International Journal of Engineering Research and General Science Volume 3, Issue 6, November-December, 2015
ISSN 2091-2730

Bio-Computational Characterization of *Wolbachia* surface protein in different species of Drosophila

Amrutha.K.M, Dr. Mahesh Pattabhiramaiah and Anusha.P.K

Centre for Applied Genetics, Department of Zoology, Jnanabharathi, Bangalore University,

Bangalore - 560056

Email:reply2mahesh@gmail.com

Ph no: 91-9916130942

Abstract - *Wolbachia* surface protein (WSP) is an eight beta-barrel, trans-membrane structure, which participates in host immune response, cell proliferation, pathogenicity and controlled cell death program. The current study employs the bio-informatics tool to unravel the structural and functional properties of the WSP infecting drosophila. The present study was focused on sequence analysis, insilico prediction of the secondary and tertiary structure of wsp sequence. Sequence analysis and physiochemical properties revealed that this protein is highly stable, negatively charged and having more hydrophobic regions. SOPMA was used to predict secondary structure of wsp, which revealed that the protein contains more of random coils and extended strands than alpha helix and beta sheets. SVM prot analysis revealed the functionality of protein including the details of metal binding sites. Multiple sequence alignment was performed using MUSCLE server which revealed highly conserved regions. The RNA structure was predicted by using Genebee service software, a set of homologous sequences as the stems with their free energy. The protein orientation and trans-membrane region was predicted using TMpred, which will be useful for drug designing. The predicted 3D model was analyzed using Swiss Model which will be helpful for further structure based studies.

Keywords - *Wolbachia* Surface Protein, Drosophila, SOPMA, SVM prot, ProtParam, MUSCLE, TMpred and Swiss Model.

INTRODUCTION

The genus *Wolbachia* (Rickettsiaceae) is a group of intracellular, gram-negative and endosymbiontic bacteria that belong to the order Rickettsiales in the α-subdivision of the class Proteobacteria [35]. The genus *Wolbachia* comprises a group of maternally inherited intracellular bacteria that have been identified in a wide range of arthropod hosts. They are parasitic bacteria of invertebrates including insects, chelicerates, crustaceans, nematodes and dipterans.

Although, *Wolbachia* usually is vertically transmitted, there are cases of horizontal transmission even across host species [36 & 19]. *Wolbachia* infection commonly causes reproductive disorders such as feminization, parthenogenesis and cytoplasmic incompatibility, but also direct morbidity such as neural tissue invasion and destruction [21].

The fruit fly *Drosophila melanogaster*, offers the ideal opportunity to investigate the host-parasite interaction. Not only is it an extremely powerful model host for investigating all aspects of infection and immunity [6 & 23], it is also one of the best-developed model systems for behavioral ecology and genetics [30 & 24]. *Wolbachia* is present in most natural population of *D.melanogaster*, although with variable frequencies of infection [25].

Wolbachia surface protein (WSP), an abundantly expressed protein of Wolbachia, was identified in the endobacteria of Drosophila spp, and has been characterized for *Wolbachia* residing in *D. immitis* [5 & 3]. WSP is a low-molecular-weight protein of 22 kDa. WSP belongs to pfam0617, primarily defined by antibody recognition.

Wolbachia surface protein (WSP) is an eight beta-barrel transmembrane structure which participates in host immune response, cell proliferation, pathogenicity and controlled cell death program. Recombination has a large impact on diversity of this protein including positive selection, which is major constraint on protein evolution. In *Wolbachia*, increased recombination is observed in ankyrin proteins, surface proteins and in some hypothetical proteins.

The wealth of *Wolbachia* surface protein sequence information that has been made publicly available in recent years requires the development of high-throughput proteomics approaches for its analysis. Characterization of proteins of interest from a particular biological study requires the application of suitable bioinformatics tools to process, annotate and prioritise the data in order to gain maximum benefit from the results generated.

From a protein function standpoint, transfer of annotation from known proteins to a novel target is currently the only practical way to convert vast quantities of raw sequence data into meaningful information. New bioinformatics tools now provide more sophisticated methods to transfer functional annotation, integrating sequence, family profile and structural search methodology.

Present study explored the physicochemical nature, three dimensional structure and detail of interactions and functions of the wsp structure. The structural characterization of WSP would provide the clues to its biological function, physiological role and is a prerequisite for the development of new drug targets. The physicochemical and the structural properties of the proteins are well understood with the use of computational tools by through insilico analysis. The statistics about a protein sequence such as number of amino acid, frequency is predicted by CLC work bench. Simple Modular Architecture Research Tool (SMART) is a biological database that is used in the identification and analysis of protein domains within protein sequences [29 & 17]. Sequence length, and

the physico-chemical properties of proteins such as molecular weight, extinction coefficient, GRAVY, aliphatic index, instability index, etc., can be computed by ProtParam. The TMpred program makes a prediction of membrane-spanning regions and their orientation. The algorithm is based on the statistical analysis of TMbase, a database of naturally occurring transmembrane proteins. The prediction is made using a combination of several weight-matrices for scoring [15]. MUSCLE stands for Multiple **S**equence **C**omparison by **L**og- **E**xpectation. The protein 3D model and its characteristics can be predicted by Swiss model server [32]. Protein homology modeling and analogy recognition is made through Phyre2 online server. Further Computer-aided techniques for the efficient identification and optimization of novel molecules with a desired biological activity have become a part of the drug discovery process.

Bioinformatics has revolutionized in the field of molecular biology. The raw sequence information of proteins and nucleic acid can convert to analytical and relative information with the help of soft computing tools. Prediction of protein function is important application of bioinformatics. The amino acid sequence provides most of the information required for determining and characterizing the molecule‘s function, physical and chemical properties. Sequence analysis and physicochemical characterization of proteins using biocomputation tools have been done by many researchers and reported [2, 18, 26, 20, & 34].

The extensive knowledge available on Drosophila provides a solid base on which to test new hypothesis on host–*Wolbachia* interactions. The main objective of this study was to perform physiochemical characterization of the wsp in Drosophila, which are very much necessary in understanding the *Wolbachia* -host interactions with special reference to parasitic *Wolbachia*, which causes different reproductive anomalies in arthropod hosts. The structural characterization would provide the clues to its biological function, physiological role and is a prerequisite for the development of integrated pest managements.

Materials and Methods:

Protein sequence retrieval : The Protein Sequences of *wolbachia* surface protein(15 sequences) were retrieved in FASTA format from NCBI database (Table1) .

Amino acid Composition: The amino acid composition of selected proteins were computed using the tool CLC free workbench (www.clc.bio.com/..../clc-main-workbench), tabulated in (Table-2).

Primary structure analysis: Counts of hydrophobic and hydrophilic residues were calculated from the primary structure analysis by CLC workbench (Table-3).

Physiochemical parameters: The physiochemical parameters such as theoretical isoelectric point (Ip), molecular weight, total number of positive and negative residues, extinction coefficient, instability index [8] aliphatic index [7] and grand average hydropathy (GRAVY) [16] were computed using the Expasy‘s ProtParam server (http://web.expasy.org/protparam/) [22], and tabulated in (Table-4).

Secondary structure prediction: The secondary structure was predicted by self-optimized prediction method with alignment by SOPMA server **(**https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.plpage=npsa_sopma.html) [2] (Table-5).

Domain architecture analysis: Domain organization and domain composition was analyzed using Simple Modular Architecture Research Tool (SMART) (http://smart.embl-heidelberg.de) (Table-6).

SVM prot analysis: The protein function prediction and classification of proteins were analyzed using SVM Prot (http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi.) (Table-7).

Transmembrane region prediction: Transmembrane helices were predicted by the TMpred software. (http://www.ch.embnet.org/software/TMPRED_form.html) (Fig-1) (Table-8).

Sequence Homology Analysis: The sequence homology was analyzed by MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/) (Fig-2).

Tertiary structure Prediction: Tertiary structure prediction (Fig-3) of *wolbachia* surface protein was performed using bioinformatics tool Phyre2 (www.sbg.bio.ic.ac.uk/phyre2/index.cgi).

RNA structure prediction: The protein sequences were reverse transcribed to DNA using Sequence manipulation suite (SMS) (http://www.bioinformatics.org/sms2/rev_trans.html). The reverse transcribed DNA was converted to RNA using transcriptional and translational tool (http://www.attotron.com/cybertory/analysis/trans.html). RNA structure was predicted using (http://www.genebee.msu.su/services/rna2_reduced.html) (Fig-4).

Swiss model: SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server or from the program Deep View (Swiss Pdb-Viewer) (http://swissmodel.expasy.org/). The purpose of this server is to make Protein Modeling accessible to all biochemists and molecular biologists worldwide. (Fig-5)

(Table 1) - *Wolbachia* Surface Protein Sequences Retrieved from NCBI

Sl No	Species	ID	Length	Protein Sequence
1	*Drosophila sechellia*	ABD75492.1	189	VRLQYNGEILPLFTKVDGATGAKKKTADTDTTTDLYKASFMAGGGAFGYK DDIRVDVEGLYSQLSKDTLDVAPTPAIADSLTAFSGLVNVYYYDIAIEDMPITP YVGVGVGAAYISTPLATAVSSQNGKFAFAGQARAGVSYDITPEIKLYAGAR YFGSFCAHFDKDTAAASKDKGELKVLYSTVGAEA
2	*Drosophila sturtevanti*	ABD75491.1	186	VRLQYNGEILPLFTKIDGIQKTKGKEKDSPLKASFVAGGGAFGYKMDDIRVD VEGLYSWLNKDADVVGDTVADNLTAISGLVNVYYYDVAIEDMPITPYIGVGI GAAYISTPLKTAVNEQNSKFGFAGQVKAGVSYDVTPEIKLYAGARYFGSYG AHFDKSEEVDKAVGGKETKVTKDAYKVLYSTV
3	*Drosophila nikananu*	ABD75490.1	189	VRLQYNGEFLPLFTKIDGITNATGKEKDSPLKASFIAGGGAFGYKMDDIRVD VEGLYSQLSKDTTIINTSEENVADSLTAFSGLVNVYYYDIAIEDMPITPYVGVG VGAAYISTPLKPAINEQNSKFGFAGQVKAGVSYDVTPEIKLYAGARYFGSYG AHFDKSEEVDKAGGGKETKVTKDAYKVLYSTV
4	*Drosophila mauritiana*	ABD75488.1	183	VRLQYNGEVLPFKTRIDGIEYKKGTEVHDPLKASFMAGGAAFGYKMDDIRV DVEGLYSQLNKNDVSGATFTPTTVANSVAAFSGLVNVYYDIAIEDMPITPYV GVGVGAAYISNPSEASAVKDQKEFGFAYQAKAGVSYDVTPEIKLYAGARYF GSYGASFNKEAVSATKEINVLYSAVGAEA
5	*Drosophila ananassae*	ABD75484.1	189	VRLQYNGEFLPLFTKVDGITYKKDKSDYSPLKPSFIAGGGAFGYKMDDIRVD VEGVYSYLNKNDVKDVTFDPANTIADSVTAISGLVNVYYDIAIEDMPITPYIG VGVGAAYISTPLEPAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYF GSYGANFDGKKTDPKDSTRQVTDAGAYKVLYSTV
6	*Drosophila arawakana*	ABD75483.1	189	VRLQYNGEFLPLFTKVDGITYKKDKSDYSPLKPSFIAGGGAFGYKMDDIRVD VEGVYSYLNKNDVKDVTFDPANTIADSVTAISGLVNVYYDIAIEDMPITPYIG VGVGAAYISTPLEPAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYF GSYGANFDGKKTDPKDSTRQVTDAGAYKVLYSTV
7	*Drosophila simulans*	ABD75482.1	189	VRLQYNGEFLPLFTKVDGITYKKDKSDYSPLKPSFIAGGGAFGYKMDDIRVD VEGVYSYLNKNDVKDVTFDPANTIADSVTAISGLVNVYYDIAIEDMPITPYIG VGVGAAYISTPLEPAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYF GSYGANFDGKKTDPKNSTGQAADAGAYKVLYSTV
8	*Drosophila melanogaster*	ABD75481.1	189	VRLQYNGEFLPLFTKVDGITYKKDKSDYSPLKPSFIAGGGAFGYKMDDIRVD VEGVYSYLNKNDVKDVTFDPANTIADSVTAISGLVNVYYDIAIEDMPITPYIG VGVGAAYISTPLEPAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYF GSYGANFDGKKTDPKNSTGQAADAGAYKVLYSTV
9	*Drosophila tropicalis*	ABD75479.1	189	VRLQYNGEFLPLFTKVDGITYKKDKSDYSPLKPSFIAGGGAFGYKMDDIRVD VEGVYSYLNKNDVKDVTFDPANTIADSVTAISGLVNVYYDIAIEDMPITPYIG VGVGATYISTPLEPAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYFG SYGANFDGKKTDPKDSTRQVTDAGAYKVLYSTV
10	*Drosophila willistoni*	ABD75478.1	189	VRLQYNGEFLPLFTKVDGITYKKDKSDYSPLKPSFIAGGGAFGYKMDDIRVD VEGVYSYLNKNDVKDVTFDPANTIADSVTAISGLVNVYYDIAIEDMPITPYIG VGVGATYISTPLEPAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYFG SYGANFDGKKTDPKDSTRQVTDAGAYKVLYSTV
11	*Drosophila pseudotakahashii*	ABD75474.1	181	VRLQYNGEILPLFTKVDGITYKKDNSDYSPLKASFIAGGGAFGYKMDDIRVD VEGVYSYLNKNDVTDAKFTPDTIADSLTAISGLVNVYYDIAIEDMPITPYIGV GVGAAYISTPLKDAVNDQKSKFSFAGQVKAGVSYDVTPEVKLYAGARYFGS FGAHFDKDAAAGKDKGELKVLYSTV
12	*Drosophila bicornuta*	ABD75473.1	187	VRLQYNGEVLPLFTKVDNMKIKKGTDDVDPFKASFIGGGAAFGYKMDDIRV DIEGLYSQLNKNVNNDEVLTPDTVAGSLTAISGLVNVYYDIAIEDMPITPYVG VGVGAAYISTPLKDAVNDQKSKFGFAGQVKAGVSYDVAPEVKLYAGARYF GSYGANFDKSGGEKDKGGHTVLYSTVGAEAGVA
13	*Drosophila fumipennis*	AAU95644.1	187	SYYVRLQYNGEVLPLFTKVDNMKIKKGTDDVDPFKASFIGGDAAFGYKMD DIRVDIEGLYSQLNKNVNNGEALTPDIVAGSLTAISGLVNVYYDIAIEDMSITP YVGVGVGAAYISAPLNDAVNGQKSKFGFAGQVKAGVSYDVTPEVKLYAGA RYFGSYGANFDKSSGEKNKGGHTVLYSTVGAEA

14	*Drosophila suzukii*	AFP86012.1	182	VRLQYNGEILPLFTKIEGIEYKKATDIHNPLKASFIAGGGAFGYKMDDIRVDV EGLYSQLNKNDVTGAAFNPDTVADSLTAISGLVNVYYYDIAIEDMPITPYVGV GVGAAYISTPLKDAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYFG SFGAHFDKDTAAASKDKGELKVLYSTV
15	*Drosophila nigrocirrus*	AFP85988.1	184	VRLQYNGEFLPLFTKVDGITNATGKEKDSPLKASFIAGGGAFGYKMDDIRVD VEGLYSWLNKDADVVGDTVADNLTAISGLVNVYYYDVAIEDMPITPYIGVGV GAAYISTPLKTPINDQKSKFGFAGQVKAGVSYDVTPEIKLYAGARYFGSYGA NFDGKKTDPKDSTKQVTDAGAYKVLYSTV

(Table 2)- REPRESENTATION OF FREQUENCY OF AMINO ACIDS of WSP

Sl No	Amino acid	ABD75492.1	ABD75491.1	ABD75490.1	ABD75488.1	ABD75484.1	ABD75483.1	ABD75482.1	ABD75481.1	ABD75479.1	ABD75478.1	ABD75474.1	ABD75473.1	AAU95644.1	AFP860112.1	AFP85988.1
1	Alanine (A)	0.143	0.091	0.09	0.131	0.085	0.085	0.095	0.095	0.079	0.079	0.099	0.096	0.096	0.11	0.092
2	Cysteine (C)	0.005	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Aspartic Acid (D)	0.085	0.081	0.069	0.055	0.095	0.095	0.09	0.09	0.095	0.095	0.099	0.086	0.075	0.082	0.092
4	Glutamic Acid (E)	0.032	0.048	0.058	0.06	0.026	0.026	0.026	0.026	0.026	0.026	0.028	0.037	0.037	0.038	0.027
5	Phenylalanine (F)	0.048	0.038	0.048	0.049	0.048	0.048	0.048	0.048	0.048	0.048	0.05	0.043	0.043	0.049	0.043
6	Glycine (G)	0.101	0.113	0.111	0.098	0.101	0.101	0.106	0.106	0.101	0.101	0.099	0.128	0.123	0.104	0.114
7	Histidine (H)	0.005	0.005	0.005	0.005	0	0	0	0	0	0	0.006	0.005	0.005	0.011	0
8	Isoleucine (I)	0.048	0.059	0.063	0.049	0.053	0.053	0.053	0.053	0.053	0.053	0.061	0.048	0.053	0.06	0.054
9	Lysine (K)	0.069	0.097	0.085	0.066	0.085	0.085	0.085	0.085	0.085	0.085	0.088	0.08	0.075	0.082	0.087
10	Leucine (L)	0.069	0.059	0.058	0.044	0.048	0.048	0.048	0.048	0.048	0.048	0.061	0.059	0.059	0.066	0.06
11	Methionine (M)	0.016	0.011	0.011	0.016	0.011	0.011	0.011	0.011	0.011	0.011	0.011	0.016	0.016	0.011	0.011

12	Asparagine (N)	0.016	0.032	0.037	0.044	0.037	0.037	0.042	0.042	0.037	0.037	0.033	0.048	0.059	0.038	0.038
13	Proline (P)	0.037	0.032	0.037	0.038	0.053	0.053	0.053	0.053	0.053	0.053	0.039	0.037	0.032	0.038	0.043
14	Glutamine (Q)	0.021	0.022	0.021	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.017	0.021	0.021	0.022	0.022
15	Arginine (R)	0.021	0.016	0.016	0.022	0.021	0.021	0.016	0.016	0.021	0.021	0.017	0.016	0.016	0.016	0.016
16	Serine (S)	0.063	0.054	0.069	0.071	0.063	0.063	0.063	0.063	0.063	0.063	0.066	0.053	0.07	0.055	0.054
17	Threonine (T)	0.085	0.059	0.069	0.049	0.063	0.063	0.058	0.058	0.069	0.069	0.055	0.048	0.043	0.055	0.071
18	Valine (V)	0.074	0.108	0.085	0.104	0.106	0.106	0.101	0.101	0.106	0.106	0.094	0.112	0.102	0.093	0.098
19	Tryptophan (W)	0	0.005	0	0	0	0	0	0	0	0	0	0	0	0	0.005
20	Tyrosine (Y)	0.063	0.07	0.069	0.077	0.085	0.085	0.085	0.085	0.085	0.085	0.077	0.064	0.075	0.066	0.071

(Table 3)-Hydrophilic and hydrophobic residues computed by CLC WORK BENCH

Accession Number	Counts of hydrophilic residues	Counts of hydrophobic residues
ABD75492.1	48	101
ABD75491.1	44	96
ABD75490.1	50	95
ABD75488.1	48	97
ABD75484.1	51	95
ABD75483.1	51	95
ABD75482.1	51	97
ABD75481.1	51	97
ABD75479.1	52	94
ABD75478.1	52	94
ABD75474.1	45	93
ABD75473.1	44	101
AAU95644.1	50	98
AFP86012.1	43	97
AFP85988.1	47	96

(TABLE 4) – Physiochemical Parameters Computated by Expasy ProtParam

Accession Number	PI	Mol.Wt	-R	+R	EC	II	AI	Gravy
ABD75492.1	4.83	19919.4	22	17	17880	17.51	81.16	0.011
ABD75491.1	5.25	20060.7	24	21	24870	21.83	86.45	-0.151
ABD75490.1	4.96	20331.8	24	19	19370	23.41	81.01	-0.185
ABD75488.1	4.91	19578.9	21	16	20860	27.06	79.45	-0.066
ABD75484.1	4.99	20503.0	23	20	23840	22.68	78.36	-0.216
ABD75483.1	4.99	20503.0	23	20	23840	22.68	78.36	-0.216
ABD75482.1	4.97	20344.8	22	19	23840	20.59	77.88	-0.194
ABD75481.1	4.97	20344.8	22	19	23840	20.59	77.88	-0.194
ABD75479.1	4.99	20533.0	23	20	23840	22.68	77.83	-0.229
ABD75478.1	4.99	20533.0	23	20	23840	22.68	77.83	-0.229
ABD75474.1	4.98	19549.0	23	19	20860	17.21	84.59	-0.112
ABD75473.1	4.87	19814.3	23	18	17880	16.89	83.90	-0.112
AAU95644.1	4.98	19931.3	21	17	20860	17.78	82.89	-0.124
AFP86012.1	5.16	19525.0	22	18	17880	21.78	87.36	-0.065
AFP85988.1	4.97	19727.2	22	19	24870	18.03	82.12	-0.154

(TABLE 5)- Representation Of helix, sheets, turns, coils by Garnier peptide Analysis through online tool by SOPMA

Accession Number	Helix(H) Residue Totals	Percentage (%)	Sheet(E) Residue Totals	Percentage (%)	Turns (T) Residue Totals	Percentage (%)	Coils Residue Totals	Percentage (%)
ABD75492.1	45	23.81	59	31.22	23	12.17	62	32.80
ABD75491.1	52	27.96	58	31.18	22	11.83	54	29.03
ABD75490.1	39	20.63	60	31.75	22	11.64	68	35.98
ABD75488.1	61	33.33	49	26.78	24	13.11	49	26.78
ABD75484.1	26	13.76	68	35.98	23	12.17	72	38.10
ABD75483.1	26	13.76	68	35.98	23	12.17	72	38.10
ABD75482.1	31	16.40	65	34.39	21	11.11	72	38.10
ABD75481.1	31	16.40	65	34.39	21	11.11	72	38.10
ABD75479.1	24	12.70	68	35.98	23	12.17	74	39.15
ABD75478.1	24	12.70	68	35.98	23	12.17	74	39.15
ABD75474.1	32	17.68	61	33.70	27	14.92	61	33.70
ABD75473.1	38	20.32	59	31.55	24	12.83	66	35.29
AAU95644.1	36	19.25	58	31.02	24	12.83	69	36.90
AFP86012.1	45	24.73	60	32.97	25	13.74	52	28.57

AFP85988.1	32	17.39	71	38.59	19	10.33	62	33.70

(TABLE 6) - SMART ANALYSIS

NAME	START	END	E-VALUE
ABD75492.1	14	34	N/A
ABD75491.1	103	166	536
ABD75490.1	No domain	No domain	No domain
ABD75488.1	4	86	1640
ABD75484.1	No domain	No domain	No domain
ABD75483.1	No domain	No domain	No domain
ABD75482.1	No domain	No domain	No domain
ABD75481.1	No domain	No domain	No domain
ABD75479.1	No domain	No domain	No domain
ABD75478.1	No domain	No domain	No domain
ABD75474.1	No domain	No domain	No domain
ABD75473.1	No domain	No domain	No domain
AAU95644.1	No domain	No domain	No domain
AFP86012.1	53	89	2260
AFP85988.1	No domain	No domain	No domain

(TABLE 7)- SVMPROT ANALYSIS

Accession Number	Protein Family							
	Metal Binding		Zinc Binding		All Lipid Protein		Outer Membrane	
	R	P	R	P	R	P	R	P
ABD75492.1	1.0	58.6	1.4	71.3	-	-	-	-
ABD75491.1	1.0	58.6	-	-	1.3	68.5	1.0	58.6
ABD75490.1	1.1	62.2	3.7	97.0	-	-	1.0	58.6
ABD75488.1	1.0	58.6	3.2	95.2	-	-	-	-
ABD75484.1	1.0	58.6	-	-	1.5	73.8	-	-
ABD75483.1	1.0	58.6	-	-	1.5	73.8	-	-
ABD75482.1	1.0	58.6	1.3	68.5	1.5	73.8	-	-
ABD75481.1	1.0	58.6	1.3	68.5	1.5	73.8	-	-
ABD75479.1	1.0	58.6	-	-	1.6	76.2	-	-
ABD75478.1	1.0	58.6	-	-	1.6	76.2	-	-
ABD75474.1	1.0	58.6	1.2	65.4	1.2	65.4	-	-
ABD75473.1	1.0	58.6	1.6	76.2	-	-	-	-
AAU95644.1	1.0	58.6	1.4	71.3	-	-	-	-
AFP86012.1	-	-	1.7	78.4	2.0	83.9	1.0	58.6
AFP85988.1	1.0	58.6	2.0	83.9	1.0	62.2	1.0	58.6

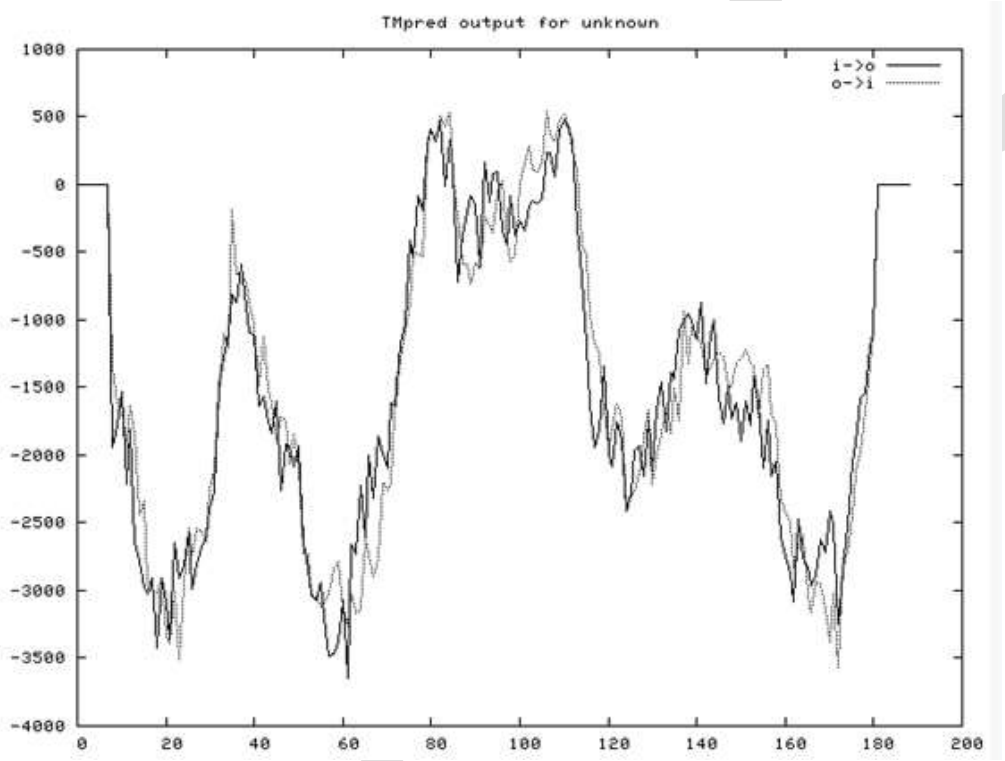
(TABLE 8)- Transmembrane region scoring showing helices (ABD75481.1)

	Inside-Outside	Outside-Inside
Helices	(73- 91 (19) 466)	(76- 94 (19) 538)
	(103- 121 (19) 303)	(98- 115 (18) 427 +)

(TABLE 9)- RNA structure stems with free energy (ABD75481)

Stem no	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Free energy(Kkal/mol)	-16.500000 Kkal/mol	-14.600000 Kkal/mol	-13.500000 Kkal/mol	-13.200000 Kkal/mol	-13.200000 Kkal/mol	-12.700000 Kkal/mol	-11.400000 Kkal/mol	-11.300000 Kkal/mol	-10.000000 Kkal/mol	-9.400000 Kkal/mol	-8.900000 Kkal/mol	-8.900000 Kkal/mol	-8.400000 Kkal/mol	-8.300000 Kkal/mol	-7.300000 Kkal/mol	-7.000000 Kkal/mol	-6.700000 Kkal/mol	-6.600000 Kkal/mol	-6.100000 Kkal/mol

(Fig-1) Transmembrane region prediction (ABD75481.1)



(Fig-2) MUSCLE-Multiple sequence alignment of wsp drosophila

Conserved sequences for hierarchical clustering, primary constructions, identity percentage strong and weakly similar sequences is predicted.

```
gi|89515329|gb|ABD75488.1|        YISNPSEASAVKDQKEFGFAYQAKAGVSYDVTPEIKLYAGARYFGSYGASFN------KE
gi|89515337|gb|ABD75492.1|        YISTPLATAVSSQNGKFAFAGQARAGVSYDITPEIKLYAGARYFGSFCAHFD------KD
gi|89515299|gb|ABD75473.1|        YISTPLKDAVNDQKSKFGFAGQVKAGVSYDVAPEVKLYAGARYFGSYGANFD--------
gi|53854523|gb|AAU95644.1|        YISAPLNDAVNGQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYFGSYGANFD--------
gi|89515335|gb|ABD75491.1|        YISTPLKTAVNEQNSKFGFAGQVKAGVSYDVTPEIKLYAGARYFGSYGAHFD--KSEEVD
gi|89515333|gb|ABD75490.1|        YISTPLKPAINEQNSKFGFAGQVKAGVSYDVTPEIKLYAGARYFGSYGAHFD--KSEEVD
gi|400365113|gb|AFP85988.1|       YISTPLKTPINDQKSKFGFAGQVKAGVSYDVTPEIKLYAGARYFGSYGANFDGKKTDPKD
gi|89515317|gb|ABD75482.1|        YISTPLEPAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYFGSYGANFDGKKTDPKN
gi|89515315|gb|ABD75481.1|        YISTPLEPAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYFGSYGANFDGKKTDPKN
gi|89515321|gb|ABD75484.1|        YISTPLEPAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYFGSYGANFDGKKTDPKD
gi|89515319|gb|ABD75483.1|        YISTPLEPAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYFGSYGANFDGKKTDPKD
gi|89515311|gb|ABD75479.1|        YISTPLEPAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYFGSYGANFDGKKTDPKD
gi|89515309|gb|ABD75478.1|        YISTPLEPAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYFGSYGANFDGKKTDPKD
gi|89515301|gb|ABD75474.1|        YISTPLKDAVNDQKSKFSFAGQVKAGVSYDVTPEVKLYAGARYFGSFGAHFD------KD
gi|400365161|gb|AFP86012.1|       YISTPLKDAVNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYFGSFGAHFD------KD
                                  *** *   .   :: :*.** *..******::**:***********: * *:
```

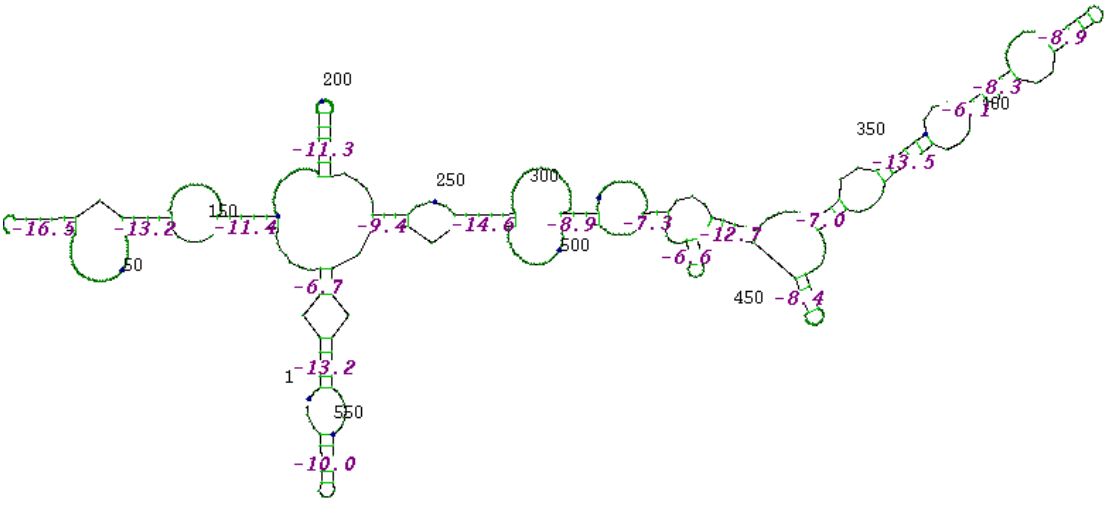
Alignment data;

Primary Construction. KKYMFYPDFLCKQPSE2C, Alignment length:

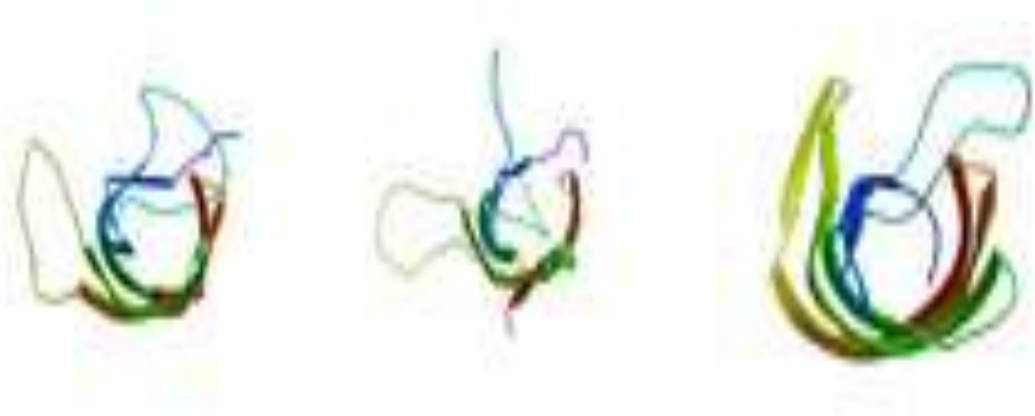
Identity (*): Strongly similar (:): Weakly similar (.)

(Fig- 3) RNA Structure prediction (ABD75481.1)

Free Energy of Structure = -107.4 kkal/mol



(Fig- 4) Protein Homology/analogy recognition by Swiss model (ABD75481.1)

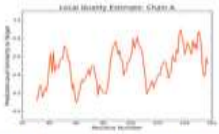


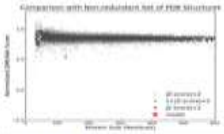
Model 1 **Model 2** **Model 3**

(Fig-5) Model building using SWISS model of (ABD75481.1)

Model #01	File	Built with	Oligo-State	Ligands	GMQE	QMEAN4
	PDB	ProMod Version 3.70	MONOMER	None	0.44	-5.94

QMEAN4	-5.94
Cβ	-3.62
All Atom	-2.73
Solvation	-3.08
Torsion	-4.15





Template	Seq Identity	Oligo-state	Found by	Method	Resolution	Seq Similarity	Range	Coverage	Description
2maf.1.A	26.19	monomer	HHblits	NMR	NA	0.33	31 - 157	0.67	Opacity protein opA60

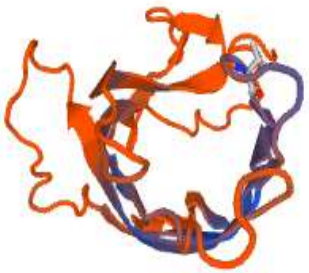
The template contained no ligands.

```
Target    VRLQYNGEFLPLFTKVDGITYKKDKSDYSPLKPSFIAGGGAFGYKMDDIRVDVEGV-YSYLNKNDVK----DVTF-----
2maf.1.A  ------------------------------IRTRSVHPRVSVGYDFGGWRIAADYARYRKWNNNKYSVNIENVRIRKENG

Target    ---DP--ANTIADSVTAI-SGLVNVYYDIAIEDMPITPYIGVGVGAAYISTPLEPA------------------------
2maf.1.A  IRIDRKTENQENGTFHAVSSLGLSAIVDFQIND-KFKPYIGARVAYGHVRHSIDSTKKTIEVTTVPSNAPNGAVTTYNTD

Target    -----VNDQKSKFGFAGQVKAGVSYDVTPEVKLYAGARYFGSYGANFDGKKTDPKNSTGQAADAGAYKVLYSTV
2maf.1.A  PKTQNDYQSNSIRRVGLGVIAGVGFDITPKLTLDAGYRYHNW--------------------------------
```

(Fig - 6) 3D View Of The Structure With Ligand Binding: (ABD75481.1)



DISCUSSION:

Annotation of protein function is one of the key problems in post genomic era. This demands bioinformatics and computational biology to predict the function of unannoated hypothetical proteins by using various efficient tools and web servers. In our study, the analysis of *Wolbachia* surface protein was done using various bio-informatics tools and servers. The sequence and structural features of wsp and its complexities was annotated. The results of which allows for the designing of desired drugs. The results are discussed under following heads;

Amino acid composition:

The physiochemical analysis of amino acids of wsp in drosophila was analyzed by CLC workbench, which revealed the sequence length of all amino acids which is found to be 180-189 amino acids, tabulated in (Table 1). The most abundant amino acids present in this protein were Glycine, Alanine and Valine which are tabulated (Table 2).This abundancy of these amino acids in the protein reveals that the protein is hydrophobic in nature since these amino acids has side chains and has small dipole moments.

Residues of cysteine and tryptophan are absent which predicts the absence of Sulphide Bridge in the protein.

Primary sequence analysis:

The primary sequence analysis predicted by CLC work bench revealed that the *Wolbachia* surface protein are hydrophobic in nature as its percentage is more which ranges between 93-101 (Table 3).

Physiochemical Analysis:

Physiochemical properties of WSP by ProtParam tools are presented in (Table 4)**.** Results show that wsp has a molecular weight of 20533 Daltons, which has highest sequence length of 186 aa (ABD75479.1) and least was found to be 19525 Daltons with its sequence length of 182 aa (AFP86012.1).

Isoelectric point is the pH at which the surface of a protein is covered with charge but net charge of protein is zero. The computed PI value reveals that the protein is basic in nature, due to its least soluble property. Computed isoelectric point of proteins > 7 soluble is acidic buffers. Isoelectric point is predicted which ranges from 4.25 - 5.25 (Table 4).Useful for developing buffer system for purification of proteins and separating the protein on a polyacrylamide gel by isoelectric focusing.

Extinction co-efficient of wsp at 280nm is ranging from 17880-24870M^{-1} Cm^{-}. This infers that the protein can absorb the light at 280nm.The extinction co-efficient can be used to calculate the concentration of a protein in solution.

Stability of wsp was studied by analyzing the values for instability index, aliphatic index and Grand average of hydropathicity (GRAVY) index.

A protein whose instability index is below 40 is predicted as stable, and a value above 40 leads to structural instability. Here the instability index ranges from 16.89 to 27.06 thereby classifying the protein as stable. The aliphatic index refers to the relative volume of a protein that is occupied by aliphatic side chains and contributes to the increased thermo stability of protein. Aliphatic index of wsp was 77.83-87.36 which indicates that the proteins are thermo stable. GRAVY index indicates the solubility of proteins, GRAVY index of wsp was -0.011 to -0.229. A negative GRAVY value for wsp describes it to be hydrophobic in nature. This indicates interaction with water. Though it can play a role in substrate recognition. Here the protein sequences showing negative value indicate the stability of protein. In particular, hydrophobic amino acids can be involved in binding/recognition of ligands.

Secondary Structure Prediction:

SOPMA was employed for calculating the secondary structural features of the selected protein sequences considered in this study. This method calculates the content of α-helix, β-sheets, turns, random coils and extended strands. SOPMA is a neural network based methods; global sequence prediction may be done by this sequence method [22].

High percentage of helices in the structure makes the protein more flexible for folding, which might increase protein interactions. Moreover the predicted secondary structural information of wsp was considered to improve the target-template alignment and for building 3D model of the wsp.

The secondary analysis showed that wsp contain more of random coils and extended strands (range: 20-40%) than alpha sheets and beta sheets. High percentage of random coils and extended strands in the structure makes the protein more flexible and which might increase protein interactions.

Being hydrophobic, glycine prefers to be buried in protein hydrophobic cores. It also shows a preference for being within extended strand more so than in beta strands. The very high coil structural content of wsp is due to the rich content of more flexible glycine and hydrophobic. Proline has a special property of creating links in polypeptide chains and disrupting ordered secondary structure. The consequence in which most of the amino acid side chains of trans membrane segments is non-polar (*e.g. Gly*, Ala, Val ,Asp, Serinr, leu, Phe) and the very polar CO-NH groups (peptide bonds) of the polypeptide backbone of trans membrane segments which participates in hydrogen bonding (H-bonds) in order to lower the cost of transferring them into the hydrocarbon interior. This H-bonding is most easily accomplished with alpha-helices for which all peptide bonds are H-bonded internally. The distribution of amino acids in WSP complex protein (Table 5) shows random distribution of alpha helix, beta turns, extended strands and more distribution of random coils from all these organisms.

Domain Analysis:

SMART Analysis-

Many proteins are multidomain in character and possess multiple functions that often are performed by one or more component domains. A Web-based tool (SMART) has been designed that makes use of mainly public domain information to allow easy and rapid annotation of signaling multidomain proteins. The tool contains several unique aspects, including automatic seed alignment generation, automatic detection of repeated motifs or domains, and a protocol for combining domain predictions from homologous subfamilies. The ability of SMART to annotate single sequences or large datasets is exemplified by the cases described in wsp. Expect value (E) a parameter that counts the number of hits one can "expect" to see by chance for a database of a particular size. It decreases exponentially as the Score (S) of the match increases. Here it is in the range 536, 1640, 2260 (ABD75491.1, ABD75488.1, and AFP86012.1) respectively (Table 6).

SVM prot Analysis-

Support vector machines method for the classification of proteins with diverse sequence distribution. SVMProt shows a certain degree of capability for the classification of distantly related proteins and homologous proteins of different function and thus may be used as a protein function prediction tool that complements sequence alignment methods. It has been employed in protein studies including protein–protein interaction prediction, fold recognition, solvent accessibility and structure prediction. The prediction accuracy ranges from 1.1 to 99% in this study. Thus SVM classification of protein functional family, a potentially developed into a protein function prediction tool to complement methods based on sequence similarity and clustering. In wsp metal binding sites, revealed through this server was Zinc Binding, All Lipid protein and outer membrane protein family was found in only four species of drosophila which was 1.0 is the R value and 58.6 is the P value.(ABD75491.1, ABD75490.1, AFP86012.1 and AFP85988.1) (Table 7).

Based on the Classification of proteins of our interest and its values, we predict that , these proteins may act as drug targets , Metal binding sites, bonding involved ligation and integrated pest management is concerned.

Transmembrane region prediction

The difference between the value of sequence scored and the threshold indicates the possibility of the protein being an outer membrane protein. TMbase is a database of transmembrane proteins and their helical membrane- spanning domains. Possible transmembrane helices, of the accession number (ABD75481.1), the sequence positions inside to outside 2 helices is found and outside to inside 2 helices is found. Transmembrane topology suggestions are purely speculative and should be used with extreme caution since they are based on the assumption that all transmembrane helices have been found. In most cases, the prediction plot (Fig 1) that is created should be used for the topology assignment of unknown proteins.

The sequence positions in brackets dominate the core region. Only scores above 500 are considered significant (Table 8). So, looking at these values we can interpret that they are insignificant because the score is below 500.

These results showed that *wolbachia* surface protein has two trans-membrane domains and it is a cytoplasmic protein having two transmembrane regions (Table 8). This shows that wsp has membrane binding properties and can be involved in transport of materials across the cell membrane.

Sequence Homology Analysis-

The identification of catalytic residues is a key to understanding the function of enzymes. MUSCLE server was used for multiple sequence alignment of WSP of various drosophila species (Fig 2).With the information from other functionally similar sequences with known crystallographic structures, we can identify the key catalytic residues. The compared sequences varied in length but essentially conserved the key catalytic residues which have been highlighted with an asterisk (*) symbol.

Multiple sequence alignment of wsp sequences revealed significant conserved (glutamine, threonine and aspartic acid) and semi conserved regions (valine, alanine and aspartic acid) are represented as strongly similar (:): Weakly similar (.) as shown in (Fig 2). Multiple Sequence alignment is widely accepted method which provides the researchers for strain typing [10].

RNA Structure Prediction

RNA is now appreciated to serve numerous cellular roles, and understanding RNA structure is important for understanding a mechanism of action. This contribution discusses the methods available for predicting RNA structure (Fig 3). Secondary structure is the set of the canonical base pairs, and secondary structure can be accurately determined by comparative sequence analysis. Secondary structure can also be predicted. The most commonly used method is free energy minimization. The free energy of 19 stems are tabulated (Table 9).The accuracy of structure prediction is improved either by using experimental mapping data or by predicting a structure conserved in a set of homologous sequences. Additionally, tertiary structure, the three-dimensional arrangement of atoms, can be modeled with guidance from comparative analysis and experimental techniques. New approaches are also available for predicting tertiary structure.

Tertiary Structure Analysis:

The tertiary structure of the wsp was annotated using Swiss Modell server. The predicted complex structure was observed in Swiss Model which shows steriochemical rotation of torsion angles. The identification of active site amino acids present in WSP complex structure was predicted using Swiss Model. A greater number of variable active sites are present in these organisms and these protein structures can be used for drug binding sites.

SWISS MODELL RESULTS: The 3D structure analysis of wsp were done by using SWISS-MODEL automated modeling server, the three models are shown (Fig 4 & 5). Template selection, alignment and model building are done completely automated by the server of the ID number**.** Predicting the protein 3D structures by this method are used which implements the four steps of the homology modeling approach.

A. Template searching to identify the structure homology: Template search with Blast and HHBlits has been performed against the SWISS-MODEL template library (SMTL, last update: 2015-09-23, last included PDB release: 2015-09-18).

The target sequence was searched with BLAST [1] against the primary amino acid sequence contained in the SMTL. A total of 10 templates were found.

An initial HHblits profile has been built using the procedure outlined in [27], followed by 1 iteration of HHblits against NR20. The obtained profile has then been searched against all profiles of the SMTL. A total of 126 templates were found.

B. Template selection: For each identified template, the template's quality has been predicted from features of the target-template alignment. The templates with the highest quality have then been selected for model building.

C. Model building: Models are built based on the target-template alignment using Promod-II. Coordinates which are conserved between the target and the template are copied from the template to the model. Insertions and deletions are remodeled using a fragment library. Side chains are then rebuilt. Finally, the geometry of the resulting model is regularized by using a force field. In case loop modeling with ProMod-II [9] does not give satisfactory results, an alternative model is built with MODELLER [28].

D. Model quality estimation: The global and per-residue model quality has been assessed using the QMEAN scoring function [4]. For improved performance, weights of the individual QMEAN terms have been trained specifically for SWISS-MODEL.

E. Ligand Modeling

Ligands present in the template structure are transferred by homology to the model when the following criteria are met (Gallo - Casserino, to be published):

(a) The ligands are annotated as biologically relevant in the template library.

(b) The ligand is in contact with the model.

(c) The ligand is not clashing with the protein.

(d) The residues in contact with the ligand are conserved between the target and the template.

If any of these four criteria is not satisfied, a certain ligand will not be included in the model. The model summary includes information on why and which ligand has not been included.

It provides the details of ligand binding sites of wsp in drosophila. The results which were annotated by Swiss model server make the protein modeling accessible to all biochemists and molecular biologists worldwide. A greater number of variable active sites are present in these organisms and these protein structures can be used for drug binding sites.

Conclusion: *Wolbachia* are a diverse group of intracellular bacteria, that show impressive adaptations within invertebrate cells and in manipulating the biology of drosophila. The mechanisms of *Wolbachia*-host interaction are important in understanding of the bacterial life and their pathogenicity as well as generation of new drug targets. In the vinegar fly, genus drosophila, *Wolbachia* causes an egg mortality phenotype known as cytoplasmic incompatibility (CI). CI has been studied in several taxa including *D. simulans* [13, 14, & 12] and *D. melanogaster* [11 & 31]. CI is manifest as severe egg mortality (up to 95%) when an infected male mates with an uninfected female [35]. This property of the drosophila and *Wolbachia* relationship attributes for the researchers to study the parasitic and mutualistic interactions.

Using the Drosophila model system, it should be possible to examine the cues affecting male killing. Annotation of the wsp in drosophila allows the comparative studies into the mechanism of male killing. Beyond this, comparison of the mechanism of *Wolbachia* male killing to that cytoplasmic incompatibility and parthenogenesis induction can be undertaken.

Present interpretation of the sequence and structural analysis of *Wolbachia* surface protein explored the physicochemical nature, three dimensional structure and detail of interactions with wsp and drosophila. Insilico analysis of the wsp in drosophila species provides the information of the mutualistic behavior which allows for further studies and in management of pests. Based on the findings, it could be concluded that further characterization of *wolbachia* in drosophila is novel and will be important for evaluating the functionality of the wsp protein which paves the path in strain typing.

The Biophysical characterization of the protein would provide the clues to its biological functions and physiological role. The conformation of these requires crystal structure and complete functional description to elucidate the effects of non-synonymous substitutions on protein structure and functionality. The comparative analysis will highlight the multiple roles of *Wolbachia* proteins which extend to cytoplasmic incompatibility, feminization of genetic males, parthenogenesis induction and male killing. Further, Structural characterization of WSP will be a break through towards gaining information on *Wolbachia* induced different phenotypes and further this would be used for applied research such as i) Management of (Arthropod) pest that cause major damage in agriculture industry. ii) Control of important vectors that cause the major diseases. iii) Enhancing the fitness and efficacy of Bio control agents. IV) Development of new drug targets [33].

Insilico analysis of wsp in various drosophila species provides the details of the protein's functionality and structural elucidation details which help in the structural based studies. Further, the Insilco predictions of *wolbachia* surface protein complexes will help in revealing the interactions with the host which greatly enhances the hypothesis and experimental investigations.

ACKNOWLEDGEMENT:
My heartful thanks to our Co-ordinator, Prof. Dr. M. S. Reddy, for the support and encouragement.

REFERENCES:

[1] Altschul S.F, Madden T.L, Schaffer A.A, Zhang J, Zhang Z, Miller W, and Lipman D.J. "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs". Nucleic Acids Res, 25, 3389-3402, 1997
[2] Ashokan K.V, Mundaganur D.S, and Mundaganur, Y.D. "Catalase: Phylogenetic characterization to explore protein cluster". Journal of research in Bioinformatics. 1:001-008, 2011
[3] Bazzocchi C, Jamnongluk W, O'Neill S.L, Anderson T.J, Genchi C, and Bandi C. "WSP gene sequences from the Wolbachia of filarial nematodes". Curr. Microbiol. 41:96, 2000
[4] Benkert P, Biasini M, and Schwede T. "Toward the estimation of the absolute quality of individual protein structure models". Bioinformatics, 27, 343-350, 2011
[5] Braig H.R, Zhou W, Dobson S.L, and O'Neill S.L. "Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont Wolbachia pipientis". J. Bacteriol. 180:2373, 1998
[6] Buchon N, Silverman N, and Cherry S. "Immunity in Drosophila melanogaster — from microbial recognition to whole-organism physiology". Nature Reviews Immunology, 14:796–810, 2014
[7] Eisenhaber F, Imperiale F, Argos P and Froemmel C. "Prediction of secondary structural content of proteins from their amino acid composition alone. I. New analytic vector decomposition methods". Proteins: Struct. Funct. Design.25:157-168, 1996
[8] Gill S.C. and Von Hippel P.H. "Calculation of protein extinction coefficients from amino acid sequence data". Anal. Biochem.182: 319-326, 1989
[9] Guex N and Peitsch M.C. "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling". Electrophoresis, 18, 2714-2723, 1997

[10] Guruprasad N. M, Harish B.M, Jalali S.K, and Puttaraju H.P. "Characterization of Wolbachia cell division protein (ftsz) gene for potential management of Uzifly Exorista sorbillans (Diptera: Tachinidae)". Journal of Entomology and Zoology Studies; 3 (2): 57-61, 2015
[11] Hoffmann A.A.A, Clancy D.J, and Merton E. "Cytoplasmic incompatibility in Australian populations of Drosophila melanogaster". Genetics 136: 993–999, 1994
[12] Hoffmann, A, A., and M, Turelli. 1988. Unidirectional incompatibility in Drosophila simulans: inheritance, geographic variation and fitness effects. Genetics 119: 435–444.
[13] Hoffmann A.A, Turelli M, Simmons G.M. "Unidirectional incompatibility between populations of Drosophila simulans". Evolution, 40, 692–701, 1986
[14] Hoffmann A.A, Turelli M, Harshman L.G. "Factors affecting the distribution of cytoplasmic incompatibility in Drosophila simulans". Genetics 126: 933–948, 1990
[15] Hofmann K and Stoffel W. "TM base - A database of membrane spanning proteins segments" Biol. Chem. Hoppe-Seyler **374,**166, 1993
[16] Kitchen D.B, Decornez H, Furr J.R, and Bajorath J. "Docking and scoring in virtual screening for drug discovery: methods and applications". Nature reviews Drug discovery. 3(11):935-949, 2007
[17] Letunic.I, Doerks.T and Bork.P "SMART 6: recent updates and new developments". Nucleic Acids Res. 37: 229/32, 2009
[18] Madhu S and Mahesh P. "Sequence analysis of Semaphorin in tumor progression: An Insilico approach". International Journal of Asian Academic Research of Multidisciplinary .1(29): 407-423, 2015
[19] Mahesh P, Brueckner D, and Reddy M.S. "Horizontal transmission of Wolbachia in the honeybee subspecies Apis mellifera carnica and its ectoparasite Varroa destructor". International journal of environmental sciences. Volume 2, No 2,526-535, 2011
[20] Mahesh P, Divya P, Akshatha M, Prathima R and Lava Kumar C. "Insilico characterization and phylogenetic analysis of novel probiotic bacteria in honey bees". International Journal of Asian Academic Research of Multidisciplinary .1(32): 337-357, 2015
[21] McGraw E.A, and O'Neill S.L. "Wolbachia pipientis: intracellular infection and pathogenesis in Drosophila". Curr Opin Microbiol 7:67–70.3, 2004
[22] Mugilan A, Ajitha M. C, Devi and Thinagar. "In silico Secondary Structure Prediction Method (Kalasalingam University Structure Prediction Method) using Comparative Analysis". Trends in Bioinformatics. 3(1):11-19, 2010
[23] Neyen C, Bretscher A.J, Binggeli O, and Lemaitre B. "Methods to study Drosophila immunity". Methods (San Diego, Calif.) 68:116–128, 2014
[24] Nichols C.D, Becnel J, and Pandey U.B. "Methods to Assay Drosophila Behavior". Journal of Visualized Experiments, 2012
[25] Nunes M.D.S, Nolte V, Schlötterer C. "Nonrandom Wolbachia infection status of Drosophila melanogaster strains with different mtDNA haplotypes". Mol Biol Evol 25: 2493–2498, 2008
[26] Praveen Kumar K.S and Mahesh Pattabhiramaiah "Sequence analysis of basic phospholipase A2 (neurotoxin) as a potential drug target: an in silico approach". International Journal of Engineering Research and General Science Volume 3, Issue. 1057-1067, 2015
[27] Remmert M, Biegert A, Hauser A. and Soding J. "HHBlits: lightning-fast iterative protein sequence searching by HMM-HMM alignment". Nat Methods, 9, 173-175, 2012
[28] Sali A and Blundell T.L. "Comparative protein modeling by satisfaction of spatial restraints". J Mol Biol, 234, 779-815, 1993
[29] Schultz J, Milpetz F, Bork P and Ponting CP. "SMART, a simple modular architecture research tool: identification of signaling domains". Proc. Natl. Acad. Sci. *U.S.A.* 95 (11): 5857/64, 1998
[30] Sokolowski M.B. "Drosophila: Genetics meets behavior". Nature Reviews Genetics 2:879–890, 2001
[31] Solignac M, Vautrin D, Rousset F. "Widespread occurrence of the proteobacteria Wolbachia and partial cytoplasmic incompatibility in Drosophila Melanogaster". Cr Acad Sci Iii-Vie 317: 461–470, 1994
[32] Tsetlin V.I and Hucho F. Snake and snail toxins acting on nicotinic acetylcholine receptors. Fundamental aspects and sciences, Bioinformatics. 4: 53-62, 2004
[33] Uday J, SampathKumar S, Huchesh C.H, Chethana V.C, Puttaraju H.P. "Insilco Analysis of Wolbachia Surface Protein in Wolbachia Endosymbiont of D. Melenogaster". B I O M I R R O R an Open Access Journal. Vol.5: 24-29, 2014
[34] Vishwanath K.V, Mahesh P, and Keerthi R. "Bio computational analysis of protein sequence of sickle cell anemia". International Journal of Engineering Research And Generic Science. Vol 1: Issue 1: 63-73, 2015
[35] Werren J.H. "Biology of Wolbachia". Ann Rev Ent 42: 587–609, 1997
[36] **Werren J. H, Windsor D, and Guo L.R**. "Distribution of Wolbachia among neotropica arthropods". Proc. R. Soc. Lond. B Biol. Sci. 262**:**197-204, 1995