Chapter-II
Proteomic analysis of midgut of female
*Anopheles stephensi* using Mass Spectrometry
2.1 Introduction

Malaria is caused by Protozoan parasite known as *Plasmodium*, which is transmitted by the infective bite of an *Anopheles* mosquito. Mosquitoes play a central role in the development and *transmission* of malaria parasites. *Anopheles stephensi* Liston is one of the major vectors of malaria in urban areas of India responsible for the transmission of 12-15% of total cases (Sumodan et al., 2004; Adak et al., 2005; Kumar et al., 2006). The species is not a sibling species complex but has three well recognized variants, *Anopheles stephensi* sensu stricto, (Type form), which is an urban form and *Anopheles stephensi* *mysorensis*, a rural form besides a third variant, *Anopheles stephensi* intermediate form (Subbarao et al., 1987). In India, *Anopheles stephensi* type and intermediate forms are potent malaria vectors in urban and peri-urban areas respectively. Vector control is central to the disease control. Preventive measures utilized to control malaria have been hampered due to the development of multiple insecticide resistance in mosquitoes and drug resistance in parasites. New strategies therefore need to be developed in order to achieve effective and sustained control of malaria. Such efforts require detailed molecular studies to understand vector biology and behavioral responses of mosquitoes.

Vector genomics has already laid foundation for understanding vector population structure and function at genetic level. The genome sequence of *Anopheles stephensi* (Indian strain) and SDA-500 strain from Pakistan has been recently published (Jiang et al., 2014; Neasefy et al., 2015). The genome assemblies of these two strains have been made available through public repository (https://www.vectorbase.org/). Mosquito midgut is the first point of contact between malaria parasites and vector host as the ingested blood meal from a *Plasmodium* infected person is obtained. Midgut presents the first physical barrier for the parasite during its life cycle in vector. The transmission of malaria to a new vertebrate host depends on the successful development of the malaria parasite in the mosquito midgut. In order to understand interactions between vector and pathogen, it is important and interesting to have knowledge on protein composition of the mosquito midgut. Characterization of the mosquito midgut proteins will identify proteins involved in the invasion and survival of the parasites inside the host vector.
Mass spectrometry-based proteomic analysis has emerged as a powerful tool, not only to perform large-scale analysis of proteins present in the complex tissues but also to characterize and understand protein-protein interactions. Many mosquito proteins are known to be involved in the parasite development and invasion of the host cells. Several potentially important proteins have been identified in vector mosquitoes, particularly in *Anopheles gambiae* (Vlachou et al., 2005; Dana et al., 2006; Mikolajczak et al., 2008; Gonzalez-Lazaro et al., 2009; Serrano-Pinto et al., 2010; Sreenivasamurthy et al., 2013). Proteomics studies on *An. gambiae* have provided numerous insights into the development, innate immunity, vectorial capacity of this important malaria vector and its susceptibility to different control strategies (Prevot et al., 2003; Pskewitz and Shi, 2005; Dinglasan et al., 2009; Chaerkady et al., 2011). However, a very little information is available on the protein composition and their possible functions in *An. stephensi* midgut, which limits understanding on attributes associated with vectorial capacity and susceptibility to various vector control methods. The first study on midgut proteins of blood and sugar fed *Anopheles stephensi* lines having different susceptibility to *Plasmodium falciparum* infection were carried out using two-dimensional gel analysis approach (Prevot et al., 1998). Around 29 proteins were identified exclusively from *P. falciparum* susceptible lines as compare to the less susceptible lines. This study has identified a limited number of proteins from the midgut tissue. Therefore, a more in-depth proteomic analysis was needed using advanced techniques.

A concerted effort in profiling midgut proteome of this important vector mosquito would provide a molecular platform to understand fundamental processes involved in the vector-parasite interaction and thus vectorial capacity. In the present study, proteomic analysis of midgut of sugar fed female *Anopheles stephensi* (Type form) was carried out using multiple fractionation strategies, high resolution mass spectrometry and multiple search algorithms in order to maximize proteome coverage and identification. This proteomic analysis led to the identification of thousands of proteins found to be involved in different biological and molecular functions, immunity and vector parasite interaction.
Female mosquitoes encounter certain challenges, which are never confronted by male primarily being blood digestion and possible *Plasmodium* infection when blood meal is obtained from an infected person in the wild. Studies have been undertaken to characterize sex-biased gene expression in *An. gambiae* (Hahn and Lanzaro, 2005), *Anopheles anthropophagus* (Geng et al., 2009), but such a study has not been carried out in *An. stephensi* thus far. There are a number of biological differences between male and female mosquitoes, such as much longer life span and much larger body size of the females compared to the males, which feed exclusively on flower nectar and sweet juices of plant tissues, while the females feed both on sugars for energy and blood meal for the development of their eggs.

To understand molecular mechanisms underlying, comparative proteomic approach was employed to investigate the differences in the midgut proteomes of male and female *Anopheles stephensi*. In this study, comparative proteomic analysis of male and female midgut proteins was carried out using iTRAQ based quantitative proteomic approach to identify the differentially expressed proteins between the two sexes. The relative quantitative values of proteins expressed in female midgut were compared to male and vice versa. This analysis led to the identification of many proteins which are more abundant in midgut of female tissue as compared to that of male and vice versa. The functional role of these differentially expressed proteins was also studied in detail. The information generated in this study may be useful in devising malaria transmission blocking strategies.

### 2.2 Materials and Methods

#### 2.2.1 Establishment of *Anopheles stephensi* colony

*Anopheles stephensi* larvae were collected from various man made breeding habitats such as curing waters, overhead tanks, wells, etc. in Goa. These larvae were transferred in the labeled plastic containers along with water from respective habitats and transported carefully to the insectary of National Institute of Malaria Research, Field Station at Goa for further development of larvae. For this, the larvae were transferred to the plastic trays containing tap water and reared under laboratory conditions at temperature of 27±2°C,
relative humidity of 70±5% and a photoperiod of 12h of day and night cycle. A pinch of food (mixture of fish food and Cerelac powder™) was given to the growing larvae once in a day till the development of pupal stage. Pupae were collected in plastic bowls containing tap water and then kept inside a closed net cage for emergence of adults. The adult mosquitoes emerged from these pupae were given feed of 10% glucose soaked in a cotton pad. After two days, mosquitoes were allowed to feed on blood using a glass feeder in which blood was maintained at 37°C with the help of water bath. Bowls with water was kept inside the cage for oviposition. In this fashion, a continuous cyclic colony of the An. stephensi was maintained in the insectary.

2.2.2 Dissection of midgut
Midguts were dissected from the 1,000 adult male and female An. stephensi. The dissection was carried out in 0.65% normal saline under dissecting microscope and immediately transferred to the PBS buffer and stored at -80°C until used. Figure 2.1 shows dissected midgut from a female Anopheles stephensi.

Figure 2.1: Dissected midgut from a female Anopheles stephensi.
2.2.3 Protein extraction and quantification
The midguts collected from 1,000 male and female mosquitoes were homogenized in 200 µl of 0.5% SDS using ultrasonication. The extracted suspension was centrifuged at 14,000 rpm for 10 min at 4°C. Supernatant was collected without disturbing the pellet. The protein concentration was determined using BioRad BCA assay kit. The total amount of extracted proteins was calculated and divided into three approximately equal parts for performing three different methods of fractionation. Both protein and peptide level fractionations were attempted to increase the coverage of the protein and peptide identifications. From the extracted protein about 50µg of midgut protein and 5µg of BSA were taken to run on SDS-PAGE. Profile of proteins obtained from midgut of male and female An. stephensi can be seen in (Figure 2.2).

![SDS PAGE profile of proteins](image)

**Figure 2.2:** SDS PAGE profile of proteins obtained from the midgut of male and female *Anopheles stephensi*. Lane 1: Bovine serum albumin (BSA), Lane 2: Protein from male midgut tissue, Lane 3: Protein from female midgut.
2.2.4 Protein fractionation and digestion

2.2.4.1 In-gel digestion

About 200µg of protein from female midgut sample was resolved on 10% SDS-PAGE gel. The gel was stained using Colloidal Coomassie 33 stain (Invitrogen, Carlsbad, CA). Excess stain was removed by giving multiple washings with 10% methanol. The protein lanes were cut into 22 gel pieces depending on the protein profile of sample and then in-gel trypsin digestion was carried out as described in detail below:

1. The destained gel was first rinsed with water and then carefully placed onto a clean and sterilize glass plate. A clean scalpel blade was used to cut the visible bands.
2. Protein bands were excised into small equal sized pieces and transferred into micro-centrifuge tubes containing destaining solution (40mM ammonium bicarbonate, pH- 8, 40% acetonitrile).
3. Tubes were placed on rocker and gently agitated to aid in destaining. The solution was discarded and this procedure was repeated until the gel pieces were completely destained.
4. Then 500ul of 100% acetonitrile was added and incubated for 10-15 min until the gel pieces were completely dehydrated. The remaining acetonitrile was removed as much as possible using gel loading tips.
5. Sufficient amount of freshly prepared reducing agent (5mm DTT in 40mM ABC; prepare shortly before use) was added to completely cover the gel pieces which were then incubate at 60°C for 30 min on a shaker.
6. The tubes were allowed to cool down to room temperature and the reducing agent was discarded using micropipette.
7. The freshly prepared Alkylation solution was added (20mM Iodoacetamide (IAA) in 40mM ABC; prepare shortly before use). Then tubes were incubated for 15 min at room temperature in the dark and quick shake every 5 minutes.
8. Alkylation solution was removed after short spin and gel pieces were dehydrated by adding 100% acetonitrile (volume of alkylation solution should be sufficient to cover gel pieces).
9. Acetonitrile was completely removed after the gel pieces were completely dehydrated and then tubes were kept on ice.

10. Sequencing grade trypsin (Promega) was suspended in chilled 40mM ammonium bicarbonate (2 ml 40 mM ABC add to one vial of trypsin contains 20 ug trypsin). And then trypsin solution was added to the tubes containing gel pieces.

11. Tubes were left on ice for 45mins to 1hr and once the gel pieces were completely rehydrated (more trypsin solution was added if necessary)

12. Once the gel pieces were completely rehydrated, excess trypsin was remove and replaced with 40mM ammonium bicarbonate just enough to cover the gel pieces. The tubes were incubated at 37°C overnight. After the process of overnight in-gel digestion, the digested peptides were extracted from the gel pieces by the following process.

13. Peptide extraction: Tubes were cooled to room temperature and 100 µl of 5% formic acid was added. Tubes were spun and aqueous phase was collected in different set of tubes labeled according to fraction number.

14. 100 µl of extraction buffer (5 % formic acid; 40% acetonitrile) was added to each tube and the tubes were kept on shaker for 30 min. Aqueous phase was collected in the tubes containing supernatant from the earlier step. This second step was repeated once or twice.

15. Final extraction was carried out by adding 100% ACN. After gel pieces completely dehydrated, supernatant was collected in the new tubes. Dehydrated gel pieces can be discarded. Peptide extracts were dried using speedvac and stored for further analysis at -80°C. Dried peptides appeared as translucent yellow to cream color film at the bottom of the tubes.

2.2.4.2 In-solution digestion

400 µg of protein from female midgut tissues were subjected to in-solution trypsin digestion. In-solution digestion was carried out as follows.

1. The proteins were extracted using 8M urea solution. The extracted proteins were then quantified according to Lowry’s method (using Bio-Rad DC Protein assay kit).
2. About 400 µg of proteins were taken in a 1.5 ml microcentrifuge tube. Reducing agent dithiothreitol solution (50 mM) was added to the protein solution so that final concentration of dithiothreitol will be 5 mM and the tube was kept at 60°C for 40 minutes after gentle mixing.

3. The tube was brought to room temperature and alkylating agent iodoacetamide solution (100 mM) was added to the solution so that the final concentration is 10mM. The tube was kept in dark for 15 minutes.

4. After reduction and alkylation, protein solution was diluted with 40mM ammonium bicarbonate in order to reduce the final concentration of urea to 2M from 8M. This is done as trypsin action will be hindered in the presence of 8 M urea.

5. Sequencing grade modified trypsin (Promega, USA) was added at the ratio of 1:20. And reaction was allowed to take place for up to 12-16 hours at 37°C.

6. Formic acid was added to the final concentration of 5% to stop the reaction. An aliquot equivalent to 20 µg of trypsin digested protein was taken out to run on SDS-PAGE to check for digestion efficiency (complete digestion is indicated by absence of bands after staining the gel). The tryptic peptides obtained was desalted and stored at -20°C for future use.

7. These tryptic digested peptide mixtures were divided into two equal amounts for further separation by using off-gel fractionator and reverse-phase liquid chromatography.

2.2.5 Cleaning/desalting of peptide preparation

Peptides obtained after in-solution digestion were cleaned using Sep-Pak columns packed with reverse phase C18 material (Waters) as follows.

1. Tryptic digested peptide solution was acidified to reach the pH range of 2-3 using 10% trifluoroacetic acid (TFA). The acidified solution was centrifuged at 1000 rpm to remove precipitate or particulate matter which can block the column.

2. In a 15 ml syringe C18 Sep-Pak column was fitted. Column was pre-wetted using 100% acetonitrile. Solvent A (0.1% TFA) is passed through the column to remove acetonitrile.
3. Tryptic digest solution was loaded on top of the column and mobile phase is allowed to trickle down under gravity.
4. Column was washed with 12ml of solvent A thrice, applied as 1 + 5 + 6 ml.
5. Peptides were eluted with Solvent B (0.1% TFA, 40% acetonitrile) twice applied as 3 + 3 ml.
6. Eluted peptides were vacuum dried and stored at -80°C until mass spectrometry analysis.

2.2.6 Off-gel fractionation of trypsin digested proteins
The trypsin digested peptides were fractionated pI-based Off-gel electrophoresis using an Off-gel fractionator (Agilent 3100) as per manufacturer’s instructions. Peptides were separated using IPG strip (pH 3-10) by focusing for 50 kVh with maximum current of 50 µA and maximum voltage set to 4000 V. After fractionation, a total of 12 fractions were collected and acidified using 1% TFA and stored at -80 °C until LC-MS/MS analysis.

2.2.7 Reverse-phase liquid chromatography
The digested peptides were fractionated by basic pH reverse-phase liquid chromatography method using solvent B (10 mMtriethyl-ammonium bicarbonate, pH 8.5 in 95% Acetonitrile) with a gradient of 5% to 60% and 1 ml flow rate per minute for over 60 minutes. The profile of chromatograph is show in Figure 2.3. A total of 96 fractions were collected using automatic fraction collector, which were further concatenated to 24 fractions, vacuum dried and stored in -80 °C freezer until further LC-MS/MS analysis.
2.2.8 Cleaning of peptides obtained from Off-gel and bRPLC fractions using C\textsubscript{18} stage tip

Individual bRPLC and Off-gel fractions were cleaned/desalted using C\textsubscript{18} stage tips. Solvent A/washing solution was 0.1% TFA and solvent B/eluting agent was 0.1% TFA, 60% acetonitrile.

2.2.8.1 Cleaning procedure followed was as follows

1. C18 stage tips were prepared by packing two ~2mm diameter discs punched out from 3M Empore C18 discs in 10µl micropipette tips. For punching out the discs pipetting needle (18X4") was used and syringe attached at the back of the needle in used to release the punched discs into the 10µl tip. Gel loader tip or silica capillary was used to slide the material and fix it at the end of the tip.

Figure 2.3: Typical elution profile of tryptic digest of female *Anopheles stephensi* midgut protein in bRPLC chromatography.
2. Fractions were acidified by adding 10% TFA, so that final concentration of TFA in the sample will be 1%.

3. C18 material in stage tips was pre-wetted by 100% acetonitrile followed by washing with solvent A.

4. Acidified peptide sample was loaded on the C18 material using gel loader tip and pushed through the material using 10 ml syringe attached at the back of the stage tip.

5. Two washes of solvent A, each of 50 µl are given in the similar way.

6. Peptides were eluted by adding 30 to 50 µl solvent B and passing it through the C18 material using syringe and slower speed. The stage tips can be incubated at room temperature for 10 minutes after pushing little bit of solvent B through the material to ensure efficient elution of the peptides.

7. Eluted peptides were vacuum dried and stored at -80°C until they were analyzed on mass Spectrometry.

2.2.9 Quantitative proteomic analysis of male and female midgut

Midgut tissues from both male and female mosquito samples were suspended in 0.5% SDS and sonicated. Suspension was centrifuged at 10,000 rpm for 10 min at 4°C to remove cell debris. Supernatant was collected without disturbing the pellet. The extracted proteins were quantified according to Lowry method using Bio-Rad DC Protein assay kit. Equal amount of protein obtained from each respective midgut tissue was loaded on the SDS-PAGE for normalization. The SDS-PAGE profile of proteins from male and female midgut tissue is shown in Figure 2.2.

2.2.10 iTRAQ (isobaric tag for relative and absolute quantitation) labeling

Equal amount of protein samples from each male and female midgut tissue was taken for iTRAQ labeling. Briefly 100µg of protein samples from each male and female midguts of An. stephensi were treated with 2µl of reducing solution (tris(2-carboxyethyl) phosphine) and incubated at 60°C for 1h and alkylated with 1µl of methyl methane thio-sulfonate (MMTS)) at room temperature for 10 min. Proteins were then digested using trypsin (Promega) (1:20) overnight at 37°C. Peptide samples were labeled using iTRAQ reagents
(Applied Biosystems Catalog No. 4352135) containing reporters “iTRAQ 116” and “iTRAQ 117” for male and female, respectively. iTRAQ labeling of peptides was carried out according to manufacturer’s protocol. Labeling was carried out for 2 h at room temperature. After labeling, the peptides from the two samples were pooled in a single tube and then fractionated using strong cation exchange chromatography. The workflow illustrating the steps involved in iTRAQ labeling of midgut proteins from male and female *Anopheles stephensi* (Figure 2.4).

**Figure 2.4:** Outline of the quantitative proteomic strategy. Midgut tissues from male and female *An. stephensi* were used to extract proteins and digested with trypsin. Peptides obtained from the male and female midgut tissue were labeled with iTRAQ reagents followed by LC-MS/MS analysis.
2.2.11 Strong cation exchange chromatography (SCX)

The fractionation of iTRAQ labeled peptides was carried out using strong cation exchange chromatography (SCX) having polysulfoethyl A column (PolyLC, Columbia, MD) (100 × 2.1 mm, 5 μm particles with 300 Å pores) using a nLC Packing HPLC system connected to a probot fraction collector. Fractions were collected at a flow rate of 0.2 ml/min using 80-min gradient of KCl from 0 to 350 mM concentration in 10 mM potassium phosphate buffer and 25% acetonitrile (pH 2.85). The fractions were dried and reconstituted in 10 μl of 2% tri-fluoroacetic acid before mass spectrometric analysis. The profile of chromatograph is show in Figure 2.5.

![Figure 2.5: Typical elution profile of tryptic digest of iTRAQ labeled male and female midgut protein in SCX chromatography.](image)

2.2.12 Mass spectrometry analysis

In this study, we performed a total of 80 LC-MS/MS analyses. 24 bRPLC fractions from midgut, were performed on LTQ-Orbitrap Elite (Thermo Scientific, USA) mass spectrometer interfaced with Easy- nanoLC II nano flow liquid chromatography system (Thermo Scientific), while the remaining 56 fractions (including in-gel, offgel and iTRAQ male and female midgut fractions) were analyzed on LTQ-Orbitrap Velos mass spectrometer interfaced with Proxeon Easy nLC system (Thermo Scientific, Bremen,
The peptides from each fraction were reconstituted in 0.1% formic-acid and loaded on to a pre-column (75μ x 2cm) packed with magic C18 AQ (Michrom Bioresources, USA) 5μ particle and 100Å pore size at flow rate of 5µl per minute. Peptides were resolved at 250nl/min flow rate using a linear gradient of 10 to 35% solvent B (0.1% formic acid in 95% Acetonitrile) over 75 minutes on an analytical column, of 75 μ x 60 cm, 5 μ particle and 100 Å pore size for Elite and 75 μ x 15 cm, 3 μ particle and 100 Å pore size for Velos was packed using nitrogen pressure cell at 2,500 psi. To reduce the back pressure 60 cm analytical column was operated in a heated insulator at 60 ºC temperature using butterfly column heater (Phoenix S&T, Inc. PA, USA) and was fitted on flex ion source that was operated at 2.5 kv voltage (Only for Elite). The analysis on mass spectrometry was carried out in a data dependent manner with a full scans in the range of m/z 350 to 2000. Both MS and MS/MS were acquired and measured using Orbitrap mass analyzer. Full MS scans were measured at a resolution of 120,000 for Elite and 30,000 for Velos at m/z 400. For iTRAQ labelled peptides, both MS and MS/MS scans were acquired in Orbitrap mass analyzer after fragmentation by higher-energy collisional dissociation (HCD) mode at 45% normalized collision energy. Fifteen to twenty most intense precursors were selected for MS/MS and were fragmented using higher-energy collisional dissociation (HCD) method and detected at a mass resolution of 30,000 for Elite and 15,000 for Velos at m/z 400. Automatic gain control for full MS was set to 1 million ions and for MS/MS was set to 0.05 million ions with a maximum ion injection time of 100 and 200 ms respectively. Dynamic exclusion was set to 30 sec and singly charged ions were rejected. Internal calibration was carried out using lock mass option (m/z 445.120025) from ambient air (Olsen et al., 2005). Figure 2.6 illustrates the steps involved in the proteomic analysis of midgut tissue using mass spectrometry.
Figure 2.6: Workflow explaining the steps involved in the proteomic analysis of midgut tissue.

2.2.13 Data analysis

The data obtained was processed using Proteome Discoverer software (version 1.4, Thermo Fisher Scientific, Bremen, Germany) workflow and searched using Sequest and Mascot search algorithm against Vector Base protein database Astel2.0. The search parameters included trypsin as the proteolytic enzyme allowing up to two missed cleavages for protein database searches; methionine oxidation was set as a dynamic modification while carbamido-methylation at cysteine was set as static modification.

In addition, for iTRAQ data, iTRAQ labeling at peptide N-terminus and lysine side chain was given as static modification. Peptide mass error tolerance and fragment mass error tolerance were set to 20 ppm and 0.1 Da, respectively. The protein and peptide data were extracted with search result parameters as peptide rank one and peptide confidence set as high. For the entire data set, false discovery rate (FDR) was calculated by enabling the peptide sequence analysis using a decoy database and a cut-off of 1% was used for identifications.
2.2.14 **Determination of relative abundance of proteins using label free quantification method**

Relative abundance of proteins was carried out using spectral counting based label-free quantitation. Briefly, the normalized spectral abundance factor (NSAF) was calculated for each protein identified in this study. NSAF value for a protein was calculated as described by Paoletti *et al.*, (2006). The protein with highest NSAF value was termed as most abundant while the one with lowest NSAF was termed as least abundant. Different bioinformatics tools were used to find out the domain structure, sequence homology and ortholog information. SMART prediction algorithm was used to determine the domain structure of the hypothetical proteins identified. BLASTP was carried out to find out the protein sequence similarity between closely related mosquito species. In-paranoid tool was used to identify the orthologous proteins in closely related species as previously described by Remm *et al.*, (2001). The best reciprocal hits of proteins between given two species with more than 70% identity were considered. Further, these identified proteins were analyzed for cellular localization and functional categorization based on biological processes and molecular functions using gene ontology annotations available for *An. stephensi* in VectorBase database.

2.2.15 **Relative quantification of proteins identified from male and female midgut**

Relative quantitation of proteins was calculated by using reporter ions quantifier node in Proteome Discoverer software. In brief, relative intensities of reporter ions 116 and 117 released during MS/MS fragmentation of peptides were calculated. Relative intensities of the two reporter ions for each peptides identified were used to determine relative quantity of a protein in male and female midgut tissue as described by Gautam *et al.*, (2012). The unique peptides identified against each protein were used to calculate relative protein content in the two samples. The over-expressed proteins identified in male and female midgut were further analyzed for cellular localization and functional categorization based on biological processes and molecular functions using gene ontology annotations available for *An. stephensi* in VectorBase database.
2.2.16 Data availability
The mass spectrometry data generated in this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD001128 (http://proteomecentral.proteomexchange.org).

2.3 Results and Discussion
2.3.1 Overview of the mass spectrometry data
The goal of the study was to achieve a deep coverage of midgut proteome of *Anopheles stephensi* which is one of the primary vectors of malaria in India and other neighbouring countries in Asia. To this end, proteins from midgut tissue of this vector were isolated and protein profile was observed on SDS-PAGE (Figure 2.2). The protein extract from midgut tissue was either digested directly in-solution with trypsin and then fractionated or first resolved by gel electrophoresis and subsequently digested with trypsin. For in-solution digested peptides, Off-gel and bRPLC was used as the fractionation strategy. In addition, each fraction thus obtained was further resolved on a LC and then analyzed on high resolution mass spectrometer. In total 58 LC-MS/MS runs were performed using high resolution mass spectrometry. In total, 7,16,932 MS/MS spectra were generated from all the runs. The resulting mass spectral data was searched against the predicted protein database of *Anopheles stephensi* Indian strains available on VectorBase (release Astel 2.0 Oct 2013), using the two different search algorithms i.e. Mascot and Sequest. Peptide spectrum matches were filtered with 1% false discovery rate (FDR).
A total of 5,43,028 peptide spectra matches (PSMs) resulted in the identification of approximately 63,525 peptides and 5,434 proteins. The distribution of mass error in parts per million for the entire peptide data set was calculated. Figure 2.7 shows that nearly 95% of the peptides were within ±5 ppm mass error, confirming the high accuracy of peptide data obtained from the mass spectrometer. Orbitrap analyzer was used for both MS and MS/MS acquisition, which has resulted in minimal mass error.
A total of 5,434 proteins were identified from midgut tissue comprising ~ 46% of the total proteome of the *An. stephensi*. As shown in Figure 2.8, only 615 proteins i.e. 11.3% of the proteins were identified with single PSM evidence. On the other hand, 88.6 % of the proteins identified had two or more PSMs. A large fraction of the identified proteins had more than 10 PSMs and this increased the confidence level in the peptides identified through mass spectrometer.
Multiple fractionation strategies used enabled to achieve a broad coverage of peptides. In addition, bRPLC fractions were analyzed on Orbitrap Elite mass spectrometer, which has comparatively higher resolution and scan speed. Distribution of proteins and peptides identified from the various fractionation methods is shown in Figure 2.9 A and B. In brief, 27,672 peptides were identified specifically from bRPLC fractions, 5,393 from In-gel digested bands and 2,589 from Off-gel fractionation. Analyzing bRPLC fractions on Orbitrap Elite instrument resulted in identification of 2,193 proteins, which were not identified in other two fractionation strategies i.e. In-gel & Off-gel. Similarly, 189 unique proteins were identified from in-gel bands and 118 proteins from Off-gel fractions. This shows the advantage of using multiple fractionation strategies and advanced mass spectrometry platform.
2.3.2 Relative abundance of the proteins identified and their biological roles
To check the relative abundance of the proteins, a label free based quantitation was carried out. For all the proteins identified through LC-MS/MS analysis, normalized spectral abundant factor (NSAF) values were calculated. The quantitation is based on the spectral count identified for each protein and the variability in the length of the proteins was accounted to normalize the abundance values. The most abundant proteins identified were Actin, F-type H+-transporting ATPase subunit beta, Sodium/potassium-transporting ATPase subunit alpha, Glyceraldehyde 3-Phosphate Dehydrogenase, Histone H4, ATP carrier protein, Galectin (GAL6), Calmodulin, and Sterol carrier protein-2 (Table 2.1). These proteins were identified with high coverage, multiple unique peptides and PSMs. A representative MS/MS spectrum of some of the highly abundant proteins identified in this study is shown in Figure 2.10 A and B.
### Table 2.1: Partial list of most abundant proteins identified in this study based on normalized spectral abundant factor (NSAF) values.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Accession Id</th>
<th>Protein Description</th>
<th>( \Sigma ) Coverage</th>
<th>( \Sigma ) Unique Peptides</th>
<th>( \Sigma ) Peptides</th>
<th>( \Sigma ) PSMs</th>
<th>#AAs</th>
<th>MW [kDa]</th>
<th>NSAF Values</th>
<th>Domain/ Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ASTEI07742-PA</td>
<td>Actin (similar to <em>An. gambiae</em>)</td>
<td>76.86</td>
<td>14</td>
<td>42</td>
<td>10120</td>
<td>376</td>
<td>41.8</td>
<td>2691.49</td>
<td>Actin/actin-like conserved site</td>
</tr>
<tr>
<td>2</td>
<td>ASTEI11539-PA</td>
<td>F-type H+-transporting ATPase subunit beta (similar to <em>An. gambiae</em>)</td>
<td>88.56</td>
<td>9</td>
<td>36</td>
<td>6227</td>
<td>437</td>
<td>47.0</td>
<td>1424.94</td>
<td>ATP synthase subunit alpha-like domain</td>
</tr>
<tr>
<td>3</td>
<td>ASTEI00450-PA</td>
<td>Sodium/potassium-transporting ATPase subunit alpha (similar to <em>An. gambiae</em>)</td>
<td>65.57</td>
<td>42</td>
<td>97</td>
<td>14396</td>
<td>1031</td>
<td>114.4</td>
<td>1396.31</td>
<td>Cation-transporting P-type ATPase, C-terminal</td>
</tr>
<tr>
<td>4</td>
<td>ASTEI04722-PA</td>
<td>Glyceraldehyde 3-phosphate dehydrogenase (similar to <em>An. gambiae</em>)</td>
<td>98.19</td>
<td>40</td>
<td>40</td>
<td>3944</td>
<td>332</td>
<td>35.4</td>
<td>1187.95</td>
<td>Glyceraldehyde 3-phosphate dehydrogenase, NAD(P) binding domain</td>
</tr>
<tr>
<td>5</td>
<td>ASTEI04162-PA</td>
<td>Histone H4 (similar to <em>Cx. quinquefasciatus</em>)</td>
<td>62.14</td>
<td>15</td>
<td>15</td>
<td>1176</td>
<td>103</td>
<td>11.4</td>
<td>1141.75</td>
<td>TATA box binding protein associated factor (TAF)</td>
</tr>
<tr>
<td>6</td>
<td>ASTEI11772-PA</td>
<td>Histone H2B (similar to <em>An. gambiae</em>)</td>
<td>100.00</td>
<td>2</td>
<td>13</td>
<td>663</td>
<td>67</td>
<td>7.4</td>
<td>989.55</td>
<td>Histone H2B</td>
</tr>
<tr>
<td>7</td>
<td>ASTEI08607-PA</td>
<td>ADP,ATP carrier protein 1 (similar to <em>An. gambiae</em>)</td>
<td>79.22</td>
<td>38</td>
<td>47</td>
<td>2951</td>
<td>332</td>
<td>36.3</td>
<td>888.86</td>
<td>Adenine nucleotide translocator 1</td>
</tr>
<tr>
<td>8</td>
<td>ASTEI0351-PA</td>
<td>Galectin (GALE6) (similar to <em>An. gambiae</em>)</td>
<td>76.22</td>
<td>10</td>
<td>12</td>
<td>1161</td>
<td>143</td>
<td>16.0</td>
<td>811.89</td>
<td>Concanavalin A-like lectin/glucanase domain</td>
</tr>
</tbody>
</table>
### Table 2.1 contd: Partial list of most abundant proteins identified in this study based on normalized spectral abundant factor (NSAF) values.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Accession Id</th>
<th>Protein description</th>
<th>Coverage</th>
<th>Unique Peptides</th>
<th>Peptides</th>
<th>PSMs</th>
<th>AAs</th>
<th>MW [kDa]</th>
<th>NSAF Values</th>
<th>Domain/ Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>ASTEI09025-PA</td>
<td>Calmodulin (similar to <em>An. gambiae</em>)</td>
<td>81.32</td>
<td>12</td>
<td>12</td>
<td>728</td>
<td>91</td>
<td>10.4</td>
<td>800.00</td>
<td>EF-Hand 1, calcium-binding site</td>
</tr>
<tr>
<td>10</td>
<td>ASTEI01359-PA</td>
<td>Elongation factor 1-alpha (similar to <em>An. gambiae</em>)</td>
<td>84.02</td>
<td>21</td>
<td>51</td>
<td>3357</td>
<td>463</td>
<td>50.4</td>
<td>725.05</td>
<td>Elongation factor, GTP-binding domain</td>
</tr>
<tr>
<td>11</td>
<td>ASTEI08379-PA</td>
<td>Hypothetical protein</td>
<td>57.00</td>
<td>25</td>
<td>25</td>
<td>2730</td>
<td>407</td>
<td>45.0</td>
<td>670.76</td>
<td>No conserved domain</td>
</tr>
<tr>
<td>12</td>
<td>ASTEI05637-PA</td>
<td>F-type H+--transporting ATPase subunit d (similar to <em>An. gambiae</em>)</td>
<td>97.67</td>
<td>23</td>
<td>23</td>
<td>1122</td>
<td>172</td>
<td>19.5</td>
<td>652.33</td>
<td>ATPase, F0 complex, subunit D, mitochondrial</td>
</tr>
<tr>
<td>13</td>
<td>ASTEI0732-PA</td>
<td>Fructose-bisphosphatealdolase, class I (similar to <em>An. gambiae</em>)</td>
<td>93.96</td>
<td>33</td>
<td>47</td>
<td>2361</td>
<td>364</td>
<td>39.5</td>
<td>648.63</td>
<td>Aldolase-type TIM barrel</td>
</tr>
<tr>
<td>14</td>
<td>ASTEI09499-PA</td>
<td>Hypothetical protein</td>
<td>70.32</td>
<td>97</td>
<td>101</td>
<td>8560</td>
<td>1385</td>
<td>150.9</td>
<td>618.05</td>
<td>Insulin-like growth factor binding protein, N-terminal</td>
</tr>
<tr>
<td>15</td>
<td>ASTEI01665-PA</td>
<td>Sodium/potassium-transporting ATPase subunit beta (similar to <em>An. gambiae</em>)</td>
<td>59.69</td>
<td>25</td>
<td>27</td>
<td>1948</td>
<td>320</td>
<td>36.2</td>
<td>608.75</td>
<td>Sodium/potassium-transporting ATPase subunit beta</td>
</tr>
<tr>
<td>16</td>
<td>ASTEI01631-PA</td>
<td>Enolase(similar to <em>An. gambiae</em>)</td>
<td>94.92</td>
<td>50</td>
<td>50</td>
<td>2476</td>
<td>433</td>
<td>46.5</td>
<td>571.82</td>
<td>Enolase</td>
</tr>
<tr>
<td>17</td>
<td>ASTEI08310-PA</td>
<td>Myosin heavy chain (similar to <em>An. gambiae</em>)</td>
<td>80.73</td>
<td>104</td>
<td>203</td>
<td>7502</td>
<td>1318</td>
<td>151.4</td>
<td>569.20</td>
<td>IQ motif, EF-hand binding site</td>
</tr>
<tr>
<td>18</td>
<td>ASTEI02421-PA</td>
<td>60S ribosomal protein L18a (partial) (similar to <em>An. gambiae</em>)</td>
<td>71.05</td>
<td>23</td>
<td>27</td>
<td>1045</td>
<td>190</td>
<td>22.8</td>
<td>550.00</td>
<td>50S ribosomal protein L18Ae/60S ribosomal protein L20 and L18a</td>
</tr>
<tr>
<td>19</td>
<td>ASTEI02293-PA</td>
<td>Dihydropyrimidinase(similar to <em>An. gambiae</em>)</td>
<td>32.08</td>
<td>4</td>
<td>4</td>
<td>580</td>
<td>106</td>
<td>12.1</td>
<td>547.17</td>
<td>no conserved domain</td>
</tr>
<tr>
<td>20</td>
<td>ASTEI09967-PA</td>
<td>14-3-3 protein epsilon (similar to <em>An. gambiae</em>)</td>
<td>82.03</td>
<td>31</td>
<td>37</td>
<td>1376</td>
<td>256</td>
<td>29.2</td>
<td>537.50</td>
<td>14-3-3 domain</td>
</tr>
</tbody>
</table>
Figure 2.10 A: MS/MS spectra of peptides for the most abundant proteins in *An. stephensi* that were identified in this study. (A) Peptide sequence LVLEVAQHLGENTVR identified against F-type H+-transporting ATPase subunit beta; (B) Peptide sequence TVTAMDVVYALKR identified against protein Histone H4.
Figure 2.10 B: MS/MS spectra of peptides for the most abundant proteins in *An. stephensi* that were identified in this study. (C) Peptide sequence VFDKDGNGFISAAELR identified against Calmodulin; (D) Peptide sequence LEIQGDATLAALK identified against Sterol carries protein-2.
Large numbers of proteins in *Anopheles stephensi* are still listed as hypothetical proteins. Their orthologs were identified in other mosquito species (*An. gambiae, Culex quinquefasciatus* and *Ae. aegypti*) based on sequence similarity. Interestingly, a hypothetical protein identified in this study was found similar to *Aedes aegypti* sterol carrier protein-2 with 553 PSMs and 72% sequence coverage. This protein is involved in cholesterol uptake in mosquito larvae. Knockdown and gene silencing of sterol carrier protein-2 in *Aedes aegypti* mosquito causes high mortality in developing adults and also reduced egg viability (Blitzer *et al.*, 2005; Kim *et al.*, 2005). Another protein identified was found similar to Histone H4 of *Culex quinquefasciatus* with 100% similarity. Two proteins were found to be similar to galecin and calmodulin of *An. gambiae* with 91 and 100% sequence similarity respectively. In sandfly, *Phlebotomus papatasi*, galecin acts as a binding receptor for parasite *Leishmania major* for its survival in the sandfly. It is also involved in embryonic development and immunity against pathogens (Kamhawi *et al.*, 2004). Calmodulin is a Ca2+ binding protein found in all eukaryotes. It belongs to CaM family which is a major class of calcium sensor proteins involved in intracellular signaling. In *Plasmodium gallinaceum*, it controls morphological differentiation from zygote to ookinete formation (Silva-Neto *et al.*, 2002; Moreno and Docampo, 2003). Calmodulin also regulates odorant receptor function in *Drosophila* (Mukunda *et al.*, 2014).

In this study, several housekeeping proteins which act as cytoskeletal proteins like actin, chitin binding, cuticular protein, myosin, tubulin, laminin, annexin, tropomysin, dynein light chain, dynactin, connectin and cadherin were identified. These proteins are found to be highly abundant in midgut tissue and suggest that these proteins play a significant housekeeping role in the *Anopheles stephensi* midgut. We have identified three members of ERM family proteins such as ezrin, radixin and moesin. These EMR proteins act as molecular linkers between the actin cytoskeleton and transmembrane receptors. Activation of ERM proteins mediates attractive growth cone guidance through the regulation of actin and adhesion receptors. These proteins also play a role in the Rho and Rac signaling pathways (Bretscher, 1999; Marsick *et al.*, 2012). Another important
protein such as actin-depolymerizing factor (ADF)/cofilins as also identified which regulates assembly and disassembly of actin filaments. This protein is expressed in all insects specially *Drosophila melanogaster* and involved in several cellular processes such as motility, cell-polarity during migration and chemotaxis (Blair, 2006; Bamburg, 2010).

In most of the blood sucking insects both endo and exo peptidases are required for complete digestion of blood proteins. Several digestive proteases (like aminopeptidase, carboxypeptidase A & B, trypsin, cathepsin D) and glycosidases (like alpha-amylase and maltase) were identified. These are known digestive proteases found in midgut tissue and involved in blood and sugar digestion in mosquitoes. The glycosidases are involved in sugar digestion, and the proteolytic enzymes are involved in host-specific proteolysis events that occur during the blood-feeding process such as clot prevention and coagulation cascades, and also involved in the digestion of extracellular matrix components. Several other proteases such as catalase, clip-domain serine protease, serine protease inhibitor (serpin) SRPN10, SRPN4, SRPN6, PAS domain containing serine/threonine kinase, serine/threonine-protein kinase PRP4, serine/threonine-protein kinase SRPK1 are also identified. These proteases are related largely to blood digestion and immunity and have been demonstrated in earlier studies in different mosquito species (Koutsos *et al*., 2007; Warr *et al*., 2007; Neira *et al*., 2008). MS/MS spectra of some of the proteolytic enzymes identified in this study are shown in Figure 2.11 A and B.
Figure 2.11 A: Representative MS/MS spectra of proteases identified in this study. (A) Peptide sequence AQLVDDVANLAR identified against Aminopeptidase N; (B) Peptide sequence NCFGIDGNR identified against Carboxypeptidase B.
Figure 2.11 B: Representative MS/MS spectra of proteases identified in this study. (C) Peptide sequence DPSAPEGGEIIFGGSDSK identified against Cathepsin D; (D) Peptide sequence TTLPIIMDLDVCR identified against Serine protease.
Jiang et al., (2014) have reported 94 genes which play an important role in immune responses in *Anopheles stephensi*. Of these, 49 proteins have been identified in this study, which included ubiquitin-conjugating enzyme E2 N, GTP-binding protein Rheb precursor, NFkappaB essential modulator, FAS-associated factor 1, Pellino, TOLL pathway signalling NF-kappaB Relish-like transcription factor, Insulin-like peptide 3 precursor, Leucine-rich immune protein (Coil-less), Long leucine-rich immune protein (LRIM1), TOLL pathway signaling, TORC1 growth control complex subunit Kog1 and IMD pathway signaling IKK-beta (ird5) (Table 2.2). These immune response proteins are well studied in *An. gambiae* midgut and *Drosophila melanogaster* (Grewal, 2009; Laplante and Sabatini, 2012; Garver et al., 2012). They are thought to be either directly involved in anti-microbial defenses or anti-Plasmodium activity. Till date, very limited knowledge is available about the role of these immune genes in *An. stephensi* midgut. Representative MS/MS spectra of some of the immune responses proteins identified in this study are shown in Figure 2.12 A and B.
Table 2.2: List of immunity related proteins identified in *Anopheles stephensi* midgut.

<table>
<thead>
<tr>
<th>S. No</th>
<th>Accession ID</th>
<th>Name</th>
<th>Σ Coverage</th>
<th>Σ# Unique Peptides</th>
<th>Σ# Peptides</th>
<th>Σ# PSMs</th>
<th># AAs</th>
<th>MW [kDa]</th>
<th>NSAF values</th>
<th>Protein description as per <em>An. gambiae</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ASTEI092500-PA</td>
<td>Bendless</td>
<td>57.24</td>
<td>8</td>
<td>8</td>
<td>121</td>
<td>152</td>
<td>17.4</td>
<td>79.61</td>
<td>Ubiquitin-conjugating enzyme E2 N</td>
</tr>
<tr>
<td>2</td>
<td>ASTEI02599-PA</td>
<td>Uev1A</td>
<td>53.15</td>
<td>9</td>
<td>9</td>
<td>103</td>
<td>143</td>
<td>16.3</td>
<td>72.03</td>
<td>Ubiquitin-conjugating enzyme E2 variant</td>
</tr>
<tr>
<td>3</td>
<td>ASTEI09147-PA</td>
<td>Grb2</td>
<td>52.75</td>
<td>10</td>
<td>10</td>
<td>70</td>
<td>182</td>
<td>21.0</td>
<td>38.46</td>
<td>Growth factor receptor-binding protein 2</td>
</tr>
<tr>
<td>4</td>
<td>ASTEI01675-PA</td>
<td>4EBP1</td>
<td>54.10</td>
<td>3</td>
<td>4</td>
<td>25</td>
<td>122</td>
<td>13.2</td>
<td>20.49</td>
<td>Eukaryotic translation initiation factor 4E-binding protein 1</td>
</tr>
<tr>
<td>5</td>
<td>ASTEI08720-PA</td>
<td>Rheb</td>
<td>27.47</td>
<td>5</td>
<td>5</td>
<td>30</td>
<td>182</td>
<td>20.8</td>
<td>16.48</td>
<td>GTP-binding protein Rheb precursor</td>
</tr>
<tr>
<td>6</td>
<td>ASTEI03567-PA</td>
<td>Effetev2</td>
<td>9.62</td>
<td>3</td>
<td>3</td>
<td>32</td>
<td>364</td>
<td>41.8</td>
<td>8.79</td>
<td>Ubiquitin-conjugating enzyme E2D</td>
</tr>
<tr>
<td>7</td>
<td>ASTEI06370-PA</td>
<td>kenny</td>
<td>17.77</td>
<td>10</td>
<td>10</td>
<td>34</td>
<td>546</td>
<td>61.2</td>
<td>6.23</td>
<td>NFkappaB essential modulator</td>
</tr>
<tr>
<td>8</td>
<td>ASTEI03681-PA</td>
<td>Pellino</td>
<td>14.04</td>
<td>4</td>
<td>5</td>
<td>18</td>
<td>399</td>
<td>44.1</td>
<td>4.51</td>
<td>Pellino</td>
</tr>
<tr>
<td>9</td>
<td>ASTEI07221-PA</td>
<td>Casr</td>
<td>14.10</td>
<td>8</td>
<td>8</td>
<td>35</td>
<td>787</td>
<td>87.4</td>
<td>4.45</td>
<td>FAS-associated factor 1</td>
</tr>
<tr>
<td>10</td>
<td>ASTEI0386-PA</td>
<td>LRIM26</td>
<td>9.30</td>
<td>4</td>
<td>4</td>
<td>17</td>
<td>398</td>
<td>43.6</td>
<td>4.27</td>
<td>tRNA pseudouridine13 synthase</td>
</tr>
<tr>
<td>11</td>
<td>ASTEI04964-PA</td>
<td>Hrsv1</td>
<td>16.54</td>
<td>8</td>
<td>9</td>
<td>28</td>
<td>780</td>
<td>86.7</td>
<td>3.59</td>
<td>hepatocyte growth factor-regulated tyrosine kinase substrate</td>
</tr>
<tr>
<td>12</td>
<td>ASTEI03826-PA</td>
<td>REL1</td>
<td>7.04</td>
<td>6</td>
<td>9</td>
<td>46</td>
<td>1350</td>
<td>146.4</td>
<td>3.41</td>
<td>TOLL pathway signalling NF-kappaB Relish-like transcription factor</td>
</tr>
<tr>
<td>13</td>
<td>ASTEI02289-PA</td>
<td>Insulin receptor</td>
<td>12.30</td>
<td>5</td>
<td>7</td>
<td>24</td>
<td>740</td>
<td>84.0</td>
<td>3.24</td>
<td>Insulin receptor</td>
</tr>
<tr>
<td>14</td>
<td>ASTEI05349-PA</td>
<td>ILP1</td>
<td>12.00</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>150</td>
<td>16.9</td>
<td>2.67</td>
<td>Insulin-like peptide 3 precursor</td>
</tr>
<tr>
<td>15</td>
<td>ASTEI07671-PA</td>
<td>Cactus</td>
<td>8.66</td>
<td>3</td>
<td>3</td>
<td>10</td>
<td>381</td>
<td>41.4</td>
<td>2.62</td>
<td>TOLL pathway signalling.</td>
</tr>
<tr>
<td>16</td>
<td>ASTEI05261-PA</td>
<td>CYLDv1</td>
<td>10.22</td>
<td>11</td>
<td>11</td>
<td>33</td>
<td>1291</td>
<td>140.1</td>
<td>2.56</td>
<td>Ubiquitin thioesterase CYLD</td>
</tr>
<tr>
<td>17</td>
<td>ASTEI08192-PA</td>
<td>Spt6</td>
<td>11.00</td>
<td>15</td>
<td>15</td>
<td>46</td>
<td>1828</td>
<td>209.2</td>
<td>2.52</td>
<td>Transcription elongation factor SPT6</td>
</tr>
</tbody>
</table>
Table 2.2 contd: List of immunity related proteins identified in *Anopheles stephensi* midgut.

<table>
<thead>
<tr>
<th>S. No</th>
<th>Accession ID</th>
<th>Name</th>
<th>Coverage</th>
<th># Unique Peptides</th>
<th># Peptides</th>
<th># PSMs</th>
<th># AAs</th>
<th>MW [kDa]</th>
<th>NSAF values</th>
<th>Protein description as per An. gambiae</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>ASTEI11789-PA</td>
<td>Slmb</td>
<td>12.50</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>80</td>
<td>9.7</td>
<td>2.5</td>
<td>F-box and WD-40 domain protein 1/11</td>
</tr>
<tr>
<td>19</td>
<td>ASTEI07471-PA</td>
<td>Slmb</td>
<td>9.23</td>
<td>5</td>
<td>5</td>
<td>12</td>
<td>509</td>
<td>55.5</td>
<td>2.36</td>
<td>F-box and WD-40 domain protein 1/11</td>
</tr>
<tr>
<td>20</td>
<td>ASTEI08241-PA</td>
<td>wispyv2</td>
<td>3.81</td>
<td>3</td>
<td>3</td>
<td>14</td>
<td>814</td>
<td>92.9</td>
<td>1.72</td>
<td>Not found in gambiae</td>
</tr>
<tr>
<td>21</td>
<td>ASTEI05393-PA</td>
<td>LRIM18</td>
<td>7.16</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>405</td>
<td>44.2</td>
<td>1.48</td>
<td>Leucine-rich immune protein (Coil-less)</td>
</tr>
<tr>
<td>22</td>
<td>ASTEI05348-PA</td>
<td>ILP3</td>
<td>4.79</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>167</td>
<td>18.8</td>
<td>1.20</td>
<td>Insulin-like peptide 3 precursor</td>
</tr>
<tr>
<td>23</td>
<td>ASTEI01297-PA</td>
<td>p70S6K</td>
<td>7.03</td>
<td>5</td>
<td>5</td>
<td>11</td>
<td>925</td>
<td>104.8</td>
<td>1.19</td>
<td>p70 ribosomal S6 kinase</td>
</tr>
<tr>
<td>24</td>
<td>ASTEI05202-PA</td>
<td>ird5</td>
<td>4.84</td>
<td>3</td>
<td>3</td>
<td>9</td>
<td>785</td>
<td>90.6</td>
<td>1.15</td>
<td>IMD pathway signalling IKK-beta (ird5 orthologue)</td>
</tr>
<tr>
<td>25</td>
<td>ASTEI09274-PA</td>
<td>LRIM6</td>
<td>8.81</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>352</td>
<td>40.1</td>
<td>1.14</td>
<td>Leucine-rich immune protein (Short)</td>
</tr>
<tr>
<td>26</td>
<td>ASTEI05746-PA</td>
<td>PDK1</td>
<td>4.92</td>
<td>5</td>
<td>5</td>
<td>14</td>
<td>1261</td>
<td>141.7</td>
<td>1.11</td>
<td>3-phosphoinositide dependent protein kinase-1</td>
</tr>
<tr>
<td>27</td>
<td>ASTEI03908-PA</td>
<td>Myopic</td>
<td>5.75</td>
<td>7</td>
<td>8</td>
<td>18</td>
<td>1757</td>
<td>192.9</td>
<td>1.02</td>
<td>Protein-tyrosine phosphatase</td>
</tr>
<tr>
<td>28</td>
<td>ASTEI03942-PA</td>
<td>Raptor</td>
<td>7.33</td>
<td>7</td>
<td>7</td>
<td>13</td>
<td>1365</td>
<td>151.2</td>
<td>0.95</td>
<td>TORC1 growth control complex subunit Kog1</td>
</tr>
<tr>
<td>29</td>
<td>ASTEI07728-PA</td>
<td>Target of Ramycin</td>
<td>4.76</td>
<td>8</td>
<td>8</td>
<td>21</td>
<td>2542</td>
<td>287.2</td>
<td>0.83</td>
<td>Autophagy related gene TOR</td>
</tr>
<tr>
<td>30</td>
<td>ASTEI02725-PA</td>
<td>FADD</td>
<td>6.05</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>248</td>
<td>29.5</td>
<td>0.81</td>
<td>IMD pathway signalling Fas-Associated Death Domain (FADD).</td>
</tr>
<tr>
<td>31</td>
<td>ASTEI06299-PA</td>
<td>TRAF6</td>
<td>1.33</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>527</td>
<td>59.8</td>
<td>0.76</td>
<td>TOLL pathway signalling TNF Receptor-Associated Factor (TRAF).</td>
</tr>
<tr>
<td>32</td>
<td>ASTEI02570-PA</td>
<td>LRIM11</td>
<td>4.89</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>409</td>
<td>46.9</td>
<td>0.73</td>
<td>Leucine-rich immune protein (Short)</td>
</tr>
<tr>
<td>33</td>
<td>ASTEI05624-PA</td>
<td>LRIM19</td>
<td>5.15</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>291</td>
<td>33.9</td>
<td>0.69</td>
<td>Leucine-rich immune protein (Coil-less)</td>
</tr>
<tr>
<td>34</td>
<td>ASTEI08579-PA</td>
<td>Dredd</td>
<td>3.15</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>603</td>
<td>66.8</td>
<td>0.66</td>
<td>Hypothetical protein</td>
</tr>
</tbody>
</table>
Table 2.2 contd: List of immunity related proteins identified in *Anopheles stephensi* midgut.

<table>
<thead>
<tr>
<th>S. No</th>
<th>An. stephensi VectorBase ID</th>
<th>Name</th>
<th>Σ Coverag e</th>
<th>Σ# Unique Peptides</th>
<th>Σ# Peptide s</th>
<th>Σ# PSMs</th>
<th># AAs</th>
<th>MW [kDa]</th>
<th>NSAF values</th>
<th>Protein description as per <em>An. gambiae</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>ASTEI00302-PA</td>
<td>TUBE</td>
<td>4.71</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>701</td>
<td>76.9</td>
<td>0.57</td>
<td>TOLL pathway signalling</td>
</tr>
<tr>
<td>36</td>
<td>ASTEI03429-PA</td>
<td>nnierv2</td>
<td>2.34</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>727</td>
<td>78.9</td>
<td>0.55</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td>37</td>
<td>ASTEI05215-PA</td>
<td>SOS</td>
<td>2.44</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>1846</td>
<td>203.1</td>
<td>0.43</td>
<td>Protein son of sevenless</td>
</tr>
<tr>
<td>38</td>
<td>ASTEI02186-PA</td>
<td>PI3kP60</td>
<td>4.95</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>1171</td>
<td>130.2</td>
<td>0.43</td>
<td>Phosphoinositide-3-kinase, regulatory subunit</td>
</tr>
<tr>
<td>39</td>
<td>ASTEI09290-PA</td>
<td>LRIM1</td>
<td>1.95</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>514</td>
<td>58.0</td>
<td>0.39</td>
<td>Long leucine-rich immune protein (LRIM1)</td>
</tr>
<tr>
<td>40</td>
<td>ASTEI01382-PA</td>
<td>LRIM8A</td>
<td>2.31</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>520</td>
<td>58.6</td>
<td>0.38</td>
<td>Leucine-rich immune protein (Short)</td>
</tr>
<tr>
<td>41</td>
<td>ASTEI05629-PA</td>
<td>TSC</td>
<td>1.06</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>2265</td>
<td>246.7</td>
<td>0.35</td>
<td>Tuberous sclerosis 2</td>
</tr>
<tr>
<td>42</td>
<td>ASTEI05979-PA</td>
<td>Myd88</td>
<td>3.04</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>460</td>
<td>52.1</td>
<td>0.22</td>
<td>TOLL pathway signalling</td>
</tr>
<tr>
<td>43</td>
<td>ASTEI03109-PA</td>
<td>ptp</td>
<td>0.56</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1955</td>
<td>216.8</td>
<td>0.20</td>
<td>PAX-interacting protein 1</td>
</tr>
<tr>
<td>44</td>
<td>ASTEI08789-PA</td>
<td>POSH</td>
<td>1.10</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>998</td>
<td>110.7</td>
<td>0.20</td>
<td>Conserved oligomeric Golgi complex subunit 8</td>
</tr>
<tr>
<td>45</td>
<td>ASTEI04717-PA</td>
<td>PTEN</td>
<td>1.94</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1032</td>
<td>111.2</td>
<td>0.19</td>
<td>phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase and dual-specificity</td>
</tr>
<tr>
<td>46</td>
<td>ASTEI08642-PA</td>
<td>Relish</td>
<td>0.70</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1140</td>
<td>126.0</td>
<td>0.18</td>
<td>IMD pathway signalling NF-kappaB Relish-like transcription factor</td>
</tr>
<tr>
<td>47</td>
<td>ASTEI01430-PA</td>
<td>Scrawnyv1</td>
<td>0.66</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1216</td>
<td>126.1</td>
<td>0.16</td>
<td>Ubiquitin carboxyl-terminal hydrolase 36/42</td>
</tr>
<tr>
<td>48</td>
<td>ASTEI10301-PA</td>
<td>Skd</td>
<td>0.39</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2845</td>
<td>295.2</td>
<td>0.07</td>
<td>Mediator of RNA polymerase II transcription subunit 13</td>
</tr>
<tr>
<td>49</td>
<td>ASTEI03518-PA</td>
<td>TOLL1A</td>
<td>0.32</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2187</td>
<td>248.0</td>
<td>0.05</td>
<td>TOLL-like receptor TOLL1A</td>
</tr>
</tbody>
</table>
Figure 2.12 A: Representative MS/MS spectra of immune response proteins identified in this study. (A) Peptide sequence LMTETVQVENIVR identified against IMD pathway signaling IKK-beta (ird5); (B) Peptide sequence VSPSPSAGAAGTTTYR identified against TORC1 growth control complex subunit Kog1.
Figure 2.12 B: Representative MS/MS spectra of immune response proteins identified in this study. (C) Peptide sequence LELFLPEDYPMSAPK identified against Ubiquitin-conjugating enzyme E2 N; (D) Peptide sequence INQTDYEIR identified against GTP-binding protein Rheb precursor.
Several proteins identified in this study were found to be involved in antioxidant, oxidoreductive and stress-related processes. The proteins like glutathione S-transferase, cytochrome P450, alpha esterase, carboxylesterase, acetylcholinesterase, carboxypeptidase B, heme peroxidase, Cytochrome c oxidase subunit and mitochondrial inner membrane protein. These are proteins known to be associated with insecticide resistance which helps insects to metabolize insecticides at a higher rate. Some of these proteins have been identified in An. gambiae, Aedes aegypti and several other mosquito species (Biron et al., 2005; David et al., 2005). Representative MS/MS spectra of proteins associated with insecticide resistance mechanisms are shown in Figure 2.13 A and B.

Several heat-shock proteins (HSPs) were also identified in midgut. These heat shock proteins are involved in cellular defense against external stress from various sources (Lund et al., 2001). There are several proteins identified like ferritin light and heavy chain subunits, catalase, enolase, arginine/creatine kinases, aquaporin found to be involved in various metabolic processes in mosquitoes. There are many hypothetical proteins which are also identified with unknown function. Further functional studies could provide insights into the functional significance of such proteins in Anopheles stephensi midgut.
Figure 2.13 A: Representative MS/MS spectra of proteins involved in insecticide resistance mechanism identified in this study. (A) Peptide sequence **LYQEIQDVLGADFR** identified against Cytochrome P450 (CYP4H24); (B) Peptide sequence **GMEEIAGFKE** identified against Glutathione S-transferase (GSTU1).
Figure 2.13 B: Representative MS/MS spectra of proteins involved in insecticide resistance mechanism identified in this study. (C) Peptide sequence SWQDYVDNOLLASQCVSK identified against Carboxylesterase; (D) Peptide sequence NLFSVPVDADTFEYK identified against Alpha esterase.
2.3.3 Proteins involved in parasite development/interaction

Several proteins have been identified in *Anopheles gambiae* which mediate the interaction of the *Plasmodium* parasites with mosquito tissues and also regulate their maturation inside host vectors (Sreenivasamurthy *et al*., 2013). In this study, 43 proteins were identified based on sequence similarity with *An. gambiae* proteins that are known to be involved in parasite development in mosquito. Of these, 18 proteins were found to be agonistic in nature that supports *Plasmodium* development in mosquito. Another 25 proteins were found to be antagonistic in nature that inhibits *Plasmodium* development in mosquito.

Among the eighteen proteins belonging to the agonistic group was adenine nucleotide translocator, Glutathione-S-transferase theta-1, Retinoid and fatty-acid binding glycoprotein (lipophorin or ApoII/I), Lysozyme c-1, Croquemort ortholog, Peptide-O-xylosyl transferase 1, Caspar, Cactus, Gelsolin etc. The list of agonistic proteins identified is provided in Table 2.3. These Agonistic molecules are associated with the *Plasmodium* development in mosquito *Anopheles gambiae* and their role has been confirmed using knock-down studies (Vlachou *et al*., 2005; Cheon *et al*., 2006; Dinglasan *et al*., 2007; Mendes *et al*., 2008; Garveret *et al*., 2009; Jaramillo-Gutierrez *et al*., 2009; González-Lázaro *et al*., 2009; Rono *et al*., 2010; Kalja *et al*., 2011; Oliveira *et al*., 2011; Zou *et al*., 2011; Garver *et al*., 2012; Lapcharoen *et al*., 2012). Representative MS/MS spectra of some of the agonistic proteins identified in this study are shown in Figure 2.14 A and B.

Twenty five proteins belonging to the antagonistic group includes Wiskott-Aldrich syndrome protein, Gram-negative bacteria binding protein A2, *Anopheles Plasmodium*-responsive Leucine-rich repeat protein 1B, Thioester-containing protein 1, Relish 1 etc. The list of antagonistic proteins identified is provided in Table 2.4. These antagonistic molecules prevent *Plasmodium* development in *Anopheles gambiae* and have been associated with increase in oocyst/sporozoite count upon knock-down (Meister *et al*., 2005; Vlachou *et al*., 2005; Warr *et al*., 2008; Dong *et al*., 2009; Garver *et al*., 2012; Molina-Cruz *et al*., 2012). MS/MS spectra of some of the antagonistic proteins identified in this study are shown in Figure 2.15 A and B.
Table 2.3: List of mosquito proteins involved in parasite interaction identified in *An. stephensi* midgut. **Agonistic molecules:** those that support *Plasmodium* development in mosquito midgut.

<table>
<thead>
<tr>
<th>S. No</th>
<th><em>An. stephensi</em> VectorBase ID</th>
<th>Protein name</th>
<th><em>An. gambiae</em> ID</th>
<th>Protein description as per <em>An. gambiae</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ASTEI00150-PA</td>
<td>RFABG</td>
<td>AGAP001826</td>
<td>Retinoid and fatty-acid binding glycol protein (lipophorin or ApoII/I)</td>
</tr>
<tr>
<td>2</td>
<td>ASTEI00999-PA</td>
<td>GSTT1</td>
<td>AGAP000761</td>
<td>Glutathione-S-transferase theta-1</td>
</tr>
<tr>
<td>3</td>
<td>ASTEI02883-PA</td>
<td>PRS1</td>
<td>AGAP006102</td>
<td><em>Plasmodium</em> responsive salivary 1</td>
</tr>
<tr>
<td>4</td>
<td>ASTEI01309-PA</td>
<td>LYSC-1</td>
<td>AGAP007347</td>
<td>Lysozyme c-1</td>
</tr>
<tr>
<td>5</td>
<td>ASTEI03999-PA</td>
<td>DUOX</td>
<td>AGAP009978</td>
<td>Dual oxidase</td>
</tr>
<tr>
<td>6</td>
<td>ASTEI08607-PA</td>
<td>ANT</td>
<td>AGAP006782</td>
<td>Adenine nucleotide translocator</td>
</tr>
<tr>
<td>7</td>
<td>ASTEI10836-PA</td>
<td>SCRBQ2</td>
<td>AGAP010133</td>
<td>Croquemort ortholog</td>
</tr>
<tr>
<td>8</td>
<td>ASTEI04167-PA</td>
<td>CLIPA8</td>
<td>AGAP010731</td>
<td>CLIP-domain serine protease subfamily A 8</td>
</tr>
<tr>
<td>9</td>
<td>ASTEI04270-PA</td>
<td>Gelsolin</td>
<td>AGAP011369</td>
<td>Gelsolin</td>
</tr>
<tr>
<td>10</td>
<td>ASTEI04472-PA</td>
<td>Saglin</td>
<td>AGAP000610</td>
<td>Saglin</td>
</tr>
<tr>
<td>11</td>
<td>ASTEI04537-PA</td>
<td>OXT1</td>
<td>AGAP005811</td>
<td>Peptide-O-xylosyltransferase 1</td>
</tr>
<tr>
<td>12</td>
<td>ASTEI04993-PA</td>
<td>ESP</td>
<td>AGAP010240</td>
<td>Epithelial serine protease</td>
</tr>
<tr>
<td>13</td>
<td>ASTEI056831-PA</td>
<td>CLIPA8</td>
<td>AGAP001648</td>
<td>CLIP-domain serine protease subfamily B17</td>
</tr>
<tr>
<td>14</td>
<td>ASTEI07221-PA</td>
<td>Caspar</td>
<td>AGAP006473</td>
<td>Caspar</td>
</tr>
<tr>
<td>15</td>
<td>ASTEI07671-PA</td>
<td>CACT</td>
<td>AGAP007938</td>
<td>Cactus</td>
</tr>
<tr>
<td>16</td>
<td>ASTEI08424-PA</td>
<td>IMPer</td>
<td>AGAP013327</td>
<td>Immunomodulatory peroxidase</td>
</tr>
<tr>
<td>17</td>
<td>ASTEI09198-PA</td>
<td>CLIPA5</td>
<td>AGAP011787</td>
<td>CLIP-domain serine protease subfamily A 5</td>
</tr>
<tr>
<td>18</td>
<td>ASTEI08922-PA</td>
<td>CLIPB3</td>
<td>AGAP003249</td>
<td>CLIP-domain serine protease subfamily B 3</td>
</tr>
</tbody>
</table>
Figure 2.14 A: Representative MS/MS spectra of agonistic proteins those that aid *Plasmodium* development in mosquito identified in this study. (A) Peptide sequence TAVAPIERVK identified against Adenine nucleotide translocator; (B) Peptide sequence EQVKLEQDAAYR identified against Caspar.
Figure 2.14 B: Representative MS/MS spectra of agonistic proteins those that aid *Plasmodium* development in mosquito identified in this study. (C) Peptide sequence SLIDAGANINAR identified against Cactus; (D) Peptide sequence EFTVHGLDGEWVGARDAR identified against Croquemort ortholog.
Table 2.4: List of mosquito proteins involved in parasite interaction identified in *An. stephensi* midgut. **Antagonistic molecules**: those that inhibit *Plasmodium* development in mosquito midgut.

<table>
<thead>
<tr>
<th>S. No</th>
<th>An. stephensi VectorBase ID</th>
<th>Protein name</th>
<th>An. gambiae ID</th>
<th>Protein description as per An. gambiae</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ASTEI02301-PA</td>
<td>GNBPA2</td>
<td>AGAP012409</td>
<td>Gram-negative bacteria binding protein A2</td>
</tr>
<tr>
<td>2</td>
<td>ASTEI02361-PA</td>
<td>MDL1</td>
<td>AGAP012352</td>
<td>MD2-like receptor 1</td>
</tr>
<tr>
<td>3</td>
<td>ASTEI01142-PA</td>
<td>WASP</td>
<td>AGAP001081</td>
<td>Wiskott-Aldrich syndrome protein</td>
</tr>
<tr>
<td>4</td>
<td>ASTEI01898-PA</td>
<td>ARCP41</td>
<td>AGAP008908</td>
<td>Actin related 2/3 complex 41 KDa subunit P41</td>
</tr>
<tr>
<td>5</td>
<td>ASTEI02571-PA</td>
<td>APL1B</td>
<td>AGAP007035</td>
<td>Anopheles Plasmodium-responsive Leucine-rich repeat protein 1B</td>
</tr>
<tr>
<td>6</td>
<td>ASTEI02725-PA</td>
<td>FADD</td>
<td>AGAP007173</td>
<td>Fas-Associated Death Domain</td>
</tr>
<tr>
<td>7</td>
<td>ASTEI03111-PA</td>
<td>ARCP21</td>
<td>AGAP001712</td>
<td>Actin related 2/3 complex 21 KDa subunit P21</td>
</tr>
<tr>
<td>8</td>
<td>ASTEI04066-PA</td>
<td>APOD</td>
<td>AGAP002593</td>
<td>Apo lipoprotein</td>
</tr>
<tr>
<td>9</td>
<td>ASTEI05785-PA</td>
<td>MC1</td>
<td>AGAP001297</td>
<td>Mitochondrial carrier 1</td>
</tr>
<tr>
<td>10</td>
<td>ASTEI04897-PA</td>
<td>GNBPB1</td>
<td>AGAP004455</td>
<td>Gram-negative bacteria binding protein B1</td>
</tr>
<tr>
<td>11</td>
<td>ASTEI06240-PA</td>
<td>STAT</td>
<td>AGAP000099</td>
<td>Signal Transducers and Activators of Transcription</td>
</tr>
<tr>
<td>12</td>
<td>ASTEI06370-PA</td>
<td>IKK-gamma</td>
<td>AGAP005933</td>
<td>Inhibitor of kappa B kinase gamma</td>
</tr>
<tr>
<td>13</td>
<td>ASTEI06809-PA</td>
<td>FBN6</td>
<td>AGAP011231</td>
<td>Fibrinogen domain immunolectin 6</td>
</tr>
<tr>
<td>14</td>
<td>ASTEI08335-PA</td>
<td>GPRFZ2</td>
<td>AGAP010442</td>
<td>Frizzled-2</td>
</tr>
<tr>
<td>15</td>
<td>ASTEI08432-PA</td>
<td>TEP1</td>
<td>AGAP010815</td>
<td>Thioester-containing protein 1</td>
</tr>
<tr>
<td>16</td>
<td>ASTEI08579-PA</td>
<td>CASPL1</td>
<td>AGAP011693</td>
<td>Caspase long class</td>
</tr>
<tr>
<td>17</td>
<td>ASTEI08642-PA</td>
<td>REL2</td>
<td>AGAP006747</td>
<td>Relish-2</td>
</tr>
<tr>
<td>18</td>
<td>ASTEI05239-PA</td>
<td>SRPN6</td>
<td>AGAP009212</td>
<td>Serine protease inhibitor 6 (also known as serpin 6)</td>
</tr>
<tr>
<td>19</td>
<td>ASTEI09290-PA</td>
<td>LRIM1</td>
<td>AGAP006348</td>
<td>Leucine-Rich Immune Molecule 1</td>
</tr>
<tr>
<td>20</td>
<td>ASTEI09780-PA</td>
<td>LL3</td>
<td>AGAP009053</td>
<td>LITAF-like 3</td>
</tr>
<tr>
<td>21</td>
<td>ASTEI10265-PA</td>
<td>IRSP1</td>
<td>AGAP006421</td>
<td>Infection responsive secreted peptide 1</td>
</tr>
<tr>
<td>22</td>
<td>ASTEI10301-PA</td>
<td>Skd</td>
<td>AGAP006436</td>
<td>Skuld (also known as Med13 or TRAP240)</td>
</tr>
<tr>
<td>23</td>
<td>ASTEI10972-PA</td>
<td>GNBPB3</td>
<td>AGAP002799</td>
<td>Gram-negative bacteria binding protein B3</td>
</tr>
<tr>
<td>24</td>
<td>ASTEI00224-PA</td>
<td>ApoLp-III</td>
<td>AGAP013365</td>
<td>Apolipoporphin-III</td>
</tr>
<tr>
<td>25</td>
<td>ASTEI03826-PA</td>
<td>REL1</td>
<td>AGAP009515-PA</td>
<td>Relish 1</td>
</tr>
</tbody>
</table>
Figure 2.15 A: Representative MS/MS spectra of Antagonistic proteins those that prevent *Plasmodium* development in mosquito identified in this study. (A) Peptide sequence LQAEINELQSEIQQLTNTK identified against *Anopheles Plasmodium*-responsive leucine-rich repeat protein-1B; (B) Peptide sequence SPISEQTINPIQNEIR identified against Thioester containing protein-1.
Figure 2.15 B: Representative MS/MS spectra of Antagonistic proteins those that prevent *Plasmodium* development in mosquito identified in this study. (C) Peptide sequence **ANRPSTLLTNEENDQLFR** identified against Wiskott- aldrich syndrome protein; (D) Peptide sequence **VSIISEAQATQQTQTNK** identified against Signal transducers and activators of transcription.
2.3.4 Classification of proteins based on gene ontology annotation

The identified proteins were functionally categorized based on their subcellular localization; biological processes and molecular function using gene ontology (GO) based annotation. The information for all the identified proteins was fetched from the VectorBase database. As shown in Figure 2.16 A, the majority of the proteins identified in this study were localized in the membrane (18%) followed by cytoplasm (11%), nucleus (7%), ribosome (4%) and mitochondrion (2%). A large majority (39%) of proteins are still unclassified in terms of their cellular localization. Biological process-based categorization showed that a majority of them played a role in transport (13%), metabolic process (11%), oxidation-reduction process (5%), signal transduction (4%) and translation (4%) (Figure 2.16 B). However, 25% of proteins are still unclassified in terms of their biological process. Classification based on molecular function showed that the majority of the identified proteins were involved in binding activity such as protein binding, ATP/GTP binding, nucleotide binding and calcium ion binding (45%), transferase activity (5%), oxido-reductase activity (4%), hydrolase activity (3%) and transporter activity (2%). About 15% of identified proteins are still unclassified in terms of their molecular function (Figure 2.16 C).
Figure 2.16: Gene ontology based classification of proteins identified in midgut tissue of sugar fed female An. stephensi. (A) Subcellular localization, (B) Biological process, (C) Molecular function.
2.3.5 Comparative proteomic analysis of male and female midgut
A quantitative proteomic analysis to compare male and female midgut proteome was carried out using iTRAQ-based quantitative proteomic approach, to study differentially expressed proteins between these two sexes. Midgut proteins extracted from male and female mosquitoes were digested using trypsin enzyme followed by labeling with two different iTRAQ tags and SCX fractionation. Quantitative proteomic analysis of these iTRAQ labeled peptides was carried out by employing high resolution fourier transform mass spectrometry (LTQ-OrbitrapVelos). From this, a total of 1,30,453 tandem mass spectra were acquired, which were searched using two search algorithms Sequest and Mascot against VectorBase protein database of Anopheles stephensi Indian strain (database version Astel2.0). A total of 1,06,774 peptide-spectrum matches (PSMs) led to identification of 21,311 peptides and 3,316 proteins. Of these 3,316 proteins, identification of 2,434 proteins was supported by 2 or more unique peptides, while 881 protein identifications were based on single peptide evidence with multiple PSMs.

2.3.6 Identification of differentially expressed proteins between male and female midgut
In all, 758 proteins out of 3,316 total proteins were found to be differentially expressed in female midgut with respect to male counterpart. Of these differentially expressed proteins, 489 proteins were identified by multiple peptides and 269 protein identifications were based on single peptide evidence but with multiple PSMs. Relative quantification of peptides was done on the basis of relative intensity of reporter ions 116 and 117 for male and female midgut respectively. Protein ratios were calculated as the median of all the peptide ratios corresponding to respective proteins. Fold change of ≥2 was considered as over expressed while ≤0.5 was considered as under expressed (Puttamallesh et al., 2013). Among the differentially expressed proteins, 575 proteins were found over expressed (≥2 fold) in female midgut as compared to male midgut, while the remaining 183 proteins were under expressed (≤0.5 fold) as shown in Figure 2.17.
Figure 2.17: Differential expression of proteins between male and female midgut. Differentially expressed proteins identified are plotted against the LOG2 values of their fold change in female midgut when compared to that of male midgut. Red colored dots represent the over-expressed proteins (≥ 2-fold), whereas blue dots represent the under-expressed proteins (≤ 0.5-fold) in female midgut with respect to male midgut.

2.3.7 Proteins over expressed in female midgut

In female midgut, a total of 575 proteins were identified as over-expressed by >2fold as compared to male midgut. The over-expressed proteins includes F-box and leucine-rich repeat protein 6, calcineurin beta subunit, bromodomain-containing protein 8, Sodium-independent sulfate anion transporter, serine protease, nuclear transport factor 2, elongation factor 1 alpha-like protein. Partial list of proteins over-expressed in female midgut is provided in Table 2.5 and a complete list of over-expressed proteins can be seen in Appendix Table 2 A.
The most abundant protein identified was F-box and Leucine-rich Repeat Protein 6. This protein was found to be 44.1 fold over-expressed in female midgut as compared to male. This protein belongs to a member of a family of proteins that are characterized by an F-box motif and involved in phosphorylation-dependent ubiquitination. Another six proteins were found to be involved in the glycolysis pathway including enolase, fructose-bis-phosphate aldolase, triose-phosphate isomerase, phosphoglycerate kinase, 6-phosphofructo kinase 1 and Phospho-glycerate mutase. These enzymes are over-expressed in female midgut as compared to the male mosquito, which suggests that glucose metabolism is higher in female mosquito than in the male mosquito. Several proteases like aminopeptidase N, carboxypeptidase B and serine proteases were found over expressed in female midgut by more than 2 fold than male midgut. Proteases are involved in proteolytic events that occur during the blood-feeding process in female mosquitoes. These proteases are also used by the *Plasmodium* parasites to activate its own enzymes during invasion of host cells for its survival (Shahabuddin *et al*., 1993). Over-expression of multiple ribosomal subunits and various translation initiation factors suggested enhanced protein synthesis in female midgut. Proteins involved in insecticide resistance like cytochrome p450 and glutathione S-transferase zeta class were found to be over-expressed in females. Over-expression of alkaline phosphatase was found in female midgut, this is known to serve as binding sites for *Bacillus thuringiensis* in insects (Thammasittirong *et al*., 2011). Heat shock proteins, which play a role in protein degradation, protein folding and signal transduction was found over expressed in female midgut.

In this study, 18 immunity related proteins were expressed differentially between female and male midgut reported in *An. stephensi* by Jiang *et al*., (2014). Of these, protein FADD was found 2.6 fold over expressed in female midgut, whereas, leucine-rich immune protein (Coil-less) was found 0.4 fold times under-expressed in female midgut (Table 2.6). Previous studies have demonstrated the role of these genes in mosquito defense against parasite and viruses (Garver *et al*., 2009; Dong *et al*., 2011). Representative MS/MS spectra of some of the proteins over-expressed in female midgut are shown in Figure 2.18 A and B.
Table 2.5: A partial list of proteins over-expressed in female midgut identified in this study.

<table>
<thead>
<tr>
<th>S. No</th>
<th>An. stephensi VectorBase ID</th>
<th>Protein description as per An. gambiae</th>
<th>Relative fold change female/male</th>
<th>Σ# Peptides</th>
<th>Σ# PSMs</th>
<th># AAs</th>
<th>MW [kDa]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ASTEI10274-PA</td>
<td>F-box and leucine-rich repeat protein 6</td>
<td>44.163</td>
<td>1</td>
<td>4</td>
<td>917</td>
<td>99.2</td>
</tr>
<tr>
<td>2</td>
<td>ASTEI00645-PA</td>
<td>Calcineurin beta subunit</td>
<td>13.695</td>
<td>1</td>
<td>2</td>
<td>193</td>
<td>21.7</td>
</tr>
<tr>
<td>3</td>
<td>ASTEI00772-PA</td>
<td>Bromodomain-containing protein 8</td>
<td>11.125</td>
<td>2</td>
<td>4</td>
<td>1449</td>
<td>160.7</td>
</tr>
<tr>
<td>4</td>
<td>ASTEI10988-PA</td>
<td>Piggybac-derived 2</td>
<td>10.243</td>
<td>1</td>
<td>6</td>
<td>359</td>
<td>41.3</td>
</tr>
<tr>
<td>5</td>
<td>ASTEI09499-PA</td>
<td>Hypothetical protein</td>
<td>9.948</td>
<td>6</td>
<td>54</td>
<td>1385</td>
<td>150.9</td>
</tr>
<tr>
<td>6</td>
<td>ASTEI03820-PA</td>
<td>Hypothetical protein</td>
<td>9.809</td>
<td>2</td>
<td>4</td>
<td>1723</td>
<td>197.7</td>
</tr>
<tr>
<td>7</td>
<td>ASTEI06278-PA</td>
<td>Dnaj-like protein subfamily C</td>
<td>8.575</td>
<td>1</td>
<td>2</td>
<td>551</td>
<td>62.1</td>
</tr>
<tr>
<td>8</td>
<td>ASTEI07509-PA</td>
<td>Hypothetical protein</td>
<td>7.638</td>
<td>1</td>
<td>2</td>
<td>240</td>
<td>26.4</td>
</tr>
<tr>
<td>9</td>
<td>ASTEI03729-PA</td>
<td>Hypothetical protein</td>
<td>7.346</td>
<td>4</td>
<td>27</td>
<td>636</td>
<td>72.4</td>
</tr>
<tr>
<td>10</td>
<td>ASTEI02407-PA</td>
<td>Sodium-independent sulfate anion transporter</td>
<td>7.269</td>
<td>7</td>
<td>30</td>
<td>655</td>
<td>71.6</td>
</tr>
<tr>
<td>11</td>
<td>ASTEI11298-PA</td>
<td>Hypothetical protein</td>
<td>7.207</td>
<td>2</td>
<td>5</td>
<td>130</td>
<td>14.9</td>
</tr>
<tr>
<td>12</td>
<td>ASTEI08015-PA</td>
<td>Hypothetical protein</td>
<td>7.206</td>
<td>1</td>
<td>21</td>
<td>146</td>
<td>15.5</td>
</tr>
<tr>
<td>13</td>
<td>ASTEI025761-PA</td>
<td>Nuclear transport factor 2</td>
<td>6.677</td>
<td>4</td>
<td>16</td>
<td>130</td>
<td>14.6</td>
</tr>
<tr>
<td>14</td>
<td>ASTEI05236-PA</td>
<td>DNA-directed RNA polymerase II subunit RPB11</td>
<td>6.619</td>
<td>1</td>
<td>1</td>
<td>117</td>
<td>13.5</td>
</tr>
<tr>
<td>15</td>
<td>ASTEI08033-PA</td>
<td>Serine protease</td>
<td>6.599</td>
<td>4</td>
<td>49</td>
<td>260</td>
<td>28.8</td>
</tr>
<tr>
<td>16</td>
<td>ASTEI04077-PA</td>
<td>Elongation factor 1 alpha-like protein</td>
<td>6.364</td>
<td>3</td>
<td>8</td>
<td>877</td>
<td>96.3</td>
</tr>
</tbody>
</table>
Table 2.5 contd: Partial list of proteins over-expressed in female midgut identified in this study.

<table>
<thead>
<tr>
<th>S. No</th>
<th>An. stephensi VectorBase ID</th>
<th>Protein description as per An. gambiae</th>
<th>Relative fold change female/male</th>
<th>Σ# Peptides</th>
<th>Σ# PSMs</th>
<th># AAs</th>
<th>MW [kDa]</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>ASTEI06639-PA</td>
<td>Hypothetical protein</td>
<td>6.058</td>
<td>1</td>
<td>2</td>
<td>415</td>
<td>46.2</td>
</tr>
<tr>
<td>18</td>
<td>ASTEI08034-PA</td>
<td>Serine protease</td>
<td>6.023</td>
<td>3</td>
<td>58</td>
<td>268</td>
<td>29.1</td>
</tr>
<tr>
<td>19</td>
<td>ASTEI09575-PA</td>
<td>Hypothetical protein</td>
<td>6.004</td>
<td>1</td>
<td>2</td>
<td>582</td>
<td>67.3</td>
</tr>
<tr>
<td>20</td>
<td>ASTEI11463-PA</td>
<td>Hypothetical protein</td>
<td>6.003</td>
<td>2</td>
<td>3</td>
<td>108</td>
<td>12.3</td>
</tr>
<tr>
<td>21</td>
<td>ASTEI05455-PA</td>
<td>Hypothetical protein</td>
<td>5.985</td>
<td>4</td>
<td>12</td>
<td>229</td>
<td>22.7</td>
</tr>
<tr>
<td>22</td>
<td>ASTEI03153-PA</td>
<td>Hypothetical protein</td>
<td>5.625</td>
<td>1</td>
<td>2</td>
<td>244</td>
<td>27.5</td>
</tr>
<tr>
<td>23</td>
<td>ASTEI01570-PA</td>
<td>Anaphase-promoting complex subunit 3</td>
<td>5.463</td>
<td>1</td>
<td>2</td>
<td>1197</td>
<td>132.4</td>
</tr>
<tr>
<td>24</td>
<td>ASTEI11423-PA</td>
<td>Hypothetical protein</td>
<td>5.389</td>
<td>1</td>
<td>4</td>
<td>181</td>
<td>20.7</td>
</tr>
<tr>
<td>25</td>
<td>ASTEI09278-PA</td>
<td>Ccdc109a</td>
<td>5.364</td>
<td>2</td>
<td>5</td>
<td>312</td>
<td>35.7</td>
</tr>
<tr>
<td>26</td>
<td>ASTEI02774-PA</td>
<td>Pyridoxine kinase</td>
<td>5.364</td>
<td>14</td>
<td>124</td>
<td>903</td>
<td>96.6</td>
</tr>
<tr>
<td>27</td>
<td>ASTEI07669-PA</td>
<td>Hypothetical protein</td>
<td>5.358</td>
<td>2</td>
<td>6</td>
<td>99</td>
<td>11.2</td>
</tr>
<tr>
<td>28</td>
<td>ASTEI03928-PA</td>
<td>Hypothetical protein</td>
<td>5.356</td>
<td>12</td>
<td>137</td>
<td>914</td>
<td>102.2</td>
</tr>
<tr>
<td>29</td>
<td>ASTEI02547-PA</td>
<td>Hypothetical protein</td>
<td>5.296</td>
<td>1</td>
<td>1</td>
<td>3093</td>
<td>347.7</td>
</tr>
<tr>
<td>30</td>
<td>ASTEI10087-PA</td>
<td>Hypothetical protein</td>
<td>5.260</td>
<td>1</td>
<td>4</td>
<td>64</td>
<td>7.6</td>
</tr>
<tr>
<td>31</td>
<td>ASTEI08929-PA</td>
<td>N-myc downstream regulated</td>
<td>5.207</td>
<td>10</td>
<td>87</td>
<td>330</td>
<td>36.3</td>
</tr>
<tr>
<td>32</td>
<td>ASTEI10850-PA</td>
<td>Mitochondrial ribosomal protein L40</td>
<td>5.149</td>
<td>1</td>
<td>4</td>
<td>196</td>
<td>22.8</td>
</tr>
</tbody>
</table>
Figure 2.18 A: MS/MS spectra of peptides for the proteins over-expressed in female midgut (A) Peptide sequence **ATDKFPYAGQNIATQFFGYK** identified against Infection responsive secreted peptide 1, (B) Peptide sequence **GGPPGLVTGLK** identified against Prohibitin 2.
Figure 2.18 B: MS/MS spectra of peptides for the proteins over-expressed in female midgut (C) Peptide sequence IVEMYVHEDYSGGVGPNDIAVFR identified against Serine protease; (D) Peptide sequence GYLYQGK identified against hypothetical protein.
Table 2.6: List of immunity related proteins identified and their expression level in female and male midgut tissue.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>An. stephensi VectorBase ID</th>
<th>Relative expression level male/female</th>
<th>Relative fold change female/male</th>
<th>Name</th>
<th>Protein description as per An. gambiae</th>
<th>Molecular function</th>
<th>Biological process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ASTEI02725-PA</td>
<td>0.377</td>
<td>2.652</td>
<td>FADD</td>
<td>IMD pathway signaling Fas-Associated Death Domain (FADD).</td>
<td>Protein binding</td>
<td>Signal transduction,</td>
</tr>
<tr>
<td>2</td>
<td>ASTEI092504-PA</td>
<td>0.543</td>
<td>1.840</td>
<td>PGRPLC</td>
<td>Peptidoglycan recognition protein (Long)</td>
<td>Hydrolase activity</td>
<td>Peptidoglycan catabolic process</td>
</tr>
<tr>
<td>3</td>
<td>ASTEI05746-PA</td>
<td>0.571</td>
<td>1.750</td>
<td>PDK1</td>
<td>Hypothetical protein</td>
<td>Protein kinase activity/transferase activity</td>
<td>Cellular protein modification process</td>
</tr>
<tr>
<td>4</td>
<td>ASTEI092500-PA</td>
<td>0.605</td>
<td>1.654</td>
<td>Bendless</td>
<td>Ubiquitin-conjugating enzyme E2 N</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>5</td>
<td>ASTEI08192-PA</td>
<td>0.608</td>
<td>1.643</td>
<td>Spt6</td>
<td>Transcription elongation factor Spt6</td>
<td>DNA binding</td>
<td>Cellular nitrogen compound metabolic process</td>
</tr>
<tr>
<td>6</td>
<td>ASTEI08720-PA</td>
<td>0.629</td>
<td>1.589</td>
<td>Rheb</td>
<td>GTP-binding protein Rheb precursor</td>
<td>Ion binding</td>
<td>Signal transduction/transport</td>
</tr>
<tr>
<td>7</td>
<td>ASTEI01297-PA</td>
<td>0.719</td>
<td>1.390</td>
<td>p70S6K</td>
<td>p70 ribosomal S6 kinase</td>
<td>Kinase activity</td>
<td>Cellular protein modification process</td>
</tr>
<tr>
<td>8</td>
<td>ASTEI09147-PA</td>
<td>0.805</td>
<td>1.242</td>
<td>Grb2</td>
<td>Growth factor receptor-binding protein 2</td>
<td>Protein binding</td>
<td>Unknown</td>
</tr>
<tr>
<td>9</td>
<td>ASTEI03826-PA</td>
<td>0.826</td>
<td>1.211</td>
<td>REL1</td>
<td>TOLL pathway signaling NF-kappaB Relish-like transcription factor</td>
<td>DNA binding transcription factor activity</td>
<td>Regulation of DNA transcription,</td>
</tr>
<tr>
<td>10</td>
<td>ASTEI03942-PA</td>
<td>0.902</td>
<td>1.108</td>
<td>Rator</td>
<td>TORC1 growth control complex subunit Kog1</td>
<td>Protein binding</td>
<td>TOR signaling</td>
</tr>
<tr>
<td>11</td>
<td>ASTEI09290-PA</td>
<td>0.933</td>
<td>1.072</td>
<td>LRIM1</td>
<td>Leucine-rich immune protein (Long)</td>
<td>Protein binding</td>
<td>Unknown</td>
</tr>
<tr>
<td>12</td>
<td>ASTEI04964-PA</td>
<td>1.006</td>
<td>0.994</td>
<td>Hrsv1</td>
<td>Hepatocyte growth factor-regulated tyrosine kinase substrate</td>
<td>Metal ion binding</td>
<td>Protein transport</td>
</tr>
<tr>
<td>13</td>
<td>ASTEI02325-PA</td>
<td>1.048</td>
<td>0.954</td>
<td>TOLL8</td>
<td>Toll</td>
<td>Protein binding</td>
<td>Signal transduction</td>
</tr>
<tr>
<td>14</td>
<td>ASTEI06299-PA</td>
<td>1.088</td>
<td>0.919</td>
<td>TRAF6</td>
<td>TOLL pathway signaling TNF Receptor-Associated Factor (TRAF)</td>
<td>Protein binding</td>
<td>Unknown</td>
</tr>
<tr>
<td>15</td>
<td>ASTEI07471-PA</td>
<td>1.122</td>
<td>0.891</td>
<td>Slinb</td>
<td>F-box and WD-40 domain protein 1/11</td>
<td>Protein binding</td>
<td>Unknown</td>
</tr>
<tr>
<td>16</td>
<td>ASTEI02567-PA</td>
<td>1.133</td>
<td>0.883</td>
<td>LRIM4</td>
<td>Leucine-rich immune protein (Long)</td>
<td>Protein binding</td>
<td>Unknown</td>
</tr>
<tr>
<td>17</td>
<td>ASTEI10386-PA</td>
<td>1.274</td>
<td>0.785</td>
<td>LRIM26</td>
<td>Leucine-rich immune protein (Coil-less)</td>
<td>Protein binding</td>
<td>Unknown</td>
</tr>
<tr>
<td>18</td>
<td>ASTEI05393-PA</td>
<td>2.357</td>
<td>0.424</td>
<td>LRIM18</td>
<td>Leucine-rich immune protein (Coil-less)</td>
<td>Protein binding</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
Proteins involved in vector parasite interaction were also identified. 31 proteins were identified which are known to play a role in vector-\textit{Plasmodium} interaction (Table 2.7). Of these 12 proteins are agonistic in nature while 19 are antagonistic in nature. Two proteins like Fas-Associated Death Domain (FADD) and Infection responsive secreted peptide 1(IRSP1) were found to be over-expressed by 2.6 and 2.2 times in female midgut respectively. These two proteins are known to suppress \textit{Plasmodium} infection in mosquito midgut which may be involved in mosquito immunity in response to infection (Dong \textit{et al.}, 2006; Garver \textit{et al.}, 2012).

Interestingly, three proteins such as Gram-negative bacteria binding protein B4, Beta thymosin family and Cell division cycle 42, are known to inhibit \textit{Plasmodium} infection were found to be under-expressed in female. In contrast, another protein Caspar known to support oocyst development in midgut was also found under-expressed. siRNA based gene silencing of Gram negative bacteria binding protein B4 in \textit{An. gambiae} have shown an enhanced rate of oocyst infection upon its gene silencing (Warr \textit{et al.}, 2008). Similarly, Beta thymosin family (Ciboulot) is known to prevent \textit{Plasmodium} development in mosquito (oocyst number increases upon knock-down) and was found to be 0.46 times under-expressed in female midgut (Vlachou \textit{et al.}, 2005). FAS-associated factor 1(Caspar) is known to support \textit{Plasmodium} development in mosquito midgut, gene silencing of Caspar gene results in decrease in oocyst formation in mosquito \textit{An. gambiae}. This protein was found 0.37 times under-expressed in female midgut (Garver \textit{et al.}, 2009, 2012).
Table 2.7: List of proteins involved in vector parasite interaction and their relative expression level between female and male midgut tissue.

(A). **Antagonistic molecules:** Inhibit the oocyst formation in mosquito midgut.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>An. stephensi VectorBase ID</th>
<th>Relative fold change female/male</th>
<th>Protein name</th>
<th>Protein description as per An. gambiae</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ASTEI02725-PA</td>
<td>2.652</td>
<td>FADD</td>
<td>Fas-Associated Death Domain</td>
</tr>
<tr>
<td>2</td>
<td>ASTEI10265-PA</td>
<td>2.227</td>
<td>IRSP1</td>
<td>Infection responsive secreted peptide 1</td>
</tr>
<tr>
<td>3</td>
<td>ASTEI04705-PA</td>
<td>1.887</td>
<td>Prohibit</td>
<td>Prohibitin 2</td>
</tr>
<tr>
<td>4</td>
<td>ASTEI01142-PA</td>
<td>1.510</td>
<td>WASP</td>
<td>Wiskott-Aldrich syndrome protein</td>
</tr>
<tr>
<td>5</td>
<td>ASTEI06240-PA</td>
<td>1.464</td>
<td>STAT</td>
<td>Signal Transducers and Activators of Transcription</td>
</tr>
<tr>
<td>6</td>
<td>ASTEI03826-PA</td>
<td>1.211</td>
<td>REL1</td>
<td>Relish 1</td>
</tr>
<tr>
<td>7</td>
<td>ASTEI07645-PA</td>
<td>1.207</td>
<td>Hsc-3</td>
<td>heat shock 70kDa protein 5</td>
</tr>
<tr>
<td>8</td>
<td>ASTEI01898-PA</td>
<td>1.193</td>
<td>ARC P41</td>
<td>Actin related 2/3 complex 41 KDa subunit P41</td>
</tr>
<tr>
<td>9</td>
<td>ASTEI07389-PA</td>
<td>1.184</td>
<td>JNK</td>
<td>Jun N-terminal Kinase</td>
</tr>
<tr>
<td>10</td>
<td>ASTEI03111-PA</td>
<td>1.099</td>
<td>ARC P21</td>
<td>Actin related 2/3 complex 21 KDa subunit P21</td>
</tr>
<tr>
<td>11</td>
<td>ASTEI09290-PA</td>
<td>1.072</td>
<td>LRIM1</td>
<td>Leucine-Rich Immune Molecule 1</td>
</tr>
<tr>
<td>12</td>
<td>ASTEI03480-PA</td>
<td>0.925</td>
<td>GSTT2</td>
<td>Glutathione-S-transferase theta-2</td>
</tr>
<tr>
<td>13</td>
<td>ASTEI05239-PA</td>
<td>0.790</td>
<td>SRPN6</td>
<td>Serine protease inhibitor 6 (also known as serpin 6)</td>
</tr>
<tr>
<td>14</td>
<td>ASTEI00224-PA</td>
<td>0.725</td>
<td>ApoLp-III</td>
<td>Apolipophorin-III</td>
</tr>
<tr>
<td>15</td>
<td>ASTEI08432-PA</td>
<td>0.680</td>
<td>TEP1</td>
<td>Thioester-containing protein 1</td>
</tr>
<tr>
<td>16</td>
<td>ASTEI00731-PA</td>
<td>0.589</td>
<td>Cdc42</td>
<td>Cell division cycle 42</td>
</tr>
<tr>
<td>17</td>
<td>ASTEI06691-PA</td>
<td>0.473</td>
<td></td>
<td>aquaporin</td>
</tr>
<tr>
<td>18</td>
<td>ASTEI10999-PA</td>
<td>0.460</td>
<td>Ciboulot</td>
<td>Beta thymosin family</td>
</tr>
<tr>
<td>19</td>
<td>ASTEI10972-PA</td>
<td>0.338</td>
<td>GNBPB4</td>
<td>Gram-negative bacteria binding protein B4</td>
</tr>
</tbody>
</table>
Table 2.7 contd: List of proteins involved in vector parasite interaction and their expression level between female and male midgut tissue.

(B) Agonistic molecules: support the oocyst formation in mosquito midgut.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>An. stephensi VectorBase ID</th>
<th>Relative fold change female/male</th>
<th>Protein name</th>
<th>Protein description as per An. gambiae</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ASTEI08607-PA</td>
<td>1.997</td>
<td>ANT</td>
<td>Adenine nucleotide translocator</td>
</tr>
<tr>
<td>2</td>
<td>ASTEI092504-PA</td>
<td>1.840</td>
<td>PGRPLC</td>
<td>PGN Recognition Protein LC</td>
</tr>
<tr>
<td>3</td>
<td>ASTEI07966-PA</td>
<td>1.462</td>
<td>SDR1</td>
<td>Short-chain dehydrogenases/reductases</td>
</tr>
<tr>
<td>4</td>
<td>ASTEI04504-PA</td>
<td>1.164</td>
<td>CPR</td>
<td>Cytochrome P450 reductase</td>
</tr>
<tr>
<td>5</td>
<td>ASTEI01530-PA</td>
<td>1.116</td>
<td>LANB2</td>
<td>Laminin gamma 1</td>
</tr>
<tr>
<td>6</td>
<td>ASTEI10836-PA</td>
<td>1.100</td>
<td>SCRBQ2</td>
<td>Class B Scavenger Receptor (CD36 domain). Croquemort ortholog</td>
</tr>
<tr>
<td>7</td>
<td>ASTEI02221-PA</td>
<td>0.966</td>
<td>ArgK</td>
<td>Arginine kinase</td>
</tr>
<tr>
<td>8</td>
<td>ASTEI04270-PA</td>
<td>0.937</td>
<td>Gelsolin</td>
<td>Gelsolin</td>
</tr>
<tr>
<td>9</td>
<td>ASTEI00999-PA</td>
<td>0.881</td>
<td>GSTT1</td>
<td>Glutathione-S-transferase theta-1</td>
</tr>
<tr>
<td>10</td>
<td>ASTEI07737-PA</td>
<td>0.846</td>
<td>CP</td>
<td>F-actin capping protein</td>
</tr>
<tr>
<td>11</td>
<td>ASTEI00150-PA</td>
<td>0.716</td>
<td>RFABG</td>
<td>Retinoid and fatty-acid binding glycoprotein, also known as lipophorin or ApoII/l</td>
</tr>
<tr>
<td>12</td>
<td>ASTEI07221-PA</td>
<td>0.374</td>
<td>Caspar</td>
<td>Caspar</td>
</tr>
</tbody>
</table>
2.3.8 Proteins over expressed in male midgut

In the male midgut, 183 proteins were found to be over-expressed (>2fold) than the female midgut. The over-expressed proteins included gustatory receptor, translocon-associated protein subunit gamma, sodium-coupled monocarboxylate transporter 1, cuticular protein RR-1 family (CPR24), innexin inx2, ankyrin, tubulin polymerization-promoting protein, 3-Glucan Binding Protein (BGBP), Fas-associated factor 1. Partial list of proteins over-expressed in male midgut are provided in Table 2.8 and a complete list of proteins can be seen in Appendix Table 2 B. An insect gustatory receptor protein belongs to gustatory and odorant receptors were found to be 19.4 times over-expressed in male as compared to female. In insects this protein plays a role in detection of sugar feeding sources, blood-feeding hosts, and oviposition sites (Freeman et al., 2014). Maltase was found to be over expressed by 2.4 fold than female which is a glycosidase involved in the digestion of sugars as male mosquitoes feeds on plant juices. Another protein, cuticular protein RR-1 family (CPR24) is a member of CPR family found to be 6.1 fold over-expressed in male midgut and plays a role in forming the insect exoskeleton and other cuticular structures. Representative MS/MS spectra of some of the proteins over-expressed in male midgut are shown in Figure 2.19 A and B.
Figure 2.19 A: MS/MS spectra of peptides for the proteins over-expressed in male midgut (A) Peptide sequence **DQQLAEDQQEQSLR** identified against Ankyrin, (B) Peptide sequence **SELESFR** identified against Thymosin.
Figure 2.19 B: MS/MS spectra of peptides for the proteins over-expressed in male midgut (C) Peptide sequence YSPGQGFNK identified against 3-Glucan Binding Protein (BGBP); (D) Peptide sequence LEILNLSR identified against Leucine-rich immune protein (Coil-less).
2.3.9 Bioinformatic analysis of differentially expressed proteins

2.3.9.1 Classification and functional annotation of differentially expressed proteins

The information on cellular localization, molecular functions and biological processes for the differentially expressed proteins were retrieved from the Vectorbase database (www.VectorBase.org). These proteins were classified based on Gene Ontology annotations. In order to evaluate which GO term was more pervasive in midgut of female or male *An. stephensi*, the differentially expressed proteins were divided into two groups: 1) Proteins which were over-expressed in female midgut and 2) Proteins which were over-expressed in male midgut tissue.

The subcellular localization of proteins over-expressed in female midgut indicated that 15% of the proteins were localized in the cytoplasm. Remaining proteins were predicted to be localized in membrane (12%), ribosome (10%) and nucleus (6%). Of the total, 39% of proteins are still reported as unclassified (Figure 2.20 A). Classification based on biological processes showed that majority of the over-expressed proteins were involved in biological process such as transport (10%), translation (10%), metabolic process (6%), oxidation-reduction process (4%) and RNA processing (3%) (Figure 2.20 B). Classification based on molecular function showed that majority of proteins over-expressed in female midgut were involved in molecular functions such as binding activity (41%), transferase activity (4%), catalytic activity (4%) and hydrolase activity (3%). 13% of identified proteins are still unclassified in terms of their molecular function (Figure 2.20 C).
Figure 2.20: Gene ontology based classification of over-expressed proteins identified in female midgut. A) Subcellular localization (B) Biological process (C) Molecular function.
Similarly, the subcellular localization of over-expressed proteins in male midgut indicated that 26% of the proteins were localized to membrane followed by proteins localized to cytoplasm (13%), mitochondria (6%), nucleus (2%) and ribosome (2%) (Figure 2.21 A). Biological process based categorization showed that the majority of proteins were involved in metabolic process (17%), transport (13%), oxidation-reduction process (10%), proteolysis (5%) and biosynthetic process (4%) (Figure 2.21 B). Classification based on molecular function showed that majority of the proteins were predicted to have molecular functions such as protein binding activity (30%), catalytic activity (9%), oxidoreductase activity (8%), transferase activity (5%) and transporter activity (5%). 14% of proteins are still reported as unclassified in terms of molecular function. (Figure 2.21 C).
Figure 2.21: Gene ontology based classification of over-expressed proteins identified from male midgut. Subcellular localization, biological process and molecular function of differentially expressed proteins. (A) Subcellular localization (B) Biological processes (C) Molecular function.
2.3.10 Sex-biased proteins

Further, the differentially expressed proteins identified were compared with sex-biased genes reported in An. stephensi Indian strain and found a strong correlation. Jiang et al., (2014) identified 313 and 241 genes with male and female biased expression respectively. Of these, 76 female biased proteins were identified in this study. Among these, 50 proteins were found to be over-expressed >2 fold in female midgut as compare to male, while remaining 25 proteins expression were found to be unchanged between two sexes. The female specific proteins identified includes F-box and leucine-rich repeat protein 6, serine proteases, lipase3, aminopaptidase N, alkaline phosphatase, female reproductive tract protease, Class B Scavenger Receptor (CD36 domain), Venom allergen as shown in Table 2.8. These female specific proteins were mainly involved in proteolysis and metabolic processes relevant to blood meal digestion, vector immunity and some involved in parasite interaction. Similarly, 19 proteins identified in this study were reported to be male biased (Table 2.9). Of these, eight proteins were found to be over-expressed >2fold in male midgut as compared to the female and of the remaining 12 proteins, expression was found to be unchanged between two sexes. The male specific proteins include Regucalein protein, Actin cytoplasmic, Alpha-crystallin chain B, Trypsin-alpha, Troponin C isoform 4", Leucine amino peptidase-like protein, and two hypothetical proteins. Further investigation of these sex-specific molecules may enhance the current knowledge regarding role of midgut proteins in malaria transmission and vector development.
### Table 2.8: A partial list of proteins over-expressed in male midgut tissue

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Accession Id</th>
<th>Protein description</th>
<th>Expression level male/female</th>
<th>Σ# Peptides</th>
<th>Σ# PSMs</th>
<th># AAs</th>
<th>MW [kDa]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ASTEI06544-PA</td>
<td>Gustatory receptor</td>
<td>19.453</td>
<td>1</td>
<td>1</td>
<td>791</td>
<td>92.6</td>
</tr>
<tr>
<td>2</td>
<td>ASTEI07019-PA</td>
<td>Hypothetical protein</td>
<td>18.422</td>
<td>1</td>
<td>1</td>
<td>175</td>
<td>19.5</td>
</tr>
<tr>
<td>3</td>
<td>ASTEI10790-PA</td>
<td>Hypothetical protein</td>
<td>11.219</td>
<td>2</td>
<td>7</td>
<td>184</td>
<td>21.2</td>
</tr>
<tr>
<td>4</td>
<td>ASTEI08704-PA</td>
<td>Hypothetical protein</td>
<td>9.645</td>
<td>1</td>
<td>2</td>
<td>735</td>
<td>78.9</td>
</tr>
<tr>
<td>5</td>
<td>ASTEI09506-PA</td>
<td>Translocon-associated protein subunit gamma</td>
<td>9.460</td>
<td>4</td>
<td>15</td>
<td>185</td>
<td>21.2</td>
</tr>
<tr>
<td>6</td>
<td>ASTEI10959-PA</td>
<td>Sodium-coupled monocarboxylate transporter 1</td>
<td>6.846</td>
<td>1</td>
<td>4</td>
<td>684</td>
<td>74.7</td>
</tr>
<tr>
<td>7</td>
<td>ASTEI08483-PA</td>
<td>Hypothetical protein</td>
<td>6.694</td>
<td>1</td>
<td>2</td>
<td>62</td>
<td>6.8</td>
</tr>
<tr>
<td>8</td>
<td>ASTEI05983-PA</td>
<td>Potassium voltage-gated channel, subfamily H (eag-related), member 4</td>
<td>6.472</td>
<td>2</td>
<td>4</td>
<td>805</td>
<td>90.7</td>
</tr>
<tr>
<td>9</td>
<td>ASTEI06317-PA</td>
<td>Cuticular protein RR-1 family (CPR24)</td>
<td>6.112</td>
<td>2</td>
<td>4</td>
<td>898</td>
<td>98.0</td>
</tr>
<tr>
<td>10</td>
<td>ASTEI11712-PA</td>
<td>Innexin inx2</td>
<td>5.829</td>
<td>1</td>
<td>2</td>
<td>116</td>
<td>13.6</td>
</tr>
<tr>
<td>11</td>
<td>ASTEI09441-PA</td>
<td>Transcription factor AP-1</td>
<td>5.754</td>
<td>1</td>
<td>2</td>
<td>325</td>
<td>35.2</td>
</tr>
<tr>
<td>12</td>
<td>ASTEI04365-PA</td>
<td>Phosphate carrier, mitochondrial</td>
<td>5.238</td>
<td>13</td>
<td>171</td>
<td>348</td>
<td>38.2</td>
</tr>
</tbody>
</table>
**Table 2.8 contd:** A partial list of proteins over-expressed in male midgut tissue

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Accession Id</th>
<th>Protein description</th>
<th>Expression level male/female</th>
<th>Σ# Peptides</th>
<th>Σ# PSMs</th>
<th># AAs</th>
<th>MW [kDa]</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>ASTEI09681-PA</td>
<td>FK506-binding protein 1</td>
<td>5.102</td>
<td>4</td>
<td>15</td>
<td>108</td>
<td>11.6</td>
</tr>
<tr>
<td>14</td>
<td>ASTEI04241-PA</td>
<td>Hypothetical protein</td>
<td>5.024</td>
<td>1</td>
<td>2</td>
<td>153</td>
<td>15.7</td>
</tr>
<tr>
<td>15</td>
<td>ASTEI06325-PA</td>
<td>Pyrimidine nucleoside transport protein</td>
<td>5.008</td>
<td>1</td>
<td>2</td>
<td>1205</td>
<td>131.2</td>
</tr>
<tr>
<td>16</td>
<td>ASTEI09665-PA</td>
<td>Hypothetical protein</td>
<td>4.811</td>
<td>2</td>
<td>8</td>
<td>204</td>
<td>21.4</td>
</tr>
<tr>
<td>17</td>
<td>ASTEI07626-PA</td>
<td>Glutathione S-transferase delta class</td>
<td>4.483</td>
<td>1</td>
<td>6</td>
<td>436</td>
<td>49.6</td>
</tr>
<tr>
<td>18</td>
<td>ASTEI05843-PA</td>
<td>Cytochrome P450 (CYP6AF1)</td>
<td>4.472</td>
<td>14</td>
<td>37</td>
<td>493</td>
<td>56.3</td>
</tr>
<tr>
<td>19</td>
<td>ASTEI07367-PA</td>
<td>40S ribosomal protein S21</td>
<td>4.223</td>
<td>1</td>
<td>2</td>
<td>506</td>
<td>56.6</td>
</tr>
<tr>
<td>20</td>
<td>ASTEI10118-PA</td>
<td>Hypothetical protein</td>
<td>3.980</td>
<td>2</td>
<td>14</td>
<td>394</td>
<td>42.4</td>
</tr>
<tr>
<td>21</td>
<td>ASTEI00680-PA</td>
<td>Solute carrier family 23 (nucleobase transporter)</td>
<td>3.973</td>
<td>1</td>
<td>6</td>
<td>524</td>
<td>56.5</td>
</tr>
<tr>
<td>22</td>
<td>ASTEI04034-PA</td>
<td>Ankyrin</td>
<td>3.953</td>
<td>2</td>
<td>4</td>
<td>173</td>
<td>19.1</td>
</tr>
<tr>
<td>23</td>
<td>ASTEI01741-PA</td>
<td>Fas-associated factor</td>
<td>3.931</td>
<td>2</td>
<td>5</td>
<td>439</td>
<td>50.5</td>
</tr>
<tr>
<td>24</td>
<td>ASTEI00353-PA</td>
<td>Dynein heavy chain, axonemal</td>
<td>3.787</td>
<td>1</td>
<td>4</td>
<td>4294</td>
<td>495.2</td>
</tr>
</tbody>
</table>
**Table 2.9:** List of male specific proteins identified and their expression level in male midgut tissue.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>An. stephani VectorBase ID</th>
<th>Relative fold change male/female</th>
<th>Protein description as per An. gambiae</th>
<th>Molecular function</th>
<th>Biological process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ASTEI05843-PA</td>
<td>4.47</td>
<td>Cytochrome P450 (CYP6AF1)</td>
<td>Oxidoreductase activity</td>
<td>Oxidation-reduction process</td>
</tr>
<tr>
<td>2</td>
<td>ASTEI01662-PA</td>
<td>3.39</td>
<td>Regucalcin protein</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>3</td>
<td>ASTEI03074-PA</td>
<td>2.53</td>
<td>Actin cytoplasmic</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>4</td>
<td>ASTEI02736-PA</td>
<td>2.50</td>
<td>Alpha-crystallin chain B</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>5</td>
<td>ASTEI07236-PA</td>
<td>2.36</td>
<td>Trypsin-alpha</td>
<td>Serine-type endopeptidase activity</td>
<td>Proteolysis</td>
</tr>
<tr>
<td>6</td>
<td>ASTEI09992-PA</td>
<td>2.25</td>
<td>Troponin C isoform 4''</td>
<td>Ion binding</td>
<td>Unknown</td>
</tr>
<tr>
<td>7</td>
<td>ASTEI07705-PA</td>
<td>2.16</td>
<td>Leucine aminopeptidase-like protein</td>
<td>Peptidase activity</td>
<td>Proteolysis</td>
</tr>
<tr>
<td>8</td>
<td>ASTEI09406-PA</td>
<td>1.24</td>
<td>Hypothetical protein</td>
<td>Serine-type endopeptidase activity</td>
<td>Proteolysis</td>
</tr>
<tr>
<td>9</td>
<td>ASTEI08327-PA</td>
<td>1.22</td>
<td>Dynein heavy chain 9, axonemal</td>
<td>Ion binding</td>
<td>Microtubule-based movement</td>
</tr>
<tr>
<td>10</td>
<td>ASTEI06940-PA</td>
<td>1.15</td>
<td>Juvenile hormone diol kinase</td>
<td>Ion binding</td>
<td>Unknown</td>
</tr>
<tr>
<td>11</td>
<td>ASTEI09889-PA</td>
<td>1.09</td>
<td>Unknown</td>
<td>GTPase activity</td>
<td>Protein polymerization</td>
</tr>
<tr>
<td>12</td>
<td>ASTEI07479-PA</td>
<td>0.95</td>
<td>Butyryl-CoA dehydrogenase</td>
<td>Peptidase activity</td>
<td>Proteolysis</td>
</tr>
<tr>
<td>13</td>
<td>ASTEI02126-PA</td>
<td>0.90</td>
<td>FMRFamide</td>
<td>Unknown</td>
<td>Neuropeptide signaling pathway</td>
</tr>
<tr>
<td>14</td>
<td>ASTEI07186-PA</td>
<td>0.78</td>
<td>Transglutaminase</td>
<td>Transferase activity</td>
<td>Cellular protein modification process</td>
</tr>
<tr>
<td>15</td>
<td>ASTEI11276-PA</td>
<td>0.66</td>
<td>Tubulin alpha</td>
<td>GTPase activity</td>
<td>Protein polymerization</td>
</tr>
<tr>
<td>16</td>
<td>ASTEI08863-PA</td>
<td>0.65</td>
<td>Dynein, axonemal, heavy chain 5</td>
<td>Ion binding</td>
<td>Microtubule-based movement</td>
</tr>
<tr>
<td>17</td>
<td>ASTEI10640-PA</td>
<td>0.63</td>
<td>Hypothetical protein</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>18</td>
<td>ASTEI05773-PA</td>
<td>0.46</td>
<td>Hypothetical protein</td>
<td>Unknown</td>
<td>Outer dynein arm assembly</td>
</tr>
<tr>
<td>19</td>
<td>ASTEI03866-PA</td>
<td>0.23</td>
<td>Dynein intermediate chain 2, axonemal</td>
<td>Protein binding</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
2.4 Conclusion

The aim of this study was to provide a catalog of the proteins in the midgut of female *An. stephensi*. In this study a high resolution Fourier transform mass spectrometer was used for identifying proteins expressed in the midgut of female *An. stephensi* fed exclusively on sugar solution. This analysis has led to the identification of largest 5,434 proteins in any of the mosquito species so far. This was achieved by the use of multiple fractionation methods and high-accuracy and high resolution mass spectrometer. Mass spectrometry has been considered as an efficient tool for proteomic analysis due to its sensitivity, high mass accuracy and ability to deal with complex sample mixtures. The functional characterisation of hypothetical proteins identified in midgut tissue is a step forward towards understanding the nature of physiological processes and molecular interactions which was lacking in the *An. stephensi*. The cellular components, biological processes and the molecular functions, in which these proteins are involved in, were summarized and discussed. This study demonstrated that the most of the proteins identified are likely to be involved in binding activity, metabolism and transport, and most of these are membrane bound or located intracellularly. Hypothetical proteins identified can be investigated for their role in malaria transmission by instituting gene silencing and expression studies.

In the present study, the difference in protein expression between male and female midgut tissue was also carried out in order to explore the differences in midgut protein expression in relation to sex. For the first time midgut tissue of male and female *An. stephensi* was analyzed using iTRAQ based quantitative proteomic approach. Comparative proteomic analysis of male and female midgut led to the identification of 3,316 proteins and their expression levels between two sexes. The results showed that 758 proteins were quantitatively different between male and female *An. stephensi*. The data generated also provides experimental evidence for several sex specific proteins expressed differentially between male and female mosquitos. Proteomic differences between two sexes also support the broad functional differences between the male and
female midgut. The female midgut specific proteins showed a significantly greater functional diversity than the male midgut-specific proteins, which were mostly involved in housekeeping processes. These findings are on expected lines as the female midgut is site for host blood digestion and parasite development. In comparison to this the male gut has to perform much simpler function of plant sap digestion. Hence overall the functional differences between the proteome of female and male midguts were mainly related to immunity and blood digestion. Prevot et al., (2003) also observed similar differences in An. gambiae protein expression between unfed female and male midguts by using 2-dimensional gel electrophoresis. The proteomic information generated from this study open up vistas for exploring the candidate genes responsible for the pathogen-transmitting processes in An. stephensi. Further functional analysis of these female-specific proteins is necessary for better understanding the differences between male and female An. stephensi. These studies are helpful and crucial for designing transmission blocking strategies via selective gene silencing.