# Chapter 3: Protein X and its relation to savicalin, a lipocalin in hemocytes

## 3.1 Introduction

The limited number of tick protein sequences available in databases is a drawback to the identification of tick proteins (Madden *et al.,* 2002; Oleaga *et al.,* 2007). The number of tick sequences is however growing rapidly and MS/MS ion spectra data as well as *de novo* sequences may be used in the future to search databases (Blackburn & Goshe, 2009; Shevchenko *et al.,* 2009). However, until then *de novo* sequences may be employed for the design of degenerate primers for gene cloning. This PCR-based approach will allow for further functional characterization (Lingner *et al.,* 1997, Shevchenko *et al.,* 2001).

In this chapter, this approach was used in an attempt to further characterize Protein X described in Chapter 2. This strategy failed to identify the original tick hemolymph *E. coli* binding protein, but led instead to the discovery of a novel lipocalin-like protein in hemocytes.

Lipocalins play an important role in immune response, transport of hydrophobic molecules (such as pheromones, steroids, bilins, retinoids and lipids), cancer cell interactions and allergies (Pervaiz & Brew, 1987; Cowan *et al.,* 1990; Flower *et al.,* 1991; Peitsch *et al.,* 1991; Nagata *et al.,* 1992). These molecules have been detected across all living organisms and exhibit three characteristic features, namely, an unusually low amino acid sequence similarity (typically 15-25 % between paralogs), a highly conserved protein tertiary structure, and a similar arrangement of exons and introns in the coding sequence of their genes (Mans & Neitz, 2004a, b; Mans *et al.,* 2008b). Lipocalins are single modular proteins of around 150-200 amino acids that fold tightly to form a β-barrel that winds around a central axis. Small hydrophobic molecules are bound in a central pocket (Flower, 2000; Flower *et al.,* 2000; Skerra, 2000). These proteins usually have two helices, one at the N- and one at C-terminal end. The N-terminal helix closes

off one side of the barrel and the C-terminal $\alpha$-helix packs against the outer surface of the barrel. Most lipocalins are classified based on variations observed in the length of the N- and C- terminal segments. These changes are not expected to alter the $\beta$-barrel core significantly, and could provide variation for functional diversification and specialization (Montfort *et al.,* 2000).

To date there are more than 300 lipocalin-like molecules in ticks (Tables 3.1, 3.2 and Appendix I, II). Tick lipocalins identified from saliva or salivary glands are distinct from other arthropod lipocalins and could only be assigned to the lipocalin family based on structural similarity (Paesen *et al.,* 2000; Mans *et al.,* 2003). The crystal structure of histamine-binding protein (HBP) from the hard tick, *R. appendiculatus*, established the first functional relationship of tick lipocalins and their ligands and indicated that tick lipocalins could function in an immunomodulatory capacity by scavenging histamine (Paesen *et al.,* 1999; 2000).

Since then, both hard and soft tick lipocalins have been implicated in the binding of a variety of bio-active ligands such as histamine, serotonin, leukotriene B4 ($LTB_4$), leukotriene C4 ($LTC_4$), arachidonic acid (AA), and thromboxane A2 ($TXA_2$) involved in immuno-modulation and platelet aggregation (Mans *et al.,* 2008a, b; Mans & Ribeiro, 2008a, b; Sangamnatdej *et al.,* 2002). Soft tick lipocalins have also been shown to target complement C5 as well as being associated with toxic effects (Mans *et al.,* 2002, 2003; Mans 2005; Mans *et al.,* 2008b; Mans &Ribeiro, 2008a, b; Nunn *et al.,* 2005).

**Table 3.1 Lipocalin functions from hard ticks**

| Gene accession code | Species | name | Expression | Function | Ligand | Reference |
|---|---|---|---|---|---|---|
| **18032205** | *D. reticulatus* | SHBP | Salivary gland | Suppression of inflammation during feeding | Serotonin and histamine (Two hydrophobic binding pockets) | Sangamnatdej *et al.,* 2002 |
| **PDB: 1QFV 3452093** | *R. appendiculatus* | HBP1-3 | Salivary gland (Stage and gender specific) | Suppression of inflammation during feeding | Histamine (Two hydrophobic binding pockets) | Paesen *et al.,*1999 |
| **219935276** | *I. ricinus* | LIR6 or Ir-LBP | Salivary gland | Inhibition of neutrophil chemotaxis and host inflammation, delayed LTB4 induced apoptosis, decreased activation of neutrophils | $LTB_4$ | Beaufays *et al.,* 2008a, 2008b |
| **67083329** | *I. scapularis* | IXOSC | Salivary gland | Suppression of inflammation | Serotonin | Ribeiro *et al.,* 2006; Mans *et al.,* 2008b |
| **67083266** | *I. scapularis* | IXOSC | Salivary gland | Suppression of inflammation | Serotonin | Ribeiro *et al.,* 2006; Mans *et al.,* 2008b |

**Table 3.2 Lipocalin functions from soft ticks**

| Gene accession code | Species | name | Expression | Function | Ligand | Reference |
|---|---|---|---|---|---|---|
| **114153055** | *A. monolakensis* | AM-182 | Salivary gland | Suppression of inflammation | Serotonin and histamine | Mans *et al.* ,2008a |
| **114152973** | *A. monolakensis* | AM-33 | Salivary gland | Suppression of inflammation | $LTC_4$ | Mans *et al.,* 2008a, b |
| **114152975** | *A. monolakensis* | Monotonin (AM-38) | Salivary gland | Anti-platelet aggregation | Serotonin | Mans *et al.,* 2008a, b |
| **114152935** | *A. monolakensis* | Monomine (AM-10) | Salivary gland | Suppression of inflammation | Histamine | Mans *et al.,* 2008a, b |
| **159944** | *O. moubata* | Moubatin | Salivary gland | Inhibition of collagen induced platelet aggregation | AA, $TXA_2$ | Keller *et al.,* 1993, Waxman & Connolly 1993, Mans & Ribeiro 2008b |
| **49409516** | *O. moubata* | Lipocalin (OMCI) | Salivary gland | Inhibition of complement system | Complement C5, $LTB_4$ | Nunn *et al.,* 2005 |
| | *O. moubata* | Lipocalin (OMCI, Chain A) | Recombinant expression (*Pichia methanolica*) | Inhibition of complement system | Fatty acid | Roversi *et al.* ,2007 |
| **149287030** | *O. parkeri* | Moubatin homolog 3 (OP-3) | Salivary gland | Suppression of inflammation | $LTC_4$ | Francischetti *et al* , 2008a; Mans & Ribeiro 2008b |
| **25991386** | *O. savignyi* | TSGP1 | Salivary gland | Salivary gland biogenesis | Histamine and serotonin | Mans *et al.,* 2001, 2003, 2004a; Mans & Ribeiro 2008b |
| **25991388** | *O. savignyi* | TSGP2 | Salivary gland | Suppression of inflammation, Toxic to host cardiovascular system | $LTB_4$, AA, $TXA_2$ and Complement C5 | Mans *et al.,* 2001; 2003; 2004a; Mans & Ribeiro 2008b |
| **25991390** | *O. savignyi* | TSGP3 | Salivary gland | Anti-platelet aggregation | $LTB_4$, Complement C5 | Mans *et al.* ,2001; 2003; 2004a; Mans & Ribeiro 2008b |
| **25991437** | *O. savignyi* | TSGP4 | Salivary gland | Suppression of inflammation, Toxic to host cardiovascular system | $LTC_4$, $LTD_4$, $LTE_4$, | Mans *et al.,* 2001; 2003; 2004a |

### 3.1.1 Lipocalins found in hard ticks

In *R. appendiculatus*, female specific HBPs, have been isolated from the salivary gland and were found to sequester histamine released by the host in response to tissue damage. HBPs fulfill one of the roles described for nitrophorin in hematophagous insects by reducing the immune and inflammatory host responses (Paesen *et al.,* 1999; Montfort *et al.,* 2000).

HBPs consist of two separate internal and binding sites for histamine. The high affinity and low affinity sites are lined with acidic residues, useful for binding a basic ligand. The hydrophobicity of these pockets represents another striking difference with the binding pockets of most lipocalins suited for binding hydrophobic ligands (Paesen *et al.,* 1999; 2000). The high affinity site occupies the position expected for other lipocalins, but the entrance of histamine to this site is anomalous when compared with the open side of other lipocalin pockets. The low affinity site occupies the closed end of the barrel (Paesen *et al.,* 2000).

The expression of HBPs is stage and gender specific in that HBP1 and 2 are secreted by adult females while HBP3 is secreted by larvae, nymphs and adult males. They also differ in their glycosylation and macromolecular complexes. HBP1 and HBP2 are non-glycosylated monomers, while HBP3 forms disulfide-linked dimers. The functional significance of their temporal and gender-dependent regulation and other molecular attributes are not fully understood (Paesen *et al.,* 1999, 2000).

In *D. reticulatus*, another type of lipocalin molecule known as a serotonin and histamine binding protein (SHBP) contains two internal binding sites. Binding of histamine to the high affinity site has been studied. In contrast a ligand for the low affinity site has not yet been identified. Analysis of its structure, however, suggests serotonin to be the most likely candidate ligand (Sangamnatdej *et al.,* 2002).

Lipocalins found in the hard tick, *I. ricinus,* are segregated into phylogenetic groups suggesting potential distinct functions. This was demonstrated by the lipocalin of *I. ricinus* (LIR6) later designated as *I. ricinus* lipocalin leukotriene B4 protein (Ir-LBP). As the name suggests it scavenges leukotriene B4. Other LIRs did not bind any of the other ligands tested. These included 5-hydroxytryptamine, ADP, norepinephrine, platelet activating factor, prostaglandins D2, E2 , $LTB_4$ and $LTC_4$ (Beaufays *et al.,* 2008a, b).

Ten putative lipocalin sequences were obtained from *I. pacificus* by Francischetti *et al.* (2005). The analysis of their primary sequences suggested that they are secreted proteins. So far there is no information on the structural or biochemical properties to understand their role in blood-feeding.

### 3.1.2 Lipocalins found in soft ticks

In *A. monolakensis,* 33 lipocalin-like sequences have been identified from the cDNA library of adult female salivary glands (Mans *et al.,* 2008a, b). Only 3 out of the 33 sequences have been identified as having lipocalin functions. The crystal structure of monotonin (AM-38) and monomine (AM-10) have only one single binding site rather than the two sites described for HBPs in the hard tick *R. appendiculatus*. The binding site of monotonin and monomine displays a similar low affinity binding site like that of HBP. The binding sites of monomine and monotonin are similar to the low affinity site of the female specific HBP. The interaction of the protein with the aliphatic amine group of the ligand is very similar for all of the proteins, whereas specificity is determined by interactions with the aromatic portion of the ligand. Protein interaction with the imidazole ring of histamine differs significantly between the low affinity binding site of HBP and monomine, suggesting that histamine binding has evolved independently in the two lineages (Mans *et al.,* 2008a, b). AM-33, another lipocalin-like molecule from *A. monolakensis*, is related to tick salivary gland protein 4 (TSGP4) of the soft

tick *O. savignyi*, which binds cysteinyl leukotrienes with high affinity (Mans & Ribeiro 2008a).

Four lipocalins were identified in *O. savignyi* and designated as tick salivary gland proteins (TSGPs). TSGPs have been proposed to have a role in salivary gland granule biogenesis and are stored in the secretory granules (Mans *et al.,* 2001, 2003; Mans & Neitz 2004a). They do not bind histamine nor any of the other mediators involved in the control of host response to tick bites. TSGPs do not affect the blood coagulation cascade or ADP- and collagen-induced platelet aggregation. TSGP2 and TSGP4 were identified as toxins that affect the cardiovascular system of the host and are therefore involved in the pathogenesis of toxicosis caused by the *O. savignyi* bite (Mans *et al.,* 2002, 2003). The toxicity of these lipocalins might be considered as detrimental for the feeding parasite especially for ticks that have to spend longer periods of time on the host to complete a meal. TSGPs were modeled using the known structure of HBP2 with a reasonable fit.

A lipocalin identified in *O. moubata* saliva, is moubatin. This molecule shows a similarity with HBPs and displays platelet aggregation inhibitory activity (Keller *et al.,* 1993; Waxman & Connolly, 1993). A recent study by Mans & Riberio (2008b) showed that moubatin and TSGP3 inhibit platelet aggregation by scavenging $TXA_2$ and thus act as potent inhibitors of $TXA_2$ mediated vasoconstriction. TSGP2, on the other hand, is unable to inhibit platelet aggregation due to an amino acid substitution in the lipocalin-binding cavity in position 85 (Mans & Ribeiro, 2008b). Moubatin, *O. moubata* complement inhibitor (OMCI), TSGP2 and TSGP3 scavenge $LTB_4$ which implicates them in the modulation of neutrophil function. As far as the C5 complement ligand is concerned, only TSGP2 and TSGP3 can bind, but not moubatin, in a similar manner as the OMCI. TSGP3 and moubatin have also shown high affinity toward arachidonic acid (Mans & Ribeiro, 2008a).

Recently, other lipocalin sequences have been identified from salivary gland cDNA of *O. coriaceus* and *O. parkeri* (Francishetti *et al.,* 2008a, b). Almost none of their structural and biochemical functions have been determined, except for moubatin homolog 3 (OP-3) which binds to serotonin and histamine (Mans & Ribeiro 2008b).

In general, tick lipocalins are very eccentric members of the lipocalin family, which highlights the versatility of the lipocalin fold to carry out many functions (Paesen *et al.,* 2000). Despite the structural and biochemical differences of tick salivary gland lipocalins, the resemblance to the β-barrel of standard lipocalins and the data of their gene structure show more similarities and therefore they are assigned as lipocalins.

All tick lipocalins described thus far are salivary gland derived and presumed to be involved in tick feeding (Mans *et al.,* 2008b; Valenzuela *et al.,* 2002; Francischetti *et al.,* 2008a, b; Ribeiro *et al.,* 2006). However, lipocalins have been described in other arthropods that are not involved in blood-feeding, but in processes such as development, coloration, defense mechanisms and transport of ligands (Sánchez *et al.,* 1995, 2000, 2006; Weichsel *et al.,* 1998; Andersen *et al.,* 2005; Kayser *et al.,* 2005; Mauchamp *et al.,* 2006). The possibility thus exists that lipocalins might play a much larger role in tick biology that is not limited to the feeding process alone. The current study describes such a lipocalin from the hemocytes of *O. savignyi*.

## 3.2 Hypothesis

Degenerate primers designed from a *de novo* sequence obtained from Protein X will enable its characterization as an immunoprotective agent.

## 3.3 Materials and methods

### 3.3.1 Hemolymph collection and RNA extraction

Ticks were dorsally immobilized with double-sided tape. A 30 gauge needle was used to puncture the first pair of coxae at the base of the trochanter followed by gentle pressure on the abdomen (Johns *et al.,* 1998). The exuding hemolymph was collected (~ 200 µl) from 20 ticks using a glass capillary and was immediately added to 800 µl of TRI-Reagent (Sigma-Aldrich). RNA was isolated according to the manufacturer's instructions.

### 3.3.2 Single stranded cDNA synthesis

Single stranded cDNA was prepared from total RNA (500 ng) using 7 µl of a 12 µM 5´ Smart IIA anchor primer AAG CAG TGG TAT CAA CGC AGA GTA CGC GGG, and a poly-T anchor primer GCT ATC ATT ACC ACA AAC CAC TCT TTT TT. DEPC-$H_2O$ was then added to obtain a final volume of 64 µl. The reaction mixtures were then spun briefly in a microcentrifuge and incubated at 65 $^oC$ for 2 min in a thermal cycler, which allowed for the denaturation of RNA secondary structure. The samples were immediately placed on ice for 5 min in order to prevent the reformation of RNA secondary structure. To each reaction tube, 10 µl 50 X dNTP's (10 mM) (Roche Diagnostics, Indianapolis, USA), 20 µl 5 X first strand buffer, 2 µl of 100 mM DTT, 100 U RNase inhibitor (Promega, Madison, WI, USA), 500 U Superscript III reverse transcriptase (Invitrogen, Carlsbad, CA) and 5 µl DEPC-$H_2O$ were added. The above mentioned components were available with the Superscript III reverse transcriptase kit. The reaction mixtures were mixed by gentle pipeting and then spun down using a microcentrifuge. This was followed by incubation at 42 $^oC$ for 90 min in a Perkin Elmer Gene Amp PCR system 2400. At the end of the incubation period, 2 µl (0.5 M) EDTA was added to stop the reaction. Samples were stored at -20 $^oC$ until purification.

Purification of the first strand DNA was performed using the Nucleospin Extract II PCR clean up and Gel extraction kit (Macherey-Nagel, Duren, Germany). Two volumes of buffer NT was added to 1 volume of sample. This starting buffer contains chaotropic ions which provide the ideal environment for the first strand cDNA to bind to the silica. The

samples were then loaded onto the columns and centrifuged for 1 min at 11 000 *g*. Buffer NT3 (600µl) was added to wash the column. Contaminants were removed by centrifugation at 11 000 *g* for 1 min. Buffer NT3 contains ethanol and for this reason, the columns were centrifuged for an additional 2 min at 11 000 *g* to remove excess ethanol and to dry the silica column. Tubes were then incubated at 70 $^{o}$C for 5 min in order to ensure complete removal of ethanol which would otherwise inhibit downstream reactions. Thereafter, 50 µl of elution buffer (Buffer NE: 5 mM Tris-HCl, pH 8.5) was added and columns were incubated at 37 $^{o}$C for 5 min and subsequently centrifuged at 11 000 *g* for 1 min. The concentration of the DNA using a 10 times dilution, was determined using the Gene Quant and the remaining sample was stored in a low adhesion siliconized tube at -20 $^{o}$C.

### 3.3.3 Degenerate primer design and 3´ RACE

To obtain the coding gene and 3´-untranslated region for Protein X, a degenerate primer TGG ACG GA(T/C) TA(T/C) TA(T/C) GA(T/C) (A/C)G (Integrated DNA Technologies, Coralville, USA) was designed from a *de novo* sequence WTDYYDRM, which had been previously obtained for Protein X (Table 2.2). Single-stranded cDNA, degenerate primer (50 µM) and poly-T anchor primer (10 µM) were used to optimize 3' RACE using Ex Taq polymerase (Takara, Japan). Optimized conditions consisted of an initial cDNA denaturation (94 $^{o}$C, 3 min), hot start addition of enzyme (80 $^{o}$C), followed by 35 cycles of DNA denaturation (94 $^{o}$C, 30 s), annealing (52 $^{o}$C, 25 s) and extension (72 $^{o}$C, 2 min) followed by a final extension (72 $^{o}$C, 7 min).

### 3.3.4 Cloning and sequencing of amplified cDNA

One-tenth of amplified products were analyzed by agarose gel electrophoresis and the rest of the amplified products were precipitated by adding 1/5 volume of sodium acetate (pH 5.0) and 3 volumes of 100 % ethanol. The solution was centrifuged for 45 min at 13 000 *g* at 4 $^{o}$C. This was followed by washing the DNA pellet using 300 µl of 70 % ethanol and centrifuged for 45 min at 13 000 *g* at 4 $^{o}$C. The supernatant was discarded and the pellet was washed with a final volume of 300 µl of 70 % ethanol and centrifuged

for a further 10 min at 13 000 *g* at 4 °C. Once the supernatant was removed the samples were dried in a vacuum concentrator (Bachoffer, Germany). The purified PCR products were reconstituted with 20 µl of ddd$H_2O$ (double sterilized, double distilled deionized water) and the concentration was determined using a Nanodrop spectrophotometer (Amersham Bioscience, USA).

In order to facilitate the downstream sequencing reaction the PCR product purified above was ligated into the pGEM T Easy vector system (Promega, Madison, WI, USA), using 5 µl of 2 x T4 DNA ligase buffer, 1 µl of 50 ng/µl pGEM-T easy vector. The amount of insert to be used was calculated using the formula:

$$\frac{\text{Size of insert (bp)} \times \text{Amount of vector used (ng)}}{\text{Size of vector (bp)}} \times \frac{3}{1}$$

PCR water was added to make up a final volume of 10 µl. The ligation reactions were precipitated by the addition of 1µl tRNA, 2 µl sodium acetate (pH 5.0) and 30 µl 100 % ethanol, to the 10 µl ligation reactions. The solutions were centrifuged for 45 min at 13 000 *g* at 4 °C and washed with 70 % ethanol and centrifuged for a further 10 min at 13 000 *g* at 4 °C. These steps were repeated twice to remove salts completely. Once the supernatant was removed the samples were dried in the vacuum concentrator and reconstituted in 20 µl PCR grade water.

Fresh colonies of *E. coli* (BL21) cells were used to inoculate 50 ml of SOB medium [2 % (w/v) tryptone; 5 % (w/v) yeast extract 10 mM NaCl; and 2.5 mM KCl, pH 7.0] in a 500 ml flask. Cells were proliferated with vigorous aeration overnight at 37 °C. Cells (2.5 ml) were diluted into 250 ml of SOB in a 1 L flask. They were grown for 2~3 h with vigorous aeration at 37 °C until the cells reached an $OD_{600}$ ~ 0.5. Cells were harvested by centrifugation at 2 600 *g* for 10 min. The cell pellet was washed by re-suspension in 250 ml of sterile ice-cold wash buffer [10 % glycerol, 90 % distilled water, (v/v)]. The cell suspension was centrifuged at 2 600 *g* for 15 min. Cells were washed again in 250 ml of

sterile ice-cold double distilled water and the same wash process was repeated. The cell suspension was centrifuged at 2 600 *g* for 15 min and the supernatant was poured off. The cell pellet was resuspended in wash buffer to a final volume of 1 ml. The cell suspension was divided into 0.1 ml aliquots and stored at - 70 $^{o}$C.

Electroporation cuvettes obtained from BioRad were kept at -20 $^{o}$C, 1 h prior to use. *Escherichia coli* (BL21) cells were allowed to thaw on ice. The ligation reaction (20 µl) was added to cells and mixed by gentle swirling. This reaction mixture was then pipetted into the cold cuvette and inserted into a slot in the chamber rack and pulsed at 2000 mV for 5 ms [Electroporator 2510 (Eppendorf, Germany)]. Thereafter, 100 µl of LB-glucose [0.01 % (w/v) tryptone; 0.01 % (w/v) NaCl; 0.005 % (w/v) yeast extract and 2 M D-glucose] was added to the electroporated cells in order to allow for recovery of the cells. The cells were then transferred to the remaining 800 µl LB-glucose. These cell solutions were incubated at 37 $^{o}$C for 45 min with shaking at 250 rpm of which 100 µl were plated onto LB-ampicillin-Xgal-IPTG plates (1 % (w/v) agar in LB-broth, 50 µg/ml ampicillin, 20 µl of 200 mg/ml IPTG and 4 µl of 50 mg/ml X-Gal). Plates were incubated at 37 $^{o}$C for 16 h.

Colony PCR was performed to identify plasmids with the correct inserts. Random colonies were picked with a pipette tip and dipped into a tube that contained 1 µl of 50 µM degenerate primer and 1 µl of 10 µM SP6 primer and 12.5 µl of KapaTaq Ready mix (Taq polymerase, dNTP, buffer, 25 mM $MgCl_2$) from Kapa Bioscience and 10.5 µl PCR water for a total volume of 25 µl. The tip was then dropped into a 5 ml LB-ampicillin (LB-broth containing 5 µl of 50 mg/ml ampicillin) tube and the cells were allowed to proliferate at 37 $^{o}$C in a shaking incubator at 250 rpm. Colonies containing the correct insert were grown overnight (16 h) for subsequent plasmid isolation.

Plasmid DNA was recovered using the NucleoSpin® Ready-to-use system for fast purification of nucleic acids (Macherey-Nagel,, Germany) as described in the manual. Overnight cultures (5 ml) that reached an optical density at 600 nm ($OD_{600}$) of approximately 3 were centrifuged for 10 min at 1 000 *g* in sterile test tubes. The cell

pellet was resuspended in 500 µl of resuspension buffer A1 containing RNases and vortexed. Lysis buffer A2 (500 µl) was then added and mixed by gentle inversion of the tubes 6 times. This solution was incubated at room temperature for 5 min. Neutralization buffer A3 (600 µl) was added and the solutions mixed by gentle inversion of the tubes. Since Buffer A3 is acidic it leads to the neutralization of the solutions. The solutions were then centrifuged at 11 000 *g* for 10 min at room temperature to remove cell lysates. The supernatant was loaded onto columns and centrifuged at 11 000 *g* for 1 min after which 600 µl of wash buffer A4 containing ethanol was added followed by centrifugation at 11 000 *g* for 1 min. An additional centrifugation step of 11 000 *g* for 2 min allowed the removal of excess ethanol and drying of the silica membrane. Purified plasmid was eluted with 50 µl PCR water followed by centrifugation for 1 min at 11 000 *g* after a 1 min incubation at 37 $^{o}$C. The concentration was determined by using the Nanodrop spectrophotometer.

Plasmids (550 ng in 1 µl) with correct inserts were sequenced in both directions either using 1 µl of T7 forward primer with sequence 5´-TAA TAC GAC TCA CTA TAG GG-3´ of and SP6 reverse primer with sequence 5´-TAT TTA GGT GAC ACT ATA G-3´ at concentrations of 5 µM in 3 µl of sequencing buffer, reconstituted to a final volume of 18 µl with PCR water. The reaction mixtures were spun down briefly in a microcentrifuge and incubated at 94 $^{o}$C for 2 min to allow denaturation of the plasmid DNA. The ABI Big Dye solution (2 µl) was added in a hot start addition at 80 $^{o}$C after 1 min. The reactions were then cycled: 96 $^{o}$C for 20 s, 50 $^{o}$C for 30 s and 60 $^{o}$C for 3 min for a total of 26 cycles. The sequencing reactions were precipitated by adding 4 µl of sodium acetate (pH 5.0) and 60 µl of 100 % ethanol was added to each 20 µl sequencing reaction. This solution was then centrifuged at 13 200 *g* at 4 $^{o}$C for 45 min after which the supernatant was gently aspirated. The pellet was washed using 50 µl 70 % ethanol and centrifugation at 13 200 *g* at 4 $^{o}$C for 10 min. The wash step was repeated twice. The nucleotide sequences were analyzed using an automated ABI 3130 DNA sequencer.

DNA sequences were analyzed for similarities with known sequences using the BLAST (Basic Local Alignment Search Tool) algorithm (www.ncbi.nlm.nih.gov). The BLAST algorithm (Altschul *et al.,* 1990; 1997) searches for local (as opposed to global) alignments and reports the significance of the search results as an expect value. The expect value is a parameter that describes the number of hits one can expect to see just by chance when searching a database. It essentially describes the random background noise that exists for matches between sequences. The lower the e-value of a similarity, the higher the probability that the hit is significant. Generally, an e-value of <0.0001 (1e-04) is considered highly significant. The amino acid sequence of the protein was deduced using Bioedit, while Expasy was used to predict amino acid composition, hydrophobicity profile and pI.

### 3.3.5 Sequence retrieval for multiple sequence alignments and phylogenetic analysis

Lipocalin sequences from both hard and soft ticks as listed below were retrieved from the NCBI Genebank database, by BLASTP, TBLASTN and PSI-BLAST analysis using the obtained amino acid sequence for Protein X (Altschul *et al.,* 1990, 1997). All sequences were NCBI database entries. NCBI entries are composed by a common description and Gene bank accession codes. Multiple sequence alignments of the tick lipocalins were performed using ClustalX with default parameters (Jeanmougin *et al.,* 1998; Larkin *et al.,* 2007). Sequences were manually checked and adjusted accordingly. Neighbor joining (NJ) analysis was conducted using MEGA version 4.0 (Saitou & Nei, 1987; Tamura *et al.,* 2007). Gapped positions were completely deleted so that 55 informative sites were used for analysis. Reliability of the inferred tree was evaluated by bootstrap analysis (100 000 replicates).

The sequences referred to are the following:

Lipocalin [*Amblyomma americanum*] (196476629), lipocalin [*Argas monolakensis*] (114153056), lipocalin [*Argas monolakensis*] (114153300), lipocalin [*Argas monolakensis*] (114153282), lipocalin [*Argas monolakensis*] (114153166), lipocalin [*Argas monolakensis*] (114153124), lipocalin [*Argas monolakensis*] (114153090), lipocalin [*Argas monolakensis*] (114153072), lipocalin [*Argas monolakensis*] (114153054), lipocalin [*Argas monolakensis*] (114153036), lipocalin [*Argas monolakensis*] (114152998), lipocalin [*Argas monolakensis*] (114152996), lipocalin [*Argas monolakensis*] (114152994), lipocalin [*Argas monolakensis*]

(114152990), lipocalin [*Argas monolakensis*] (114152982), lipocalin [*Argas monolakensis*] (114152974), lipocalin [*Argas monolakensis*] (114152960), lipocalin [*Argas monolakensis*] (114152958), Chain A, Crystal Structure Of Am182 Serotonin Complex [*Argas monolakensis*] (171849040), Chain A, Crystal Structure Of Monomine [*Argas monolakensis*] (171849042), Monomine [*Argas monolakensis*] (114152936), Monotonin [*Argas monolakensis*] (114152976), Chain A, Crystal Structure Of Monomine-Histamine Complex [*Argas monolakensis*] (171849043), lipocalin-like protein [*Rhipicephalus (Boophilus) microplus*] (45360102), serotonin and histamine binding protein [*Dermacentor reticulatus*] (18032205), putative secreted histamine binding protein of 25.9 kDa [*Ixodes pacificus*] (51011604), putative secreted histamine binding protein of 22.5 kDa [*Ixodes pacificus*] (51011586), IR1 [*Ixodes ricinus*] (219935266), LIR2 [*Ixodes ricinus*] (219935268), LIR3 [*Ixodes ricinus*] (219935270), LIR4 [*Ixodes ricinus*] (219935272), LIR5 [*Ixodes ricinus*] (219935274), LIR6 or Ir-LBP [*Ixodes ricinus*] (219935276), LIR7 [*Ixodes ricinus*] (219935278), LIR8 [*Ixodes ricinus*] (219935280), LIR9 [*Ixodes ricinus*] (219935284), LIR10 [*Ixodes ricinus*] (219935288), LIR11 [*Ixodes ricinus*] (219935290), LIR12 [*Ixodes ricinus*] (219935292), LIR13 [*Ixodes ricinus*] (219935294), LIR14 [*Ixodes ricinus*] (219935296), 25 kDa salivary gland protein A [*Ixodes scapularis*] (15428310), putative secreted salivary protein [*Ixodes scapularis*] (67083547), 25 kDa salivary gland protein B [*Ixodes scapularis*] (15428302), putative protein [*Ixodes scapularis*] (22164276), putative secreted protein with HBP domain [*Ixodes scapularis*] (67083737), putative salivary secreted protein [*Ixodes scapularis*] (67083439), 25 kDa salivary gland protein family member [*Ixodes scapularis*] (67083725), histamine binding protein [*Ixodes scapularis*] (67083717), histamine binding protein [*Ixodes scapularis*] (15428292), histamine binding protein [*Ixodes scapularis*] (67083721), salivary histamine binding protein [*Ixodes scapularis*] (67083719), 25 kDa salivary gland protein C [*Ixodes scapularis*] (67083485), 25 kDa salivary gland protein C, putative [*Ixodes scapularis*] (215509983), putative secreted protein [*Ixodes scapularis*] (67083637), putative salivary HBP family member [*Ixodes scapularis*] (67083407), secreted protein, putative [*Ixodes scapularis*] (215498016), secreted protein, putative [*Ixodes scapularis*] (215502003), putative 22.5 kDa secreted protein [*Ixodes scapularis*] (22164318), putative secreted salivary protein [*Ixodes scapularis*] (67083669), putative secreted protein [*Ixodes scapularis*] (67083623), secreted protein, putative [*Ixodes scapularis*] (215491831), putative salivary secreted protein [*Ixodes scapularis*] (67083682), nymphal histamine binding protein [*Ixodes scapularis*] (67083741), salivary lipocalin [*Ornithodoros coriaceus*] (172051218), salivary secreted lipocalin [*Ornithodoros coriaceus*] (172051166), salivary lipocalin [*Ornithodoros coriaceus*] (172051154), salivary lipocalin [*Ornithodoros coriaceus*] (172051146), salivary lipocalin [*Ornithodoros coriaceus*] (172051116), salivary lipocalin [*Ornithodoros coriaceus*] (172051090), moubatin-like lipocalin [*Ornithodoros coriaceus*] (172051084), salivary lipocalin [*Ornithodoros coriaceus*] (172051236), complement inhibitor precursor [*Ornithodoros moubata*] (49409517), Chain A, The Complement Inhibitor Omci In Complex With Ricinoleic Acid [*Ornithodoros moubata*] (146386434), Chain A, The Complement Inhibitor Omci In Complex With Ricinoleic Acid [*Ornithodoros moubata*] (146386433), salivary secreted lipocalin [*Ornithodoros parkeri*] (149287112), salivary lipocalin [*Ornithodoros parkeri*] (149287038), salivary lipocalin [*Ornithodoros parkeri*] (149287008), truncated salivary lipocalin [*Ornithodoros parkeri*] (149286990), salivary lipocalin [*Ornithodoros parkeri*] (149286978), salivary lipocalin [*Ornithodoros parkeri*] (149286916), salivary lipocalin [*Ornithodoros coriaceus*] (172051234), salivary lipocalin [*Ornithodoros coriaceus*] (172051222), salivary lipocalin [*Ornithodoros coriaceus*] (172051210), moubatin-like lipocalin [*Ornithodoros coriaceus*] (172051206), salivary lipocalin [*Ornithodoros coriaceus*] (172051204), salivary lipocalin [*Ornithodoros coriaceus*] (172051168), salivary lipocalin [*Ornithodoros coriaceus*] (172051114), salivary lipocalin [*Ornithodoros coriaceus*] (172051112), salivary lipocalin [*Ornithodoros coriaceus*] (172051078), salivary lipocalin [*Ornithodoros coriaceus*] (172051062), salivary lipocalin [*Ornithodoros parkeri*] (149287102), salivary lipocalin [*Ornithodoros parkeri*] (149287092), salivary secreted lipocalin [*Ornithodoros parkeri*] (149287088), salivary lipocalin [*Ornithodoros parkeri*] (149287084), salivary lipocalin [*Ornithodoros parkeri*] (149287076), salivary lipocalin [*Ornithodoros parkeri*] (149286974), salivary lipocalin [*Ornithodoros parkeri*] (149286972), moubatin-like 7 [*Ornithodoros parkeri*] (149287170), moubatin-like 4 [*Ornithodoros parkeri*] (149287126), moubatin-like 5 [*Ornithodoros parkeri*] (149287116), moubatin-like 5 variant [*Ornithodoros parkeri*] (149287118), moubatin-like 3 [*Ornithodoros parkeri*] (149287030), moubatin 1-like 2 [*Ornithodoros parkeri*] (149287000), cDNA sequence from whole ticks [*Ornithodoros porcinus*] (17510378), TSGP4 [*Ornithodoros savignyi*] (25991438), TSGP3 [*Ornithodoros savignyi*] (25991391), TSGP2 [*Ornithodoros savignyi*] (25991389), TSGP1/ lipocalin [*Ornithodoros savignyi*] (25991387), Chain A, Histamine Binding Protein From Female Brown Ear *Rhipicephalus appendiculatus* [*Rhipicephalus appendiculatus*] (7767032).

### 3.3.6 Homology modeling and quality assessment

For homology modeling of savicalin, female specific histamine binding protein (PDB ID: 1QFT, 1QFV; Paesen *et al.,* 1999) was used as a template using SWISS-PdbViewer (Guex & Peitsch, 1997). The initial model was submitted to the SWISS-MODEL automated comparative protein modeling server (Guex *et al.,* 1999). Savicalin's sequence was also submitted to the Phyre fold recognition server (Kelley & Sternberg, 2009) and analyzed using the conserved domain database (CDD) (Marchler-Bauer *et al.,* 2009).

### 3.3.7 Transcriptional profiling

Reverse transcription polymerase chain reaction (RT-PCR) was carried out to analyze gene expression of savicalin. *Bacillus subtilis* (ATCC: 13933) cells were resuspended in physiological saline to a final concentration of the 2.5 x $10^6$ cells/ml and 1 μl of the suspension was heat-inactivated and injected into 20 unfed adult female ticks. Saline was injected into the same number of ticks as a control. Total RNA was isolated as described in section 3.3.1 from the midguts, ovaries, salivary glands and hemolymph obtained from both groups of ticks 24 h post injection.

In the second part of the experiment, 20 ticks were fed artificially on heparinized cattle blood (obtained from Experimental Farm, University of Pretoria, SA) infected with *B. subtilis* (2.5 x $10^6$ cells/ml blood) as described section 2.3.4.2. Native blood (no bacteria added) was used for the control group. At 1 day and 10 days post feeding, total RNA was isolated from the same tissues as described for unfed, hemocoelic injected ticks.

First strand synthesis was performed as described in section 3.3.2. For expression analysis, first strand cDNA for each tissue (500 ng), 10 μM of both the gene specific primer TGG ACG GAT TAC TAC GAC CG and a poly-T anchor primer were used. The forward primer CAG ATC ATG TTT GAG ACC TTC AAC and reverse primer G(C/G)C CAT CTC (T/C)TG CTC GAA (A/G)TC at a concentration of 10 μM were used for the amplification of the housekeeping gene, actin. PCR reactions were performed using an

initial denaturation step (94 $^{o}$C, 3 min), hot start addition of exTaq enzyme (80 $^{o}$C) followed by 35 cycles of DNA denaturation (94 $^{o}$C, 30s), annealing (54 $^{o}$C, 25s) and extension (72 $^{o}$C, 2 min) followed by a final extension (72 $^{o}$C, 7 min).

## 3.4 Results and discussion

### 3.4.1 Sequence analysis

As described in the previous chapter, none of *O. savignyi* hemolymph proteins that recognize and bind to *E. coli* bacteria could be identified by searching the current databases with both the MS/MS ion spectra as well as the derived *de novo* sequences obtained for these proteins. In this study, a degenerate primer derived from a *de novo* sequence for one of these hemolymph proteins, was used for 3´-RACE and resulted in the amplification of a single 900 bp fragment from cDNA prepared from total hemocyte RNA. Sequencing did not reveal the sequence of Protein X but instead the full gene sequence of a non-related hemocyte protein (Fig 3.1). The sequence contains a 5'UTR, open reading frame, stop codon, poly-adenylation site and 3'UTR. The translated protein sequence has a signal peptide indicating that the hemocyte protein is targeted to the secretory pathway. The mature processed protein has a calculated pI of 4.37 and molecular mass of 21481.9 Da, that includes 10 cysteine residues predicted to be involved in disulphide bonds.

Failure to amplify Protein X from hemocyte RNA could be as a result of increased stability of the corresponding secondary mRNA structure or possibly that a primer in the degenerate mix may have a higher affinity for the lipocalin mRNA. However, in Chapter 2 the Gram-negative bacteria binding Proteins X and Y were only detected in plasma obtained from challenged ticks and not in the corresponding hemocyte extracts. Therefore, the results obtained here plus the latter results suggest that Protein X is not of hemocytic origin but is synthesized in other tissues and possibly in the fat body. Another possibility is that the HMM bacteria binding proteins observed in the plasma are polymers composed of monomeric proteins of hemocytic origin. For this reason and the fact that the fat body of argasid ticks is distributed throughout the connective tissue of the body and therefore difficult to dissect (Sonenshine, 1991), hemocytes were used in this study.

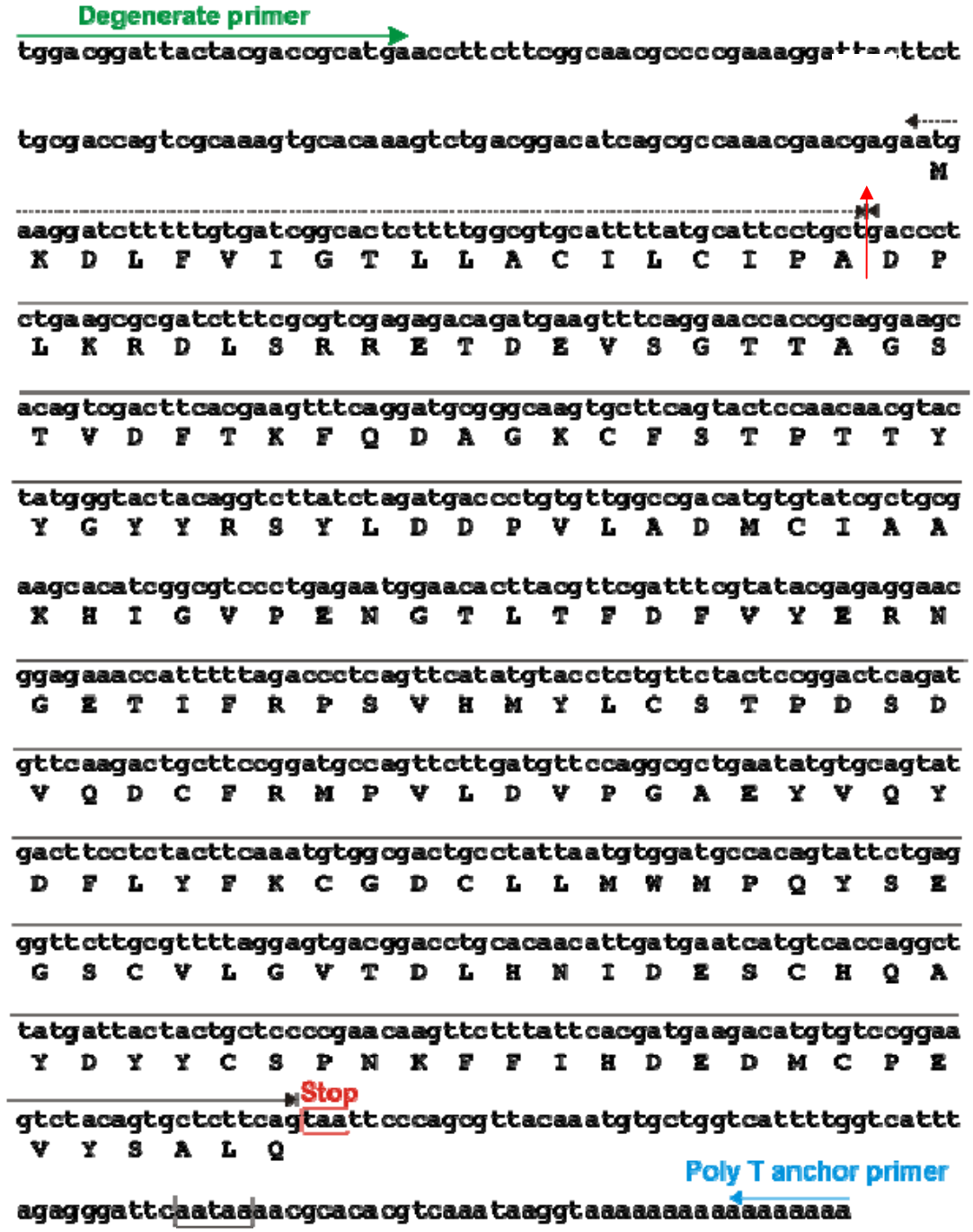



**Figure 3.1: Nucleotide and deduced amino acid sequence of the amplified cDNA.** The mature peptide is indicated by the solid line. The red arrow indicates the putative cleavage site of the signal peptide (broken line). The stop codon is marked with red box. The polyadenylation signal is indicated by the black box. Degenerate primer is indicated by the green arrow. Poly T anchor primer indicated by the blue arrow. Nucleotide sequence data has been submitted to the database (accession number: 298200310).

BLASTP analysis indicated similarity of the sequence to tick lipocalins and the name savicalin was coined for this protein. The three best hits included lipocalins from *R.* (*Boophilus*) *microplus* (E-0.004), *I. scapularis* (E-0.004) and *A. monolakensis* (E-0.003). In addition, a TBLASTN query of the non-redundant EST database retrieved EST sequences (5E-14) from a whole body cDNA library of the closely related tick, *O. porcinus*. The translated EST sequences showed 22 % sequence identity to savicalin. In addition, savicalin was submitted to the CDD and Phyre servers in order to confirm that it belongs to the lipocalin fold. In the case of the CDD analysis savicalin was assigned to the His-binding superfamily, which essentially describes all tick lipocalins. The top 4 hits obtained with the Phyre server were all tick lipocalins for which structures were previously solved and in all cases the estimated precision was greater than 95%, while all other hits corresponded to lipocalins from other organisms. As such, savicalin was assigned to the lipocalin family using three different algorithms that preferentially selected the lipocalin fold from a variety of known sequences and folds. This increased the confidence that savicalin belongs to the lipocalin fold, even if it is divergent.

### 3.4.2 Multiple sequence alignments of tick lipocalins

Multiple sequence alignments with these proteins as well as tick lipocalins that have been functionally characterized indicated that savicalin shows overall less than 20 % sequence identity to these tick lipocalins (Fig 3.2).

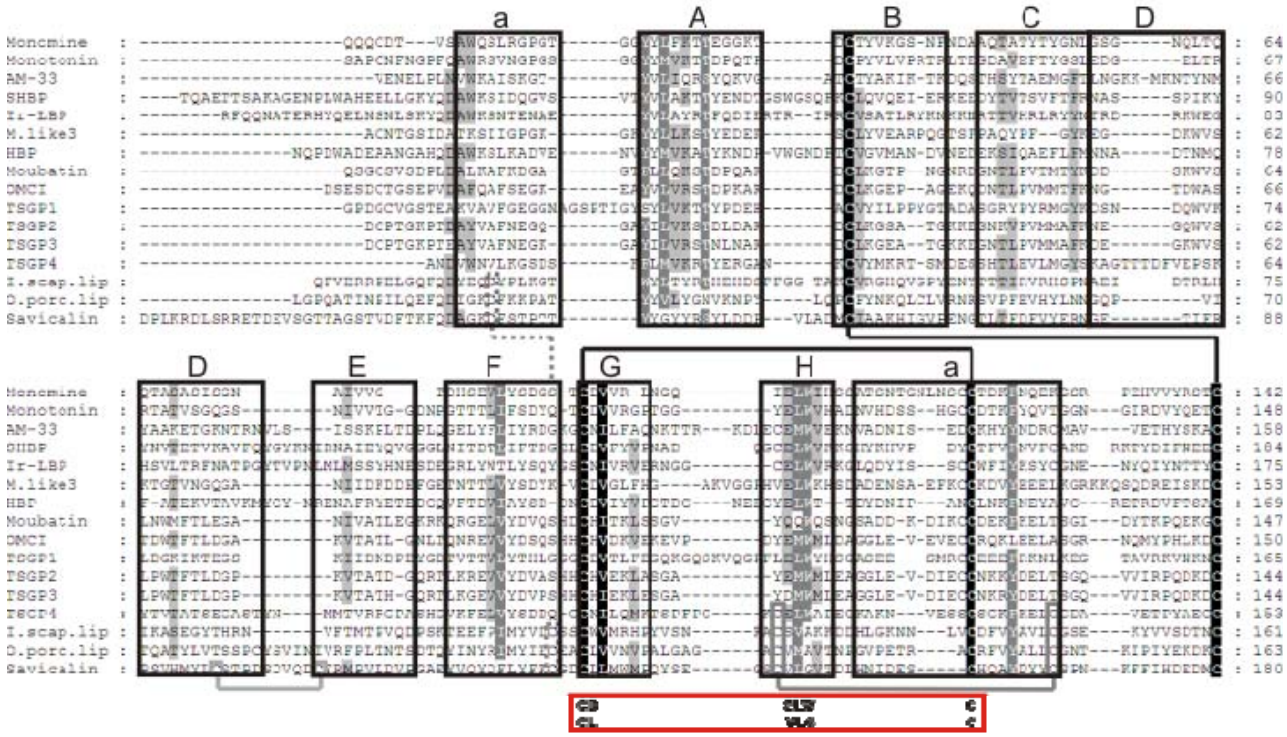



**Figure 3.2: Multiple sequence alignments of tick lipocalins.** Alignment of savicalin with the following molecules and their accession codes: Monomine (114152936), Monotonin (114152976) and Am-33 (114152974) from the soft tick, *A. monolakensis;* SHBP (18032205) from the hard tick, *Dermacentor reticulatus*; Ir-LBP (219935277) from the hard tick, *Ixodes ricinus;* M.like3 (Moubatin like 3, 149287030) from the soft tick, *O. parkeri;* HBP (7767032) from the hard tick, *R. appendiculatus;* Moubatin (159945) and OMCI (49409517) from the soft tick, *O. moubata;* TSGPs (TSGP1: 25991387, TSGP2: 25991389, TSGP3: 25991391, TSGP4: 25991438) from the soft tick, *O. savignyi;* O.porc.lip *(O. porcinus* lipocalin, 29779506) from the soft tick, *O. porcinus*; I.scap.lip *(Ixodes scapularis* lipocalin, 241679301) from the hard tick, *I. scapularis*. A secondary structure based on the SWISS model of savicalin is boxed in black as α-helices (a) and β-sheets (A~H). Conserved cysteine residues found in both hard and soft ticks with predicted disulphide bonds are indicated with solid black line. The solid grey line indicates shared TSGP 4 fold disulphide bond. The dotted grey line indicates disulphide bond shared with the *I. scapularis* lipocalin. The light grey line indicates disulphide bond unique only to savicalin. Red rectangular box indicates biogenic amine binding motif compared to CL-VLG-C sequence obtained from savicalin.

Conserved features include two disulphide bonds found in both hard and soft ticks (Cys59-Cys180; Cys132-Cys159) (Mans *et al.,* 2003). A third disulphide bond (Cys145-Cys167) is shared with TSGP4, the serotonin and histamine binding protein from *D. reticulatus*, and an *I. scapularis* sequence, that is characteristic of the leukotriene C4 binding clade of soft ticks (Mans & Ribeiro, 2008a). A fourth disulphide bond is shared with the same *I. scapularis* sequence (Cys35-Cys129). The fifth disulphide bond is unique and links the β strands D and E (Cys96-Cys106) (Fig 3.2). Savicalin lacks the biogenic amine-binding (BAB) motif (CL[L]X(11)VL[G]X(10)C vs CD[VIL]X(7,17)EL[WY]X(11,30)C),

and would therefore not bind biogenic amines (Mans *et al.,* 2008b). In addition, the residues proposed to be involved in leukotriene binding and complement C5 interaction (Mans & Ribeiro, 2008b), are not conserved in savicalin either, suggesting that it will lack these functions as well.

### 3.4.3 Structural modeling of savicalin

The molecular model obtained presents all secondary structure features associated with lipocalins (Fig 3.3). This includes the eight stranded anti-parallel +1 beta-barrel, the N-terminal helix that closes off the barrel and the C-terminal alpha-helix that packs against the barrel (Flower 2000; Flower *et al.,* 2000; Skerra 2000). This indicated that the overall features of the model fits well with the proposed lipocalin structure and supports the inclusion of savicalin into this family. All cysteine residues are spatially organized to form intact disulphide bonds in the model, supporting the proposed disulphide bond pattern for savicalin (Fig 3.2; Fig 3.3).

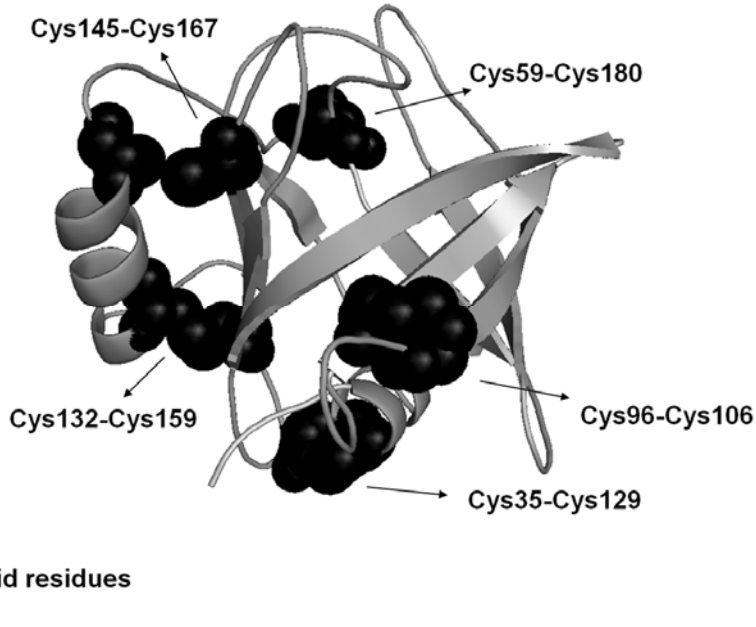
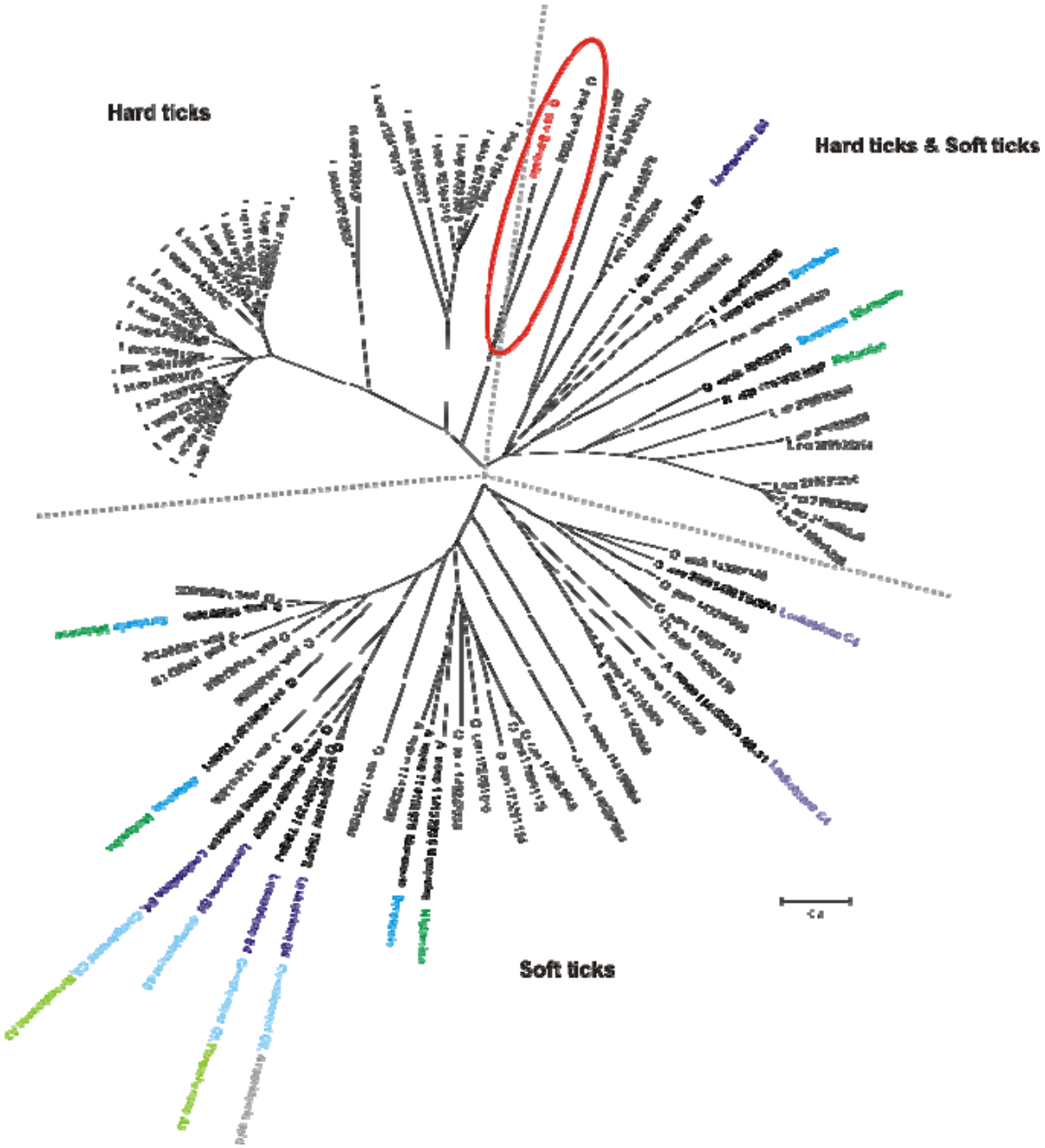



**Figure 3.3: Structural modeling of savicalin.** Cysteine residues are indicated with space fill spheres with their corresponding disulphide bonds. The root mean square deviation (RMSD) value compared with the modeling template HBP-2 is also indicated.

### 3.4.4 Phylogenetic analysis

Phylogenetic analysis using the sequence set from the multiple sequence alignments indicated that savicalin does not group with any of the known functionally characterized clades, implying that it will lack these functions (Fig. 3.4).




**For figure legend please see following page (Page 68).**

**Figure 3.4: Phylogenetic analysis of the tick lipocalin family.** Lipocalins were retrieved from non-redundant database by BLAST analysis of savicalin. Neighbor joining (NJ) analysis was conducted using MEGA version 4.0. Reliability of the inferred tree was evaluated by bootstrap analysis (100 000 replicates). Sequences are described by a species designation (A. mon: *A. monolakensis*; Am. amer: *A. americanum*; B. micro: *R.* (*Boophilus). microplus*; D. recti: *D. reticulatus*; I. paci: *I. pacificus*; I. scap: *I. scapularis*; I. rici: *Ixodes ricinus*; O. cori: *O. coriaceus*; O. park: *O. parkeri*; O. moub: *O. moubata*; O. porc: *O. porcinus*; O. sav: *O. savignyi*; *R*. app: *R.s appendiculatus*). Savicalin is shown in bold red. Red circle indicate lipocalins that have more than 8 cysteine residues. Grey dotted line divides classes of ticks. Bold black: Functions of lipocalins have been determined experimentally with ligands indicated in different colours (Blue: Serotonin; Green: Histamine; Light purple: Leukotriene C4, Purple: Leukotriene B4; Light blue: Complement C5; Light Green: Thromboxane A2).

Savicalin groups within a clade formed by the three best hits obtained by BLASTP analysis and the translated EST sequence from *O. porcinus*. The support for this clade is quite high, but does not necessarily imply that these proteins are orthologous, as the expected species relationships for *Ornithodoros*, *Argas*, *Ixodes* and *Rhipicephalus* are not recapitulated. It is of interest though that *Ixodes*, *Rhipicephalus* and possibly the *O. porcinus* sequences derive from non-salivary gland tissues and could suggest that their orthologous/ paralogous relationships date back to a split between salivary and non- salivary gland derived sequences.

### 3.4.5 Tissue expression profile of savicalin

Expression profiling by mRNA level detection showed that savicalin was up-regulated in hemolymph of unfed ticks upon hemocoelic bacterial challenge as well as ten days after feeding (Fig 3.5). Down-regulation occurred, however, 1 day after feeding. In contrast, savicalin was not up-regulated in midgut and ovaries irrespective of bacterial challenge or 1 day after feeding, but seems to be constitutively expressed. It is, however, down-regulated 10 days after feeding. Savicalin was also up-regulated in midgut of unfed ticks. In contrast, no transcription was detected in salivary glands.

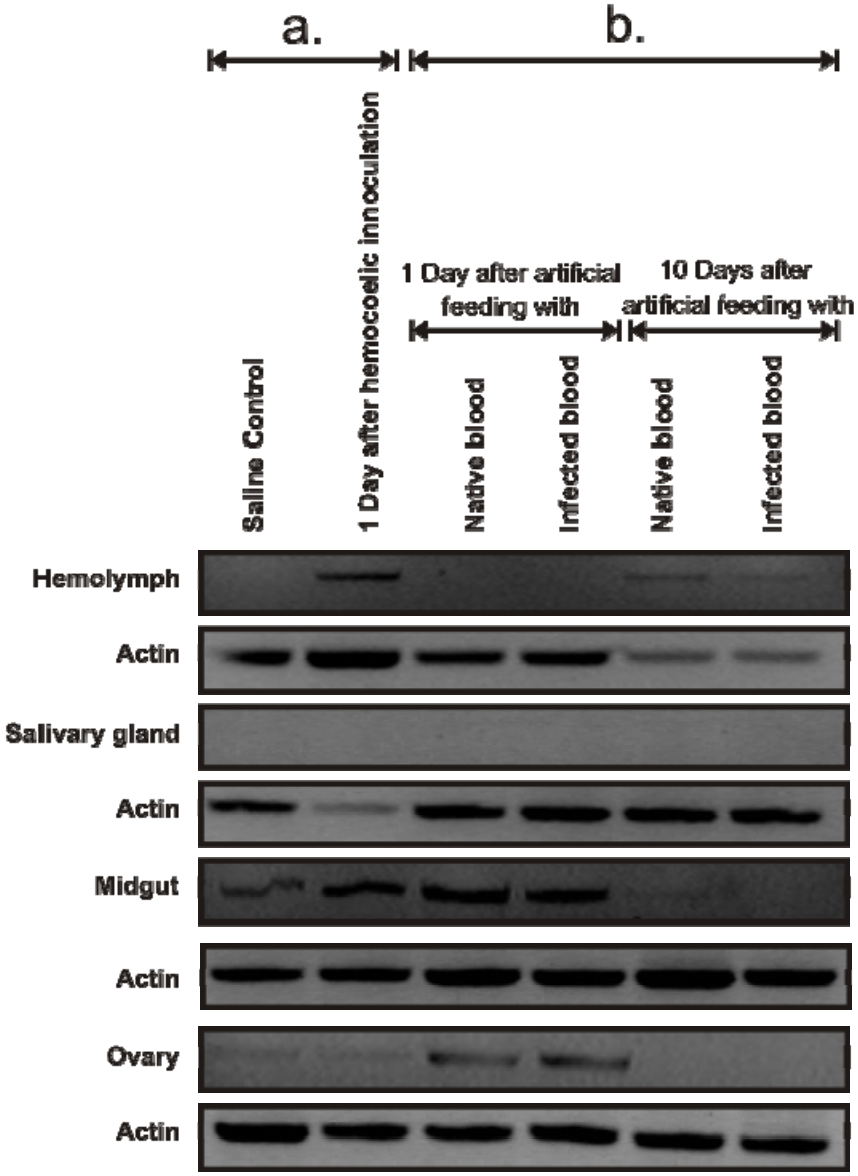



**Figure 3.5 Transcriptional profile of savicalin.** Tissue distribution was analyzed by RT–PCR in the hemolymph, salivary gland, midgut and ovary of the ticks (a) 1 day after hemocoelic inoculation with heat killed *B. subtilis* or saline and (b) 1 day/10 days after artificial feeding with either native or *B. subtilis* infected blood. Actin is shown as the internal standard

The fact that savicalin is absent from salivary gland, suggests that it does not function at the tick-host interface, as found for other salivary gland derived lipocalins. Expression patterns do, however, indicate that its expression is up-regulated during feeding and this could suggest a role in the post-feeding development of the tick. This could include processes such as blood-digestion,

molting and embryogenesis (Sonenshine, 1991). Orthologs for this lipocalin could also be discovered once more lipocalins are described that are not expressed in the salivary gland. The function for savicalin has not yet been determined, but given its tissue distribution, expression patterns and the fact that it is lipocalin, it would be likely that it can act as a scavenger or transporter of bio-active molecules involved in post-feeding development of soft ticks or have an antimicrobial role.