



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)



### Data Article

# Dataset on fat body proteome of *Anopheles stephensi* Liston



Manish Kumar<sup>a,b</sup>, Ajeet Kumar Mohanty<sup>c</sup>, Gourav Dey<sup>a,b</sup>,  
Sreelakshmi K. Sreenivasamurthy<sup>a,b</sup>, Ashwani Kumar<sup>c,\*</sup>,  
Keshava Prasad<sup>a,d,\*\*</sup>

<sup>a</sup> Institute of Bioinformatics, International Technology Park, Bangalore 560066, Karnataka, India

<sup>b</sup> Manipal Academy of Higher Education, Manipal 576104, Karnataka, India

<sup>c</sup> ICMR-National Institute of Malaria Research, Field Station, Goa 403001, India

<sup>d</sup> Center for Systems Biology and Molecular Medicine, Yenepoya Research Center, Yenepoya (Deemed to be University), University Road, Mangalore 575018, India

#### ARTICLE INFO

##### Article history:

Received 25 October 2018

Received in revised form

24 December 2018

Accepted 7 January 2019

Available online 11 January 2019

#### ABSTRACT

Fat body from *Anopheles stephensi* female mosquitoes were dissected and processed for proteomic analysis. Both SDS-PAGE and basic Reverse Phase Liquid Chromatography-based fractionation strategies were used to achieve a broad coverage of protein identification. The fractionated peptides were then analyzed on a high-resolution mass spectrometer. Searching the raw data against the protein database of *An. stephensi* resulted in identification of 4535 proteins, which is, to our knowledge, the largest catalog of fat body proteome in any mosquito vector species reported so far. Bioinformatics analysis on these fat body proteins suggested the enrichment of biological processes including carbon and lipid metabolism, amino acid metabolism, signal peptide processing and oxidation-reduction. In addition, using proteogenomic approaches, 43 novel proteins were identified, which were not listed in the annotated gene annotations of *An. stephensi*. The data used in the

\* Corresponding author.

\*\* Corresponding author at: Center for Systems Biology and Molecular Medicine, Yenepoya Research Centre, Yenepoya (Deemed to be University), University Road, Mangalore 575018, India.

E-mail addresses: [ashwani07@gmail.com](mailto:ashwani07@gmail.com) (A. Kumar), [keshav@yenepoya.edu.in](mailto:keshav@yenepoya.edu.in) (K. Prasad).

URL: <http://csbmm.yenepoya.edu.in/> (K. Prasad).

analysis are related to the article 'Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes' (Prasad et al., 2017).

© 2019 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Specifications table

Subject area	Biology
More specific subject area	Mosquito proteomics
Type of data	Tables and figures
How data were acquired	LTQ-OrbitrapVelos ETD and Orbitrap Elite mass spectrometer (Thermo Scientific, Bremen, Germany)
Data format	Analyzed output data
Experimental factors	Fat body tissues were dissected and processed from sugar fed <i>Anopheles stephensi</i> female mosquitoes
Experimental features	Fat body tissues from <i>An. stephensi</i> female mosquitoes were processed for mass spectrometry-based proteomic analysis using multiple fractionation methods
Data source location	Goa and Bangalore, India
Data accessibility	Tables and figures are available in this article. Raw files and search results have been deposited to ProteomeXchange Consortium ( <a href="http://proteomecentral.proteomexchange.org">http://proteomecentral.proteomexchange.org</a> ) via the PRIDE partner repository with the dataset identifier <a href="https://proteomecentral.proteomexchange.org/dataset/PXD006677">PX006677</a>
Related research article	Title: Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes Author list: Prasad TS, Mohanty AK, Kumar M, Sreenivasamurthy S, Dey G, Nirujogi RS, Pinto SM, Madugundu AK, Patil AH, Advani J, Manda SS, Gupta MK, Dwivedi SB, Kelkar DS, Hall B, Jiang X, Peery A, Rajagopalan P, Yelamanchi SD, Solanki HS, Raja R, Sathe GJ, Chavan S, Verma R, Patel KM, Jain AP, Syed N, Datta KK, Khan AA, Dammalli M, Jayaram S, Radhakrishnan A, Mitchell CJ, Na CH, Kumar N, Sinnis P, Sharakhov IV, Wang C, Gowda H, Tu Z, Kumar A, Pandey A. Status: Published

## Value of the data

- The data define the fat body proteome of *Anopheles stephensi*.
- The data provide experimental evidence at the level of protein expression for the computationally predicted gene annotations for the genome of *An. stephensi*.
- The data also provide 43 novel genes with protein-level evidence in *An. stephensi*.

## 1. Data

Fat body tissue was dissected from *Anopheles stephensi* female mosquitoes and processed for mass spectrometry-based proteomic analysis. In brief, protein was extracted from the dissected fat body samples. Both protein and peptide-level fractionation strategies were employed using SDS-PAGE and

basic Reverse Phase Liquid Chromatography (bRPLC), respectively (Fig. 1). In all, 44 fractions were independently analyzed on high-resolution mass spectrometer. Peptides present in each fraction were further separated on nano-LC analytical column, interfaced with the mass spectrometer. Peptides obtained from in-gel digested protein bands were analyzed on LTQ-Orbitrap Velos. To achieve a broad coverage, bRPLC fractions were analyzed on Orbitrap Elite hybrid mass spectrometer.

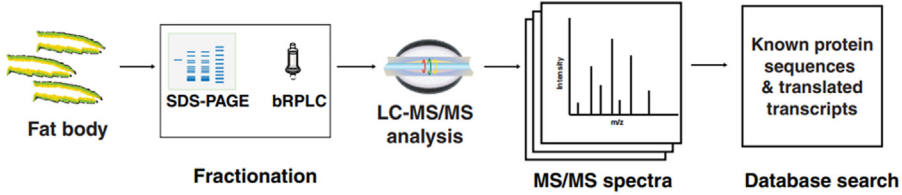


Fig. 1. Overall workflow illustrating the steps involved in proteomic analysis of fatbody of female *An. stephensi*.

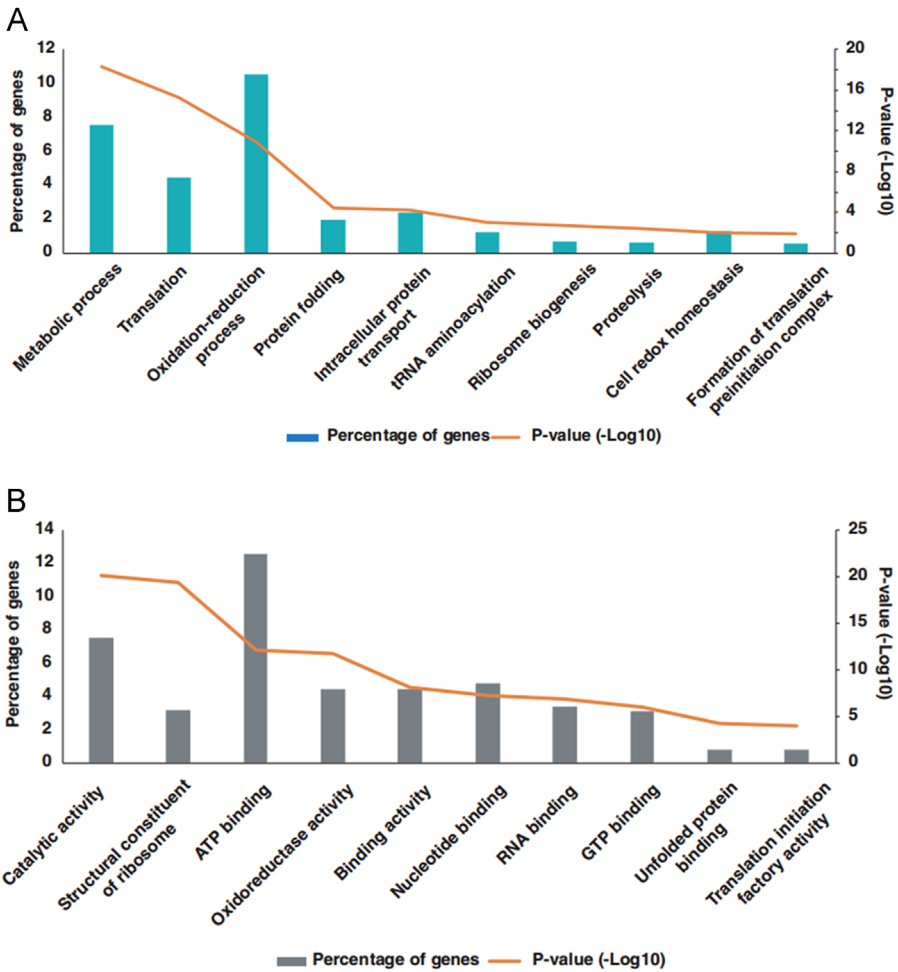


Fig. 2. Gene ontology analysis on the proteins identified in the study. A) and B) represents the biological processes and molecular functions identified to be enriched in the data, respectively.

The raw data were searched using both Sequest and Mascot search engines using Proteome Discoverer software suite (Supplementary Table 1). Bioinformatics analysis was carried out to identify the biological processes and molecular functions that were enriched in the fat body of *An. stephensi* (Fig. 2, Supplementary Table 2). In addition, 45 immunity-related genes were identified in the analysis and a subset of these is known to be involved in vector–pathogen interactions (Supplementary Table 3) [2]. MS/MS spectra that did not map to the known protein sequences were further searched against 3-frame translated transcripts obtained from the fat body and six-frame translated genome database [3]. Novel peptides thus identified were used for the construction of novel gene models from the *An. stephensi* genome (Supplementary Table 4).

## 2. Experimental design

### 2.1. Rearing of mosquitoes and dissection of fat body

Colonies of *An. stephensi* mosquitoes used in this study were maintained at ICMR-National Institute of Malaria Research (NIMR), Field Station, Goa. Ambient conditions (photoperiod:scotoperiod of 12:12 h, temperature of  $27 \pm 2^\circ\text{C}$  and  $70 \pm 5\%$  humidity) were used to raise the mosquitoes. Adult mosquitoes were provided with 10% glucose soaked in a cotton pad. For dissection of the fat body, *An. stephensi* female mosquitoes post 3–4 days of eclosion were used. Abdomen region of adult mosquitoes was separated from the thorax and other organs such as midgut, ovaries and Malpighian tubules were carefully taken out of the abdomen under stereomicroscope. Dissected fat body tissues were then transferred to a separate 2 ml tube containing 0.65% saline solution.

### 2.2. Tissue lysis and extraction of proteins

Dissected fat body samples were suspended in lysis buffer (4% SDS, 100 mM DTT, 100 mM Tris, at pH 7.5), homogenized using mortar and pestle and sonicated at 40% amplitude for 10 cycles. The homogenized lysate was centrifuged at 12,000 rpm for 30 min and supernatant containing the proteins was collected in a separate tube. Protein amount was estimated using Bicinchoninic Acid assay (Pierce, Waltham, MA) as per the manufacturer's protocol.

### 2.3. Protein level fractionation

Two-hundred micrograms of proteins from fat body were resolved on 10% SDS-PAGE. Protein bands were stained with colloidal Coomassie blue stain (Sigma) and entire lane was excised in 20 distinct bands for in-gel digestion as described earlier [1]. In brief, 40% Acetonitrile (ACN) and 40 mM ammonium bicarbonate (ABC) solution was used to destain the gel bands and proteins were reduced and alkylated using 10 mM Dithiothreitol (DTT; Sigma) and 10 mM Iodoacetamide (IAA; Sigma) respectively. Gel pieces were dehydrated and saturated with sequencing grade trypsin (Promega) solution. Post trypsin digestion, peptides were extracted, dried in speedvac and cleaned using C18 resin.

### 2.4. Peptide level fractionation

Proteins were reduced using 10 mM DTT and alkylated with 10 mM IAA prior to trypsin digestion. Trypsin was added to the sample (1:20, enzyme to substrate concentration) and kept at  $37^\circ\text{C}$  for 12 h. Lysate equivalent to 30 micrograms of protein was subjected to SDS-PAGE to confirm protein digestion. Peptides were lyophilized and processed further for bRPLC fractionation. Peptides were reconstituted in solvent A (10 mM triethylammonium bicarbonate, pH 8.5) and resolved on Xbridge C18 5  $\mu\text{m}$  250  $\times$  4.6 mm column (Waters Corporation, Milford, MA, USA) using Agilent 1100 binary pump (Agilent Technologies, Santa Clara, CA, USA). To resolve the peptides, a gradient of 5% to 60% solvent B (10 mM triethylammonium bicarbonate, pH 8.5 in 95% Acetonitrile) was used. The resolved

peptides were finally pooled in 24 fractions, which were dried and stored in  $-80^{\circ}\text{C}$  until LC-MS/MS analysis.

### 2.5. LC-MS/MS analysis

Peptides extracted from SDS-PAGE in-gel bands were analyzed on LTQ-Orbitrap Velos mass spectrometer, which was interfaced with Proxeon Easy nLC system (Thermo Scientific, Bremen, Germany). Each fraction was reconstituted in  $20\ \mu\text{l}$  of 0.1% formic-acid and loaded on to a pre-column ( $75\ \mu \times 2\ \text{cm}$ ) packed in-house with magic C18 AQ (Michrom Bioresources, Auburn, CA, USA). A 15 cm long analytical column ( $75\ \mu\text{m}$ ,  $3\ \mu\text{m}$  particle size) was used to resolve the peptides using a linear gradient of 10–35% solvent B (0.1% formic acid in 95% Acetonitrile) over 75 min. Precursor ion scans were acquired in Orbitrap with resolution of 30,000 at  $400\ m/z$ . Twenty most abundant precursor ions were selected for fragmentation using higher-energy collisional dissociation (HCD) method. MS/MS scans were acquired in Orbitrap mass analyzer using a mass resolution of 15,000 at  $400\ m/z$ . Lock mass option ( $m/z\ 445.1200025$ ) from ambient air was enabled for internal calibration.

For a broad coverage of identification, bRPLC peptide fractions were acquired on LTQ-Orbitrap Elite mass spectrometer (Thermo Scientific, San Jose CA, USA) interfaced with Easy-nano LC II nano flow liquid chromatography system (Thermo Scientific, Odense, Southern Denmark). A 60 cm long analytical column ( $75\ \mu\text{m}$ ,  $5\ \mu\text{m}$  particle and  $100\ \text{\AA}$  pore size) was used to resolve the peptides. Both MS and MS/MS spectra were acquired in Orbitrap with mass resolution of 120,000 and 30,000, respectively.

### 2.6. Data analysis

Raw data were searched against the known and annotated protein database of *An. stephensi* using Sequest and Mascot search engines. The protein database was downloaded from VectorBase (version Astel2.2) and consisted of 11,789 protein sequences. Oxidation of methionine was set as variable modification, whereas, carbamidomethylation of cysteine and acetylation of protein N-terminus was selected as static modification. Precursor and fragment ion mass tolerance of 10 ppm and 0.05 Da was selected. MS/MS spectra that did not map to the known protein sequences were searched against 3-frame translated transcript sequences and six-frame translated genome sequence. Percolator node was used to compute false discovery rate (FDR) and data were filtered with 1% FDR at PSM level. Novel peptides that were identified against the translated transcripts and genome sequence were manually inspected for their MS/MS spectra quality. The respective open reading frames (ORFs) from the transcripts were used to identify the homolog proteins in other insect species.

Gene ontologies related to *An. stephensi* proteins were downloaded from VectorBase resource. The gene ontology files were then loaded on Funrich tool and used as background database to calculate enrichment of biological processes and molecular functions [4].

### 2.7. Submission of the data to public repository

The mass spectrometry-derived raw data and search results were deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository with the dataset identifier [PXD006677](https://doi.org/10.26434/chemrxiv-2019-06-06).

## Acknowledgments

This study is funded by the joint research project to NIMR and IOB entitled “Characterization of Malaria mosquito vector *Anopheles stephensi* Proteome and Transcriptome” (EMR/2014/000444) from the Science and Engineering Research Board (SERB), Government of India. The funders had no role in the experimental design, implementation, analysis and/or reporting of the present work. It bears the approval of NIMR Publication Committee (No. 04/2018). We thank Karnataka Biotechnology and Information Technology Services (KBITS), Government of Karnataka for the support to the Center for Systems Biology and Molecular Medicine at Yenepoya (Deemed to be University) under the

Biotechnology Skill Enhancement Programme in Multiomics Technology (BiSEP GO ITD 02 MDA 2017). We also thank the Department of Biotechnology (DBT), Government of India, for research support to the Institute of Bioinformatics.

### **Transparency document. Supporting information**

Transparency document associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2019.01.016>.

### **Appendix A. Supporting information**

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.dib.2019.01.016>.

### **References**

- [1] T.S. Prasad, A.K. Mohanty, M. Kumar, S.K. Sreenivasamurthy, G. Dey, R.S. Nirujogi, S.M. Pinto, et al., Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes, *Genome Res.* 27 (1) (2017) 133–144 (Jan).
- [2] S.K. Sreenivasamurthy, G. Dey, M. Ramu, M. Kumar, M.K. Gupta, A.K. Mohanty, H.C. Harsha, et al., A compendium of molecules involved in vector-pathogen interactions pertaining to malaria, *Malar. J.* 12 (2013) 216 (Jun 26).
- [3] S.K. Sreenivasamurthy, A.K. Madugundu, A.H. Patil, G. Dey, A.K. Mohanty, M. Kumar, K. Patel, et al., Mosquito-borne diseases and omics: tissue-restricted expression and alternative splicing revealed by transcriptome profiling of *Anopheles Stephensi*, *OMICS J. Integr. Biol.* 21 (8) (2017) 488–497 (Aug).
- [4] M. Pathan, S. Keerthikumar, C.S. Ang, L. Gangoda, C.Y. Quek, N.A. Williamson, D. Mouradov, et al., Funrich: an open access standalone functional enrichment and interaction network analysis tool, *Proteomics* 15 (15) (2015) 2597–2601 (Aug).