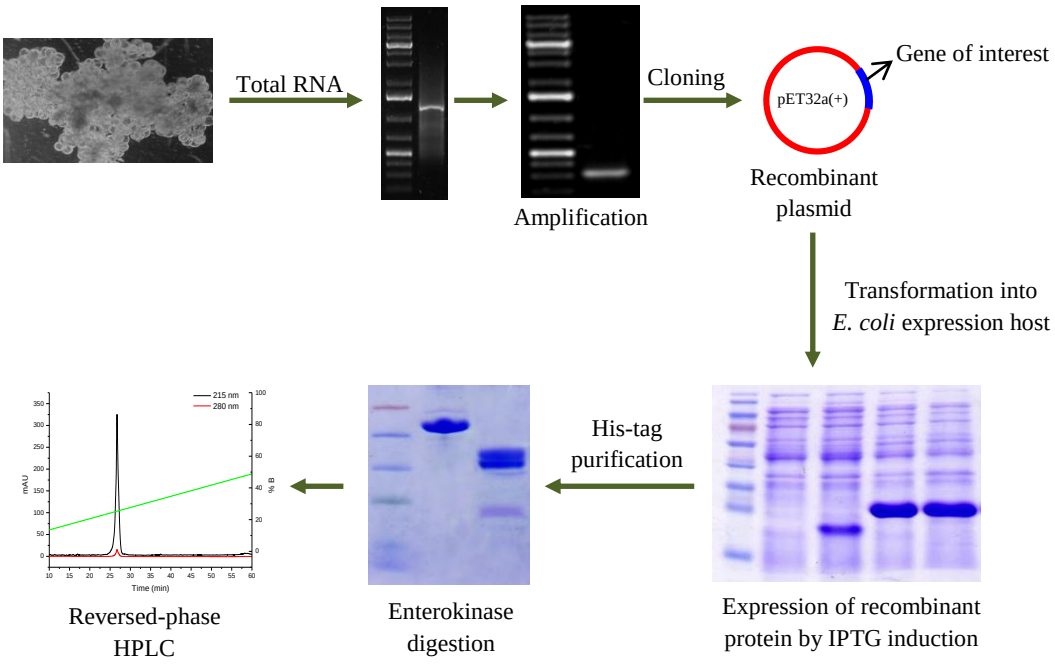
Chapter 4

Isolation, cloning and over-expression of anti-thrombin protein from *H. bispinosa*

Graphical abstract



4.1 Introduction

Ticks have evolved, over a period of several hundred million years, to avoid the host hemostatic system through a repertoire of inhibitors of the coagulation proteases [275]. Several inhibitors have been identified from various hematophagous animals including ticks, however, only few are in clinical use. The current class of antithrombotic therapy for prevention of cardiovascular diseases is limited and has several disadvantages. There is possibility of identifying novel antithrombotic molecules from the saliva of these blood hungry animals. Besides, these molecules present a good target for development of new vaccines for control of ticks and tick-

borne diseases. Therefore, an attempt has been made to identify anti-coagulant proteins from *Haemaphysalis bispinosa* salivary gland which showed potent anti-coagulant property (discussed in chapter 3). In this chapter, isolation and characterization of an anti-thrombin from *H. bispinosa* salivary gland using genomic approach has been described. The cDNAs coding for two isoforms of anti-thrombin protein were isolated from the salivary gland of *H. bispinosa* by PCR amplification. The proteins belong to I53 family of inhibitors and had calculated molecular mass of 6690.1 Da and 6709.1 Da, respectively. The cDNA coding for the matured peptides were cloned into bacterial expression vector and transformed into prokaryotic expression host. The recombinant proteins were purified using metal-affinity chromatography and reversed-phase HPLC methods and were characterized.

4.2 Materials and methods

4.2.1 Materials

SMART™ cDNA library construction kit was from Clontech Laboratories Inc. (CA, USA). RNeasy mini kit was from Qiagen (Hilden, Germany). *E. coli* BL21(DE3)pLysS and pET32a(+) vector were procured from Novagen® (Darmstadt, Germany). 1 kb Plus DNA molecular weight marker and Pageruler Plus protein molecular weight marker were from Thermo Scientific (Pittsburgh, USA). All other chemicals and reagents used were of analytical grade and purchased from Sigma (MO, USA), Merck (Mumbai, India) and Himedia (Mumbai, India).

4.2.2 Isolation of salivary glands

Partially fed adult female *Haemaphysalis bispinosa* ticks were collected from cattle. Salivary glands were carefully dissected out using sterile forceps under microscope and washed 3-4 times with Tris buffered saline (TBS) to prevent contamination of other tissues. The glands were either stored in RNAlater or directly used for isolation of total RNA using RNA isolation kit following manufacturer's instructions.

4.2.3 Isolation of total RNA from salivary gland

Total RNA was isolated from 8-10 pairs of salivary glands using RNeasy mini kit. Briefly, salivary glands were resuspended in 350 µl buffer RLT and homogenized using a rotor-stator tissue homogenizer (IKA, Bangalore, India). The lysate was centrifuged at full speed for 3 min and the supernatant collected. 1 volume 70% ethanol was added to the clear lysate and the mixture was applied to spin column, followed by centrifugation at 10,000 rpm for 30 s. The flow-through was discarded, 700 µl buffer RW1 added to the column and centrifuged at 10,000 rpm for 30 s. The column was then washed with 500 µl buffer RPE by centrifugation at 10,000 rpm for 2 min. The bound RNA was then eluted using 30-50 µl RNase-Free water in a fresh tube by centrifugation at 10,000 rpm for 1 min. The integrity of isolated RNA was analyzed on 1.1 % agarose gel. The concentration of total RNA was measured using NanoDrop 2000 (DE, USA).

4.2.4 Salivary gland cDNA synthesis

First-strand synthesis was carried out using SMARTScribe reverse transcriptase at 42 °C for 1 h in the presence of 1 µl SMART IV and 1 µl CDS III (3′) primers using 0.5 µg of RNA in a 10 µl reaction. Second-strand synthesis was performed using a long distance (LD) PCR-based protocol, using Advantage™Taq polymerase mix in the presence of the 5′ PCR primer and the CDS III (3′) primer. PCR conditions were as follows: 95°C for 1 min; 21 cycles of 95°C for 5 s, 68°C for 6 min. A small portion of the cDNA obtained by PCR was analyzed on a 1.1% agarose gel to check the quality and range of cDNA synthesized.

4.2.5 Amplification of gene coding for thrombin inhibitors by PCR

Primers were designed to amplify the transcript coding for anti-thrombin from the salivary gland cDNA. The cDNA encoding for the protein was amplified from the cDNA pool by PCR using the forward primer HbTI-F, designed from the 5′ untranslated region (UTR) of madanins (5’-TTTGACCGCAATGAAGCAC-3’) and two different reverse primers HbTI-R1 (5’-CTTCCAGCCTACAACATCAC-3’) and HbTI-R2 (5’-TCTATAACCTACCGACGGC-3’), designed from the 3′ UTR of Madanin 1 and Madanin 2, respectively. A total of 0.2 µM of the primer sets and

about 200 ng of template DNA were used in a 30 µl PCR reaction mixture. PCR was performed as follows: one cycle of 94ºC for 2 min; 30 cycles of 94ºC for 20 s, 55ºC for 30 s, 72ºC for 30 s; and final extension of 72ºC for 10 min. The amplified DNA was electrophoresed on 1.1 % agarose gel and visualized under UV light.

4.2.6 Sequencing and analysis of the thrombin inhibitor genes

The cDNA was sequenced using BigDye Terminator Cycle Sequencing kit (Applied Biosystems, MA, USA) in an automated sequencer (as described in section 2.2.8). The sequences were submitted to GenBank of National Center for Biotechnology Information (NCBI) using BankIt submission tool (https://www.ncbi.nlm.nih.gov/genbank/) as haemathrin 1 (KM086726) and haemathrin 2 (KM086725). The putative amino acid sequences were deduced using GENERUNNER. The sequences were searched for sequence similarity using BLAST progam at NCBI. The predicted molecular weight and theoretical pI of the protein was calculated using ProtParam server program (http://web.expasy.org/protparam/). The signal peptide and the cleavage site for the mature protein was predicted using SignalP (Version 4) server program (http://www.cbs.dtu.dk/services/SignalP-4.0/) at the Center for Biological Sequence Analysis (CBS) [276]. Conserved domains were identified and analyzed by the Conserved Domain Database (CDD; http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The nucleotide and protein sequences of haemathrin 1 and 2 were aligned with sequences obtained from the NCBI database by ClustalW2 program (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

4.2.7 Preparation of Escherichia coli BL21(DE3)pLysS competent cells

Glycerol stock of *E. coli* BL21(DE3)pLysS was revived on Luria-Bertani Agar (LBA) plate. The starter culture of the bacterium was prepared and 2 ml of it was inoculated into 100 ml Luria-Bertani (LB) broth and grown at 37ºC till OD_{600} reached 0.6-0.8. The culture was aliquoted in 50 ml centrifuge tube and kept in ice for 10 min. The tubes were centrifuged at 3000 rpm at 4°C for 10 min, supernatant discarded and the pellet was re-suspended in 50 ml of 100 mM $CaCl_2$. The suspension was centrifuged at 3000 rpm for 10 min at 4°C and the pellet was suspended in 6 ml

of 100 mM $CaCl_2$ and 2 ml 50% glycerol. Aliquots of 50 µl was transferred into sterilized pre-marked pre-chilled eppendorf tubes and stored at -80°C for later use.

4.2.8 Isolation of plasmids from bacterial cells

Single colonies of bacterial cells were picked and inoculated in 5 ml LB broth with Ampicillin and incubated with shaking at 37°C for overnight. Plasmids were isolated using QIAprep Spin Miniprep kit (Qiagen) according to the manufacturer's instruction. Briefly, the overnight bacterial culture was centrifuged and the pellet was resuspended in 250 µl Buffer P1, RNase added and transferred to a microcentrifuge tube. 250 µl Buffer P2 was added and mixed thoroughly, followed by addition of 350 µl Buffer N3. The tube was centrifuged at 10,000 rpm for 10 min and the supernatant applied to QIAprep spin column and centrifuged for 30-60 sec. The flow-through was discarded and the column washed with wash buffer, Buffer PE. The flow-through was discarded and centrifuged for additional 1 min to remove the additional wash buffer. The DNA was eluted with 30-50 µl of Buffer EB (10 mM Tris-Cl, pH 8.5) and stored at -20ºC until use. The isolated plasmids were analyzed on 0.8% agarose gel.

4.2.9 Transformation of plasmids into competent cells

Plasmids (1-5 µl) was added to the *E. coli* BL21(DE3)pLysS competent cells and mixed gently. The tubes were incubated in ice for 45 min, followed by heat shock treatment at 42°C for 90 s in a heat block. 500 µl of LB broth without Ampicillin was added and incubated at 37°C for 1 h with shaking. The cells were plated on LB Agar plate with Ampicillin and incubated at 37°C for 16 h. The plates were observed for single colonies.

4.2.10 Cloning of cDNA coding for haemathrins

The DNA fragments encoding the mature peptide of haemathrin 1 and 2 were amplified (as described in section 2.2.7) using gene-specific primers (HbMl1-F: 5`-ATCCATGGCATACCCGGAGAGAGA-3` and HbMl1-32R: 5`-ATCTCGAGTCAAGCATTCTTTCGTCC-3`). The amplified product and pET32a(+) were digested with 5 U each of restriction enzymes *Xho*I and *Nco*I at 37°C for 1-3 h. The digested insert and vector were gel extracted using commercial kit (as described

in section 2.3.7). The digested insert and vector (3:1) were ligated using T4 DNA ligase at 16°C for overnight. The ligation product was transformed into *E. coli* BL21(DE3)pLysS competent cells by heat-shock method as described in previous section (4.2.9).

4.2.11 Screening of clones by colony PCR

The clones were confirmed for insertion of the gene by colony PCR. Briefly, colonies were picked using sterilized tooth-pick, suspended in 10 µl distilled water and heated at 95°C for 2 min. The lysate was centrifuged at 10,000 rpm for 2 min and 1 µl of the supernatant was used as template for PCR. PCR was performed as described in section 2.2.6. The PCR products were analyzed on 1.1% agarose gel.

4.2.12 Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was performed as described in section 3.2.4. After electrophoresis at constant current of 20-30 mA in Tris-Glycine-SDS buffer (25 mM Tris, 192 mM Glycine, 0.1% SDS), the gel was stained with staining solution containing 0.25% Coomassie Brilliant Blue R-250 in methanol, acetic acid and water (40:10:50). The gel was de-stained with methanol, acetic acid and water (40:10:50) till the background was clear. The gels were photographed and documented.

4.2.13 Tricine – SDS – Polyacrylamide Gel Electrophoresis

Tricine-SDS-PAGE was performed as described. Briefly, 18% resolving gel was prepared by mixing 3.6 ml 40% Acrylamide/bisacrylamide (19:1), 2 ml 3 M Tris-Cl buffer (pH 8.45), 30% ethylene glycol, 0.04% APS and 12 µl TEMED, to a final volume of ~8 ml. For 5% stacking gel, 0.5 ml 40% Acrylamide/bisacrylamide was added to a final volume of ~4 ml, containing 1 ml 3 M Tris-Cl buffer, 0.04% APS and 16 µl TEMED. The samples were prepared by adding sample buffer and boiling for 2 min. The samples were electrophoresed at a constant voltage of 150-200 V till the tracking dye reached the bottom of the gel and stained with Coomassie Brilliant Blue R-250 staining solution as described in previous section and subsequently documented after de-staining.

4.2.14 Expression of recombinant haemathrins (rHaemathrins)

Single colonies of *E. coli* BL21(DE3)pLysS transformed with recombinant pET-32a(+) (rpET-32a(+)) were picked and inoculated in 4 ml Luria Bertani (LB) broth with Ampicillin (Amp) and incubated with shaking at 37°C for overnight. The starter culture was then inoculated in 100 ml LB with Amp and grown at 37°C with shaking till OD_{600} reached 0.5 – 0.8. Cells transformed with pET-32a(+) was taken as control. IPTG was added to the culture to a final concentration of 0.1 mM and incubated at 37 °C for 3 hours. Before addition of IPTG 5 ml of the culture was kept aside which served as uninduced cells. The cells were collected by centrifugation at 5,000 rpm for 10 min and lysed by heating at 95°C for 5 min in gel loading buffer, followed by loading on 12.5% SDS-PAGE gel. For optimization of the expression conditions, the cells were induced at different temperatures (16°C, 25°C, and 37°C), for different time intervals (2 and 4 h) and using different IPTG concentrations (0.05, 0.1, 0.5 and 1 mM). Cells were pellet down at 5,000 rpm for 10 min. The cell pellet was suspended in 4 ml lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM Imidazole, pH 8.0), and sonicated for 4 min (75% amplitude, 60% cycle duty, 3 min) in a Labsonic M Ultrasonic homogenizer (Sartorius Group, Bangalore, India). The cell lysate was centrifuged at 10,000 rpm at 4°C for 10 min. Pellet and supernatant were collected and run on 12.5% SDS-PAGE gel. The SDS-PAGE gels were stained and visualized as described in section 4.2.12.

4.2.15 Mass culture of recombinant protein for His-tag purification

The starter culture was inoculated to 1 L LB broth (with Amp) and grown at 37°C till OD_{600} reached 0.5-0.8. The cells were induced with 0.05 mM IPTG and grown at 37°C for 4 h. Cells were collected by centrifugation at 5000 rpm for 10 min. The cell pellet was suspended in 40 ml lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM Imidazole, pH 8.0) and sonicated for 4 min (60% cycle duty). The cell lysate was centrifuged at 10,000 rpm at 4 °C for 20 min. Pellet and supernatant were collected and run on 12.5% SDS-Page gel.

4.2.16 His-tag purification of rHaemathrins

Fifteen ml Ni-NTA (Nickel-nitrilotriacetic acid) agarose slurry was taken and washed with distilled water and equilibrated with lysis buffer containing 10 mM imidazole (pH 8.0). The buffer was removed and 20 ml of the lysate was added to it. The mixture was incubated in ice with constant shaking for 2 h and then gently applied into column while keeping the outlet closed. The unbound fraction was collected in a fresh tube, followed by washing with 2 column volume of lysis buffer containing 10 mM or 20 mM imidazole. The bound recombinant protein was then eluted with 2 column volume of lysis buffer containing 100 mM imidazole. The fractions were collected and analyzed on 12.5% SDS-PAGE gel.

4.2.17 Dialysis of partially purified rHaemathrins

The partially purified rHaemathrins were dialyzed against (a) 25 mM NaH_2PO_4, 200 mM NaCl, 5 mM Imidazole (pH 7.4), (b) 10 mM NaH_2PO_4, 100 mM NaCl, 2 mM Imidazole (pH 7.4), (c) 5 mM NaH_2PO_4, 100 mM NaCl, 1 mM Imidazole (pH 7.4), and (d) finally the protein was dialyzed in 20 mM Tris-Cl (pH 7.4) for 8 h at 4°C using SnakeSkin™ Pleated Dialysis Tubing (Thermoscientific, MA, USA). The dialyzed sample was analyzed on 12.5% SDS-PAGE gel.

4.2.18 Peptide mass fingerprinting of rHaemathrins

For ESI-LC-MS/MS analysis of recombinant haemathrins, the recombinant proteins were digested trypsin and run on mass spectrometer. Briefly, about 50 µg of protein was dissolved in 50 µl of MilliQ water, followed by addition of 41.5 µl of 50 mM of ammonium bicarbonate, 1 µl of 0.5 M DTT and 2 µl of 1% Protease Max (Promega, WI, USA). The reaction was incubated for 20 min at 56°C in a heating block, followed by addition of 2.7 µl of 0.55 M IAA and incubation for 15 min (protected from light). To this 1.8 µl of trypsin solution and 1 µl of 1% Protease Max were added and was incubated at 37°C for 3 h. The reaction was stopped by addition of 0.5 µl of 100% trifluoroacetic acid (TFA). After centrifugation at 12,000 rpm for 10 min, it was loaded onto Accela LCQ Fleet Ion Trap Mass Spectrometer (Thermo Scientific, MA, USA). About 40 µl of the sample was separated in a Zorbax 300SB-C18 column (1 x 150 mm, 3.5 µm, Thermo Scientific) pre-equilibrated with 0.1%

formic acid at a flow rate of 200 µl min^{-1} with a linear gradient of 100% acetonitrile. The eluent was directly analyzed on the mass spectrometer with a positive ionization mode and spectra were obtained in MS/MS mode with scan range of 500 to 2000 m/z. The MS/MS spectra were analyzed by the Proteome Discoverer 3.1 using Sequest program for offline identification of the peptides.

4.2.19 Cleavage of rHaemathrins by enterokinase (EK)

For removal of the fusion tag, 0.4 mg of recombinant protein was digested using 40 unit of enterokinase in a 1 ml reaction at 21°C for 16 h. The cleaved rHaemathrins were analyzed on 18% Tricine SDS-PAGE gel. The cleaved peptides were recovered by chromatographic separation using Jupiter C18 column (4.6 mm x 250 mm, Phenomenex, CA, USA) pre-equilibrated with 0.1% TFA using a gradient of 0-80% Acetonitrile.

4.2.20 ESI-MS analysis of recombinant haemathrin 1 and 2

The molecular masses of the peptides were determined on Accela LCQ Fleet™ Mass Spectrometer (Thermo Scientific, CA, USA). The mass spectrometer was equipped with an electrospray ion source (ESI). Data were acquired in positive ion mode and scanned from m/z 500–2000. Peptides were diluted with MiliQ water to a final concentration of 10 µM and injected into an automated sampler. Xcalibur™ software (Thermo Scientific) was used to generate the intact mass spectra for the peptides and later deconvoluted for intact molecular weight determinations.

4.2.21 Circular Dichroism (CD) measurements of haemathrins

Far-UV CD spectra (260 – 190 nm) of 20 µM peptides in 20 mM Sodium phosphate buffer pH 7.4 were recorded in a 0.2 mm path-length quartz cuvette at 20°C with a 0.2 nm resolution, a bandwidth of 2 nm, and a scan speed of 50 nm min^{-1} using a JASCO J-815 spectropolarimeter (Jasco, Tokyo, Japan), with a Peltier system to control cell temperature. The CD intensities were expressed as molar ellipticity, [θ], with the unit deg cm^2 $dmol^{-1}$.

4.3 Results

4.3.1 Isolation of total RNA from H. bispinosa salivary gland

Total RNA was successfully isolated from salivary gland of *H. bispinosa* as observed in 1.1% agarose gel (Fig 4.1). The presence of bright band at about 1500 bp and 750 bp corresponding to 28S rRNA and 18S rRNA and smear indicates the quality of the total RNA isolated. The concentration of RNA was measured using NanoDrop 2000 and found to be 146 ng μl^{-1} with an A260/A280 of 1.952.

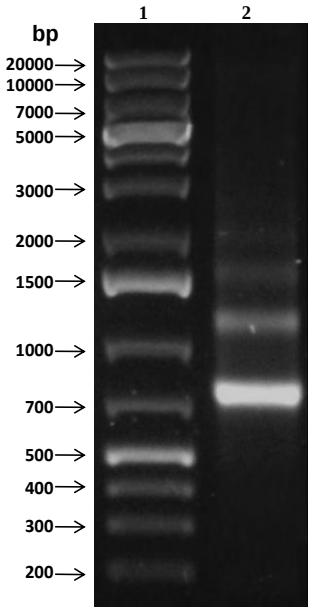


Figure 4.1 1.1% agarose gel profile of total RNA isolated from salivary gland. Lane 1: 1 kb plus DNA ladder; lane 2: total RNA.

4.3.2 cDNA synthesis from salivary gland total RNA

cDNA synthesized using Clontech cDNA library construction kit, appeared as a smear ranging from ~0.5 to ~4.5 kb on 1.1% agarose gel (Fig. 4.2), with few bands showing abundant mRNA at about 700-1500 kb. The concentration of synthesized cDNA was measured to be 377.2 ng μl^{-1} with A260/A280 ratio of 1.63.

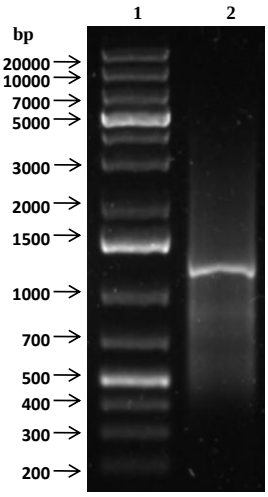


Figure 4.2 1.1% agarose gel profile of synthesized cDNA from salivary gland total RNA. Lane 1: 1 kb plus DNA ladder; lane 2: cDNA.

4.3.3 Amplification of thrombin inhibitors from H. bispinosa

cDNA coding for thrombin inhibitors were amplified from the cDNA synthesized from salivary gland total RNA of partially fed female ticks. Two amplicons of ~300 bp were obtained using the gene-specific primers (Fig. 4.3). The amplicon obtained using primer HBTI-F and HBTI-R1 was named as amplicon 1 and that using HBTI-F and HBTI-R2 names as amplicon 2. The concentration of amplicon 1 and amplicon 2 was calculated to be 52.5 ng μl^{-1} and 78.75 ng μl^{-1} respectively.

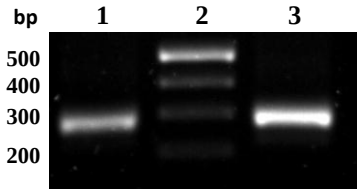


Figure 4.3 1.1% agarose gel profile of amplified gene. Lane 1: Amplicon 1 amplified using HbTI-F and HbTI-R1 primers; lane 2: 1 kb plus DNA ladder; lane 3: Amplicon 2 amplified using HbTI-F and HbTI-R2 primers.

4.3.4 Sequencing and analysis of haemathrin 1 and 2

Sequencing of the amplified products using BigDye Terminator reagents reveals the nucleotide sequence of 331 bases. This nucleotide sequence includes the 5`and 3`UTRs and the open reading frame. The amino acid translation of the cDNA

sequences had coding sequence of 234 bp encoding a protein of 78 amino acid residues and was devoid of cysteine residues (Fig. 4.4). Similarity searches of the two translated proteins were performed using BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Blast search results indicated that haemathrins were similar to thrombin inhibitors, madanins (acc. no. AAP04349 and AAP04359) and madanin-like proteins (acc. no. BAE00175 and BAE00067) isolated from *Haemaphysalis longicornis* and had the Inhibitor I53 superfamily putative conserved domain (Figure 4.5a, 4.5b). The genes were named as haemathrin 1 and haemathrin 2. The first 19 amino acids were predicted to be signal peptide by SignalP program at CBS, Technical University of Denmark (http://www.cbs.dtu.dk/services/SignalP/). The predicted mature peptide of the proteins had calculated molecular weight of 6690.1 Da and 6709.1 Da for haemathrin 1 and 2, respectively.

Haemathrin 1

GGACGAAAAGATATTTTGACCGCAATGAAGCACTTCGCAATTTTTATTCTTGCTGTTGTG 60

M K H F A I F I **L A V V**

*GCCAGTGCCGTGGTGATGGCATAC*CCGGAGAGAGATTCAGCGAAGGAAGGCAACAAAGGG 120

A S A V V M A Y P E R D S A K E G N K G

CAAAAGAGAGCTCGGCTAGTTAATGTACAAGAACGTTCAGGTGAAACTGACTATGATGAA 180

Q K R A R L V N V Q E R S G E T D Y D E

TATGAAGAAAATGAAAACACTCCTACTCCGGATCCAAGTGCACCGACGGCGAGACCACGG 240

Y E E N E N T P T P D P S A P T A R P R

CTTGGACGAAAGAATGCTTGAATCAATGGTGCTCTTGATTTCTATAACCTACCGATGGCG 300

L G R K N A *

GTGATGTTGTAGGCTGGAAGA*AAACATCTCC* 331

Haemathrin 2

GGACGAAAAGATATTTTGACCGCAATGAAGCACTTCGCAATTTTTATTCTTGCTGTTGTG 60

M K H F A I F I L A V V

*GCCAGTGCCGTGGTGATGGCATAC*CCGGAGAGAGATTCAGCAAATAGAGGCAGCCAAGAG 120

A S A V V M A Y P E R D S A N R G S Q E

AAAGAGAGAGCTCTGCTAGTTAAAGTACAAGAACGTTCTAGCCAAGATGACTACGATGAA 180

K E R A L L V K V Q E R S S Q D D Y D E

TATGATGCAGATGAGACCACTCTTTCTCCGGATCCAGATGCACCCACTGCCAGACCACGG 240

Y D A D E T T L S P D P D A P T A R P R

CTCGGACGAAAGAATGCTTGAATCAATGGTGCTCTTGATTTCTATAACCTACCGATGGCG 300

L G R K N A *

GTGATGTTGTAGGCTGGAAGA*AAACATCTCC* 331

Figure 4.4 Nucleotide sequences and deduced amino acid sequences of haemathrin 1 and haemathrin 2. The first 19 amino acids (**bold**) were predicted to be signal peptide for both the peptides and the *italicized* nucleotide base are untranslated region (UTR).

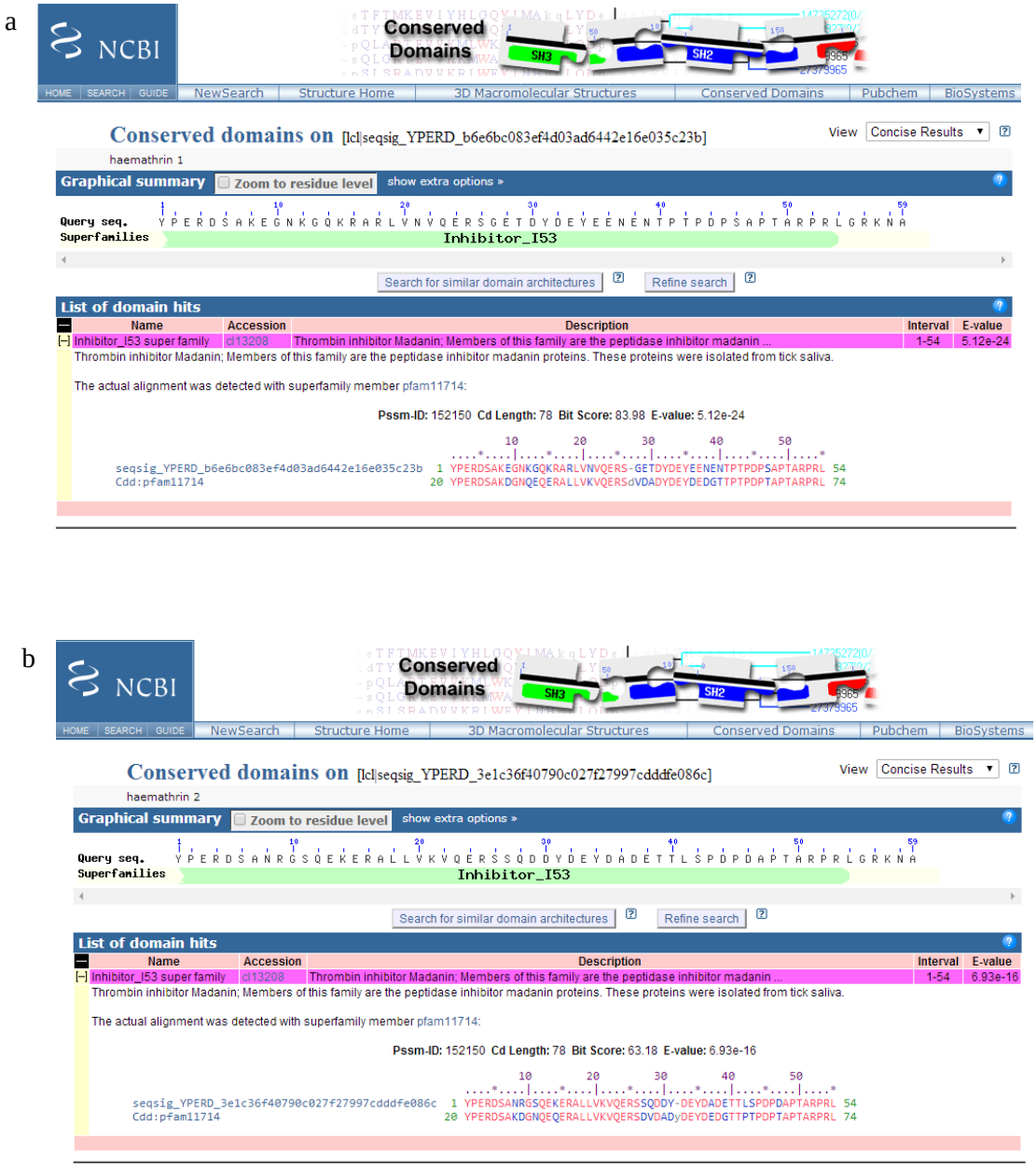


Figure 4.5 Conserved domains search of (a) haemathrin 1 and (b) haemathrin 2 showed that these belong to I53 superfamily of inhibitors.

4.3.5 Nucleotide sequence alignment of haemathrin 1 and 2

The nucleotide sequence of haemathrin 1 showed 90.63% similarity (Fig. 4.6a) to that of haemathrin 2, while madanin 1 and 2 nucleotide sequences were about 85% similar. Haemathrin 1 showed 78.22% homology to madanin 1 (Fig. 4.6b) and haemathrin 2 showed 80.88% to madanin 2 (Fig. 4.6c).

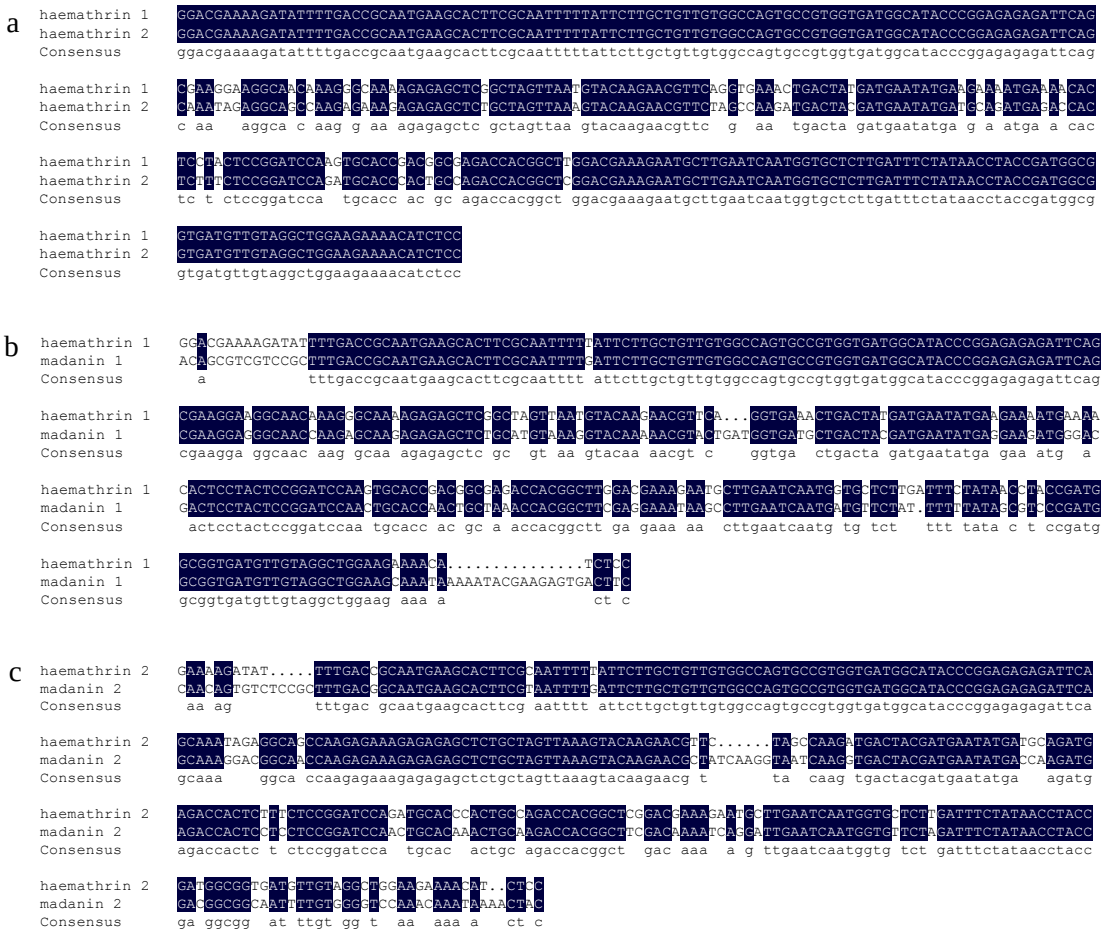


Figure 4.6 Nucleotide sequence alignment of haemathrins and madanins performed using DNAMAN. (a) Sequence alignment of haemathrin 1 and 2; (b) Sequence alignment of haemathrin 1 and madanin 1; (c) Sequence alignment of haemathrin 2 and madanin 2. The shaded bases represent identical bases, while the dots represent gaps.

4.3.6 Amino acid sequence alignment of haemathrin 1 and 2

The deduced amino acid sequence of mature peptide of haemathrin 1 showed 67.8 % similarity to that of haemathrin 2 (Fig. 4.7), which was ~3% less than the similarity between madanins (70.49 %). Haemathrin 1 showed 65% similarity to madanin 1, while haemathrin 2 was 70.9% similar to madanin 2 (Fig. 4.7). Hence, haemathrins are madanin-like isoforms expressed in the salivary gland of *H. bispinosa*.

```
Haemathrin 1 YPERDSAKEGNKGQKRARLVNVQERSGETDYDEYEENENTPTPDPSAPTARPRLGRKNA 59
Haemathrin 2 YPERDSANRGSQEKERALLVKVQERSSQDDYDEYDADETTLSPDPDAPTARPRLGRKNA 59
             *******:.*.: ::** **:*****.: *****: :*.* :***.************

Haemathrin 1 YPERDSAKEGNKGQKRARLVNVQERS-GETDYDEYEENENTPTPDPSAPTARPRLGRKNA 59
Madanin 1    YPERDSAKEGNQEQERALHVKVQKRTDGDADYDEYEEDGTTPTPDPTAPTAKPRLRGNKP 60
             ***********: *:** *:**:*: *::********: .******:****:*** ::.

Haemathrin 2 YPERDSANRGSQEKERALLVKVQERS--SQDDYDEYDADETTLSPDPDAPTARPRLGRKNA 59
Madanin 2    YPERDSAKDGNQEKERALLVKVQERYQGNQGDYDEYDQDETTPPPDPTAQTARPRLRQNQD 61
             *******: *.**************   .*.****** **** .*** * ****** :::
```

Figure 4.7 Alignment of amino acid sequence of mature peptides of haemathrins and madanins.

4.3.7 Cloning of cDNA coding for haemathrins

The nucleotide sequence of the mature peptide was re-amplified using gene-specific primers flanked with *Nco*I and *Xho*I restriction sites (Fig. 4.8). The double digested product was successfully sub-cloned into the *Nco*I and *Xho*I restriction sites of pET32a(+) expression vector. The recombinant plasmids were transformed into *E. coli* BL21(DE3)pLysS competent cells. Colony PCR of the bacterial colony confirmed the insertion of the gene of interest. Three clones each of haemathrin 1 and haemathrin 2 were found to be positive (Fig. 4.9 and 4.10). Insertion of the correct open reading frame (ORF) of the genes into the expression vector was further confirmed by sequencing of the recombinant plasmids using T7 promoter and T7 terminator universal primers. .

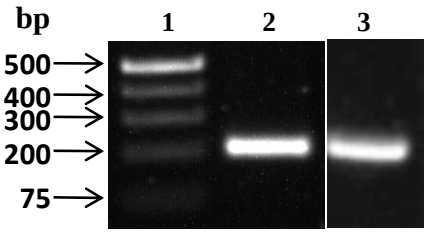


Figure 4.8 1.1% agarose gel profile of PCR products of gene coding for mature peptides of haemathrin 1 and 2. Lane 1: 1 kb plus DNA ladder; lane 2: haemathrin 1 PCR product; lane 3: haemathrin 2 PCR product.

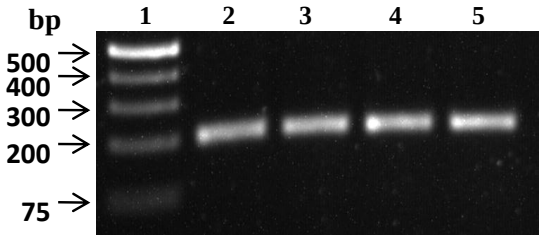


Figure 4.9 1.1% agarose gel profile of colony PCR products (haemathrin 1). Lane 1: 1 kb plus DNA ladder; lane 2: Positive control; lane 3: clone 1; lane 4: clone 2; lane 5: clone 3.

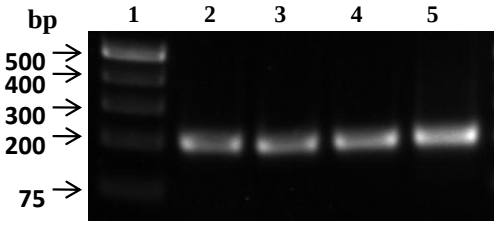


Figure 4.10 1.1% agarose gel profile of colony PCR products (haemathrin 2). Lane 1: DNA molecular weight marker; lane 2: Positive control; lane 3: clone 1; lane 4: clone 2; lane 5: clone 3.

4.3.8 Expression of recombinant haemathrins (rHaemathrins)

E. coli BL21(DE3)pLysS transformed with recombinant vectors were induced with IPTG and the cell lysate was analyzed on SDS-PAGE. Bands of about 25 kDa were observed on coomassie-stained SDS-PAGE gel which was absent in uninduced cells. This is the expected size of rHaemathrins with the fusion tag (Fig. 4.11, lane 4 and 5). The molecular mass of the fusion tag is about 18 kDa (Fig. 11, lane 3) while that of haemathrins is about 6.7 kDa, which sums up to about 25 kDa.

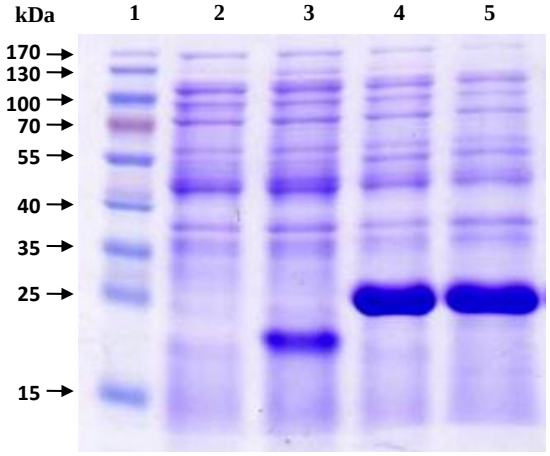


Figure 4.11 12.5% SDS-PAGE gel profile of expressed rHaemathrins. Lane 1: Protein molecular weight marker; lane 2: cell lysate of uninduced cells; lane 3: cell lysate of induced cells with pET32a; lane 4: cell lysate of induced cell with recombinant plasmid (haemathrin 1); lane 5: cell lysate of induced cells with recombinant plasmid (haemathrin 2).

4.3.9 Expression of rHaemathrin at different temperatures

E. coli BL21(DE3)pLysS cells transformed with rHaemathrin 2 were grown at 37ºC till the mid log phase. For induction, 0.1 mM IPTG (final concentration) was added to the culture and grown at different temperatures (16ºC, 25ºC and 37ºC). It was observed that there was no significant difference in expression of rHaemathrins at different temperatures (Fig. 4.12).

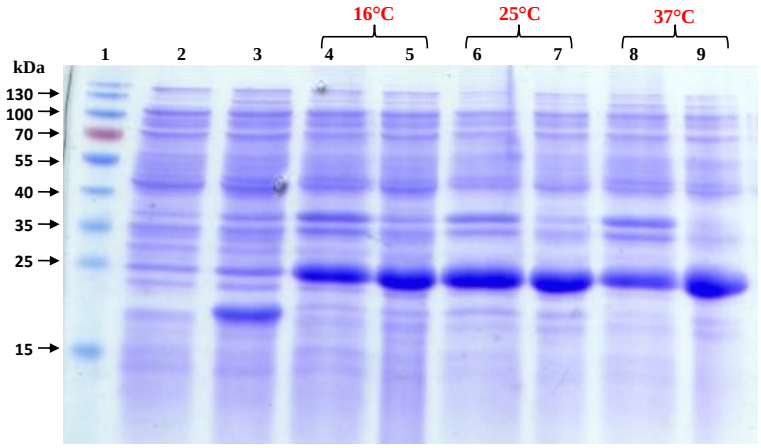


Figure 4.12 12.5% SDS-PAGE gel profile of expression of rHaemathrin 2 at different temperatures. Lane 1: Protein molecular weight marker; lane 2: cell lysate of uninduced cells; lane 3: cell lysate of induced cells with pET32a; lane 4: cell pellet of induced cell at 16ºC; lane 5: cell lysate of induced cells at 16ºC; lane 6: cell pellet of induced cell at 25ºC; lane 5: cell lysate of induced cells at 25ºC; lane 8: cell pellet of induced cell at 37ºC; lane 9: cell lysate of induced cells at 37ºC.

4.3.10 Expression of rHaemathrin 2 with different IPTG concentrations

Varying concentration of IPTG was used for over-expression recombinant haemathrins keeping the temperature constant at 37ºC. It was observed that 0.05 mM IPTG was adequate for expression of rHaemathrin (Fig. 4.13). Higher concentration of IPTG did not have any effect on the expression of the recombinant protein.

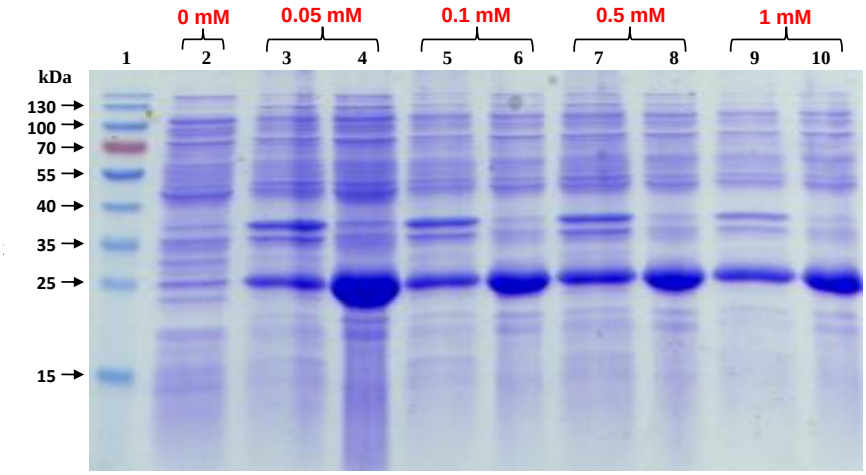


Figure 4.13 12.5% SDS-PAGE gel profile of expression of rHaemathrin 2 using different concentration of IPTG. Lane 1: Protein molecular weight marker; lane 2: cell lysate of uninduced cells; lane 3: cell pellet of cells induced by 0.05 mM IPTG; lane 4: cell lysate of cells induced by 0.05 mM IPTG; lane 5: cell pellet of cells induced by 0.1 mM IPTG; lane 6: cell lysate of cells induced by 0.1 mM IPTG; lane 7: cell pellet of cells induced by 0.5 mM IPTG; lane 8: cell lysate of cells induced by 0.5 mM IPTG; lane 9: cell pellet of cells induced by 1 mM IPTG; lane 10: cell lysate of cells induced by 1 mM IPTG.

4.3.11 Expression of rHaemathrins for different time intervals

Over-expression of the recombinant proteins was also tested for different time interval at 37ºC. It was observed that the recombinant proteins were expressed as soluble and insoluble fraction at 2 h and 4 h incubation time with no visible difference in expression level (Fig. 4.14, lane 4 and 6).

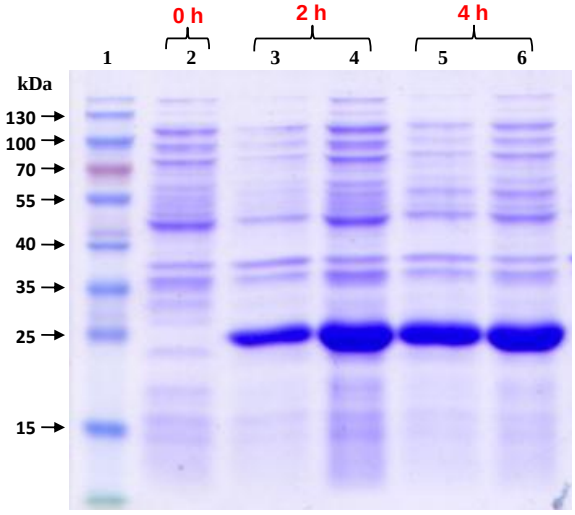


Figure 4.14 12.5% SDS-PAGE gel profile of expression of rHaemathrin 2 using different time intervals of induction. Lane 1: Protein molecular weight marker; lane 2: cell lysate of uninduced cells; lane 3: cell pellet of cells induced for 2 h; lane 4: cell lysate of cells induced for 2 h; lane 5: cell pellet of cells induced for 4 h; lane 6: cell lysate of cells induced for 4 h.

4.3.12 Mass culture of recombinant protein for His-tag purification

The recombinant proteins were produced in 1 L culture media using standard conditions (0.05 mM IPTG induction, 37ºC incubation temperature and 4 h incubation time). The recombinant fusion proteins of about 25 kDa (Fig. 4.15, lane 3 and 4) were produced on induction with 0.05 mM IPTG.

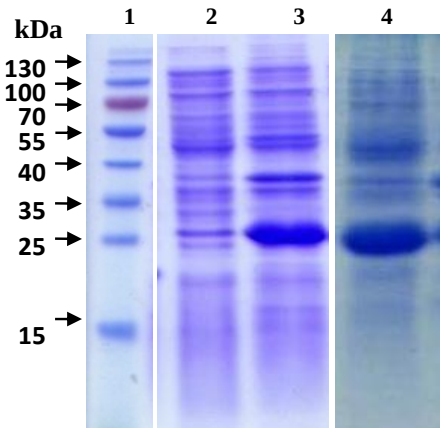


Figure 4.15 12.5% SDS-PAGE gel profile of expression of rHaemathrins using optimized conditions. Lane 1: Protein molecular weight marker; lane 2: cell lysate of uninduced cells; lane 3: cell lysate of cells (rHaemathrin 1); lane 4: cell lysate of cells (rHaemathrin 2).

4.3.13 His-tag purification of rHaemathrins

The soluble fraction obtained from the mass culture was used for purification of the recombinant protein using Ni-NTA agrose beads. The bound recombinant proteins were eluted using 100 mM imidazole from the Ni-NTA agrose beads. Bands of ~25 kDa of partially purified rHaemathrins were observed when eluents was analyzed on SDS-PAGE gel (Fig. 4.16, lane 5 and Fig. 4.17, lane 7).

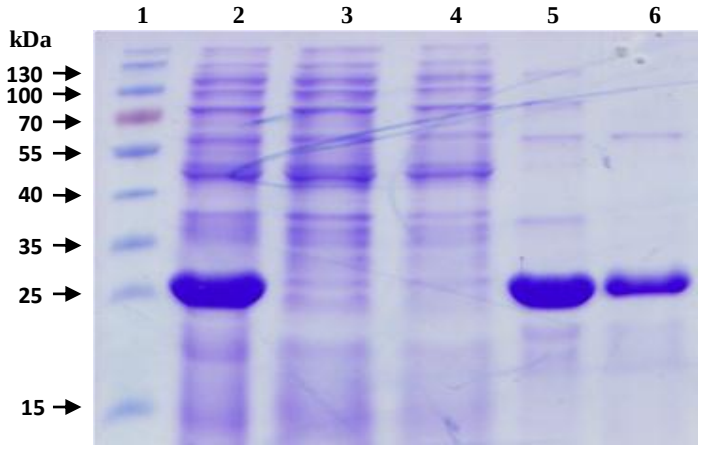


Figure 4.16 12.5% SDS-PAGE gel profile of His-tag purification of rHaemathrin 1. Lane 1: Protein molecular weight marker; lane 2: cell lysate; lane 3: unbound fraction; lane 4: wash fraction 1 (20 mM imidazole); lane 5: elute fraction 1; lane 6: elute fraction 2.

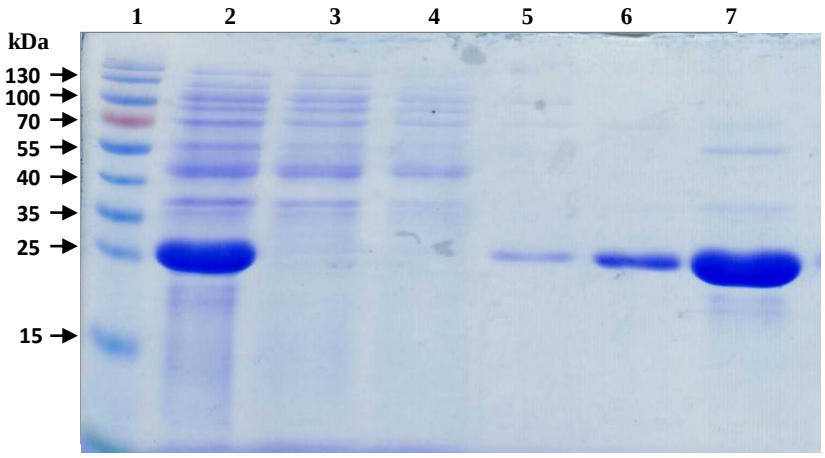


Figure 4.17 12.5% SDS-PAGE gel profile of His-tag purification of rHaemathrin 2. Lane 1: Protein molecular weight marker; lane 2: cell lysate; lane 3: unbound fraction; lane 4: wash fraction 1 (10 mM imidazole); lane 4: wash fraction 2 (20 mM imidazole); lane 5: elute fraction 1; lane 6: elute fraction 2; lane 7: elute fraction 3.

4.3.14 Dialysis of partially purified rHaemathrins

The partially purified rHaemathrins were dialyzed using SnakeSkin™ pleated dialysis tubing to remove salts. Dialyzed rHaemathrin 1 and rHaemathrin 2 were centrifuged and analyzed on 12.5% SDS-PAGE, which showed a prominent bands at ~25 kDa which confirmed that the protein did not aggregate during this process (Fig. 4.18).

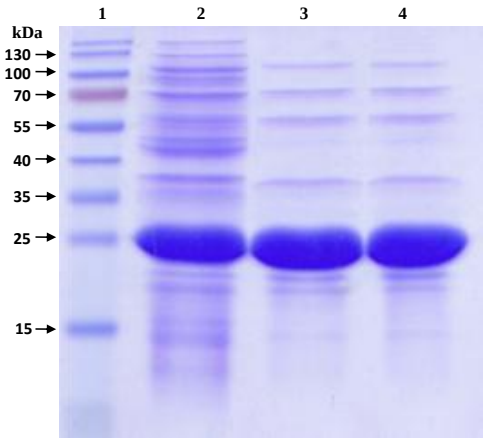


Figure 4.18 12.5% SDS-PAGE gel profile of dialyzed rHaemathrins. Lane 1: Protein molecular weight marker; lane 2: cell lysate; lane 3: dialyzed rHaemathrin 1; lane 4: dialyzed rHaemathrin 2.

4.3.15 Peptide mass fingerprinting of rHaemathrins

The recombinant proteins were subjected to tryptic digestion followed by ESI-LC-MS/MS. Analyses showed that the recombinant proteins contained the trx-tag with His_6-tag and the haemathrin amino acid residues (63-70% coverage; Fig. 4.19). This proved that the expressed proteins are the recombinant proteins of interest.

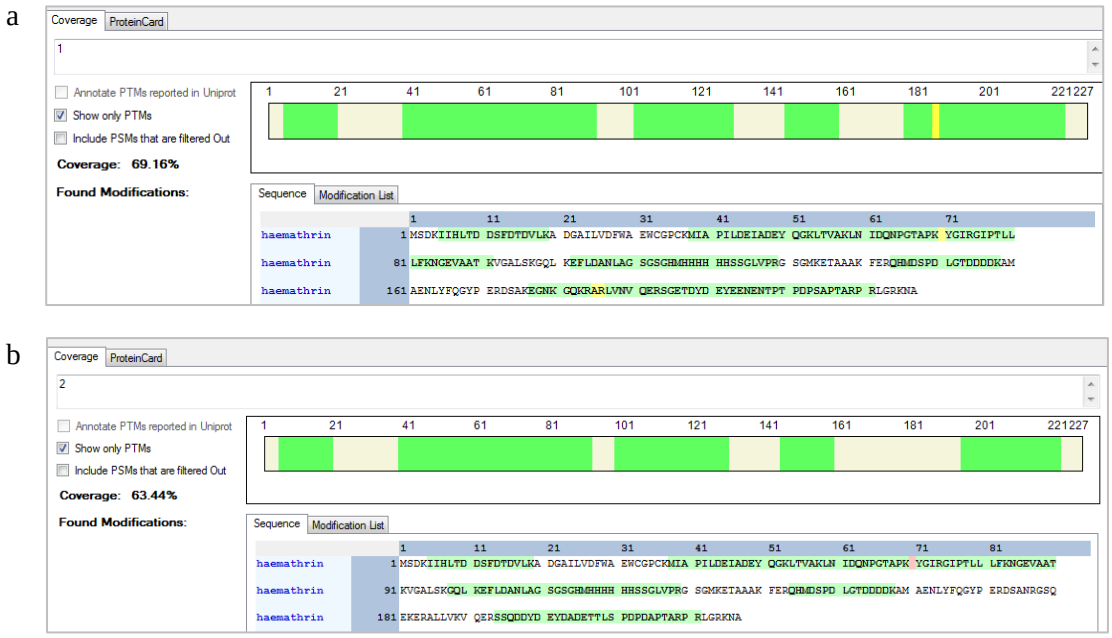


Figure 4.19 MS/MS analysis of recombinant haemathrins of (a) rHaemathrin 1 and (b) rHaemathrin 2. Peptides were identified using Sequest program in offline mode.

4.3.16 Cleavage of rHaemathrins by enterokinase (EK)

The fusion protein was digested with enterokinase which revealed the release of peptidic products of ~7 kDa (Circled, red) and fusion partner of ~18 kDa on Tris-tricine SDS-PAGE gel (Fig. 4.20a and 4.20b). This corresponds to rHaemathrin 1 and rHaemathrin 2 and the fusion tag, respectively. The cleaved peptides were further purified using RP-HPLC. Single peaks corresponding to rHaemathrin 1 (Fig. 4.21a) and rHaemathrin 2 (Fig. 4.21b) were obtained showing the homogeneity of the purified peptides.

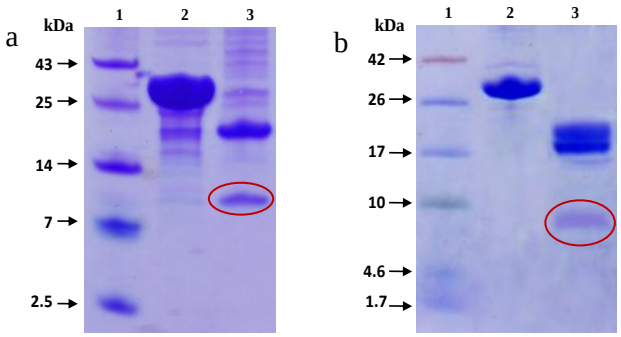


Figure 4.20 18% tricine-SDS-PAGE gel profile enterokinase digested (a) rHaemathrin 1 and (b) rHaemathrin 2. Lane 1: Protein molecular weight marker; lane 2: undigested protein; lane 3: enterokinase digested protein (peptides of interest are circled).

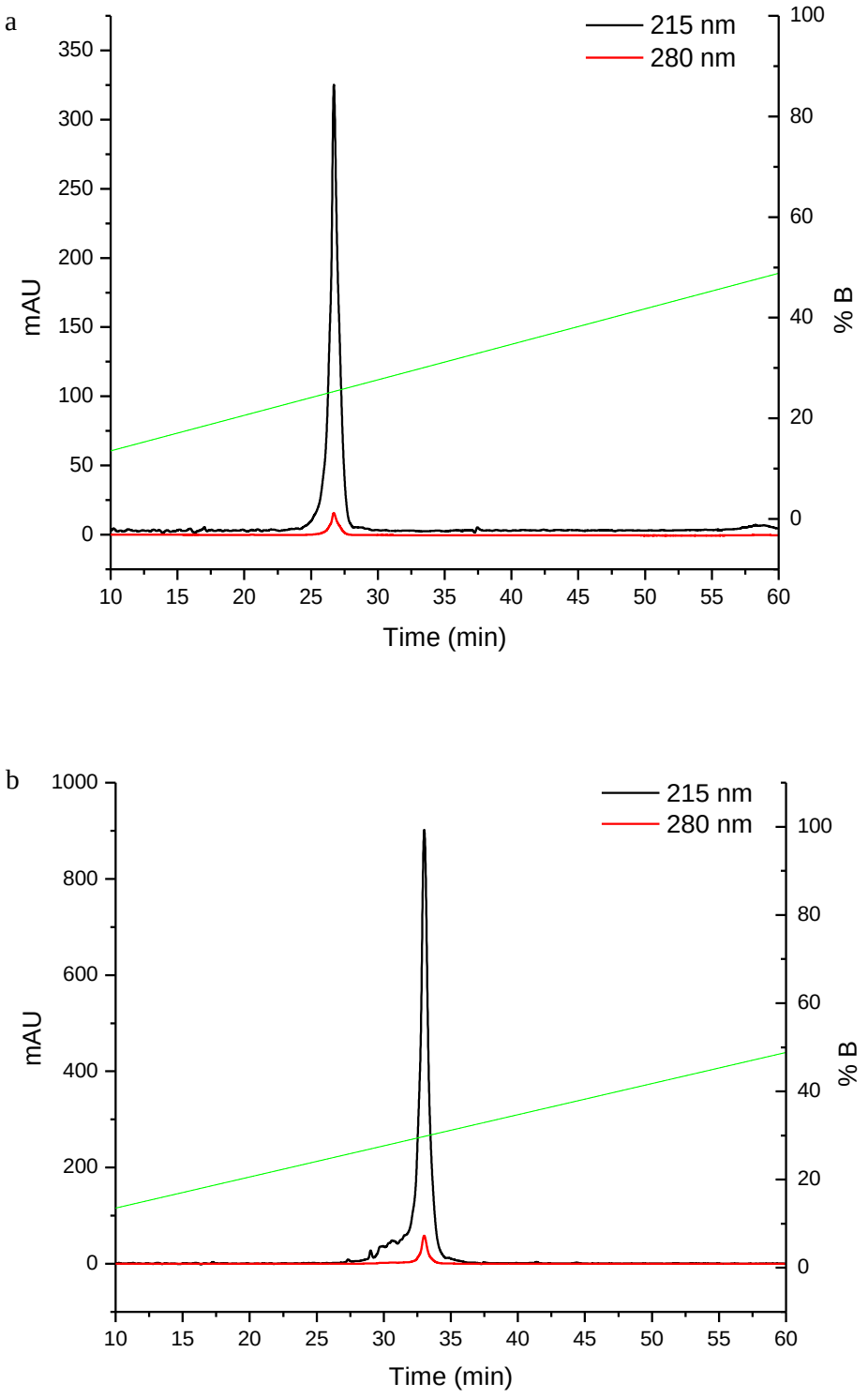


Figure 4.21 Chromatogram showing peaks corresponding to (a) rHaemathrin 1 and (b) rHaemathrin 2.

4.3.17 ESI-MS analysis of recombinant haemathrin 1 and 2

The integrity and molecular masses of the recombinant haemathrins were verified by mass spectrometry. The ESI-MS spectra of recombinant haemathrin 1 and 2 revealed a deconvulated mass of 6690.3 Da and 6709.1 Da, respectively (Fig. 4.22), which were in agreement with the predicted molecular masses of the proteins (6690.1 Da for haemathirn 1 and 6709.1 for haemathrin 2).

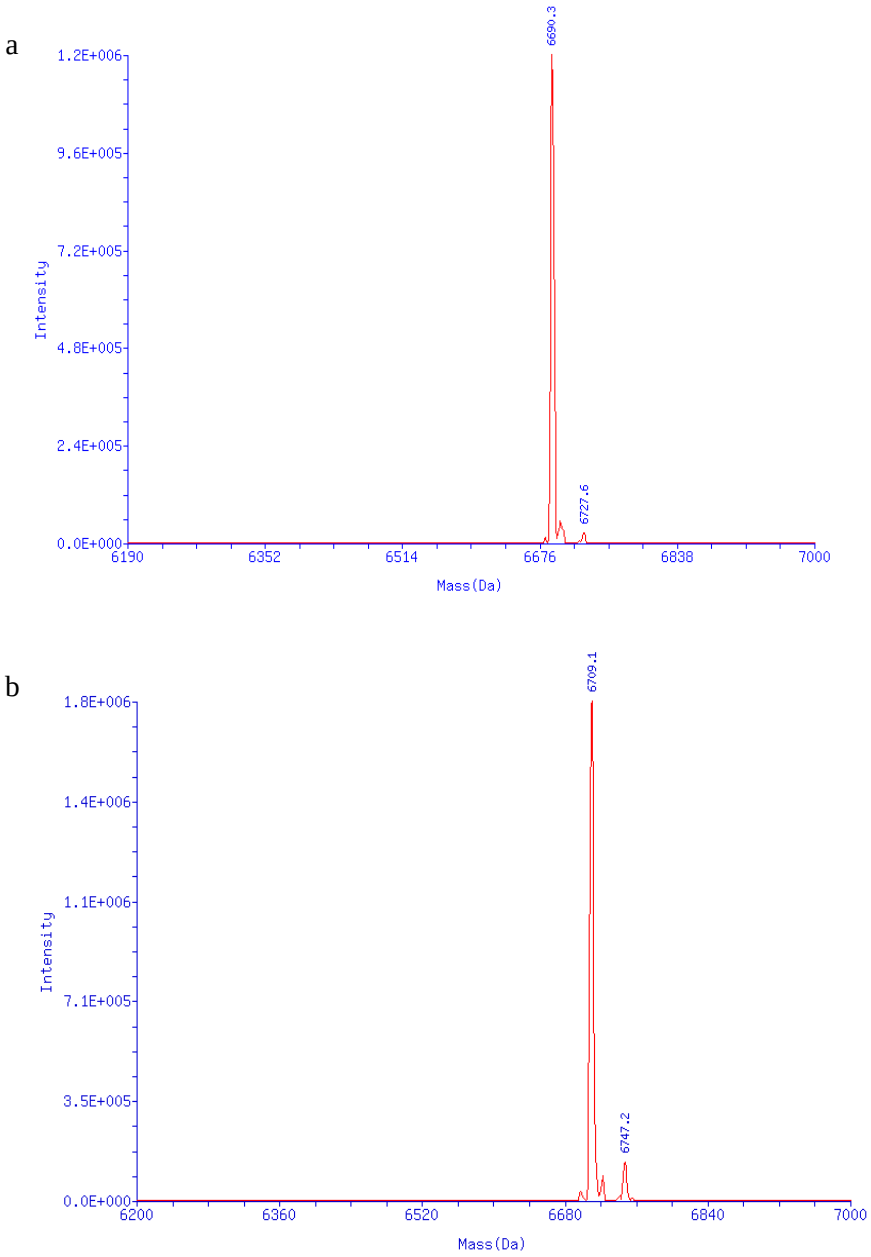


Figure 4.22 Deconvulated ESI-MS spectra of (a) rHaemathrin 1 & (b) rHaemathrin 2.

4.3.18 Circular Dichroism (CD) measurements of rHaemathrins

rHaemathrins were found to lack ordered secondary structure, as shown by their CD spectra that are characteristic of random coils (Fig. 4.23). This indicates that the recombinant peptides are intrinsically disordered in solution.

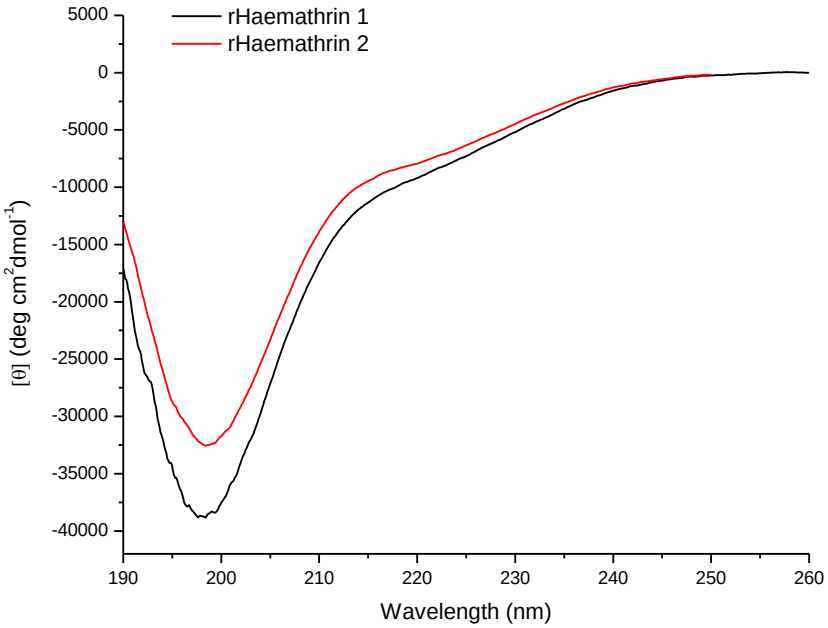


Figure 4.23 CD spectra of purified rHaemathrin 1 (black) and rHaemathrin 2 (red) were recorded in the far-UV region (190-260 nm).

4.4 Discussions

Blood coagulation cascade, a physiological response to vascular injury, is activated by sequential proteolysis of circulating zymogens of serine proteinases resulting in fibrin clot formation [277]. Within the cascade, thrombin plays a central role interacting with most of the zymozens, co-factors and receptors through its catalytic site and anion-binding exosite 1 and 2 [278]. Therefore, thrombin is one of the most targeted proteinases for inhibitors from haematophagous animals [279,280] owing to its role in hemostasis. About 17 tick thrombin inhibitors have been identified, of which the sequence information of seven is known [107]. Most of the thrombin inhibitors characterized from ticks are kunitz-type proteinase inhibitors. These are found in both soft (ornithodorin from *Ornithodoros moubata*, savignin form *O. savignyi*, monobin from *Argas monolakensis*) and hard ticks (boophilin from

R. (B.) microplus, amblin from *Amblyomma hebraeum*, hemalin from *Haemaphysalis longicornis*, variegin from *Amblyomma variegatum*) [109-116]. These thrombin inhibitors are potential targets for therapeutic drugs against thromboembolic disorders [281,282]. There are several drugs developed from thrombin inhibitors from natural sources, but there are still no drugs developed from thrombin inhibitors or other salivary components from ticks.

This chapter reports the isolation, cloning and expression of genes coding for cysteine-less thrombin inhibitors of Inhibitor family I53 from *Haemaphysalis bispinosa*. Amplified products of 331 bp each of cDNAs coding for madanin like proteins from *H. bispinosa* were obtained. BLAST analysis showed that these are similar to madanins and madanin-like proteins from *H. longicornis*. Madanins are unique class of cysteine-less peptidic thrombin inhibitors isolated and characterized from the bush tick *H. longicornis* and are the founding members of the MEROPS inhibitor family I53 [117]. The *MEROPS* database is a manually curated information resource of proteolytic enzymes, their inhibitors and substrates. The hierarchical classification for peptidase inhibitors was established in 2004 and these are clustered into families, which are in turn clustered into clans [119,120]. Each clan, family or holotype inhibitor is assigned to an identifier 'I'. Each family is assigned to an identifier with the letter 'I' and a serial number (e.g. I1, I12). The family I53 is a new family of inhibitors of the serine-type endopeptidase thrombin (Identifier created on 4 August 2008 – MEROPS 8.2). It includes a novel thrombin inhibitor, madanin, isolated from *Haemaphysalis bispinosa* [118] and no known homolog of the inhibitor exists from other species. These are cleavable competitive inhibitors of thrombin and exert its function by binding to the enzyme's active site [117]. The isolated genes were christened as haemathrin (***Haema****physalis* **thr**ombin **in**hibitor). Haemathrins coded for mature peptides of about 6.7 kDa and showed about 60-70% similarity to madanins from *H. longicornis*. These were found to be cysteine-less thrombin inhibitor like madanins, indicating that the proteins have no three-dimensional structures stabilized by disulfide bonds. The theoretical isoelectric point was comparatively lower for haemathrin 2 (pI=4.65) than haemathrin 1 (pI=5.39). The grand average of hydropathicity (GRAVY) of both the proteins was negative suggesting that these are water soluble. The number of negatively charged amino acid

residues was higher (Asp + Glu = 14) in haemathrin 2 than haemathrin 1 (Asp + Glu =12). The percentage of Glu (13.56%) was 2-fold of that of Asp (6.78%) in haemathrin 1, while the percentage of Glu (10.17%) was less than that of Asp (13.56%) in haemathrin 2. The number of Asp was double in haemathrin 2 to that in haemathrin 1. It was also found that the ratios of the negatively charged amino acid residues and their amides were same in both the proteins, but were reverse. The presence of high number of negatively charged amino acids in haemathrins may account for their interaction with thrombin. The positions of the aromatic amino acid Trp were conserved in both the proteins.

The cDNA coding for the mature peptides of haemathrin 1 and 2 were successfully cloned into pET32a(+) expression vector and the recombinant proteins were over-expressed using *E. coli* BL21(DE3)pLysS expression host. As pET32a(+) vector was used the peptides were expressed with a thioredoxin (trx-tag™) fusion protein of about 18 kDa. Expression of a protein with the trx-tag™ may increase the yield of soluble proteins in the cytoplasm [283]. The fusion protein has a His_6-tag which facilitates the purification of the recombinant protein by metal affinity chromatography. To determine the solubility of recombinant protein, the expression of recombinant haemathrins (rHaemathrins) was optimized using different culture condition, i.e., induction at different temperatures, for different time intervals and by different IPTG concentrations. It was observed that there was no significant difference of expressed rHaemathrins when these were expressed at different temperature, under different concentrations of IPTG induction or for different time interval (Fig. 12, 13 14). rHaemathrins were expressed as soluble and insoluble fraction in all of these expression conditions. However, more than 50% of the rHaemathrins was expressed in the soluble fraction which is good enough for purification of the recombinant proteins. Minimum IPTG concentration of 0.05 mM and 37°C of culture condition was sufficient for over-expression of the recombinant protein. Thus rHaemathrins can be overexpressed in simple laboratory condition for downstream processing. Biophysical analysis of the recombinant protein reveals there is no change in the mass of the peptides; however, secondary structure analysis reveals they are intrinsically disordered in solution which is a characteristic of random coils. The absence of

disulfide bridges in haemathrins might be responsible for this disordered state in solution.

Though this is the first report of expression of haemathrins, but several thrombin inhibitors characterized from ticks have been produced using recombinant methods. Isolation of anti-hemostatic proteins from saliva or salivary gland extracts of ticks is difficult and cumbersome, because of the method of extractions and the amount of saliva that can be extracted from the ticks and the time-consuming and pain-staking dissection of the salivary gland from the tiny ticks. Therefore, identification and isolation of these proteins using genomic and recombinant approach is an easy and reproducible method. For example, Madanin 1 and 2 (6770.9 Da and 7122.42 Da, respectively) and chimadanin (7472.1 Da) from *H. longicornis* and monobin (13678 Da) from *Argas monolakensis* have been expressed in prokaryotic expression system [112,118,283], boophilin (20 kDa) from *Rhipicephalus boophilus*, amblin (18 kDa) from *Amblyomma hebraeum* and ixophilin (15kDa) from *Ixodes scapularis* were produced in eukaryotic expression system [125,284], while hemalin from *H. longicornis* was expressed in both prokaryotic and eukaryotic systems (40 and 22 kDa, respectively) [223].

To conclude, two isoforms of a thrombin inhibitor was successfully isolated, characterized and over-expressed in prokaryotic expression system. The characterization of the purified protein is discussed in the succeeding chapter.