratory data analysis. *PLOS ONE* **6**(5): 1-15.

Christensen, T., Trabbic-Carlson, K., Liu, W. & Chilkoti, A. (2007). Purification of recombinant proteins from *E.coli* at low expression levels by inverse transition cycling. *Analytical Biochemistry* **360**(1): 166-168.

Geerts, S. & Gryseels, B. (2000). Drug resistance in human helminths: current situation and lessons from livestock. *Clinical Microbiology Reviews* **13**(2): 207-222.

Geldhof, P., Murray, L., Couthier, A., Gilleard, J.S., McLauchlan, G., Knox, D.P. & Britton, C. (2006). Testing the efficacy of RNA interference in *Haemonchus contortus*. *International Journal for Parasitology* **36**: 801-810.

Gharamah, A.A., Rahman, W.A. & Siti Azizah, M.N. (2012). Morphological variation between isolates of nematode *Haemonchus contortus* from sheep and goat populations in Malaysia and Yemen. *Journal of Helminthology* **1**: 1-7.

Greenbaum, D., Medzihradszky, K.F., Burlingame, A. & Bogyo, M. (2000). Epoxide electrophiles as activity – dependent cysteine protease profiling and discovery tools. *Chemistry & Biology* **7**: 569-581.

Hashmi, S., Britton, C., Liu, J., Guiliano, D.B., Oksov, Y. & Lustigman, S. (2002). Cathepsin L is essential for embryogenesis and development of *Caenorhabditis elegans*. *The Journal of Biological Chemistry* **277**(5): 3477-3486.

Jackson, F. & Coop, R.L. (2000). The development of anthelmintic resistance in sheep nematodes. *Parasitology* **120**: 95-107.

Knox, D.P., Redmond, D.L., Skuce, P.J. & Newlands, G.F.J. (2001). The contribution of molecular biology to the development of vaccines against nematode and trematode parasites of domestic ruminants. *Veterinary Parasitology* **101**: 311-335.

Knox, D.P., Smith, S.K. & Smith, W.D. (1999). Immunization with an affinity purified protein extract from the adult parasite protects lambs against infection with *Haemonchus contortus*. *Parasite Immunology* **21**: 201-210.

Kwan, S.H., Baie, S. & Ismail, M.N. (Profiling of proteins and post translational modifications of *Channa striatus* dried meat. *Current Proteomics* **13**: 9-19.

Linding, R., Russell, R.B., Neduva, V. & Gibson, T.J. (2003). GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Research* **31**(13): 3701-3708.

Livingstone, C.D. & Barton, G.J. (1993). Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Computer Applications in the Biosciences* **9**(6): 745-756.

Lustigman, S., McKerrow, J.H., Shah, K., Lui, J., Huima, T., Hough, M. & Brotman, B. (1996). Cloning of a cysteine protease required for the molting of *Onchocerca volvulus* third stage larvae. *The Journal of Biological Chemistry* **271**(47): 30181-30189.

Muleke, C.I., Yan, R., Sun, Y., Zhao, G., Xu, L. & Li, X. (2007). Vaccination of goats against *Haemonchus contortus* with a recombinant cysteine protease. *Small Ruminant Research* **73**: 95-102.

Murray, L., Geldhof, P., Clark, D., Knox, D.P. & Britton, C. (2007). Expression and purification of an active cysteine protease of *Haemonchus contortus* using *Caenorhabditis elegans*. *International Journal for Parasitology* **37**: 1117-1125.

Park, C. & Zhang, J. (2011). Genome – wide evolutionary conservation of N-glycosylation sites. *Molecular Biology and Evolution* **28**(8): 2351-2357.

Redmond, D.L. & Knox, D.P. (2006). Further protection studies using recombinant forms of *Haemonchus contortus* cysteine proteinases. *Parasite Immunology* **28**: 213-219.

Ritz, A., Shakhnarovich, G., Salomon, A.R. & Raphael, B.J. (2009). Discovery of phosphorylation motif mixtures in phosphoproteomics data. *Bioinformatics* **25**(1): 14-21.

Robichon, C., Luo, J., Causey, T.B., Benner, J.S. & Samuelson, J.C. (2011). Engineering *Escherichia coli* BL21 (DE3) derivative strains to minimize *E. coli* protein contamination after purification by immobilized metal affinity chromatography. *Applied and Environmental Microbiology* **77**(13): 4634-4646.

Smith, T.S. & Munn, E.A. (1990). Strategies for vaccination against gastro-intestinal nematodes. *Revue Scientifique et Technique (International Office of Epizootics)* **9**(2): 577-595.

Ultaigh, S.N.A., Carolan, J.C., Britton, C., Murray, L. & Ryan, M.F. (2009). A cathepsin L-like protease from *Strongylus vulgaris*: an orthologue of *Caenorhabditis elegans* CPL-1. *Experimental Parasitology* **121**: 293-299.

Wahab, A.R. & Suhaila, A.H. (2007). Morphological characterization of *Haemonchus contortus* in goats (*Capra hircus*) and sheep (*Ovis aries*) in Penang, Malaysia. *Tropical Biomedicine* **24**(1): 23-27.

Wahab, A.R., Suhaila, A.H. & Ahmad, S.O. (2007). Genetic characterization of *Haemonchus contortus* in goats (*Capra hircus*) and sheep (*Ovis aries*) in Penang, Malaysia. *Jurnal Biosains* **18**(2): 45-56.

Waller, P.J. & Chandrawathani, P. (2005). *Haemonchus contortus*: parasite no. 1 from tropics – polar circle. Problems and prospects for control based on epidemiology. *Tropical Biomedicine* **22**(2): 131-137.

Xiang, X., Qiu, D., Hegele, R.D. & Tan, W.C. (2001). Comparison of different methods of total RNA extraction for viral detection in sputum. *Journal of Virological Methods* **94**: 129-135.