Summary

&

Future Prospects
Summary

Midgut of mosquito is a vital tissue, involved in blood digestion and constitutes the first physical barrier to Plasmodium invasion. Therefore, molecular understanding of midgut proteome is necessary to appreciate its role in vector-parasite interaction and blood digestion. The present investigation was carried out in order to characterize midgut proteome of female *Anopheles stephensi* using mass spectrometry. In this study multiple fractionation methods as well as high-resolution and high mass accuracy mass spectrometer were used. The three different protein/peptide fractionation strategies were used namely, In-gel digestion, bRPLC and Off-gel fractionation which resulted in reduction of sample complexity and facilitated deep proteome coverage of midgut tissues. The fractionated samples were analyzed on high resolution mass spectrometer (LTQ-Orbitrap Elite and LTQ-OrbitrapVelos) which generated over 7,16,932 high resolution tandem mass spectra from 58 LC-MS/MS runs. Two search engines i.e. Sequest and Mascot were used for searching the mass spectrometry derived data against the protein database of Indian strain of *An. stephensi* available on VectorBase, which resulted in identification of 5,43,028 peptide spectral matches (PSMs) corresponding to 63,525 high quality peptides. These peptides provided experimental evidence for 5,434 computationally predicted proteins, out of 11,785 total numbers of proteins, which comprises ~ 46% of the total reported proteome of the malaria vector *An. stephensi*. This is the largest protein catalogue of midgut in any mosquito species reported so far. This proteomic profiling of midgut tissue has provided translation evidence for thousands of genes for which there were no experimental evidence thus far.

For the entire peptide data set, the distribution of mass error in parts per million was calculated. This analysis shows that nearly 95% of the peptides were within ±5 ppm mass error, confirming high accuracy of peptide data obtained from the mass spectrometer. To obtain relative expression of these 5,434 identified proteins within in the tissue, intensity based quantitative expression was carried out. The quantitation was based on spectral count identified for each protein and the variability in the length of the proteins was accounted for to normalize the abundance values. This analysis shows that most of the house-keeping genes including cytoskeleton proteins, glucose metabolism and proteins
involved in translation and embryonic development were among the most abundant found in midgut tissue. These proteins were identified with high coverage, multiple unique peptides and multiple number of PSMs. Qualitative and quantitative information of protein expression will lead to better understanding of the vector life cycle. Further the identified proteins were categorized according to the subcellular localization, their involvement in different biological processes and molecular functions based on gene ontology annotation. For all the identified proteins, ortholog information from the closely related mosquito species An. gambiae was fetched using in-paranoid tool. This analysis resulted in identification of 43 known proteins similar to Anopheles gambiae. These proteins mediate the interaction of the Plasmodium parasites with mosquito tissues and regulate their maturation within their host vectors (Sreenivasamurthy et al., 2013). Another, 49 proteins have been identified in this study which play an important role in immune responses in Anopheles stephensi, as previously reported by Jiang et al., (2014).

For the proteins, whose ortholog information was missing in closely related species, domain analysis was carried out using SMART domain prediction in order to predict their function in midgut tissue.

Comparative proteomic analysis of male and female midgut revealed differential expression of proteins between two sexes. By employing iTRAQ based quantitative approach, a total of 3,316 proteins were identified. Further relative quantitation of these identified proteins was achieved by using reporter ions quantifier node in Proteome Discoverer software to determine relative quantity of a protein in male and female midgut tissue. This analysis resulted in identification of 758 proteins found to be differentially expressed in female midgut with respect to the male counterpart. As expected, proteins involved in blood digestion, glucose metabolism and detoxification processes were found to be over-expressed in female midgut. Thirty two proteins involved in vector parasite interaction were also identified. Further, the differentially expressed proteins identified in this study were compared with sex-specific genes of An. stephensi Indian strain and found a strong correlation (Jiang et al., 2014). Fifty female specific proteins identified in this study were found to be over-expressed by >2 fold in female midgut as compare to male. These female specific proteins were mainly involved in proteolysis and metabolic
processes relevant to blood meal digestion, vector immunity and some are involved in parasite interaction. Another, 19 male specific proteins identified in this study were reported earlier as male biased by Jiang et al., (2014). This quantitative proteomic information generated clearly explains the functional diversity between the male and female mosquito, and also explains why female mosquitoes are the vector for this debilitating disease while a male mosquito does not possess vectorial capability. Further functional analyses of these female-specific proteins are necessary for better understanding the molecular differences between male and female An. stephensi.

Comparative proteomic analysis of midgut of sugar, blood fed and Plasmodium berghei infected mosquitoes provides information about the proteins differentially expressed in response to blood meal and parasite infection in midgut tissue of female An. stephensi. This analysis led to identification of 2,327 proteins. Relative quantification of proteins identified between blood fed versus sugar fed and infected blood fed versus blood fed was carried out on the basis of relative intensity of reporter ions for each of the peptides identified. This analysis showed that, 435 proteins expressed differentially between blood fed and sugar fed midguts. Of these, 247 proteins were up-regulated by >1.5 fold while 188 proteins were down-regulated by <0.66 fold in blood fed as compare to sugar fed midgut tissue. These proteins were found to be involved in vector development process, survival, blood digestion or in the synthesis of the peritrophic matrix that surround the blood meal, transport and storage of iron within cells, vector immunity and defense against microorganisms ingested along with the blood meal (Nolan et al., 2011; Rono et al., 2010). Similarly, 231 proteins were identified expressed differentially between Plasmodium berghei infected and blood fed midgut. Of these 231 proteins, 55 proteins were found to be up-regulated by >1.5 fold while 176 proteins were down-regulated by <0.66 fold in Plasmodium berghei infected as compare to blood midgut tissue. Interestingly one agonistic protein, oxidation resistance protein (OXR1) was found to be 2.01 times up-regulated in P. berghei infected midgut as compare to blood fed mosquitoes. Another protein, Wiskott - Aldrich syndrome which is a antagonistic protein was found < 0.41 times down-regulated in P. berghei infected midgut as compared to blood fed mosquitoes. Among the 55 up-regulated proteins, 15 hypothetical proteins were
identified for which orthologous information was missing in closed related mosquito species *An. gambiae*. These hypothetical proteins are anticipated to be of vital importance to *Anopheles stephensi* and can be used for the targeted expression of anti-Plasmodium effector molecules. These proteins can be explored for functional studies on mosquito-parasite interaction by transgenic expression of these, which can coincide with midgut stages of the Plasmodium parasite. Further the role of these proteins needs to be confirmed by conducting functional analysis using gene knock down.

The proteogenomic analysis was carried out using the peptides identified from the midgut tissue. The novel peptides, which could not be mapped to the known proteins, were searched against 3 frame translated transcript and 6 frame translated genome database. This analysis led to identification of 1,077 genome search specific peptides (GSSPs). The proteogenomic analyses of GSSPs lead to the identification of 104 novel gene models that were missed during genome assembly and correction of over 140 existing gene models through discovery of 40 novel exons, 75 protein extensions, 10 novel protein start sites, 5 novel translational frames and 10 events of joining of adjacent genes as a single gene. The similar proteogenomic analysis was carried out against the first genome assembly of *An. stephensi* (AsteI1) submitted to VectorBase in June 2012. Analysis against first predicted protein and translated genome database (AsteI1) led to identification of over 31 novel protein coding regions. These novel gene models were validated by performing RT-PCR and subsequent sequencing of cDNA to support the existence of novel gene models discovered in this analysis. The cDNA sequences thus obtained were submitted as ESTs to NCBI Genbank. All the 31 novel genes identified previously were annotated in the next revised database i.e. AsteI2. This shows the significance of proteomic data in refinement of genome annotation.

This is the first in-depth proteogenomic analysis of *An. stephensi* Indian strain since its genome sequence has become available. This study suggests that the genome annotation process should comprise of both computational and experimental methods. Several of the novel genes identified in the study are interesting candidates for intervention studies. This study provides a platform for further investigations especially those focused on preventing malaria transmission and vector development.
Future Prospects

In the present study tandem mass spectrometry-based proteomic profiling of *An. stephensi* was carried out to achieve broad coverage of midgut proteome. This study provides valuable information on sex-specific protein expression in female and male midgut. Over-expressed proteins in the female midgut are attractive candidates for targeted studies to test their role in vector development, malaria transmission or inhibition process.

Quantitative proteomic analysis of *Anopheles stephensi* midgut infected with *Plasmodium berghei* allowed us to obtain the biologically relevant midgut proteins that are specifically induced in response to malaria parasite infection. Attractive candidates among these can be further pursued for their role in parasite invasion, inhibition or survival using dsRNA interference based gene silencing assays, paving way for candidates for the development of malaria transmission blocking vaccines. The transmission-blocking vaccines (TBVs) can be explored for interruption of transmission of parasites via suppression of parasite development in vector mosquitoes.

The major highlight of the study has been the use of high-resolution and high mass accuracy mass spectrometer for more accurate annotation of the protein coding genes in *Anopheles stephensi*. This study clearly demonstrates that mass spectrometry derived proteome data can be used to accurately annotate protein-coding genes and can be effectively used to validate hypothetical proteins. Novel genes identified exclusive to midgut can be further investigated for their role in malaria transmission by gene silencing and expression studies.