Comparative Genome Analysis and Localization of Hypothetical Conserved Plasmodial Proteins

THESIS

Submitted in partial fulfilment
of the requirements for the degree of
DOCTOR OF PHILOSOPHY

by

ISHA PANDEY

Under the Supervision of

Prof. Ashis Kumar Das

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

2017

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

PILANI, RAJASTHAN, INDIA

CERTIFICATE

This is to certify that the thesis entitled **"Comparative Genome Analysis and Localization of Hypothetical Conserved Plasmodial Proteins"** submitted by Isha Pandey, ID. No. 2010PHXF0413P for award of Ph.D. Degree of the institute embodies original work done by her under my supervision.

Signature (Supervisor) :

Name (Supervisor) : Ashis Kumar Das, PhD

Designation : Professor
Department of Biological Sciences
BITS, Pilani, Pilani Campus

Date :

Abstract

In humans, malaria is caused by six species of *Plasmodium* of which *Plasmodium falciparum* is considered the most virulent and is responsible for much of the mortality and morbidity. The parasite has a complex lifecycle due to which it is challenging to combat malaria. The major cause of the pathogenicity of the parasite is its inherent property of getting acclimatized to the host environment by expressing virulent genes at different time-points in its asexual and sexual stages of the life-cycle, the other reasons could be the genomic variations, like insertions and deletions of the small and large segment of the chromosomes, inversions, single point mutations and allelic variations that can affect gene and protein expression. It has been reported that the host environment is entirely different from the *in-vitro* culture environment, leading to the differences in gene and protein expression.

The study reports an investigation of genomic differences between parasite isolates causing complicated malaria or uncomplicated malaria from the same geographical location. For this purpose, a tiling array-based Comparative Genomic Hybridization approach using 2 x400k Agilent platform was used. The platform contains approximately 418,577, 60 mer custom-designed probes with average probe spacing of 56 bp. Since none of the isolates were sequenced, the probes were designed using genome assembly of *Plasmodium falciparum* 3D7 isolate. The regions showing genomic variations have been seen in the telomeric and sub-telomeric regions of the chromosome, mostly representing the var genes. Major variations have also been seen in chromosomal region other than sub-telomeric regions; where a continuous amplification of 15.90 kb is present in the chromosome-12 that cover 6 genes, one of which is GTP- cyclohydrolase-1 (GCH 1), the 1st rate-limiting enzyme of the folate biosynthesis pathway. Other than the variations in the coding regions, we have seen variations in intergenic regions also. However, the implications of amplifications and deletions in the coding regions of the parasite isolate causing complicated disease remaining unanswerable at this level of the study.

Another aspect of the study involves the cloning and expression of the proteins (hub genes) that were extracted from a co-expression based differentially regulated systems network created from parasite RNA extracted from isolates from patients showing uncomplicated or complicated disease. One gene from the up-regulated module

(PFD0300W) and one gene from the steady-state module (PFL2320W) was selected as a hub genes from each of the module. The modules used for the extraction of the genes were enriched in the translational biological process and oxidation and reduction process, which are important for the parasite survival. Due to the presence of rare codon in the aforementioned genes, the gene sequences were optimized according to *E. coli* heterologus machinery that was used for the expression study of the proteins, however, even after codon optimization of the genes, the protein expression was not achieved. In-silico analysis of the proteins suggests that the proteins are disordered in nature, and does not contain any of the signal peptide sequences or transmembrane domains. The intrinsically disordered nature of the proteins possibly makes it important for the parasite, a fact that emphasizes the point that both are hub genes. It has been reported that the important proteins of the *Plasmodium* parasite like pre-erythrocytic antigens (CSP, LSA, TRAP) and erythrocytic antigens (MSP1, MSP2, MSP3, EBA175, AMA1) are intrinsically disordered proteins and are leading vaccine candidates.

Another facet of the study was the localization of the two proteins in the parasite. Since the expression of the total protein was not achieved in an *E. coli* based system, the B-cell epitopes were predicted for both the protein for localization purposes and anti-sera raised in mice against peptides. The localization study of one of the protein PFD0300W suggests the cytoplasmic localization in all the asexual and sexual stages of the parasites, whereas for the other protein PFL2320W localization was not achieved with the available reagents although the presence of protein was detected.

In conclusion, by exploring the genomic variation in the parasite isolate causing complicated manifestation in comparison with the parasite isolate causing uncomplicated manifestation, in addition to those already reported, we have been able to detect some putative and novel variations. However, the differences in the parasite isolates could also be inherently natural differences between two parasite isolates without any direct correlation with the disease conditions.

The in-silico analysis of the selected proteins suggests the disorderd nature of the proteins. Previous reports on *Plasmodium* pre-erythrocytic antigens have also reported the importance of disordered proteins as potential vaccine candidates. The peptide based localization of the PFD0300W in the cytoplasm of the parasite in both sexual and asexual stages could make it a good drug target for future intervention studies.

Declaration

I hereby declare that this Ph.D. thesis entitled "**Comparative Genome Analysis and Localization of Hypothetical Conserved Plasmodial Proteins"** is my original work which was carried out for the degree of Doctor of Philosophy under the supervision of Prof. Ashis. K. Das at Birla Institute of Technology and Science, Pilani, Pilani campus, India. I have duly acknowledged all the sources of information used in the work presented in the thesis.

Signature:

Name: Isha Pandey

Date:

Dedicated to all those who lost their life because of malaria and my family members

Acknowledgements

Happiness does not come from doing easy work, but from the after glow of satisfaction that comes after the achievement of a difficult task that demanded our best. By the grace of the almighty, i am able to successfully achieve one part of my life in the form of Ph.D., which would have never been completed without the continuous support of wonderful people in my life.

"A mentor and a teacher burn himself like a candle to lighten up the student life", I would like to express my sincere gratitude to my mentor and supervisor ***Prof. Ashis Kumar Das,*** *Department of Biological sciences, who is a passionate scientist with strong imagination, for his continuous support and encouragement during my Ph.D study, for his patience, motivation and immense knowledge. I also thank Prof. Das for believing in me and guiding me to improve weaker areas, I am thankful to Prof. Das for offering me an opportunity by accepting me as his PhD student.*

I am thankful to ***Prof. Vishal Saxena*** *and* ***Dr. Shilpi Garg****, Members of my doctorol Advisory Committee (DAC) for their insightful comments and encouragement, but also for being critical towards my experiments, which incented me to widen my research from various perspectives.*

My sincere thanks to ***Prof. Souvik Bhattacharyya****, Vice Chancellor BITS Pilani,* ***Prof. Ashok Kumar Sarkar****, Director BITS Pilani,* ***Prof. S. K, Verma,*** *Dean Academic Research Division,* ***Prof. Hemant R Jadv****, Associate Dean, Academic and Research Division, and nucleus members of BITS Pilani, Pilani Campus for Academic, Administrative supports.*

The thesis would not have been successfully completed without the close collaboration of the Sardar Patel Medical College, Bikaner. I am thankful to ***Prof. Dhanpat K Kochar****, Former Head, Cerebral Malaria Research Center, S.P. Medical College, Bikaner and Visiting Professor-Medical Research, Rajasthan University Health Sciences, Jaipur, and* ***Prof. Sanjay Kumar Kochar,*** *head of Respiratory division; Malaria research centre, Department of medicine, S.P medical college. Bikaner for providing invaluable malaria infected blood samples.*

I am grateful to the Head of the Department of biological sciences, ***Prof. Prabhath Nath Jha,*** *DRC convener* ***Prof. Jitendra Panwar****, for helping me throughout by encouraging and solving all my academic related problems. I would like to thank all the faculty members of the Department of Biological sciences, BITS Pilani, Pilani campus, for being helpful in every situation. I would also like to thank staff of the Department*

of Biological sciences ***Subhash ji, Kamlesh soni ji, Mukesh bhaiya, Naresh bhaiya, Ajay ji, Parmeshwar ji*** *and* ***Mahendar ji*** *for their help in my experiments and office work.*

I express my gratitude to ***Prof. Agam Pratap Singh****, one of the immense scientists of the National Institute of Immunology, for his time, support and patience in helping me to properly analyse and localize the proteins in parasite slides.*

I would also like to acknowledge ***Dr. Neena valecha (Director NIMR), Dr. Abhinav sinha, Dr. C. R. Pillai*** *and the parasite culture facility of NIMR (National Institute of Malaria Research) for providing paraformaldehyde fixed slide of P. falciparum culture strain* ***RKL-9****, used for immunolocalization studies for both the proteins.*

I would also like to thank ***Dr. Asif Mohammad ICGEB, New Delhi*** *for providing E.R (Endoplasmic reticulum) tracker.*

I would like to express my special gratitude to ***Dr. Sushil Yadav****, in charge of animal facility BITS Pilani, Pilani campus, for providing mice and helping me in animal handling, thanks to staff of animal facility for taking care of animals and their timely help.*

It is said there should be a hand always on your head, after the supervisor if some people with whom you can share all your problems and get inspired, are seniors, I would like to thanks my laboratory seniors, ***Dr. Narayan Kumar Jalan, Dr. Deepak Paklapati, Dr. Amit kumar Subudhi, Dr. P.A. Bhoopathi****, for there critical comments and for teaching me work ethics*

I would also like to thank, non laboratory seniors ***Dr. Prakash, Dr. Sachi, Dr. Navin Jain, Dr. Purva*** *and* ***Dr. Garima*** *for their continuous help.*

It is hard to find a group of people, who challenge you, inspire you, spend lot of time with them and it will change your life, on a similar note i would like to thank the people with whom i met in Pilani and without whom my journey in Pilani would be really tough, thanks to ***Dr.Arpit Bhargava, Dr. Kuldeep Gupta and Dr. Gagandeep Singh Saggu,*** *more than my seniors they were like pillars, they helped me in every way be it emotional or in experiments, my BITS family* ***Dr. Gurpreet Kaur Sidhu, Dr. Zana Pala, Panchsheela Nogia, Poonam Singh, Sandeep, Vikram Pareek, Shradha Mishra*** *and* ***Nidhi Bub*** *they always stood beside me.*

Thanks to my colleagues and friends ***Rini Dhawan, Shobha, Monica. M Jhangir, , Vidushi, Manohar, Jyothi, Shubhra, Leena Frageria, Vikas Lamba, Tania Pal Chaudhary, Pareek Kakani, Parva Sharma, Rajneesh, Shahid, Pinki Dhatrewal, Zaiba Khan, Heena Saini, Astha, Neelam****.*

I would like to thank to the people who brings smile on my face even in toughest situation; Thanks to ***Ritika Lakhotia, Naveen Singh, Saif Ismail, Surabhya Balasubramanium, Akshat Mishra, Deeptanshu Agarwal, Imran Meghani, Surbhit, Sai bhardwaj and Anand.***

I am grateful to ***Ramandeep Kaur****, who is more than a junior, who always makes me laugh, argued with me, lighten my mood and helped me in my experiments and was my companion during the half way of my Ph.D tenure.*

Last but never the least, I would like to thank ***my family, my parents,*** *who supported me and allowed me to stay away for so long and letting me to achieve my goal.*

Words are not enough to thank this one person my husband ***Vineet Malaviya****, for encouraging me even in my failures, for believing in me and for standing beside me even after being so far.*

Isha Pandey

Contents		
S.No	**Description**	**Page No**

List of Tables

List of Figures

Abbreviations

No.	Caption
µg	Microgram
µl	Microlitre
ADM-2	Aberration detection method II
AMA	Apical Membrane Antigen
AMADID	Agilent Microarray Design ID
APS	Ammonium persulphate
AQ	Autoclave mili-Q water
ARDS	Acute Respiratory distress syndrome
ARF	Acute renal failure
ATP	Adenosine Tri phosphate
BC	Before Christ
BCP	Bromochloropropane
BPB	Bromophenol blue
BSA	Bovine Serum Albumin
CAI	Codon Adaptation index
CBB	Comassie Briliant Blue
CDD	Conserved Domain database
CGH	Comparative Genomic Hybridization
CNV	Copy Number Variation
CSP	Circumsporozoite protein
CSF	Cerebral spinal fluid
DAPI	4',6-Diamidino-2-Phenylindole
DEPC	Di-ethyl pyrocarbonate
dL	DeciLitre
DNA	Deoxyribonucleic acid
EBL	Erythrocyte Binding like protein
EDTA	Ethylene Diamine tetra acetic acid
ELISA	Enzyme linked immunosorbant assay
EPCR	Endothelial protein C receptor
ER	Endoplasmic Reticulum
GAA	Glacial Acetic Acid
GCH-1	GTP-cyclohydrolase-1
GEO	Gene expression omnibus
gm	Gram
GPA	Glycophorin Binding Antigen
GTP	Glycosyl Tri Phosphate
Hb	Haemoglobin

HCL	Hydrochloric acid
HIS	Histidine tag
HRP	Horse radish peroxidase
HSP	Heat Shock Protein
HSPGS	Heparan Sulphate Proteoglycans
ICAM-1	Intracellular Adhesion Molecule-1
IMTECH	Institute of Microbial Technology
IPTG	Iso Propyl ThioGalactoside
I-TASSER	Iterative Threading ASSEmbly Refinement.
IU-PRED	Intrinsically Unstructured protein Prediction
JAU	Jaundice
Kb	Kilobase pair
Kcl	Potassium chloride
Kg	Kilo gram
LB	Lauria Bertni
MB	Mega Basepair
Mg	Milligram
Ml	Millilitre
mmHg	Millimeter of mercury
mmol	MilliMole
MSP	Merozoite surface protein
Mt	Mitochondria
N	Normality
NA	Not available
NaCl	Sodium Chloride
NaOH	Sodium hydroxide
NCAM-1	Neural Cell Adhesion Molecule
NCBI	National Centre of Bioinformatics
NEB	New England Biolabs
nm	Nano meter
NTA	Nickel N-TA
OD	Optical Density
PAGE	Poly-Acrylamide Gel Electrophoresis
PBMC	Peripheral Blood Mononuclear Cell
PBS	Phosphate Buffered Saline
PBST	Phosphate Buffer Saline-Tween-20
PCR	Polymerase Chain Reaction
PF-C	*Plasmodium falciparum* Complicated
PfEMP-1	*Plasmodium falciparum* Erythrocyte Memberane protein-1
PHYRE	Protein Homology/analogY Recognition Engine

PI	Propidium Iodide
Plasmo DB	*Plasmodium* Database
PONDR	Predictor Of Natural Disordered Region
PTEX	*Plasmodium* Translocon Exported proteins
PV	Parasitophorus Vacuole
PVM	Parasitophorus Vacuole Memberane
q.s.	Quantiity sufficient
RBC	Red Blood Cells
RF	Renal Failure
RNA	Ribonucleic Acid
RON	Rhopteries protein
RPM	Revolution Per Minute
RT	Reverse Transcriptase
RT	Room Temperature
SDS	Sodium Do-decyl Sulphate
SMA	Severe Malaria Anaemia
SNP	Single Nucleotide Polymorphism
TBS	Tris-Buffer Saline
TBST	Tris-Buffer Saline- Tween 20
TE	Tris and EDTA
TGM	Tris-Glycine Methanol
TGS	Tris Glycine SDS
THR	Thrombocytopenia
TMHMM	Transmembrane Helices; Hidden Markov Model
TRAP	Thrombospondin related Anonymous Protein
UCSC	University of South California
UP	Up-regulated
UV	Ultra Violet
VSA	Variant Surface Antigen
WGCNA	Weighted Gene Co-expression Network Analysis
WHO	World Health Organization

Keywords

S.No	Keywords
1	*Plasmodium falciparum*
2	Complicated
3	Uncomplicated
4	Malaria
5	Field isolates
6	Array CGH
7	Tiling array
8	Whole genome
9	Indian isolates
10	Conserved *Plasmodium* protein
11	Comparative Genomic Hybridization
12	Protein expression
13	Codon optimization
14	RARE Codon
15	PFD0300W
16	PFL2320W
17	Immunolocalization
18	RKL-9
19	Cytoplasm
20	Apicoplast

Chapter I

Introduction and Review of Literature

1.1- Malaria: Historical Perspective:

In the present day world, an appreciable amount of mortality and morbidity is caused by vector borne infectious diseases. Prominent among these are Dengue, Chikungunya, yellow fever, lymphatic filariasis, Leishmaniasis, Lyme disease, Malaria, Chagas disease, etc. In spite of much modern scientific advancement, malaria continues to remain among the very important vector borne diseases and hence has been studied from ancient times. Malaria symptoms were discussed rigorously in a Chinese document from about 2700BC, Clay Tablets from Mesopotamia from 2000 BC, Egyptian papyri from 1570 BC and Hindu texts from the sixth century (Reviewed in Francis EG Cox, 2010). The word malaria derives from the Italian word mal'aria or malicious air, this name was originally given because it was thought that this disease is caused by noxious air. Later, after the discovery of bacteria by Antoni van Leeuwenhoek in 1676 and the development of the germ theory by Louis Pasteur and Robert Koch in 1878 -1879 it was implied that microorganisms are the cause of infectious diseases. With this development, the search for the cause of malaria intensified. The first discovery of the malaria parasite was done

by Charles Louis Alphonso Laveran when he discovered a pigment in malaria patients with an enlarged spleen. Laveran observed the crescentic bodies in all the malaria cases and not in the patients without malaria, he also noted that quinine removed these stages from the blood, and a breakthrough resulted with the discovery of a parasitic protozoan which he called *Osciallaria malarie.* By 1890 it was known that malaria is caused by protozoan parasites that invade and multiply in red blood cells. Later there were three species discovered with specific periodicities which differed in specific characteristics and were considered to be responsible for benign tertian (*Haemamoeba vivax*), malignant tertian (*Laverania malariae*), and quartan malaria (*Haemoamoeba malariae*) which are now known as *P. vivax, P. falciparum* and *P. malariae*. (Hempelmann, & Krafts., 2013), subsequently, in 1897, the *Anopheles* mosquito as a vector for malaria was discovered by Ronald Ross.

1.2-*Plasmodium* species causing malaria in humans:

Malaria in humans is caused by 4 different *Plasmodium* species, namely *P. falciparum, P. vivax, P. malariae,* and *P. ovale*. Of these *P. falciparum* and *P. vivax,* are the major cause of morbidity and mortality. These *Plasmodium* species are thought to have originated from different families of Apes, viz, the origin of *P. falciparum* is from gorilla (Liu, et al., 2010) and *P. vivax* is of African origin where infection is mostly found in Gorillas and Chimpanzees (Liu, et al., 2014). In recent years, it has been found that two of the monkey malaria parasites can also be transmitted to humans; these are *P. knowlesi* and *P. cynomolgi*. However, the transmission by *P. cynomolgi* between humans has not yet been observed. Naturally acquired human infection by *P. knowlesi* was thought to be extremely rare until 2004, since then human cases have been documented in Southeast Asian countries and is now recognized as a fifth human malaria parasite (Singh & Daneshvar., 2013). Although these species are originated from apes and found to infect a human host, they are distinctly different in their characteristics of infection (**table 1.1**).

1.3-Global status of malaria:

According to the world malaria report 2017, an estimated 216 million new malaria cases were reported worldwide, ranging from 148-304 million of which the African region accounted for most of the global cases of malaria (90%), followed by the Southeast Asian region (7%) and the Eastern Mediterranean region (2%). It has also been reported that between 2010-2015, incidence of new malaria cases decreased by 21% globally as well as in the African regions. During the same period, malaria mortality rate decreased by an estimated 29% globally and by 31% in the African regions (**Figure 1.1**). In spite of a reduction in the mortality and the morbidity threat of malaria still, persists due to insecticide resistance among the vector mosquitoes and drug resistance in parasites.

Characteristics	***P. falciparum***	***P. knowelesi***	***P. malarie***	***P. ovale***	***P. vivax***
Pre-Erythrocytic stage (days)	5-7	8-9	14-16	9	6-8
Pre-Patent Period (days)	9-10	9-12	15-16	10-14	11-13
Erythrocytic cycle (days)	48	24	72	50	48
Erythrocyte infection	Immature and mature RBC	Immature and mature RBC	Mature erythrocytes	Reticulocytes	Reticulocytes
Febrile paroxysm (hours)	16-36 or longer	8-12	8-10	8-12	8-12
Severe malaria	Yes	Yes	No	No	Yes
Relapses from liver forms	No	No	No	Yes	Yes
Recurrences	Yes(treatment failure)	Yes	Yes	No	Yes (treatment failure)

Table 1.1: Characteristics of infection of five *Plasmodium* species infecting humans. (Antinori, et al., 2012)

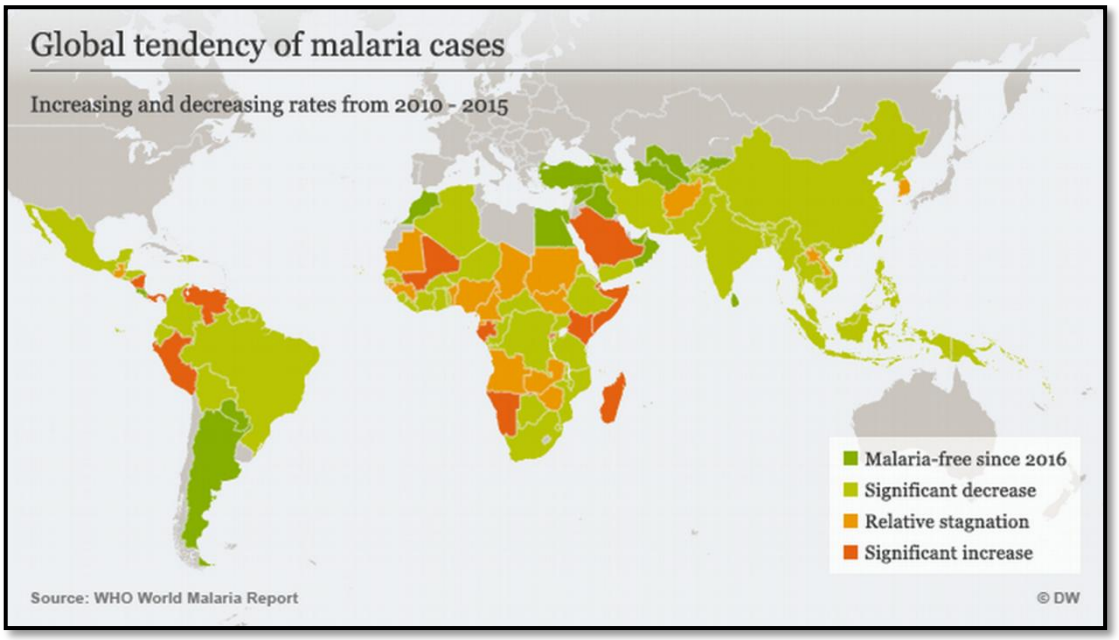


Figure 1.1: Schematic representation of the Global Status of malaria depicting increasing and decreasing rates from 2010-2015, (WHO, World Malaria Report, 2016)

1.4- ***Plasmodium falciparum*** **as the most virulent human malaria parasites:**

P. falciparum is usually considered to be the most virulent among all the five species of the *Plasmodium* parasite infecting humans, although *P. vivax* has also been reported to cause disease complication (Kochar, et al., 2005; Pinzón, et al., 2013; Aashish & Manigandan, 2015). India is a populous country and most of the population lives in the area of higher transmission of malaria infection suffering from infection of both *P. falciparum* and *P. vivax* and in the Indian scenario, approximately 63% of malaria cases is caused by *P. falciparum* or mixed. However, *P. vivax* accounts for only 35% of cases in India (WHO, world malaria report 2017) **(Figure-1.2).**

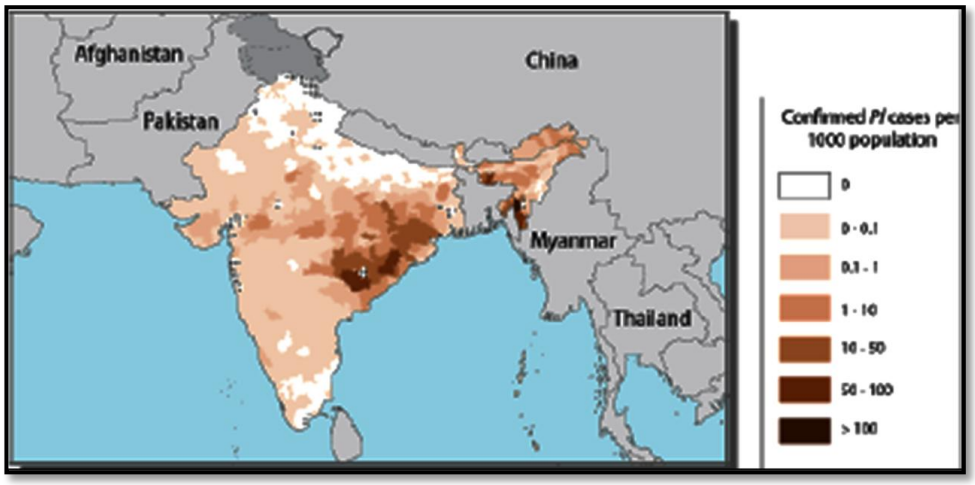


Figure:1.2-Schematic representation showing, the prevalence of *P. falciparum* infection in different regions of India, the colour representation is as follows, increase in the intensity of colour representing increasing *P. falciparum* infection. The white colour represents 0 or states without any *P. falciparum* infection, the dark brown colour represents states with >100 *P. falciparum* infection, (WHO, World Malaria Report, 2016)

The majority of mortality is due to *P. falciparum* infection because it can infect a higher percentage of the red blood cells not being preferential to reticulocytes like *P. vivax*. The erythrocytes infected with the *P. falciparum* can interact with different host endothelial cell receptors due to surface protein expressed on the infected RBC. This is the major cause for pathogenesis seen in human malaria. The infected parasite during its life cycle in host expresses different adhesion molecules on the surface of infected erythrocytes to escape from the host immune system, which leads to the complicacy of the disease.

1.5- Lifecycle of *Plasmodium falciparum:*

P. falciparum completes its life cycle in two hosts: a female *Anopheles* mosquito, which serve as a vector and humans. The parasite completes its asexual part in human host which is further divided into i.e Exo-erythrocytic stage (Liver) and erythrocytic stage (blood) while the sexual stages which, though present in the blood of the vertebrate host exflagellates, fuses, and develops to the infected sporozoite stage in the mosquito (**Figure 1. 3).**

1.5.1-Exo-erythrocytic stage:

The asexual life cycle of *P. falciparum* begins with the bite of the infected female *Anopheles* mosquito during its blood meal from the vertebrate host. While taking up the blood, the female *Anopheles* mosquito releases approximately 15-120 sporozoites along with the saliva (containing an anticoagulant) into the host. Most of the sporozoites are injected into the dermis of the skin and not directly to the circulatory system. The sporozoites move through the dermis by gliding movement till they come in contact with the blood vessels and move into the circulatory system, which allows them to travel to the liver (Vanderberg & Fervert, 2004; Amino, et al., 2006). Once sporozoites reach into the liver of the host, HSPGS (Heparin Sulfate Proteoglycans) present on the surface of hepatocytes are the main receptors responsible for the attachment of sporozoites. HSPGs interacts with the CSP (Circum Sporozoite Protein) present on the surface of sporozoites, which eventually leads to the sporozoite sequestration (Prudêncio, et al., 2006). HSPGs have an important role in binding and internalizing sporozoites as they interact with both CSP and TRAP (Thrombospondin Related Anonymous Protein) (Jethwaney, et al., 2005). Following sequestration and migration sporozoites leads to the final invasion, with the formation of the parasitophorous vacuole. These sporozoites multiply inside the hepatocytes and generate thousands of merozoites, which are released by the formation of merozoite-filled vesicles called merosomes that bud off from the infected hepatocytes into the lumen of the liver sinusoids (Sturm, et al., 2006). The merosomes plasma membrane are of host cell origin that is why it is not recognized by host dendritic cells and kupffer cells and migrate safely to the host bloodstream. In *P. vivax and P. ovale,* sporozoites may not follow the same pattern and remain in a dormant state (hypnozoite) in the liver that may activate after a long time leading to the relapses. The exoerythrocytic stage of the parasite is non-pathogenic and does not cause any symptoms.

1.5.2-Erythrocytic phase:

Merozoites once released into the bloodstream invades the red blood cells (RBC's), this is the start of the erythrocytic phase of the life cycle; the first stage of invasion is the ring stage that evolves into a trophozoite. The trophozoites are not directly able to digest the haem, haem is toxic so it is converted into hemozoin and parasites digest the globin which is used as a source of amino acid for their reproduction (Jani, et al., 2008). The

next stage is the erythrocytic schizont. Each of the mature schizonts gives birth to the new merozoites, After RBC's ruptures merozoites are released to infect other RBC's this is when parasitemia increases and clinical manifestations appear, the exoerythrocytic stage occurs only once where erythrocytic phase occurs multiple times.

1.5.3-Gametocytic Stage:

Gametocytogenesis begins when some of the merozoites instead of following asexual replication, develops into micro (male) and macro (female) gametocyte in red blood cells. Infected erythrocytes in fine capillaries of bone marrow only develop into the gametocyte (Paul, 2000). Over 1-2 weeks the parasite develops into (I-V) distinct stages in *P. falciparum* gametocytogenesis. The mosquito takes up only the mature stage-V of gametocyte that circulates within the host bloodstream, making them available for the mosquito, with the other stages (I-IV) sequestering away from the peripheral circulation (Gardiner, et al., 2015). When mature gametocytes are ingested by the mosquito the transformation of the gametocytes to the gametes occur and fertilization of macro and microgamete takes place in the lumen of the mosquito midgut. Within 24 hrs, the non-motile zygote differentiates into a motile ookinete that penetrates the midgut wall of the mosquito, where it encysts to form an oocyst. Inside the oocyst, the ookinete nucleus divides to produce thousands of sporozoites. Oocysts rupture and release sporozoites, only a few hundreds of the sporozoites are able to enter into the salivary gland of the mosquito. When again the infected mosquito bites a healthy host the life cycle of parasites starts (Perlmann & Troye-Blomberg, 2002).

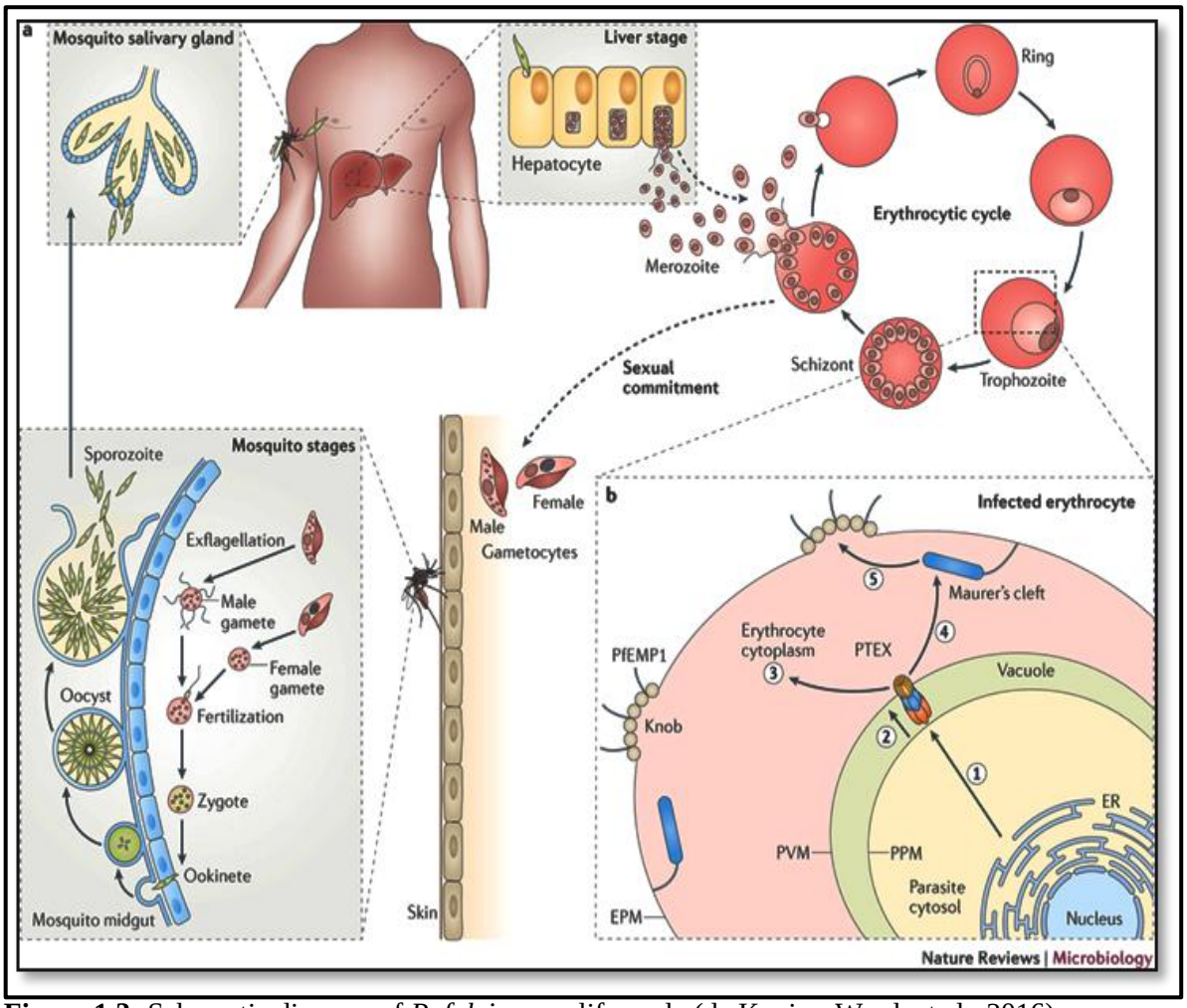


Figure 1.3: Schematic diagram of *P. falciparum* life cycle (de Koning-Ward, et al., 2016)

1.6-Cellular mechanism of *Plasmodium falciparum* invasion in erythrocytes:

The invasion mechanism of parasite starts with the egress of the merozoite, which is the smallest cell in the parasite life cycle and the motile life cycle stage of the parasite. The biology of merozoite is exquisitely adapted for the invasion of erythrocytes, it is composed of several merozoite surface proteins and proteins in secretory organelle each of which play an essential role in the successful invasion of erythrocytes (Patarroyo, et al., 2008). The merozoite possesses an apical complex of secretory organelles (microneme, rhoptries and dense granules), mitochondrial, apicoplast and nucleus. The invasion process is divided into four distinct phases. Adhesion, Reorientation, Tight junction formation, Ingress **Figure: 1.4**

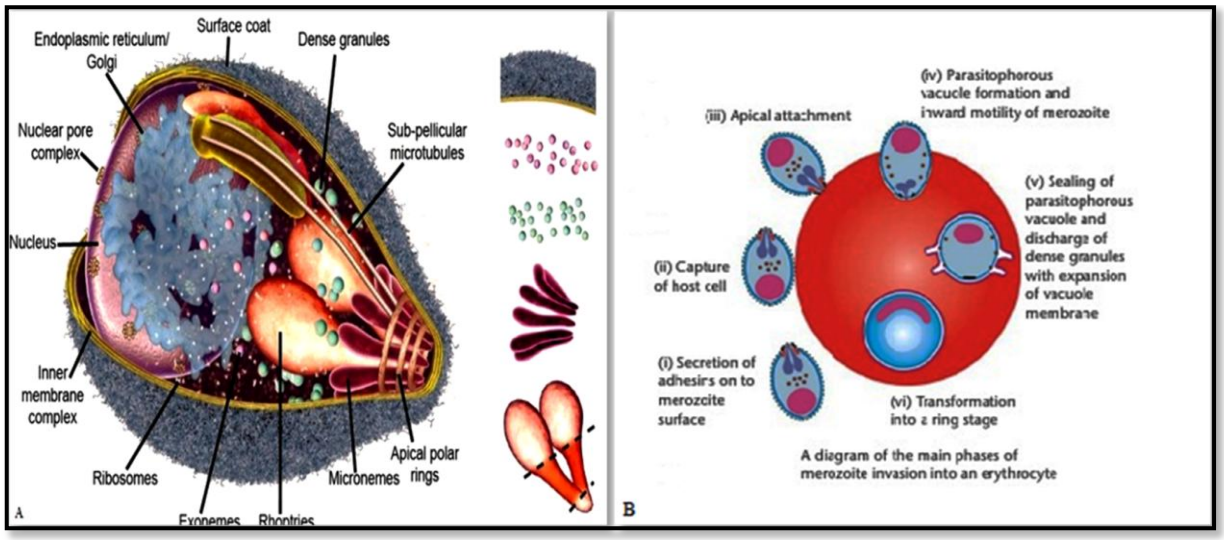


Figure-1.4: Merozoite sectioned structure and mechanism of invasion: A) A sectioned Structure of invasive merozoite highlighting important organelles along with specific surface proteins expressed on the surface of organelles that are important for parasite invasion.(Cowman, et al; 2012) (B) Mechanism of merozoite invasion into erythrocytes.(Bannister & Sherman, 2009).

1.6.1-Adhesion and reorientation:

After the egress of merozoites from infected hepatocytes, these merozoites migrate with the flow of blood as they lack any of the motile filaments like flagella and cilia. It takes about 5s for merozoites to make contact with erythrocytes. The adhesion of merozoites to erythrocytes takes place with the help of merozoite surface proteins (MSPs) that bind to the erythrocyte plasma membrane receptors and cause parasite adhesion to the host cell surface. MSP-1 is a Glycosylphosphatidyl inisitol binding protein, considered an essential parasite surface molecule that is responsible for adhesion of merozoites to the surface of erythrocytes (Baldwin, et al., 2015) showed that MSP-1 plays a crucial role in the invasion of merozoites to erythrocytes by binding of an N - terminal subunit of MSP-1 to a specific polypeptide region of Glycophorin binding antigen (GPA) during the invasion of human RBC. In second phase the process of reorientation takes place during which merozoites orient their apical complex towards the erythrocytes, In this process the type-1 membrane proteins are involved, like Erythrocyte binding-like proteins (EBLs) and reticulocyte binding proteins (PfRhS). These proteins are receptor-specific proteins, in which EBL proteins bind with specific receptors, including Glycophorin A, B, C and PfRh protein bind to complement receptor-1 and basigin present on the surface of the erythrocyte (Mohandas and An, 2012).The third phase involves tight junction formation where AMA-1(Apical Membrane protein-1), a microneme protein forms a complex with RON (Rhoptries

protein) forming link between parasite and erythrocyte. Formation of junction releases the rhoptry bulb that provides the proteins and nutrients required for the formation of the parasitophorous membrane and the parasitophorous vacuole. This makes space for merozoite to easily move and invade, the erythrocyte and finally, the invasion of merozoite into erythrocytes takes place.

1.7-Classification of malaria, according to clinical manifestation:

Based on the clinical features, infection by malaria parasites is broadly classified into two types of disease states. Uncomplicated malaria: which is the usual stage of infection of *Plasmodium* in the host. The complicacy of infection depends upon the host immunity, drug resistance and multiplicity of infection, which eventually leads to the seriousness of the disease and is categorized in complicated malaria based on clinical symptoms like severe anaemia, renal failure, jaundice, cerebral malaria etc. All these features are described in detail in subsequent paragraphs.

1.7.1-Uncomplicated malaria:

The clinical manifestations of malaria depend upon the hosts previous immune history and the species of parasite which is causing the infection. The first symptoms of malaria are common in all malarial parasites, these symptoms are non-specific and mimic flu-like syndrome. The patients suffering from malaria are at first diagnosed with fever, up to two days before the fever, prodromal symptoms, such as malagia, anorexia, lassitude, dizziness, backache, nausea, vomiting and sense of chilliness may be experienced. The fever is usually irregular at first and suddenly temperature rises with shivering and mild chills. After some day fever tends to become periodic depending upon the synchronized schizogony (Bartoloni & Zammarch, 2012).

1.7.2- Severe malaria/Complicated malaria caused by *Plasmodium falciparum*:

Most of the severe form of malaria is caused in the people with the weak immunity, that involves central nervous system (cerebral malaria), pulmonary symptoms (respiratory failure), renal system (acute renal failure) and hematopoietic systems (severe anaemia), jaundice and hepatic dysfunction. Much of these severe malaria symptoms are seen with *P. falciparum* infection. However, in last decade, reports of *P. vivax* causing complicated disease have also been documented (Kochar, et al., 2005; Kochar, et al., 2009; Price, et al., 2009; Howes, et al., 2016)

1.7.2.1-Cerebral malaria:

The term cerebral malaria has been restricted to the syndrome in which there is altered consciousness, associated with malaria infection. A child with loss of consciousness after febrile convulsions should not be considered to have cerebral malaria unless coma persists for more than 1 hour after the convulsion. According to the WHO definition, Glasgow coma scale for cerebral malaria should be less than 11 (Glasgow coma scale<11, Blantyre coma scale<3). A patient suffering from cerebral malaria may exhibit various forms of abnormal posturing like decerebrate rigidity (extensor posturing with arms and legs extended), decorticate rigidity extensor decorticate rigidity (extensor posturing, with arms flexed and legs extended), and opisthotonus (Warrell, 1997).

1.7.2.2-Acute respiratory distress syndrome:

This symptom is mostly seen in the patient with *P. falciparum* infection, with a 70% mortality rate. The most severe presentation of pulmonary dysfunction is Acute Respiratory Distress Syndrome (ARDS), characterized by diffuse endothelial injury with increased capillary permeability. Clinical features consist of sudden dyspnea, coughing and severe hypoxia, and is frequently accompanied by agitation and disorientation. (Mohan et al., 2008; Mazhar & Haider, 2016).

1.7.2.3-Acute renal failure:

The effect of infected red blood cells on the microcirculation of the patient is the main cause of the Acute Renal Failure. Changes in hemodynamic and immunological status increase creatinine and blood urea nitrogen levels, these metabolic changes can cause ARF. There are reports of presence of iron positive pigment in the epithelium of tubules and interstitial cells, suggesting that hyper parasitization leads to the release of hemozoin pigment, which eventually leads to the release of inflammatory mediators causing damage to the structure and function of the organ like kidney (Nayak, et al., 2014).

1.7.2.4-Severe Malaria Anemia (SMA):

Many factors may cause severe anaemia, including erythrocyte sequestration, erythropoietic changes, and blood loss. Severe anaemia is defined as hematocrit below 15% or a haemoglobin concentration below 5g/Dl. Low Hb level in SA is due to intravascular and extravascular haemolysis and reduced erythropoietic response. Previous studies have suggested that several chemokine and cytokines reduced the level of HSP-70

and Gln level in patients suffering from SMA it also suggests that the suppression of HSP-70 due to *P. falciparum* hemozoin suppressed by glutamine supplementation that can lead to SMA (Lamikanra, et al., 2015, Kempaiah, et al., 2016).

1.7.2.5-Jaundice and hepatic dysfunction:

It is one of the common severe manifestations of *P. falciparum* malaria. Jaundice in *falciparum* malaria could be due to intravascular haemolysis of pRBCs, glucose 6 phosphate dehydrogenase deficiency, drug-induced hemolysis etc. Laboratory findings suggests that in patients suffering from jaundice due to malaria the hyperbilirubinaemia is mostly unconjugated type. (Mishra, et al., 2003), (Anand & Puri, 2005). Conditions for determining complicated malaria are presented in a **Table:1.2**

Feature	Characteristics
Cerebral Malaria	**Coma not ascribed to other causes, Glasgow≤ 11 in adults or a Blantyre coma scale < 3 in children**
Severe Anemia	**Haemoglobin <7g/dl or hematocrit<20% respectively in adults and Hb concentration <5g/dl or hematocrit of <15%in childerens<12 years of age with parasitaemia>10,000/µl**
Acute Renal Failure	**Serum creatinine >265 µmol/l (> 3.0 mg/dl) or blood urea >20mM**
Pulmonary oedema	**Radiologically confirmed, or oxygen saturation <92% on room air with a respiratory rate >30/min, often with chest indrawing and crepitations on auscultation**
Severe hypoglycemia	**Blood glucose < 40 mg/dL**
Shock	**Systolic blood pressure < 70 mm Hg in children or <80mmHg in adults with evidence of impaired perfusion.**
Metabolic acidosis	**A base deficit of >8 meq/l or, if unavailable, a plasma bicarbonate of <15 mM or venous plasma lactate >5mM. Severe acidosis manifests clinically as respiratory distress–rapid, deep and laboured breathing**
Significant bleeding	**Including prolonged bleeding from nose gums or vein puncture sites haematemesis or melaena**
Jaundice	**Plasma or serum bilirubin >50 µM (3 mg/dl) together with a parasite count >100 000/µl**

Table1.2: Malaria S., (2014). Severe malaria. *Tropical Medicine & International Health*.

1.8-Genome of *Plasmodium falciparum*:

Unlike other eukaryotes, malaria parasites possess a nuclear genome and two organellar genomes, namely the apicoplast and the mitochondrial genome. An international effort for sequencing of *P. falciparum* started in 1996 (Hoffman, et al., 1997). With the complete nuclear genome sequence of *P. falciparum* culture isolate, 3D7 published later in 2002 (Gardner, et al., 2002).

1.8.1-Nuclear genome:

Sequencing of the malaria parasite nuclear genome in October 2002 has provided a breakthrough and helped malaria research to enter into the post-genomic era. A whole chromosome shotgun strategy was used to sequence the *P. falciparum* 3D7 clone. The 23 Mb nuclear genome of *P. falciparum* is composed of 80.6% A+T approximately more than 90% in intergenic regions of the chromosome. It consists of 14 chromosomes that encodes for 5,300 genes, out of which 60% are hypothetical proteins. 54% of the protein-coding genes consist of introns, it has been found that the *P. falciparum* genes are longer than 4 kb compared to *S. pombe* and *S. cerevisiae* (3.0% and 3.6%) These large genes encodes for uncharacterized proteins without any signal peptides. The nuclear genome of parasite contains a set of transfer RNA (tRNA) ligase genes, and 43 tRNA's were found to bind all codons except TGT and TGC that codes for Cys. An unusual tRNA were also found in the nuclear genome of *P. falciparum* which resembles a selenocysteine tRNA. *P. falciparum* contains several single 18S-5.8S-28S rRNA units distributed among chromosomes. The expression of each rRNA unit is developmentally regulated, which results in the expression of a different set of rRNA at different stages of parasite life cycle. Genes, mostly that show antigenic variations like multigene family genes var, rifin and stevor are mostly located in the sub-telomeric regions of the chromosomes (Gardner et al., 2002).

1.8.2-Apicoplast genome:

Apicoplast, a non-photosynthetic plastid originated from the secondary endosymbiosis of red algae (Lim, et al., 2010) and is essential for the survival of malaria parasites, it is homologous to the chloroplasts of plants and algae, the genome size of *P. falciparum* apicoplast is 35 kb and is circular (Köhler, et al., 1997). The apicoplast genome encodes for the small and large ribosomal subunits, 17 ribosomal proteins and the elongation factor Tu (ef-Tu). Proteins that function in the apicoplast consist of special character i.e

bipartite leader sequence. This leader sequence consists of a secretory signal peptide, and downstream of the signal peptide there is so-called transit peptides. Most of the genes encoding predicted apicoplast anabolic functions belong to a fatty acid, haem and isopentenyl diphosphate biosynthetic pathways and Iron/sulphur cluster pathway (reviewed by Ralph, et al., 2004, Lim & McFadden, 2010).

1.8.3-Mitochondrial genome:

The mitochondrial genome of *P. falciparum* is of ~6kb in size, haploid in nature and contains three protein-coding genes, cytochrome oxidase I (cox1), cytochrome oxidase III (coxIII), and cytochrome b (*cytb*). Mt DNAs also codes for the r RNA needed for mt translation. (Feagin, et al; 2012, Tyagi, S., Pande, V., & Das, A., 2014).

1.9-Genomic variations in *Plasmodium falciparum*:

Genomic studies using high throughput microarray technology like Comparative Genomic Hybridization (CGH) facilitate the variation study of whole genome of *P. falciparum*. Comparative genomic studies of *P. falciparum* revealed the causes of genome sequence variations in the parasite like single nucleotide polymorphism (SNP), insertions, deletions of short sequences, major amplifications, translocations, inversions, and allelic variations, etc (Cheeseman, et al; 2009). These genetic variations may lead to variations in disease severity and hence studying the occurrence of these variations is important. Most of the genomic analysis of *P. falciparum* has been done in strains, which have long been established in continuous culture systems. However, analysis of genome variations from culture-adapted parasite lines remains intrinsically very different from the analysis of *in-vivo* derived parasite nucleic acid material. The latter could have a major impact on the understanding of the parasite biology (LeRoux, et al; 2009) because environmental pressures have been known to introduce changes in genomes providing organisms with an evolutionary advantage for adaptation and Copy Number Variation (CNVs) could be one such change that assists the parasite adaptation to the new environment. These CNVs could enable survival of the parasite under drug pressure. An amplification event in the GTP cyclohydrolase (GCH-1) has been frequently reported and the high copy number of GCH-1 has been shown to confer resistance to antifolates (Kidgell, et al; 2006; Nair, et al; 2008; Jiang, et al; 2008; Ribacke, et al; 2007; Heinberg , et al; 2013).

The major known parasite molecules that influences the severity of the disease are variation in the Variant Surface Antigen (VSA) family genes that are located in the subtelomeric part of the chromosome. *P. falciparum* 3D7 genome sequencing revealed the presence of ~60 var genes followed by other variant surface antigen families like *rifin/stevor* (~190), *Pfc-2TM* (~12)and surfin (~10) (Scherf, et al; 2008; French Christian and Chen Nansheng., 2013). The most studied of these, the var genes encodes for PfEMP-1 (Erythrocyte membrane protein-1) that play an important role in the pathogenesis by periodic switching and immune evasion. Along with the exported family members, these genes play an important role in cytoadhesion by binding to multiple cell receptors, including ICAM-1, CD36, E-selectin, NCAM and CD31 on microvascular endothelium which mediate sequestration that in-turns lead to the severity of the disease. Recently another receptor, Endothelial protein C receptor (EPCR) a brain-specific receptor has also been identified as a binding partner of PfEMP-1. (Wassmer, et al; 2004; Smith, et al; 2013; Wassmer & Grau.,2017). Var genes are divided into 3 major groups A, B and C and 2 intermediate groups UpsB/A and B/C, (Lavstsen, et al; 2003). The higher transcript level of var group B and var group C family genes which are CD36 binders (Bernabeu, et al; 2016) have been reported in severe malaria cases in comparison with the uncomplicated malaria cases from India (Subudhi, et al; 2015).

Genomics technologies like Microarrays and more recently Next-generation sequencing have enabled deeper analysis of various types of genomic variations. Using a reference clone/isolate, copy number variations of other isolates can be identified (Carret, et al; 2005). Array-based technologies have been used to distinguish CNV's in different disease studies reviewed by (Coughlin, et al; 2012) Such techniques have been extensively used in cancer research to identify DNA copy number variability to represent structural variations at higher resolution and increased sensitivity (Mockler, T. C., & Ecker, J. R, 2005), Przybytkowski, et al; 2011).

1.10-Comparative Genomic Hybridization (CGH):

In order to understand the genomic variation in the chromosomes without the need of cell culturing, CGH as a new chromosome analysis technique was first reported by Kallioniemi, et al in 1992. Later Manoir, et al in 1993 took tumor DNA and labelled it with a green fluorescent dye (1:1), mixed it with normal control DNA tagged with red labelled dye, and hybridized to normal human metaphase, the green and red labelled DNA fragments compete with each other for their hybridization to their locus of origin on the chromosomes.The green to red fluorescence was measured along the chromosome axis, where the ratio less than 1is considered as a loss and the ratio is more than 1 is considered as a gain. (Weiss, Hermsen, Meijer, et al; 1999) The use of this initial method was limited because of technical difficulties and very low resolutions which are limited to 10-20Mb. Later on, with this advent array, CGH was developed, by substituting the metaphase chromosomes with oligonucleotide arrayed probes on glass slides **figure 1.5.**

Oligonucleotide arrays are classified into two types whole genome arrays and targeted arrays. Whole genome arrays are more advantageous than targeted array's because of the limitation of it being used specifically for disease specific chromosomal anomalies for which it cannot fit to use for the survey of whole genomes (Aradhya and Cherry, 2007). The availability of sequenced genomes and oligonucleotide tiling microarrays has enabled in many genomic applications. Oligonucleotides in a tiling array covers the whole genome or contigs of the genome in an unbiased fashion. The Commercially available tiling array's are present with different probe lengths, probe spacing and array design characteristics for e.g. Affymetrix tiles 6 million 25 mer probes and provides one colour assay, so individual samples can hybridize to different array's, Nimblegen can tile 385,000, 50-75 mer probes, and Agilent that can tile 244,000 60 mer probes per array, the latter two platforms with longer oligonucleotide probes are more sensitive and they provide feasibility for two colour assays by which treated and control samples can be labelled differently and put on the same array for competitive hybridization. Among the SNP array and other oligonucleotide arrays tiling array is one of the robust array designs used for the study of genome-wide aberrations. A tiling array usually spans the entire genome of an organism. The probes can be non-overlapping or partially overlapping and tiled to cover the entire genome end to end or spaced at regular intervals **figure 1.4.**

1.10.1- Array CGH Tiling array:

There are two general types of high-density microarrays, (1) high-density oligonucleotide arrays with a relatively short oligonucleotide (<100mer), (2) with >6,000,000 discrete features can be made which comprises of millions of copies of distinct probe sequences. In the one with the relatively short oligonucleotides (<100), the probes are synthesized directly on the surface of the arrays by photolithography using light-sensitive synthetic chemistry and photolithography masks, an ink jet device or programmable optical mirrors. The second type of oligonucleotide array is made by mechanically printing or spotting probes, amplified PCR product, oligonucleotides, or cloned DNA fragments onto glass slides.

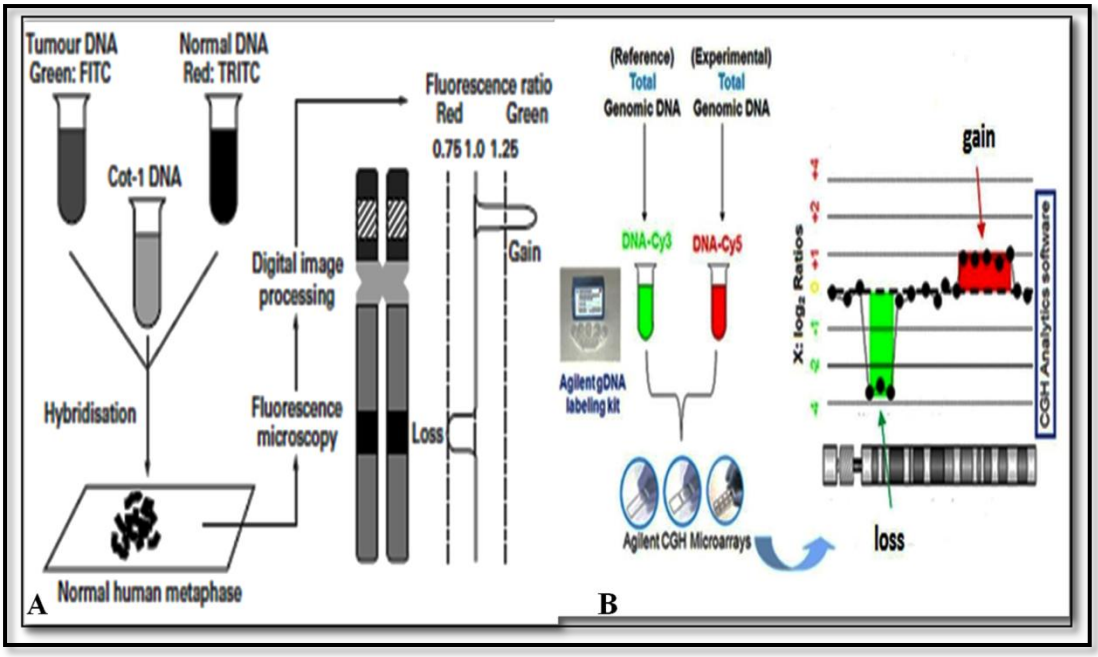


Figure 1.5: Difference between metaphase and oligonucleotide array CGH, A) representation of the CGH technique tumour DNA is tagged with green fluorochrome and normal DNA is tagged with red fluorochrome, and hybridized to normal metaphase, the ratio of the red and green signal determined. B) schematic representation of array CGH from Agilent technologies where reference genomic DNA is labelled with green dye Cy3 and Treated sample/ diseased DNA with red dye Cy5, the image is processed using Agilent genomic workbench software and the ratio of red vs green signal recorded as gain and loss. (Weiss, Hermsen, Meijer, et al; 1999, and Wozniak, et al; 2007)

High resolution CGH tiling arrays has been used in assessment of different disease conditions and possible genomic regions responsible for the cause of the disease for example a CGH tiling array with high resolution has been used to assess chromosomal breakpoints in neuroblastoma and primary tumors, which shows 58 chromosomal breakpoints that generated 45 large scale partial chromosomal imbalances (Selzer, et al; 2005), another example of use of these arrays is in determining the prominent amplifications and deletions related to chronic disease conditions like alteration in SHOX homeobox gene, which is related to skeletal dysplasias (Benito-Sanz, et al; 2016). The array CGH tiling array provides various opportunities to determine the comparative analysis between the treated and the control samples. A Custom designed high density overlapping 25 mer tilling microarray was used to detect the SNPs and small insertions and deletions (indels) and copy number variations to distinguish the three *P. falciparum* laboratory strains HB3, Dd2 and 3D7, where 3D7 was used as a reference strain and Dd2 was compared with the 3D7 strain, Deletion of 6 subtelomeric gene has been detected in chromosome num-2 of Dd2 strain in comparison with the 3D7 strain an amplification event has also been found in Dd2 on chromosome-5 surrounding *pfmdr-1* gene and also on chromosome-12 near *pfgch-1* gene.In fosmidomycin resistant mutants of Dd2, an amplification event has been found in chromosome-14 compared to the parental Dd2 strain (Dharia,et al;2009).

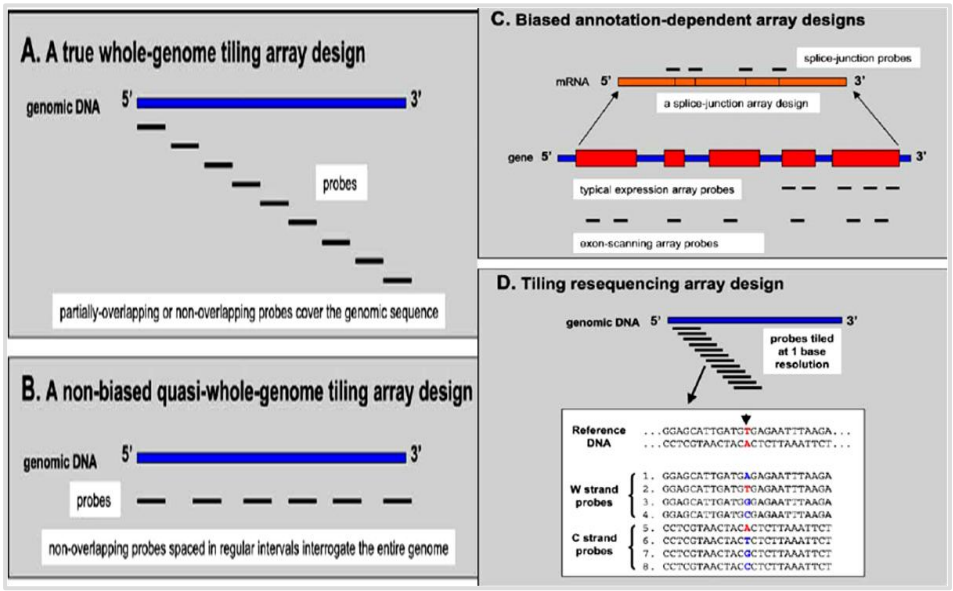


Figure1.6: Schematic representation of CGH tiling array: A) A whole-genome tiling array with partially overlapping probes, B) Biased annotation dependent tiling array design, C) whole genome non-overlapping tiling array, D) whole genome tiling array design where probes tiled at 1 base pair resolution (T.C. Mockler, J.R. Ecker, 2005).

1.11-Genome to Proteome of *Plasmodium falciparum* a systems biology approach:

With the completion of the genome sequence of *P. falciparum* in 2002, it was predicted that it would provide insights into parasite genome biology and also provide information leading to discovery of new drug and vaccine candidates. As stated before, the genome of *P. falciparum* consists of approximately 60% of the genes that codes for hypothetical proteins and many of these do not show any sequence homology with other organisms, especially the human host. One of the objectives of functional genomics and systems biology is to provide indicators of antimalarial targets, in this context transcriptomics studies have led to the discovery of differential expression of different genes in different stages of the parasite life-cycle (Le Roch, et al; 2012).

System level functionality of the genes are extensively explored nowadays by the use of gene-co-expression networks. The network construction is fairly straightforward, where genes are represented as nodes. These nodes are connected if the genes are significantly co-expressed in the diseased samples. The question remains, whether the connection between the nodes is meaningful or not? To this end, a general framework of soft thresholding has been designed by Zhang and Horvath in 2005, where each connection has been given a weight, known as weighted gene co-expression networks (WGCN). This network is more biologically relevant than unweighted gene co-expression networks. In WGCN different modules are generated in which one gene is highly connected with the other genes in that subnetwork or module, this HUB gene could be highly related to the external gene information (Zhang and Horvath, et al; 2005). These hub genes are proposed to have an important role in disease state and in biological development. In a study, it has been found that the biomarkers of the disease state periodontitis IL8, IL1ß, ICAM1 and VEGFA were hub genes and useful diagnostic markers for periodontitis(Guo, et al; 2015). In another study on Alzheimer's disease (AD), WGCNA identified many novel genes that are involved in AD, apart from some of the key hub genes like cyclin-dependent kinase-5 (CDK-5) and presenilin-1 (PSEN1) that are well characterized in AD (Miller, et al; 2008). The weighted gene co-expression networks also play an important role to identify important interacting proteins. In a study on *P. falciparum*, protein interaction network revealed the cluster of proteins that are important for regulation of the interactions important for organism survival. It has been found that many of the hub genes are connected with the

conserved *Plasmodium* proteins with unknown function. For example hub protein ATP dependent helicase, located in nuclear chromatin and involved in nucleosome assembly and regulation interact with 57 other proteins out of which many belong to conserved *Plasmodium* protein with unknown function (Bhattacharyya and Chakrabarti, 2015).

A study from our lab on construction of co-expression based systems network using microarray co-expression data from field samples revealed the vital role of hub genes and have shown the enrichment of these genes into several crucial biological processes that are important for parasite survival and pathogenicity, In this study, some of the hub genes identified were up or stably expressed genes, e.g. PF14_0150 was found to be up-regulated in the network whereas two other PF08_0085 and PF07_0083 were found to be stably expressed genes (Subudhi, et al; 2015).This data suggests that whether the genes are differentially expressed or otherwise, being hub gene the proteins coded by these genes could be essential for the parasite metabolic activities, and thus for survival. In addition to annotated genes, some of the unknown function proteins or hypothetical proteins were also reported as hub genes and could also play a major role in parasite survival. Characterization of these hub genes could lead to the discovery of novel drug target or molecules for other intervention/prevention strategies.

1.14-Gaps in existing research:

The 23.3 MB nuclear genome of *P. falciparum* is known to display vast genetic diversity. These are mostly due to genome variations, i.e. loss or gain of small or large segments of chromosomes or other types of copy number variation like insertions and deletions (indels). These variations have been analysed between *P. falciparum* culture isolates. There is scarcity of knowledge about genomic variations between field isolates causing differing disease manifestations (Uncomplicated and Complicated). This large gap restricts proper analysis of the phenomenon observed in this region, in the context of reports from other parts of the world.

Approximately 50-60% of the genome of *P. falciparum* codes for proteins with unknown functions and often nothing is known about their expression status or localization.

A weighted gene co-expression network based study from parasite transcriptome data from patients with non-severe or severe disease shows that some of the hypothetical/conserved *Plasmodium* proteins with unknown function are critical hub genes in the created network (Subudhi, et al; 2015). Due to the lack of information on these genes, there is a scope of further study to investigate their possible role in parasite biology.

1.15-Objectives of the study:

Based on the gaps in existing research, and to understand the parasite biology at genomic, transcriptomic as well as proteomics level, the objective of the study was divided into two parts.

1. Comparative Genome Hybridization analysis to identify large-scale amplifications and deletion at the genome level between parasites causing complicated and uncomplicated disease manifestations.
2. Identification and characterization of key molecules from co-expression based systems-network generated based on parasite RNA from complicated and uncomplicated patient samples. This objective is further divided into subparts:

   a) Correlation of the CGH array data with the regions containing the selected genes and analysis of minor genetic variations like SNPs in the field isolates for the selected molecules.

   b) Cloning and expression of selected molecules.

   c) In-silico based analysis of selected molecules.

   c) Peptide-based immunolocalization studies.

Chapter II

Materials and Methods

To achieve the objectives mentioned above various techniques were used based on the exhaustive literature survey. The methodology adopted for comparative genome hybridization analysis and reagents and kits used for the characterization and localization of the proteins are detailed in this chapter. The chapter starts with detailing of the methodology used for the comparative genome hybridization analysis, the results of this analysis are discussed at length in chapter-III. Following this detailing, the next section of this chapter includes a selection of the hub genes from co-expression based systems network for the study. Further, chapters are divided into two parts: The first part includes experimental details and procedures used for the cloning, expression and localization studies of the proteins which includes blood sample collection from malaria infected patients, isolation of parasite DNA/RNA, confirmation of the infection by PCR, amplification of selected hub genes, their cloning and expression trials, along with the subcellular localization studies. The result of the experiments performed are

discussed at length in chapter iv. The second part of the chapter includes *in-silico* based approach for the analysis of the selected proteins using various bioinformatics tools. This includes the analysis of conserved domain in proteins, and disordered nature of protein, functional prediction of proteins, secondary structure prediction and physicochemical nature of the selected proteins with the help of various freely available online software and peptide based immunolocalization of both the proteins. The result of the study is discussed in chapter-v. The compositions of the reagents used for the cloning and expression purposes are detailed in Appendix-A1.

2.1: Comparative Genomic Hybridization data analysis:

P. falciparum is the most virulent among the 6 *Plasmodium* parasite species that causes malaria in humans. Current efforts to eliminate the disease have not been successful because of the rapid genetic adaptation and natural selection of the parasite. The currently known causes of genome sequence variations in the parasite are single nucleotide polymorphisms (SNP), insertions, deletions of short sequences, major amplifications, translocations, inversions, and allelic variations (Cheeseman, et al., 2009). These genetic variations may lead to variations in disease severity and hence studying the basis of these variations is important. In this study, a custom designed tiling microarray has been used to investigate the overall differences between the two clinical isolates i.e. Pf-isolate-2 (Pf-2) (complicated/cerebral malaria) and Pf-isolate-1 (Pf-1) (uncomplicated malaria). The reference genome 3D7 was used to design the probes. The tiling array used in this study enables large-scale detection of regions showing copy number variations across the whole genome. The CNV's were identified in Pf-2 (complicated) in comparison to Pf-1 (uncomplicated malaria). The well-accepted ADM-2 algorithm was used for the analysis of CNV. Here we are reporting significant CNV regions seen in Pf-2 in comparison with the Pf-1. In this data, the term amplification and deletion are used to represent copy number of test strain Pf-2 compared to Pf-1. Deletions are reported where the signal intensities of the probes representing the region in the test strain are near background and represent the absence of hybridization. The microarray data discussed in this study has been deposited in the NCBI Gene Expression Omnibus (GEO) under the GEO series accession number (GSE77165).

2.1.1: Array CGH Design:

A custom *P. falciparum* 3D7 2x400K CGH microarray (AMADID: G4124A_025327) was designed on the Agilent platform by Genotypic Technology, Bangalore, India, by using Sequence information, chromosomal location, the direction of transcription of 3D7 which was retrieved from *Plasmodium* Genome Resource [Plasmo DB. V. 6.0] (Gardner, M, J.et al., 2002; Aurrecoechea, et. al., 2009)**.** The 60mer probes were designed based on the optimal GC% and Melting temperature using Genotypic_Probe_Parser.pl (Perl program developed by Genotypic Technology). The probes were synthesised *in situ* as per the standard algorithms and methodologies used by Agilent Technologies. The array design comprised of 418577 probes tiled across the whole genome with the average probe spacing of 56 bp, that covers the whole genome irrespective of intergenic or intronic sequence, distribution of probes is shown in **table 2.1.** The synthesis of probes was done by Genotypic technology, Banglore, India.

Reference Genome	**3D7**
Source	PlasmoDBV6.0
Number of sequences:	16 (14 chromosomes+1mitochondria+1plastid)
Average Probe spacing	56bp
A total number of the features:	420288
Available features:	418577
Agilent controls:	1711
Agilent control grid	IS-420288-V2_2by420K_Genomic_20080924
A total number of probes collected:	418577
Blank features:	0

Table 2.1: Probe distribution in 2x400k array

2.1.2-Genomic DNA preparation and hybridization:

1.5 µg of both test (Pf-2) and control (Pf-1) DNA sample was digested using *Alu1* and *Rsa1*. After digestion; Pf-1 and Pf-2 were labelled with Cy5 and Cy3 respectively, using Agilent Genomic DNA labelling kit (Part Number: 5190-0453). The labelled DNA was concentrated and quality assessed for yields and specific activity. 5 micrograms of the labelled samples were hybridized on a Genotypic designed *P. falciparum* 3D7 Custom 2x400K CGH Microarray, (Agilent Microarray Design ID AMADID: 025327). Hybridization of probes was carried out in Agilent's Surehyb Chambers at 65º C with rotation for 40 hours. The hybridized slides were washed using Agilent aCGH wash

buffer (Part No: 5188-5221/22) and scanned using the Agilent Microarray Scanner G2505C. The Scanned image was processed using Agilent Feature extraction software and probe intensities were normalized using the linear dye normalization method. The methodology explained in the above paragraph was performed by Genotypic technologies, Banglore, India.

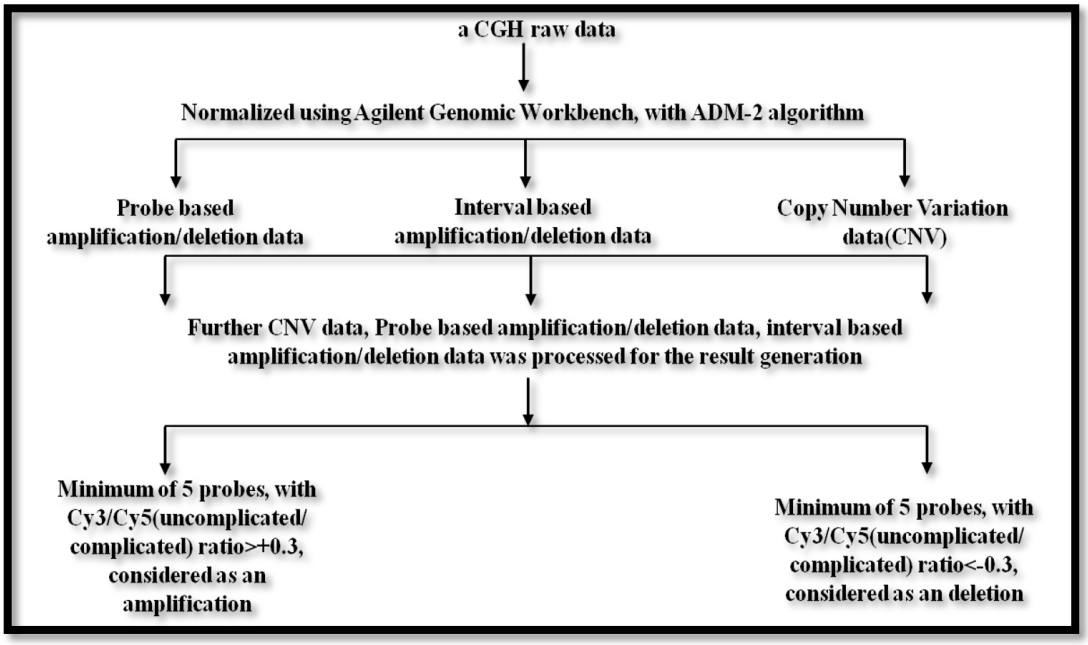
2.1.3-Comparative genome array design:

A high resolution custom CGH array was designed with 418,577 probes, representing the entire *P. falciparum* 3D7 reference genome using the e array tool by Agilent certified microarray facility of Genotypic Technology, Bangalore, India. For the same ,whole genome sequence for *P. falciparum* 3D7 strain was downloaded from NCBI genome database; the 60-mer probes were designed based on optimal GC% and melting temperature using Genotypic_Probe_Parser.pl (Perl program developed by Genotypic Technology), the chromosome sequences have been given as an input to the program. The probes were designed by tiling over the DNA sequence. The tiling resolution was 56 bp. The probe length has been set as 60bp which is the optimised length for Agilent comparative genomic hybridisation workflow. The probes were designed by calculating GC% and optimum Tm values of (>65 °C to <85°C), and specificity to the target genome.

2.1.4- Array CGH data analysis and aberration filter to detect copy number variation:

Image analysis and data normalization was done by applying the linear dye normalization method using the Agilent Feature Extraction Software. The Raw data were analysed using well-accepted ADM-2 algorithm (Aberration Detection Method II algorithm) developed by Agilent and included in the Agilent Genomic Workbench for CGH analysis. ADM-2 is a statistical algorithm which computes copy number differences between the sample and reference using an iterative procedure to identify all genomic regions from its expected value of 0 larger than a given threshold (In this case we have taken a minimum of 5 consecutive probes, with an average log ratio of 0.3). At each iteration, the region with the most significant score is reported. ADM- 2 statistical score is computed as the average normalized Cy3/Cy5 $\log_2$ ratios of all the probes in the genomic interval multiplied by the square root of the number of these probes. Deletions are reported where the signal intensities of the probes representing

the region in the test strain are near background and represent the absence of hybridization. Graphical visualization of the data for the significant aberrations was done using UCSC genome browser (Schneider,et al., 2006). The methodology used for the analysis of aCGH is explained in the following flowchart:



Flowchart of methodology used for Comparative Genomic Hybridization data analysis

The second part of the thesis includes characterization and localization of the proteins (hub genes) selected from the *in-vivo* co-expression based systems network. For the same we have selected hub genes and checked them against the CGH data to rule out their possibility of falling in the region showing variations. Two of the selected hub genes were sequenced to see if any sequence variation is present in the exonic region of the genes w.r.t the 3D7 annotated gene sequence available in the Plasmo DB database.

2.2-Shortlisting of hub genes from systems network:

A Weighted Gene Co-expression network (WGCN) generated from *in-vivo* parasite-derived transcriptome data from clinical samples of differing disease conditions (uncomplicated and complicated) was generated in our lab as described in Subudhi, et al., 2015. The microarray data of the co-expressed transcripts were used to generate a systems network, where the genes were denoted as nodes and expression values as edges. The genes that were highly connected with other nodes were categorized as hub genes or the genes that are controlling expression of other genes in the module. The

criteria for selecting the modules for the selection of hub genes from the system's network are as follows:

a) The genes present in the selected modules should be with higher K-within i.e the intramodular connectivity of the genes should be more than other modules. This is based on the assumptions that the genes with higher K-within could be the central genes involved in pathways in the parasite and could be essential for the parasite survival.

b) Two modules were selected; one with the majority of up-regulated molecules in both the complicated cases, i.e. jaundice and jaundice+renal failure, i.e., cyan module, another with more number of steady-state molecules i.e., blue module. The genes present in cyan module were enriched in the translational biological process with P=0. 02, **figure 2.1.A.** The genes belong to blue module were enriched in redox biological processes with P=0. 0001, **figure2.1.B.**

The enrichment of these two modules in specific biological processes suggests these genes to be essential for parasite survival and pathogenesis. On the basis of the above criteria, four genes were selected from cyan and two genes form blue module for further characterization of proteins as shown in **table2.2.A and B**. Out of these, three genes were successfully cloned of which two of them were annotated as "conserved *Plasmodium* protein with unknown function"(PFD0300W and PFL2320W) were taken for further studies.

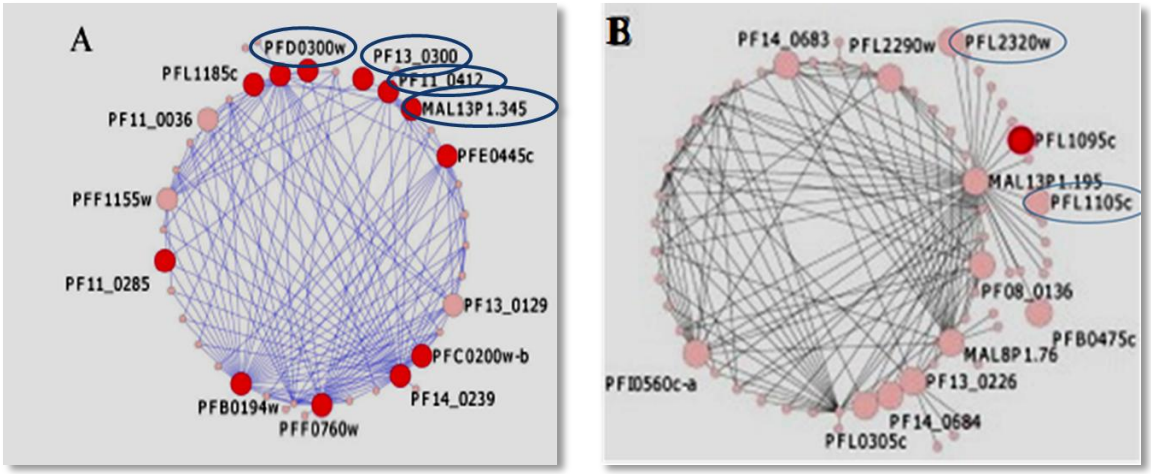


Figure 2.1: Modules selected from systems network for the selection of hub genes; (A) Cyan Module, Consist of 136 genes with 54 up-regulated genes, Module (B) Blue module consists of 523 genes, (Subudhi, et al., 2015).

Genes from Cyan module	Function	gDNA Size(bp)	cDNA size (bp)	K-within	Differential expression
PF11_0412	V-type proton ATPase subunit F, putative	1083	387	18.56	**Up-comp**
PFD0300W	Conserved *Plasmodium* protein with unknown function	1527	1527	18.506	**Up-comp**
MAL13P1.345	Conserved *Plasmodium* protein with unknown function	1525	693	14.2399	**Up- comp**
PF13_0300	Mitochondrial inner memberane translocase, putative	772	504	14.107	**Up-comp**

Table.2.2.A: Genes selected from cyan module: where gDNA represents genomic DNA, Up represents up-regulated gene in co-expression systems network.

Genes from blue module	Function	gDNA size bp	cDNA size	k-within	Differential expression
PFL2320W	Conserved *Plasmodium* Protein with unknown function	858	858bp	59.298	**Steady state**
PFL1105C	Conserved *Plasmodium* protein with unknown function	875	405bp	59.74468	**Steady State**

Table 2.2.B: Hub genes selected from modules with up-regulated and steady-state expression.

2.3-Blood sample collection of malaria infected patients:

Venous blood samples were collected (5ml) from *P. falciparum* infected patients admitted to S.P. Medical College, and associated group of hospitals from Bikaner, India. Patient's samples were collected on informed consent and after taking appropriate approvals from human ethics committee (No.F. (ACAD) SPMC/2003/2395) after clinically confirming the infection with *P. falciparum* was confirmed by microscopy and RDTs (OptiMal test; Diamed AG, cressier sur Morat, Switzerland, Falcivax test; Zephyr Biomedical System Goa, India). Blood was immediately (within 15 mins) subjected to density gradient based separation (Histopaque 1077, Sigma Aldrich, USA) to separate the peripheral blood mononuclear cells (PBMCs) from the infected blood samples following manufacturers instruction after this, both the fractions were subsequently washed with phosphate buffer saline (PBS) twice and lysed using 4 volumes of TRI –Reagent (Sigma Aldrich, USA) and stored at -80^0 C, these samples were then transported in cold chain to BITS, Pilani,

Pilani campus further used for DNA/RNA extraction. These samples were categorised as uncomplicated and complicated samples based on World Health Organization guidelines (WHO world malaria report, 2010).

2.3.1-Parasite Genomic DNA Isolation:

Parasite genomic DNA isolated from above-preserved samples as per the manufacturer's protocol using Tri-reagent. The methodology followed for isolation of g DNA and RNA from Tri-reagent preserved samples are as follows.

2.3.1.1: DNA and RNA extraction using Tri-reagent:

2.3.1.1.1: RNA isolation:

For isolation of RNA, every material that are to be used for RNA isolation was treated with DEPC treated water overnight and subsequently autoclaved separately for the removal of any trace of DEPC. All the reagents were made in autoclaved DEPC treated water.

Procedure:

2.3.1.1.2-Sample preparation: For tissue or blood, the blood samples were homogenised in Tri-reagent (1 ml of Tri/50-100 mg of tissue), the volume of the blood/tissue should not exceed 10% of the volume of the TRI Reagent.

2.3.1.1.3: For bacterial cells, Cells were centrifuged at 6000 rpm for 5 mins, then pallet was lysed in 1 ml of the Tri-reagent.

2.3.1.1.4: Phase separation:

After adding Tri-reagent to the samples, samples were allowed to stand for 5 minutes at room temperature for complete dissociation of nucleoprotein complexes. 0.1 ml of BCP (Bromo-chloropropane), was added, per ml of Tri-reagent. Samples were shaken vigorously for 15 seconds and allowed to stand for 15 mins at room temperature. Vigorous shaking of the samples allows the formation of two layers. The resulting mixture was centrifuged at 12,000x *g* for 15 mins at 4°C. Centrifugation separates the mixture into 3 phases: red organic phase (containing protein), Interphase (containing DNA), and a colourless upper aqueous phase (containing RNA). The resulting aqueous phase containing RNA was separated in fresh DEPC treated microfuge tube and 0.5ml of isopropanol per ml of TRI was added and mixed. The samples were allowed to stand for 5-10 min at room temperature. After which they were again centrifuged at 12,000 x

g for 10 mins at 4°C. The pellet was washed with 1ml of 75 % ethanol made in autoclaved DEPC water. The samples were vortexed and then centrifuged at 7,500 x g for 5minutes at 4°C. Resulting RNA pellet was air dried for 15 minutes finally RNA pellets were reconstituted in 50ul of autoclaved DEPC treated water.

2.4-cDNA preparation from total RNA of bacterial cells:

cDNA was prepared from total bacterial RNA using the Omniscript® Reverse Transcription Kit, (QIAGEN®), Using manufacturer's protocol. Before preparation of cDNA, DNAse treatment were given to remove any genomic DNA contamination for the same, RNA samples were treated with the DNAase enzyme at 37°C for 1hr after which the reaction was stopped by adding 1µl of EDTA **table 2.3.**

Reagents	Amount
Total RNA (50ng)	Variable
10x RT buffer	2µl
DNTPmix (5mM each dNTP)	2 µl
Gene-specific primer (0.1µM)	8 µl
RNase inhibitor (10units/µl)	1 µl
Omniscript reverse transcriptase	1 µl
RNase-free water	Quantity sufficient

Table: 2.3: Reaction mixture used for cDNA preparation.

2.5-Extraction of genomic DNA:

After the extraction of the total RNA the leftover mixture, containing DNA and protein aqueous layer was used to isolate DNA. For this, to the leftover mixture, 0.5ml of back extraction buffer was added per ml of Tri-reagent used. The tubes were shaken vigorously for 10-15 seconds. Samples were incubated for 10 mins at room temperature and centrifuged at 12,000 x *g* for 10 minutes. The Aqueous layer was transferred into the fresh tube and 0.5ml of isopropanol per ml of Tri-reagent was added and mixed the samples were kept at room temperature for 20 minutes for precipitation of DNA. The DNA was pelleted down by centrifuging at 12000x*g* for 10 minutes and washed twice with 70% ethanol after which it was air dried for 30 minutes at room temperature. The DNA pellet was dissolved by adding 90 µl of 1XTE buffer and kept at-80° refrigerators till further use.

2.6-PCR based method to confirm parasite infection (Diagnostic PCR):

In order to confirm the species-specific parasite infection, a PCR based method has been used which was designed on the basis of 18S rRNA (Das, et al., 1995; Kochar, et al., 2005) and 28S rRNA (Pakalapati, et al.,2013) PCR amplification.

2.6.1-Diagnostic PCR based of 18S rRNA:

A multiplex PCR designed with a single, genus-specific forward universal primer for *Plasmodium* and two species-specific reverse primers specific for *P. vivax* and *P. falciparum* was used to confirm the type of infection **table 2.4**. The reaction conditions utilised are shown in **figure 2.2**. The presence of *P. falciparum* infection is confirmed by the presence of ~1400 bp amplicon whereas the presence of *P. vivax* is confirmed by the amplicon size of ~500 bp **figure-2.3** (Das, et al.,1995; Kochar, et al.,2005).

Primer Sequence 5' to 3' direction	Specific for	Orientation
5'-ATCAGCTTTTGATGTTAGGGTATTG-3'	**Genus-specific**	**Forward**
5'-GCTCAAAGATACAAATATAAGC-3'	***P. falciparum***	**Reverse**
5'-TAACAAGGACTTCCAAGCC-3'	***P. vivax***	**Reverse**

Table 2.4 : Diagnostic Primers for 18S r RNA gene amplification by multiplex PCR

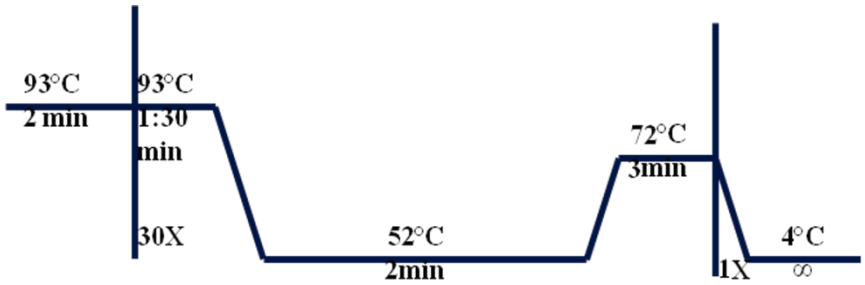


Figure 2.2: Reaction conditions used in multiplex/diagnostic PCR

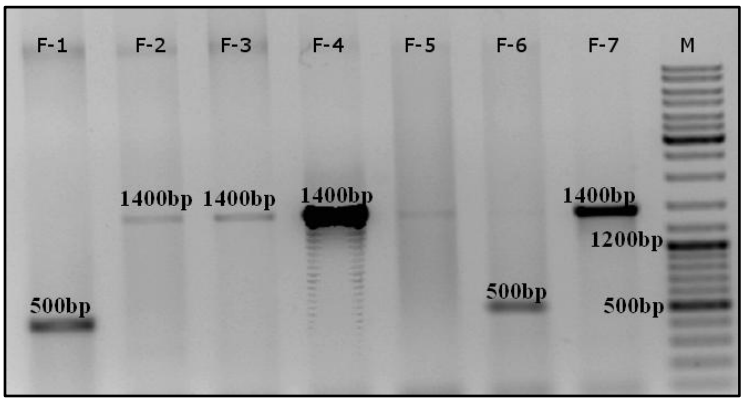


Figure2.3: Parasite infection confirmed by multiplex PCR based on 18S rRNA gene amplification [Marker=Gene ruler DNA ladder mix (SM0331Fementas). F-1 to F-8 are Parasite field isolates, F1 and F6 are *P. vivax* infected, F2, F3, F4, F5 and F7 were *P. falciparum* infected samples.

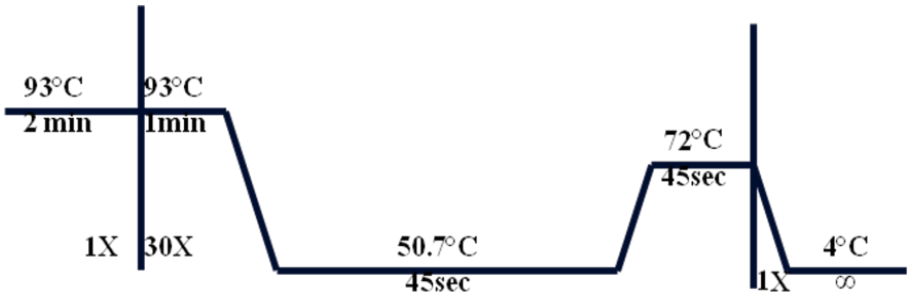
2.6.2-Nested PCR based on 28S rRNA amplification:

Nested PCR based 28S r RNA detection method is a two reaction PCR. First PCR is used to amplify a region common to *Homo sapiens*, and *Plasmodium* species (*P. falciparum* and *P. vivax*), the product size of which is ~790 bp. For the secondary reaction, the product of primary reaction is used as a template, in two separate reactions, each specific for *Pf* or *Pv* (Pakalapati, et al., 2013). The nested PCR-1 using primers NPV2 and NPVR amplifies ~ 294 bp region, which is specific for *P.vivax* 28S rRNA gene, while the nested PCR 2 amplifies ~ 286 bp region specific for *P. falciparum* 28S rRNA gene. The primers used for this primary and secondary nested PCR reaction are given in **table 2.5** while the reaction conditions utilised are shown in **figure 2.4**. The samples which showed the band at the desired position were utilised for further studies **figure 2.5**.

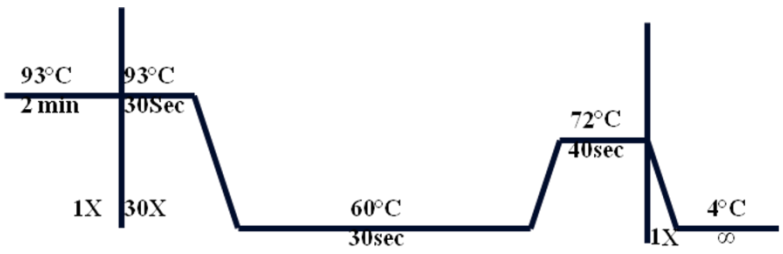
Lab ID	Primer sequence	Detection	Orientation
NUF	5’ GAT TTC TGC CCA GTG CTT TGA ATG T 3’	Human, Pf&Pv	Forward
NUR	5’AAT GAT AGG AAG AGC CGA CAT CGA A 3’	Human, Pf&Pv	Reverse
NPV2	5’ TCG GTT CGC CGG GTA TTC ATA TT 3’	*P. vivax*	Forward
NPVR	5’ CAC AGT AGG AAG ATA AAT TCC T 3’	*P. vivax*	Reverse
NPF1	5’ TAT CCT TCG GGA AGG CAT TCT G 3’	*P. falciparum*	Forward
NPF2	5’ CTA TAT GCA CAG TAG TAA GTA ATT TA 3’	*P. falciparum*	Reverse

Table 2.5 Nested PCR primers for 28S rRNA gene amplification by Nested PCR

a) Primary reaction conditions



b) Nested PCR 1 (*P. vivax* specific) reaction conditions



b) Nested PCR 2 *(P. falciparum*-specific) reaction conditions:

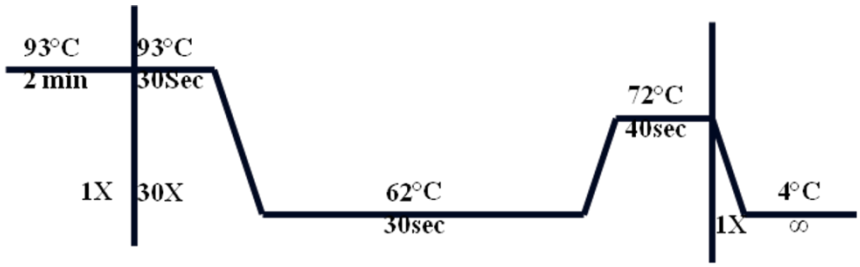


Figure 2.4: Nested PCR conditions for *P. vivax* and *P.falciparum*; PCR conditions for, a) primary reaction b) secondary nested PCR conditions specific for *P.vivax*, c) Secondary nested PCR conditions specific for *P. falciparum.*

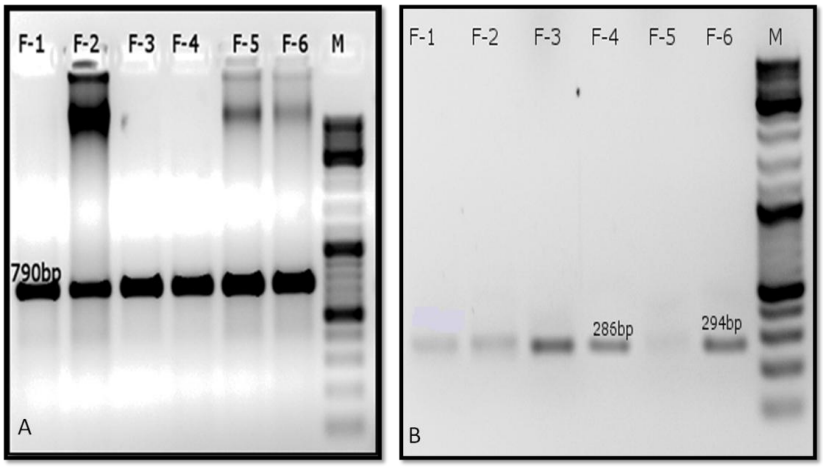


Figure 2.5: Parasite infection confirmed by 28S rRNA gene amplification (M=Gene ruler DNA ladder mix (SM0331 Fermentas); A) represents 28S rRNA gene amplification primary nested reaction of F1- F6 parasite isolates; B) Secondary nested PCR amplification where F1 and F-6 are *P. vivax* infected samples and ; F2, F3, F4, F5- *P. falciparum* Infection.

2.7- Amplification of selected genes from the co-expression based network:

2.7.1- Primer designing for the cloning and expression:

Primers were designed for genes selected from the co-expression based network by retrieving the sequences from online available *Plasmodium* database PlasmoDB (Aurrecoechea, et al., 2008) and NCBI (Altschul, et al.,1997). Primers for genes were designed using Primer-BLAST (Ye, et al., 2012) and the freely available software Gene Runner (Hasting Software's Inc., USA) Since we are using field isolates, the primers were checked against *Homo sapiens* to avoid any discrepancies in results. The presence

of any dimers and hairpin loop Tm were taken into consideration. The primers were synthesised commercially and log Tm was calculated using a standard formula-

$$\text{Log Tm} = 81.5+16.6\ [\log_{10}(Na^{+})] + [41\ (G+C)\ /n] - 675/n$$

(n = number of bases)

2.7.2-<u>Amplification of the selected genes:</u>

Amplification of PFL2320W and PFD0300W was done using genomic DNA of *P. falciparum* field isolates from complicated cases. Primers used to amplify 1.5 kb of PFD0300W, forward primer with restriction site XhoI at the 5' end of the primer and reverse primer with NcoI restriction site at the 5' end was used. To amplify PFL2320W, forward primer with restriction site of XhoI and reverse primer with restriction site of Kpn I was used. Primer sequence along with restriction sites are given in **table 2.6**.To amplify PFD0300W Phusion High fidelity polymerase from Thermo Fischer Scientific was used Denaturation and annealing temperature was as per primer pair according to the guidelines provided by the manufacturer. Amplification conditions used are set up according to the conditions given in the manual of Phusion Taq Polymerase, amplification conditions used are shown in **figure 2.6, 2.7.** Amplified fragment of genes are shown in **figure2.8.**

Gene Name	Primer sequence		R.E. sites	Size of amplicon
PFD0300W	5'ATAACTCGAGATGAAAGATCTTGTAGGAAGGTC3'	FP	XhoI	1500bp
	5'ATTCCCATGGTTATCGATCCATTTGCTTATTTGACC3'	RP	NcoI	
PFL2320W	5'ATCCCTCGAGATGAAAAAGTTATCC3'	FP	XhoI	858bp
	5'CAATGGTACCTTATGAAATGTATCCGCT3'	RP	KpnI	

Table 2.6: Primer sequence for selected genes along with the restriction sites

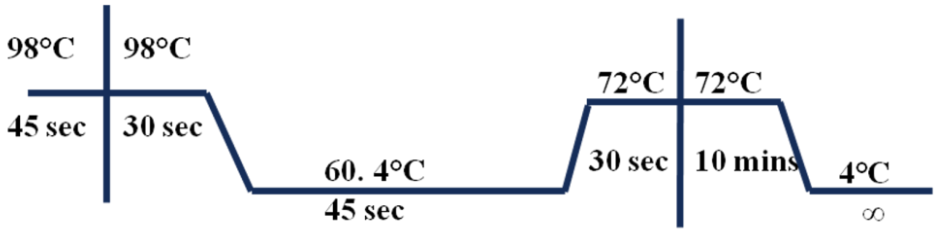


Figure 2.6: PCR reaction condition of PFD0300W

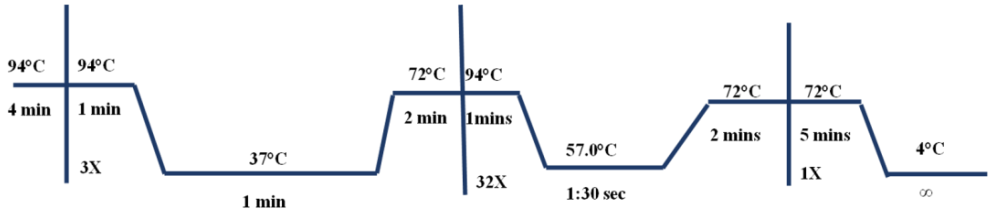


Figure 2.7: PCR condition for PFL2320W

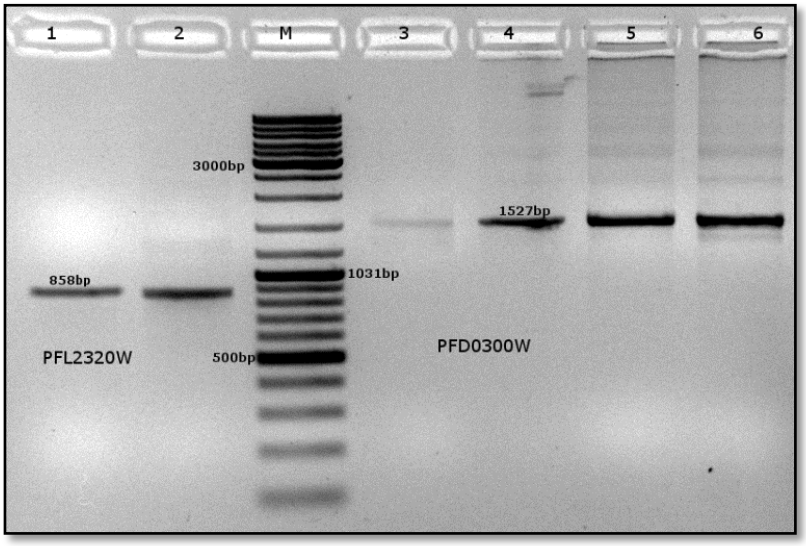


Figure 2.8: Amplified fragment of PFL2320W and PFD0300W-Lane 1 to 2 represents PFL2320W~858bp fragment, M=Marker (SM0331, Generuler DNA ladder), Lane 3 to 6 represents a PFD0300W amplification ~1527bp.

2.9-Cloning of PFD0300W and PFL2320W:

To characterise the desired proteins from the systems network, the full-length amplicon was obtained from *P. falciparum* field isolates and was cloned in cloning and expression vectors. The amplified fragment had restriction sites at their ends incorporated during primer designing according to the desired vector. Cloning is a multi-step process that requires competent cell preparation, Plasmid isolation, preparation of vector (pRSETA plasmid has been used) and insert (amplified gene) using restriction digestion, ligation, transformation and screening of recombinant clone.

2.9.1: Competent cell preparation:

Desired strains of *E. coli* host cells were streaked on an LB agar plate and grown overnight at 37°C. A single colony was picked up and inoculated in 5ml of LB broth, the culture was kept at 37°C in an incubator shaker with overnight shaking at 150 RPM. The subsequent day, overnight grown culture was inoculated in 1:40 ratio in 150 ml of fresh LB broth and kept at 37°C incubator shaker until O: D_{600} reached to 0.3-0.4 (~2 hrs). Once the required OD was achieved, the cells were harvested by centrifuging the culture in 50 ml Oakridge tubes at 6,000 RPM for 5 mins, at 4°C. The supernatant was discarded and the pellet was slowly re-suspended in 80 ml of 0.1 M calcium chloride ($CaCl_2$) and subsequently incubated on ice for 25 mins. After the incubation, the suspended pellet was again centrifuged at 10,000, RPM, at 4°C for 10 mins. The supernatant was discarded and the pellet was resuspended in 10 ml of ice-cold 0.1M $CaCl_2$ solution and 15% of 100% glycerol, subsequently, 200 µl of the complete suspension was distributed in pre-chilled microfuge tubes and kept frozen in - 80°C until further use.

2.9.2: Plasmid DNA isolation:

Single colonies from the LB agar plate were inoculated in 5 ml of LB broth with appropriate antibiotics and kept at 37°C incubator shaker for 12 -14 hrs, (standard time for bacteria to reach to late log phase), After 14 hrs, 3 ml of culture was harvested by centrifuging at 6,000 RPM for 5min at 4°C. Supernatant was discarded and the pellet was resuspended in 200 µl of solution I (GET buffer), by gentle vortexing. After the complete re-suspension, 150µl of solution II (SDS-NaOH) was added and immediately kept on ice for 2 mins, To this solution chilled solution III (Na-acetate solution) was

added and kept on ice for about 4 mins, after 4 mins the suspension was centrifuged at 10,000 RPM for 20mins. Subsequently, the supernatant was taken and transferred to a fresh microfuge centrifuge tube. RNAse treatment was given for the removal of RNA from the from the suspension by incubating it at 37 °C for 30 mins. Subsequently a traditional Phenol: Chloroform: Isomylalchol treatment was performed to remove proteins. After centrifugation at 4°C. The supernatants was taken into a fresh tube and the plasmid DNA was precipitated at -20°C overnight by adding 100% alcohol and 1/10th volume of sodium acetate (NaoAc) pH5.2. The precipitated DNA was pelleted by centrifuging at 10,000 RPM for 15 mins at 4°C. The supernatant was discarded and the pellet was washed with chilled 70% ethanol. The pellet was air dried and suspended in an adequate amount of 1XTE buffer. The integrity of plasmid DNA was checked on 1% agarose gel (Sambrook. et al., volume-1).

2.9.2.1: Preparation of Vector and Insert for Cloning:

Cloning of the desired gene requires digestion of vector and insert with proper restriction enzymes, for sticky end cloning. To clone PFL2320W the restriction enzymes XhoI and KpnI (NEB) were used and for PFD0300W XhoI and NcoI (NEB) has been used the reaction mixture is given in **table 2.7**. The vector was digested using both the enzymes and finally purified using a QIAquick gel extraction kit (QIAGEN, Germany) following the manufacturer's instructions. The insert was first cloned into a TA cloning vector (Ins TA clone PCR cloning kit by Thermo Fischer scientific) as per manufacturer's protocol. After the successful cloning in this, the insert was digested out from the vector by using selected enzymes and gel eluted from the gel successful cloning of vector and insert shown in **figure 2.9.**

Reagents	Quantity used
Template DNA (Plasmid	**500ng-µg**
Restriction enzyme	**3-5U**
Buffer (10X)	**1X buffer**
Nuclease-free water	**Quantity Sufficient**

Table2.7: Reaction mixture used for digesting template DNA

2.9.2.3: Gel elution of vector and Insert:

Purification of DNA fragments (digested plasmid and digested insert from TA-vector) was performed using a QIAquick gel extraction kit (QIAGEN, Germany) following the manufacturer's instructions. DNA fragment was excised from the gel using a clean scalpel and gel slices were taken into specific centrifuge tubes.Excised agarose slices were melted in 3 volumes of QI buffer to the weight of the gel, by keeping it at 50 degrees for 20 mins, with in-between vortexing for the proper mixing of the gel. After the proper melting of the gel, 1 gel volume of isopropanol was added and kept at room temperature for 2 mins. The suspension was then transferred to QIAquick column and spun at 13,000 RPM for 1 min. The flow-through was discarded and again 500µl of the QI buffer was added in the column and was centrifuged at 13,000 RPM for 1 min, flow-through was discarded. In the column, 750 µl of PE buffer was added and kept at room temperature for 5 mins. After 5 mins the column was centrifuged at 13,000 RPM for 1 min, flow through was discarded and empty spin was given to remove residual PE buffer in the column, Finally, the samples were eluted in 30 µ1X TE buffer by spinning the column at 13,000 RPM for 1 min, with an empty spin.

2.9.2.4: Ligation of Insert and Vector:

Purified plasmid and insert after double digestion were ligated with the help of T4 DNA ligase enzyme. An equal quantity of the vector and the insert were checked prior to the ligation, to compare the concentration. Three ligation reactions were set up, one with only vector to check the self-ligation of the vector and two reactions with double digested vector and insert at different ratios i.e. in the ratio of 1:3 and 1:6.

2.9.2.5: Ligation Mixture:

The intensity of the vector and insert were measured using UV/Visible spectrophotometer (Jasco), the nanogram concentration of vector was 11.95 ng/µl and insert intensity with 5ng/ µl. The ligation mixture was kept at 16 degrees overnight. Ligation mixture is shown in **table 2.8**

Reagents	Quantity used
Double digested vector	1.5 µl
Double digested insert	3.5 µl
Ligase buffer	2 µl
T4 DNA ligase enzyme	3 µl
Water	9.5 µl
total	20 µl

Table 2.8: Ligation mixture used for the ligation of vector and insert

2.9.2.6: <u>Screening of the Colonies for positive clones:</u>

All the colonies obtained were streaked in a master plate and simultaneously inoculated in LB broth with an appropriate antibiotic and the culture was incubated at 37°C for 12 -14 hours. After 14 hours cultures were used to isolate plasmid from the colonies as mentioned in materials and method section 2.10. All the plasmids were run on 1% agarose gel, and shifts were observed in colonies. Colonies with shifts in comparison to vector alone were restriction digested. PFL2320W were digested with XhoI and KpnI, and PFD0300W were digested with XhoI and NcoI. Digestion mixture for both the genes were kept at 37°C for 3 hours and fragments were analysed on agarose gel. Glycerol stock was made for the positive clones. Clones were then sequenced commercially to confirm the insert sequence. Recombinant construct in which clone is in proper frame were used for transformation in expression host for protein expression analysis.

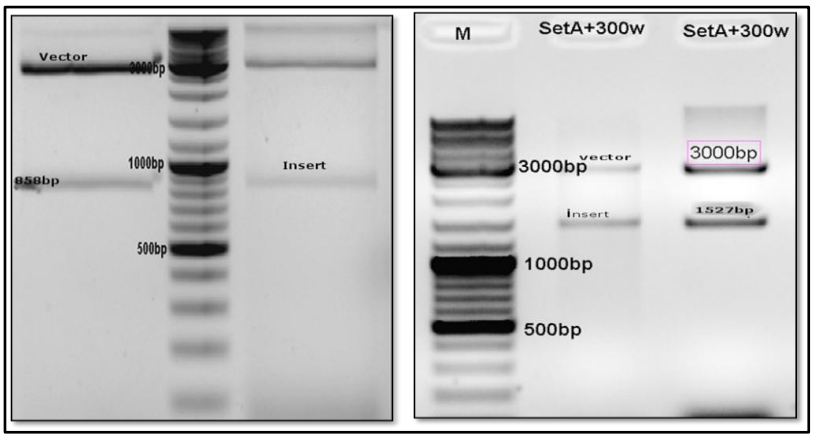


Figure2.9: **Plasmid digestion for the confirmation of successful cloning from the selected colonies**; First gel picture showing PFL2320W clone insert size 858bp and vector size pRSETA=3000bp, second gel successful cloning of PFD0300W, vector=pRSETA=3000bp and Insert=1527bp.

2.10- Protein Expression of selected genes:

Escherichia coli is one of the most considered organisms for the heterologous protein expression for the production of recombinant proteins, because of its fast doubling time even with 1/100 dilutions of starter culture cells can reach to stationary phase within few hours. Several *E.coli* strains are present that can be used for the expression of recombinant protein, *E. coli* strains used in this study with their characteristic features are shown below:

2.10.1-Protein expression using BL21D3, BL21 DE3 pLysS, Rosetta pLysS:

A single colony was picked from an LB agar plate and inoculated in LB media with specific selection marker. The culture was kept at 37°C incubator shaker for overnight with shaking speed of 200 RPM. After 14 hrs, two 20 ml flasks with fresh LB media was inoculated with primary culture in the ratio of 1:20 with the specific selection markers and incubated in an incubator shaker at 37°C for 3hrs.After 3 hrs when the OD_{600} reached at 0.7, one of the flasks with secondary culture was induced with 1mM IPTG. Samples were collected at every 2 hr, 4hr,6hr,8hr and 16 hr, samples were centrifuged at 6000 RPM for 10 mins, and pellet resuspended in 2x sample lysis buffer. The other flask of the secondary culture was used at the same time points for un-induced control. Conditions used for the expression of proteins are mentioned in **table-2.9.**

IPTG Concentration	Temp °C	Media Used	O.D at 600nm
0.5 mM	37,28,16	LB media, Terrific Broth, SOC media.	0.6-2.00
0.75 mM	37,28,16		
1 mM	37,28,16		
2 mM	37,28,16		

Table 2.9: Conditions used for the expression of both the proteins.

2.10.2: Protein expression using Arctic cells:

Arctic Express (DE3) Competent cells, derived from the high-performance Agilent BL21-Gold competent cells that enable high-level of heterologous expression of heterologous protein in *E.coli*. These cells are engineered to co-express cold-adapted chaperonins cpn10 and cpn60 isolated from psychrophilic bacterium, *Oleispira Antarctica*. These chaperonins have 74% and 54% identity with *E.coli* chaperonins, i.e GroEL and GroES respectively, and show high protein refolding activities at temperatures of 4-12°C (Agilent technologies, manual for Arctic Express Competent cells, (Gopal&Kumar., 2013; Hartinger, et al., 2010). A single colony was picked from LB agar plate and inoculated in LB media with Gentamycin sulphate (50mg/ml) and carbenicillin (35mg/ml). The culture was kept at 37°C incubator shaker for overnight at 200 RPM. After 14 hrs, 20 ml of fresh LB media was inoculated with primary culture in the ratio of 1:20 without the antibiotics and incubated in an incubator shaker at 30°for 3 hrs. After 3 hrs 1ml of culture was taken and used as a control. In the culture, antibiotics was and added subsequently the culture was kept 12°C temperature in incubator shaker for 10 mins, after which it was induced with 1mM IPTG, and again kept at 12°C in an incubator shaker for 24 hrs. After 24 hrs, the cells were harvested at 6000 RPM for 10 mins and pellet was resuspended in 2x sample lysis buffer.

2.11-Western Blotting:

A 12% SDS PAGE was run before western blotting for 3 hrs using a pre-stained marker (BR-Biochem, BMEE80500,11-245kDa),Semi-dry transfer method was used for transferring the protein bands onto the nitrocellulose membrane. For efficient transfer, all the blotting sheets and membranes were briefly soaked in chilled, transfer buffer, i.e. Tris Glycine Methanol (TGM) for at least 5-10 mins. For semi-dry transfer gel a

sandwich of blotting sheet, nitrocellulose membrane, SDS gel, blotting sheet was made, to remove the bubbles glass rod was used, the transfer was done at 15V for 40 mins. Once the transfer was completed the membrane was stained with Ponceau-S for 10 mins to check the efficiency of transfer and then washed thoroughly with autoclaved mili-Q water. The membrane was kept in blocking buffer (5% skim milk+ 0.05% tween 20) at 4°C overnight (~18 hrs). After washing 3X with TBST the primary antibody (anti His-tag antibody) in 1:200 dilution was used and kept for hybridization for 90 mins at 37°C. Again washing was repeated, after which the membrane was treated with secondary antibody (goat anti-mouse antibodies labelled with HRP) and kept at 37°C for 90 Mins. Washing was repeated again same as before, development of membrane was done by using HRP conjugate $TMBH_2O_2$ for membranes (Sigma Aldrich, USA).

2.12-Peptide immunisation and Immunolocalization:

Peptides for each of the proteins were selected using BcePred, prediction of linear B-cell epitopes, using physicochemical properties (Saha, & Raghava, 2006). Server 1.0 was used which is freely available through IMTECH (the Institute of Microbial Technology) at (www. imtech.res.in/raghava/ bcepred/). Peptide sequence is given in chapter V.

2.12.1-Peptide immunization:

To raise the polyclonal antibodies in swiss albino mice primary immunization was done using an emulsion of peptide and complete Freud's adjuvant, where each of the 6 mice was immunised with 100µg/100µl, of the peptide emulsion. Subsequently, booster doses were given on day 15, day 21, day 28 with peptide and IFA (Incomplete Freud's adjuvant) emulsion win same concentration of 100µg/100µl. After 35 days the mice were tail bled to collect the serum, the pre-bleed sera was also collected from each of the mouse as a control, which was further used for the ELISA and immunolocalization studies.

2.13-ELISA for antibody titres:

Both the peptides were coated on 96 well plate with a required 200ng that were made in 1X Coating buffer. Plates were covered with aluminium foil and kept at 4°C for 18 hours. Subsequently, coating buffer was discarded and plates were washed using 1X PBS tween 20 in ELISA washer 3 times for a total of 15 mins i.e., 5 mins for each

wash. After washing plates were again coated with 100µl of primary antibody (serum) in required dilutions in triplicates and kept at 37°C in a hybridization chamber for 2 hours. After 2 hours, washing was again performed as mentioned in step 2. Subsequently, plates were again coated this time with secondary antibody (1:1000) dilution and plates were kept at 37°C in a hybridization chamber for 2 hours. Washing step was performed again 3 times as mentioned in step 2.100µl of the 1X substrate solution was added to each well. (This step was performed in the dark as the substrate is light sensitive). Incubated at RT till the blue colour appear. 100 µl of stop solution was added to stop the reaction (the samples turned yellow in colour after adding stop solution). Reading was taken at 450nm.

2.14: Immunolocalization of the proteins with specific antibodies raised against peptides:

For immunolocalization studies using sera on the basis of antibody titre monitored using ELISA 1:200 primary antibody dilution was used. Different trackers were used for localization of two proteins in paraformaldehyde fixed slides.

2.14.1- Paraformaldehyde fixation of infected culture slides: (Bengtsson, et al., 2008)

10ul of culture was placed on a slide, the smear was made as an usual thin smear. Slides were allowed to dry for approximately 5-10 mins. Slides were fixed by covering the smear with 4% of paraformaldehyde for 10 mins. After the complete fixation, the slides were rinsed with 1x PBS just by dipping the slide in the solution**.**

2.14.2-Immunolocalization of E1 and E2 (PFD0300W):

4% PFA prefixed slides were washed with 1xPBS for 10 mins, and air dried for 5 mins. After the slides were completely air dried, smear was covered with 0.2% saponin and incubated for 15 mins at room temperature (RT). After this, the slides were washed with 1 x PBS gently for 5 mins x 3 times. Slides were blocked with 3 % BSA solution for 60 mins. Further the slides were treated with approximately 750 µl to 1 ml of primary antibody (1:200 dilution) and incubated for 90 mins at room temperature (RT), after which they were washed with 1X PBS for 10 mins at R.T with 1X PBST for 15 mins at RT, again with 1XPBS at R.T for 10 mins. Subsequently, secondary antibody (goat anti-mouse IgG antibody), Alexa-flour-594 (Invitrogen, Molecular Probes) was added in 1:1000 dilution and incubated for 90 mins at room temperature. Followed by

washing step as mentioned previously. ER tracker was used for E1-300w to rule out the possibility of protein being localised in the ER of the parasites, 0.5µM working concentration was used. Slides with ER tracker were kept for 90 mins at RT. Again 1X PBS and 1XPBST wash was given. Further, the slides were incubated in DAPI solution for 10 mins. Again washing step was repeated as mentioned previously with 1XPBS and 1XPBST. After complete washing, the glass slides were kept for air drying for 5mins, after which they were mounted with vectashield (Vector Laboratories, Inc. Burlingame, CA94010).

2.14.3- Immunolocalization of E-1A and E-3 (PFL2320W):

PFA prefixed slides were washed with 1xPBS for 10 mins, and air dried for 5 mins after the slides were completely air dried, the smear was covered with 0.2% saponin and slides were incubated for 15 mins at R.T. After 15 mins the slides were washed with 1XPBS gently for 5m ins, 3 times each. Slides were covered with 3 % BSA solution for 60 mins. Further, the slides were covered with approximately 750µl to 1 ml of primary antibody (1:200 dilution) and incubated for 90 mins at R.T. Slides were washed with 1X PBS for 10 mins at R.T with 1X PBST for 15 mins at R.T, again with 1XPBS at R.T for 10 mins. After washing, the secondary antibody (goat anti-mouse IgG antibody) FITC was added in 1:1000 dilution and incubated for 60 mins at R.T. After the incubation with secondary antibody washing step was repeated as mentioned previously. For E1A-2320w, another secondary antibody was added i.e. Qdot (Invitrogen) which is a streptavidin conjugate and binds to the biotin present in the apicoplast membrane. Further, the slides were incubated for 90 minutes. 1XPBS and 1XPBST was given. To rule the possibility of protein localised in mitochondria MitoTracker®Red FM (Invitrogen™ molecular probes®) was also used in 0.5µM concentration and slides were incubated for 90 minutes. After washing, subsequently, the slides were incubated in DAPI solution for 10 mins. 1XPBS and 1XPBST wash was given. After complete washing the glass slides were kept for air drying for 5mins, after which a drop of Vectashield (Vector Laboratories, Inc. Burlingame, CA 94010) anti-fade was placed on the stained slides and covered with the coverslip after the vectashield was spread out uniformly on the smear, the coverslip was sealed using the transparent nail paint.

2.16-In-silico analysis of the selected genes:

In-silico analysis of the proteins was done using different online available tools for deciphering the characteristics of the proteins. The in-silico analyses are as follows.

2.16.1-Conserved Domain analysis:

The protein sequences of both the proteins were submitted to NCBI CDD-BLAST tool (Marchler-Bauer, et al., 2005) (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) for searching conserved domain architecture. The domain was cross-checked with other domain database tool InterProScan (Zdobnov, et al., 2001) and Pfam (Bateman, et al., 2004).

2.16.2: Determination of Codon adaptation index:

The expression probability of the conserved *Plasmodium* protein with unknown function protein sequences was revealed by using codon adaptation index (CAI) calculated by CAI cal tool (Puigbò, et al., 2008) (http://genomes.urv.es/CAIcal/) Prediction of subcellular localisation, signal peptide and physicochemical characterization of PFL2320W and PFD0300W was performed with *Plasmodium* specific tools like PlasmoAP (Foth,et al., 2003), PLASmit and CELLO (Yu, et al., 2004; Yu, et al., 2006) (Peterson., 2011) For signal peptide prediction Signal P.4.0 was used, physicochemical characterization of proteins were done using ExPASy ProtParam tool (Gasteiger,etal., 2005) (http://web.expasy.org/cgi-bin/ protparam/ protparam).

2.16.3-Analysis of Disorderdness of Proteins:

Analysis of disordered amino acids in the protein was done using PONDR (Dunker, et al., 2002), DisEMbL™ (Linding, et al., 2003), Disopred (Ward, et al., 2004), ANCHOR (Dosztányi,et al., 2009), IU-PRED (Dosztányi, et al., 2005), that suggests that the protein is disordered in certain regions because of the presence of hydrophilic amino acid stretches.

2.16.4-Transmembrane Domain prediction:

To predict the transmembrane domain in both the proteins tools like TMHMM available at (http://www.cbs.dtu.dk/services/TMHMM/), PHOBIUS (http://phobius.

sbc.su.se/cgibin/predict.pl) (Käll, et al., 2007) and TMPRED from Expasy tool (http://www.ch.emb net.org/software/TMPRED_form.html) were used.

2.16.5-Secondary structure prediction:

The secondary structure prediction of selected proteins was done by using PsiPred (McGuffin, et al., 2000). Homology modelling of the selected proteins were not be able to achieve because of the absence of suitable homologous template of the selected proteins with sequence identity of more than 25%.

Chapter-III

Comparative genomic hybridization of Plasmodium falciparum clinical isolates

It is generally accepted that there is a difference in *in-vivo* and *in-vitro* conditions from the perspective of studying the malarial parasites. In *in-vivo* condition, the parasite is subjected to host immune and other physiological pressures, which makes it express different genes under different environmental conditions. These changes at the transcriptomic level could be related to the genomic level differences which could lead to the complicacy of the disease. However, there is limited information on whether parasites causing different disease manifestations *in-vivo* have significant differences at the genome level. Therefore, there is a need to identify genome differences in parasites isolated from exhibiting severe manifestations in comparison to those from patients showing uncomplicated malaria. This chapter focuses on the results obtained from the Comparative Genomic Hybridization (CGH) data analysis of Pf-isolate-2 (complicated) in comparison with the Pf-Isolate-1 (uncomplicated) and details about the variations (loss/gain) present

different regions of the genome of Pf-isolate-2. A detailed description of significant regions that are showing aberrations of more than 1kb, like, sub-telomeric regions and amplifications and deletions other than the subtelomeric regions are reported in this chapter.

3.1 Introduction:

The genome sequencing of *P. falciparum* 3D7 culture isolate revealed that the genome size of the parasite is about 22.8 Mb and it comprises of 14 chromosomes with approximately 5,690 genes, of these 5,362 genes are protein encoding 60% still designated as hypothetical proteins (Gardner, et al., 2002; Ramaprasad, et al., 2012). To analyze the differences in the genome of parasite isolates showing the different manifestation, aCGH tiling array was designed based on the annotated 3D7 as a reference genome. The size of each chromosome with the number of genes reported are as per PlasmoDB V.6.0 (Aurrecoechea, et al., 2008). The regions that are re-annotated in the genome assembly of *P. falciparum* 3D7 and have been updated in a recent version of PlasmoDB i.e PlasmoDB V.34 were eliminated from the study to rule out any inadvertent errors due to the re-annotation processes. The ADM-2 algorithm which is one of the robust calling algorithms (Roy & Reif., 2013) in the Agilent genomic work Bench, was used to find the genomic aberrations, using filtering criteria of a minimum of 5 probes and the average log fold change of 0.3. The probe based amplifications or deletions were analysed to report the significant genome variations. In our analysis, we have seen variation in all the 14 chromosomes of Pf-isolate-2, including the intergenic regions. In this data, the term amplification and deletion are reported as a variation of test strain Pf-2 compared to Pf-1. Deletions were the regions where the signal intensities of the probes in the test strain are near background and represent the absence of hybridization. In some of the genes, the variations are already reported and some are novel variations.

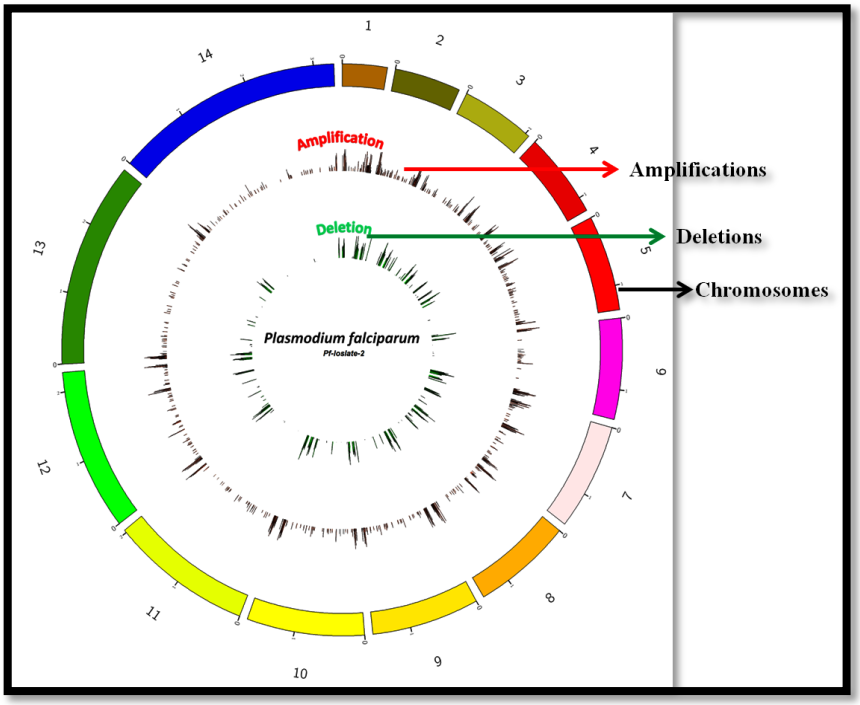


Figure-3.1- Circos plot of the whole genome variation in Pf-isolate-2: Circle of Genome variation of *P. falciparum* (Pf-2), the outer coloured region represents the number of chromosomes showing probe based chromosome wise Copy number variations. The outer red bars are showing amplifications i.e. regions with a Cy3/Cy5 ratio of ≥ 0.3 the inner green bars are showing deletions present in the region of the Chromosomes. i.e. regions with Cy3/Cy5 ratio≤-0.3. Where Cy3 represent the signal intensity of Pf-2 and Cy5 represents the signal intensity of Pf-2.

3.2-Genome wide gene Copy Number Variation in Pf-2 in comparison with Pf-1:

Regions showing variation comprises total of 687 unique genes in the genome of Pf-isolate-2. The loss/gains varied in size from 100 nucleotides to 47,000 bp including intra and intergenic regions. Out of 687 genes, amplifications was seen in 511 genes and deletions in 176 genes. There are 27 genes in the regions of Chromosome-4, 6,7,8,10,11,12,13 in Pf-2 in which there is the presence of both amplifications/gains and deletions/loss. These regions are rich in Variant Surface Antigen family genes. The deletion and amplifications covered smaller regions and did not cover more than two contiguous genes. Most of the variation was seen in the sub-telomeric and telomeric regions.

3.3- Major Variations in Variant Surface Antigens (VSA):

Our analysis revealed the presence of insertions and deletions in regions covering for 268 unique VSA gene families, out of which variation is present in 58 unique var genes in Pf-2 in comparison to Pf-1, Most of these genes showing CNVs belong to var group B and C. A previous report from our group has also reported differential expression in the transcriptome of these family genes in the isolates causing severe malaria (Subudhi, et al ., 2015). A Deletion of 16.3 kb (563139- 579488) and 11.7 kb are present at the right end of the chromosome-1 of Pf-isolate-2 that covers for 5 genes and 4 genes, respectively description of genes and segments are presented in **table 3.1**, **figure 3.2** and **figure3.3,** In previous reports, SNPs have been detected in *Plasmodium* exported protein (hyp10) (PFA0700C) in malaria parasite from Guinean and Gambian populations (Mobegi et al., 2014). A study by Mok, et al in 2008 have shown a segmental duplication of PFA0675W, PFA0685C, PFA0690W and PFA0700C in different strains and the same group have also reported the duplication of PFA0685C, PFA0690W and PFA0695C located at the right end of the chromosome number-1 of 3D7 in a clinical isolate (UAM25) (Ribacke,et al., 2007; Mok, et al., 2008). The region in which the above genes fall shows deletion in the chromosome in our clinical isolate showing the complicated manifestation (Pf-2). The g-processed signal intensity represents the background intensity of Cy3 which represents the Pf-isolate-2 hybridization signal and r-processed signal intensity represents the hybridization intensity of Pf-isolate-1. The signal intensity of r processed signal is more than g processed signal intensity but is near to background which suggests that there is no amplification in Pf-2 (test strain) as shown in **table 3.2**, **figure3.4**. A deletion of 11.9 kb region in chromosome-3 is present that includes conserved *Plasmodium* protein with unknown function (PFC0325C, PFC0330W, PFC0335C). It is interesting to note that although these genes are of unknown function their annotated GO functions suggest that these genes are involved in zinc ion binding activity, receptor activity and in ATP binding activity. Other than above, stated deletions major VSA deletions are also present in other regions of chromosomes and are represented in the tabular format. A deletion of 14.28 kb, 13.2 kb has been seen on chromosome-11 and chromosome-13 apart from this an amplification of 13.2 kb is present in the right end of the chromosome-11 **table-3.2**. Details of the mentioned regions and other regions covering VSA, are shown in **figure 3.5**.

Chromosome	Number of genes	Size of variation in kB	Gene symbols	Description	Variation
Chr-1	5	16.3	PFA0705c, PFA0710c, PFA07 5c, PFA0720w	Stevor pseudogene, *rifin* (RIF, Plasmodium exported protein (hyp7), unknown function, Plasmodium exported protein, unknown function pseudogene.	Deletion
Chr-1	4	11.7	PFA0685c, PFA0690w, PFA0695c, PFA0700c	exported protein family-3 (EPF-3) and exported protein family-4 (EPF-4), Erythrocyte Membrane Protein-1 (PfEMP1), exon2, pseudogene (VAR), Plasmodium exported protein (hyp10) of unknown function.	Deletion

Table 3.1: Deletion present in chromosome number-1, segments with VSA (Variant Surface Antigen).

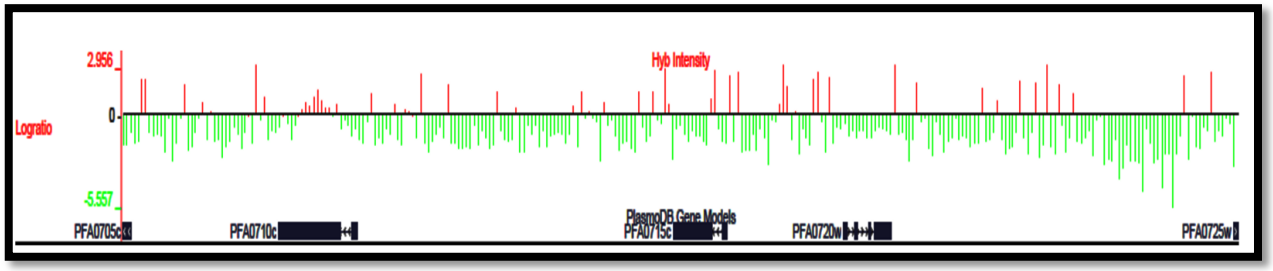


Figure 3.2; Deletion of 16.3 kb present in chromosome-1; Green bar is showing deletion i.e Cy3/Cy5 ratio<-0.3, and red bar is showing amplification i.e Cy3/Cy5 ratio> +0.3, blue bars are showing genes and arrow heads as the direction of transcription of the genes, genes present in this region are PFA0705C, PFA0710C, PFA0715C, PFA0720W, PFA0725W.

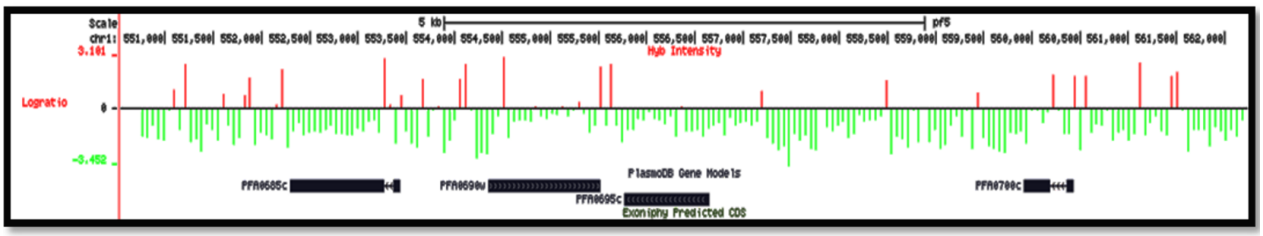


Figure 3.3: Deletion of 11.7 kb chromosome-1segment: Continous deletion is present in chromosomal segments covering genes PFA0685C, PFA0690W, and PFA0695C .The green bar is showing deletions i.e Cy3/Cy5 ratio is <-0.3, whereas regions with some of the probes showing Cy3/Cy5 ratio >+ 0.3 in the regions with red bar is showing amplification.

Chr	Number of genes	Size of Variation in kb	Gene symbol	Description	Deletion/ Amplification
Chr-3	3	11.9	PFC0325C, PFC0330W, PFC0335C	Conserved Plasmodium protein with unknown function	Deletion
Chr-11	6	14.28	PF11_0010,PF11_0011,PF11_0012,PF11_0013,PF11_0014,PF11_00115	Rifin(RIF)RESA like protein pseudogene, Plasmodium exported protein(PHISTa),unknown fucntion,pseudogene,Pfmc-2TM Maurer's cleft two transmemberane(MC-2TM),RESA like protein,pseudogene.	Deletion
Chr-13	5	13.2	MAL13P1.495,MAL13P1.500MAL13P1.505,MAL13P1.510 MAL13P1.515	Rifin(RIF),Rifin(RIF),Stevor Erythrocyte memberane protein 1(PfEMP1)exon 2, pseudogene (VAR) Rifin(RIF)	Deletion
Chr11	5	13.2	PF11_0515,PF11_0516,PF11_0517,PF11_0518,PF11_0519	Rifin(RIF),Stevor,Rifin(RIF) Rifin(RIF),pseudogene,Rifin(RIF)	Amplification

Table 3.2: Variations present in the chromosomal segment contains VSA (Variant Surface Antigen).

Continous deletion region with small amplifications of 16.3 Kb comprises of 4 genes log ratio is -ve for this region log ratio = -0.905391					
ProbeName	GeneName	gProcessedSignal	rProcessedSignal/Pf-	gProcessedBackground	rProcessedBackground/Pf-1
GT_Pfalci_Pf3D7_01_10057	chr1:563137-563196	645	2290	22.6975	54.479
GT_Pfalci_Pf3D7_01_10058	chr1:563193-563252	71.2	237	27.2107	46.7165
GT_Pfalci_Pf3D7_01_10059	chr1:563249-563308	325	677	27.8708	53.2288
GT_Pfalci_Pf3D7_01_10060	chr1:563305-563364	306	956	24.1864	51.2401
GT_Pfalci_Pf3D7_01_10061	chr1:563361-563420	530	1510	26.4336	49.9648
GT_Pfalci_Pf3D7_01_10062	chr1:563417-563476	103	23.1	28.0167	52.6667
GT_Pfalci_Pf3D7_01_10063	chr1:563473-563532	66.3	14.4	24.2863	53.5242
GT_Pfalci_Pf3D7_01_10064	chr1:563529-563588	201	399	25.7113	50.8866
GT_Pfalci_Pf3D7_01_10065	chr1:563585-563644	308	726	24.5583	49.5292
GT_Pfalci_Pf3D7_01_10066	chr1:563641-563700	1940	4310	23.2852	50.6778
GT_Pfalci_Pf3D7_01_10067	chr1:563697-563756	3300	7640	25.811	48.5118
GT_Pfalci_Pf3D7_01_10068	chr1:563753-563812	45.4	213	24.8583	51.0551
GT_Pfalci_Pf3D7_01_10069	chr1:563809-563868	45.7	44.5	23.0531	52.2345
GT_Pfalci_Pf3D7_01_10070	chr1:563865-563924	648	4340	24.175	48.2042
GT_Pfalci_Pf3D7_01_10071	chr1:563921-563980	386	1240	23.8083	52.7744
GT_Pfalci_Pf3D7_01_10072	chr1:563977-564036	45.8	13.7	22.9157	49.7791
GT_Pfalci_Pf3D7_01_10073	chr1:564033-564092	56.4	15.7	25.5858	58.4979
GT_Pfalci_Pf3D7_01_10074	chr1:564089-564148	507	2190	24.8935	49.6122
GT_Pfalci_Pf3D7_01_10075	chr1:564145-564204	179	628	25.4859	50.7791
GT_Pfalci_Pf3D7_01_10076	chr1:564201-564260	4460	8940	22.7456	50.0658
GT_Pfalci_Pf3D7_01_10077	chr1:564257-564316	45.7	14.8	25.6612	52.2612
GT_Pfalci_Pf3D7_01_10078	chr1:564313-564372	148	84.4	27.0858	53.0901
GT_Pfalci_Pf3D7_01_10079	chr1:564369-564428	301	799	24.6349	47.7024
GT_Pfalci_Pf3D7_01_10080	chr1:564425-564484	117	93.5	27.8932	52.4021
GT_Pfalci_Pf3D7_01_10081	chr1:564481-564540	414	1160	26.7946	53.4535
GT_Pfalci_Pf3D7_01_10082	chr1:564537-564596	481	1330	22.8876	50.494

GT_Pfalci_Pf3D7_01_10083	chr1:564593-564652	69.1	394	24.417	48.4332
GT_Pfalci_Pf3D7_01_10084	chr1:564649-564708	57.9	208	23.8931	48.458
GT_Pfalci_Pf3D7_01_10085	chr1:564705-564764	1200	3480	24.0952	56.9405
GT_Pfalci_Pf3D7_01_10086	chr1:564761-564820	161	258	24.6908	52.0305
GT_Pfalci_Pf3D7_01_10087	chr1:564817-564876	473	1060	26.8722	50.7895
GT_Pfalci_Pf3D7_01_10088	chr1:564873-564932	122	497	24.5627	50.6844
GT_Pfalci_Pf3D7_01_10089	chr1:564929-564988	123	241	25.4597	49.8957
GT_Pfalci_Pf3D7_01_10090	chr1:564985-565044	108	95.9	19.8281	45.1598
GT_Pfalci_Pf3D7_01_10091	chr1:565041-565100	519	1580	25.1631	50.8227
GT_Pfalci_Pf3D7_01_10092	chr1:565097-565156	125	15.3	27.9184	52.7128
GT_Pfalci_Pf3D7_01_10093	chr1:565153-565212	171	197	26.7623	51.3245
GT_Pfalci_Pf3D7_01_10094	chr1:565209-565268	108	50.1	24.0423	52.4308
GT_Pfalci_Pf3D7_01_10095	chr1:565265-565324	380	1030	25.556	54.0388
GT_Pfalci_Pf3D7_01_10096	chr1:565321-565380	435	834	23.5	53.8968
GT_Pfalci_Pf3D7_01_10097	chr1:565377-565436	1830	3590	26.2789	51.5976
GT_Pfalci_Pf3D7_01_10098	chr1:565433-565492	998	1620	24.8185	49.7339
GT_Pfalci_Pf3D7_01_10099	chr1:565489-565548	11400	10100	23.416	49.9237
GT_Pfalci_Pf3D7_01_10100	chr1:565545-565604	4650	6650	26.8739	52.7658
GT_Pfalci_Pf3D7_01_10101	chr1:565601-565660	1340	3630	23.5812	47.9025
GT_Pfalci_Pf3D7_01_10102	chr1:565657-565716	4510	6620	24.0782	44.465
GT_Pfalci_Pf3D7_01_10103	chr1:565713-565772	7320	6670	27.1552	49.6679
GT_Pfalci_Pf3D7_01_10104	chr1:565769-565828	6820	5240	24.3435	54.5573
GT_Pfalci_Pf3D7_01_10105	chr1:565825-565884	9500	5460	24.3692	47.4014
GT_Pfalci_Pf3D7_01_10106	chr1:565881-565940	22600	15400	27.3047	51.4609
GT_Pfalci_Pf3D7_01_10107	chr1:565937-565996	26200	12200	25.1916	53.0881
GT_Pfalci_Pf3D7_01_10108	chr1:565993-566052	135000	48200	24.0853	50.7016
GT_Pfalci_Pf3D7_01_10109	chr1:566049-566108	26500	13800	24.9838	50.0526
GT_Pfalci_Pf3D7_01_10110	chr1:566105-566164	15100	10700	25.075	52.0417

GT_Pfalci_Pf3D7_01_10111	chr1:566161-566220	32700	23200	24.1115	49.5962
GT_Pfalci_Pf3D7_01_10112	chr1:566217-566276	14600	14600	23.8478	49.387
GT_Pfalci_Pf3D7_01_10113	chr1:566273-566332	18000	11600	24.1144	50.4703
GT_Pfalci_Pf3D7_01_10114	chr1:566329-566388	257	457	23.7241	50.5603
GT_Pfalci_Pf3D7_01_10115	chr1:566385-566444	367	440	22.5965	51.4605
GT_Pfalci_Pf3D7_01_10116	chr1:566441-566500	73.9	111	24.2326	52.3178
GT_Pfalci_Pf3D7_01_10117	chr1:566497-566556	615	1400	23.6142	52.874
GT_Pfalci_Pf3D7_01_10118	chr1:566553-566612	1610	2740	25.0411	48.6941
GT_Pfalci_Pf3D7_01_10119	chr1:566609-566668	340	933	24.8447	52.3333
GT_Pfalci_Pf3D7_01_10120	chr1:566665-566724	527	1710	28.3254	52.6349
GT_Pfalci_Pf3D7_01_10121	chr1:566721-566780	1590	2110	24.0179	50.675
GT_Pfalci_Pf3D7_01_10122	chr1:566777-566836	123	47.7	24.508	48.868
GT_Pfalci_Pf3D7_01_10123	chr1:566833-566892	220	762	21.682	48.0335
GT_Pfalci_Pf3D7_01_10124	chr1:566889-566948	84.5	219	25.5341	53.5703
GT_Pfalci_Pf3D7_01_10125	chr1:566945-567004	411	1290	24.0868	50.8906
GT_Pfalci_Pf3D7_01_10126	chr1:567001-567060	817	1460	23.6234	48.5065
GT_Pfalci_Pf3D7_01_10127	chr1:567057-567116	558	1180	24.5598	48.5769
GT_Pfalci_Pf3D7_01_10128	chr1:567113-567172	67.8	42.5	24.8794	53.2529
GT_Pfalci_Pf3D7_01_10129	chr1:567169-567228	1080	3080	24.1731	51.1846
GT_Pfalci_Pf3D7_01_10130	chr1:567225-567284	4120	13600	24.3274	51.4126
GT_Pfalci_Pf3D7_01_10131	chr1:567281-567340	210	161	27.1174	48.6932
GT_Pfalci_Pf3D7_01_10132	chr1:567337-567396	220	188	27.5394	53.437
GT_Pfalci_Pf3D7_01_10133	chr1:567393-567452	170	156	28.9619	50.4534
GT_Pfalci_Pf3D7_01_10134	chr1:567449-567508	815	2130	22.5643	49.2075
GT_Pfalci_Pf3D7_01_10135	chr1:567505-567564	75.7	14	24.9412	54.2078
GT_Pfalci_Pf3D7_01_10136	chr1:567561-567620	149	483	22.3306	52.7959
GT_Pfalci_Pf3D7_01_10137	chr1:567617-567676	45.3	215	28.4133	54.2509
GT_Pfalci_Pf3D7_01_10138	chr1:567673-567732	502	1500	26.5	52.5984

Figure 3.4: Raw signal intensity of the continuous deletion region of size 16.3 kb; Column1; Probe set, **Column2;**Chromosome region, **Column3;** g-processed signal intensity (Pf-2),**Column4;** r-Processed signal intensity (Pf-1), **Column5;**g-processed background subtracted signal intensity(Pf-2), **Column6;** r-processed signal intensity(Pf-1).

Chr	Start	Stop	Size	#Probes	#Gains	#Losses	#Genes	Genenames	Updated Gene Id	strand	Amplificat	Deletion	pval	Description
chr1	31865	33432	1567	28	1	0	1	PFA0005w	PF3D7_0100100	forward	1.521611	0	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr1	45361	47155	1794	32	1	0	1	PFA0015c	PF3D7_0100300	reverse	2.059596	0	2.602E-43	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr1	613985	614771	786	14	1	0	1	PFA0765c	PF3D7_0115700	reverse	2.52213	0	5.508E-52	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr2	25201	26992	1791	32	1	0	1	PFB0010w	PF3D7_0200100	forward	2.724216	0	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr2	31081	33099	2018	36	0	1	1	PFB0010w	PF3D7_0200100	forward	0	-0.92443	1.962E-22	erythrocyte membrane protein 1, PfEMP1 (VAR)
						1		PFB1055c	PF3D7_0223500	reverse				erythrocyte membrane protein 1, PfEMP1 (VAR)
chr3	33937	38867	4930	88	1	0	1	PFC0005w	PF3D7_0300100	forward	1.194316	0	2.399E-63	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr3	1E+06	1032811	2130	38	1	0	1	PFC1120c	PF3D7_0324900	reverse	1.021388	0	2.468E-37	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	35057	40267	5210	93	1	0	1	PFD0005w	PF3D7_0400100	forward	1.325018	0	2.967E-88	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	53089	55835	2746	49	0	1	1	PFD0020c	PF3D7_0400400	reverse	0	-1.17578	4.482E-42	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	55945	63507	7562	135	1	0	1	PFD0020c	PF3D7_0400400	reverse	2.212176	0	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	555465	560675	5210	93	1	0	1	PFD0615c	PF3D7_0412400	reverse	0.816015	0	5.233E-58	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	569969	570531	562	10	0	1	1	PFD0625c	PF3D7_0412700	reverse	0	-1.53324	2.482E-14	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	570641	574843	4202	75	1	0	1	PFD0625c	PF3D7_0412700	reverse	0.849752	0	1.685E-53	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	586489	590187	3698	66	1	0	1	PFD0630c	PF3D7_0412900	reverse	1.040007	0	2.51E-58	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	601161	606091	4930	88	1	0	1	PFD0635c	PF3D7_0413100	reverse	0.995241	0	1.992E-75	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	941193	943995	2802	50	1	0	1	PFD0995c	PF3D7_0420700	reverse	1.159983	0	7.231E-42	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	954689	955923	1234	22	1	0	1	PFD1000c	PF3D7_0420900	reverse	0.955563	0	6.304E-20	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	967233	967739	506	9	1	0	1	PFD1005c	PF3D7_0421100	reverse	1.962016	0	5.775E-26	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	975521	980003	4482	80	1	0	1	PFD1015c	PF3D7_0421300	reverse	0.964774	0	3.416E-65	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	1E+06	1163736	4087	73	1	0	1	PFD1235w	PF3D7_0425800	forward	1.646792	0	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	1E+06	1179083	618	11	1	0	1	PFD1245c	PF3D7_0426000	reverse	2.116132	0	2.448E-16	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr4	1E+06	1183227	1850	33	1	0	1	PFD1245c	PF3D7_0426000	reverse	1.305608	0	3.489E-25	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr5	22793	23859	1066	19	1	0	1	PFE0005w	PF3D7_0500100	forward	2.264525	0	8.976E-53	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr6	7171	10867	3696	65	1	0	1	PFF0010w	PF3D7_0600200	forward	1.831645	0	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr6	724977	725595	618	11	1	0	1	PFF0845c	PF3D7_0617400	reverse	1.383315	0	2.187E-14	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr6	729233	730803	1570	28	1	0	1	PFF0845c	PF3D7_0617400	reverse	2.780359	0	7.521E-71	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr6	1356545	1357888	1343	24	0	1	1	PFF1580c	PF3D7_0632500	reverse	0	-2.13537	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr6	1361137	1363491	2354	42	1	0	1	PFF1580c	PF3D7_0632500	reverse	1.532143	0	2.757E-47	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr6	1376313	1378835	2522	45	1	1	1	PFF1595c	PF3D7_0632800	reverse	0	-1.91032	6.766E-44	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr7	30857	34216	3359	60	1	0	1	MAL7P1.212	PF3D7_0833500	reverse	1.303276	0	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr7	569465	580667	11202	200	1	1	2	PF07_0048	PF3D7_0711700	reverse	1.229069	0	1.831E-93	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr7	584977	585875	898	16	1	0	1	PF07_0049	PF3D7_0712000	reverse	2.121782	0	3.726E-41	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr7	599705	600323	618	11	0	1	1	MAL7P1.50	PF3D7_0712300	reverse	0	-1.33136	8.911E-16	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr7	602393	604187	1794	32	1	0	1	MAL7P1.50	PF3D7_0712300	reverse	1.43832	0	8.974E-38	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr7	609561	614435	4874	87	1	0	1	PF07_0050	PF3D7_0712400	reverse	3.982155	0	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr7	624849	628099	3250	58	1	1	1	PF07_0051	PF3D7_0712600	reverse	0	-2.67204	3.277E-29	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr7	640417	643275	2858	51	1	0	1	MAL7P1.55	PF3D7_0712800	reverse	1.656472	0	1.375E-94	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr7	648817	650723	1906	34	1	0	1	MAL7P1.56	PF3D7_0712900	reverse	1.10914	0	5.584E-33	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr7	1476105	1483552	7447	133	1	2	1	MAL7P1.187	PF3D7_0733000	reverse	0	-3.49465	7.131E-45	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr8	1	23187	23186	414	1	1	1	PF08_0142	PF3D7_0800100	forward	0	-1.86987	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr8	23409	24587	1178	21	1	1	1	PF08_0142	PF3D7_0800100	forward	0	-2.76965	4.884E-38	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr8	32425	39144	6719	120	1	0	1	PF08_0141	PF3D7_0800200	reverse	1.456676	0	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr8	42563	49619	7056	125	1	0	1	PF08_0140	PF3D7_0800300	forward	1.52152	0	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr8	432265	437699	5434	97	1	0	1	PF08_0107	PF3D7_0808600	forward	0.960766	0	2.366E-68	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr8	441225	444475	3250	58	1	0	1	PF08_0106	PF3D7_0808700	forward	1.805731	0	2.377E-123	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr8	462057	462675	618	11	0	1	1	PF08_0103	PF3D7_0809100	forward	0	-3.02119	1.498E-37	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr8	1392611	1411200	18589	331	0	1	1	MAL8P1.220	PF3D7_0700100	forward	0	-0.97134	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr9	22177	25483	3306	59	1	0	1	PFI0005w	PF3D7_0900100	forward	1.534178	0	3.531E-73	erythrocyte membrane protein 1, PfEMP1 (VAR)

chr9	1484785	1489155	4370	78	1	0	1	PFI1820w	PF3D7_0937600	forward	1.563857	0	1.652E-50	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr9	1489825	1492459	2634	47	0	1	2	PFI1820w	PF3D7_0937600	forward	0	-1.84393	2.48E-91	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr9	1497833	1498899	1066	19	1	0	1	PFI1830c	PF3D7_0937800	reverse	1.891545	0	6.977E-29	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr10	30129	31923	1794	32	1	0	1	PF10_0001	PF3D7_1000100	forward	1.633426	0	1.937E-52	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr10	33097	33603	506	9	1	0	1	PF10_0001	PF3D7_1000100	forward	4.251085	0	1.307E-33	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr10	1640577	1644443	3866	69	0	1	1	PF10_0406	PF3D7_1041300	reverse	0	-0.81844	4.753E-34	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr10	1644721	1648307	3586	64	1	0	1	PF10_0406	PF3D7_1041300	reverse	1.601156	0	3.985E-91	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr11	1	42843	42842	763	1	1	2	PF11_0007	PF3D7_1100100	forward	0	-1.86987	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
					1	1		PF11_0008	PF3D7_1100200	reverse				erythrocyte membrane protein 1, PfEMP1 (VAR)
chr11	2025353	2028544	3191	57	1	0	1	PF11_0521	PF3D7_1150400	forward	2.413057	0	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr11	2028771	2029499	728	12	1	0	1	PF11_0521	PF3D7_1150400	forward	2.643028	0	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr11	2030339	2031680	1341	23	1	1	1	PF11_0521	PF3D7_1150400	forward	0	-2.40462	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr11	2031739	2032184	445	7	1	0	1	PF11_0521	PF3D7_1150400	forward	2.344646	0	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr11	2032299	2032803	504	8	1	0	1	PF11_0521	PF3D7_1150400	forward	1.344341	0	0	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr11	2033305	2034259	954	17	1	0	1	PF11_0521	PF3D7_1150400	forward	1.858437	0	9.426E-26	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr12	18145	22179	4034	72	1	1	1	PFL0005w	PF3D7_1200100	forward	0	-2.1731	1.906E-21	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr12	34217	34947	730	13	1	0	1	PFL0020w	PF3D7_1200400	forward	5.032743	0	5.048E-81	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr12	36177	36739	562	10	1	0	1	PFL0020w	PF3D7_1200400	forward	3.125305	0	1.674E-23	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr12	768769	774203	5434	97	1	0	1	PFL0935c	PF3D7_1219300	reverse	0.700598	0	1.969E-42	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr12	1695457	1699603	4146	74	1	1	1	PFL1950w	PF3D7_1240300	forward	0	-2.34418	2.14E-19	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr12	1719537	1719931	394	7	1	0	1	PFL1960w	PF3D7_1240600	forward	1.806398	0	4.547E-21	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr12	1721273	1725307	4034	72	1	0	1	PFL1960w	PF3D7_1240600	forward	1.172205	0	1.678E-74	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr12	1736953	1740091	3138	56	1	0	1	PFL1970w	PF3D7_1240900	forward	1.219594	0	4.194E-60	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr12	2244033	2250976	6943	124	1	2	1	PFL2665c	PF3D7_1255200	reverse	0	-3.02971	3.88E-29	erythrocyte membrane protein 1, PfEMP1 (VAR)
chr13	21337	26491	5154	92	2	1	1	MAL13P1.1	PF3D7_1300100	forward	1.100993	0	1.234E-64	erythrocyte membrane protein 1, PfEMP1 (VAR)

Figure 3.5: Major variations present in PfEMP-1 VAR genes.

3.4-Major variations other than Variant Surface Antigens:

On analysis of the CNV data we have detected variations in all the 14 chromosomes. Apart from the subtelomeric regions of the chromosomes, variations were also seen in other locations on the chromosomes, including intergenic regions. Major amplifications are present in Chr-8, 10, 12, 13 and 14. Details of the genes and regions showing amplifications are shown in **table-3.3** contionous amplification of 28 kb is present in a region on chromsome number 8 as shown in **figure 3.6** Amplification of 15.90 kb is present in close proximity to the centromeric region of chromosome 12, **figure 3.7**, the amplification in the region cover 6 genes, namely conserved *Plasmodium* protein, unknown function (PFL1160C), mitochondrial carrier protein putative (PFL1145W), conserved *Plasmodium* protein, unknown function (PFL1135C), GTP cyclohydrolase-I (GCH1) (PFL1155W), ribosomal protein L24, putative (PFL1150C), 50S integral membrane protein, putative (PFL1140W). Amplification of the region, that covers three genes PFL1155W, PFL1150C, and PFL1145W have been seen previously in a Dd2 strain of *P. falciparum* (Dharia, et al., 2009). One of the genes showing amplification that codes for the GTP cyclohydrolase enzyme which is the 1st and the rate limiting enzyme of the folate Biosynthesis Pathway, amplification of this gene has been previously reported in parasite isolates from certain endemic countries (Kidgell, et al., 2006;

Nair, et al., 2008). Another gene, (PFL1145W) Mitochondrial carrier protein, putative, is reported as a homologue of Yhm2 which was characterised from *S. cerevisiae* as an inner membrane, DNA binding protein, it is speculated that this protein helps in the mitochondrial maintenance by tethering mitochondrial DNA to the inner membrane.

Amplifications of 3.19 kb and 3.13 kb are present in chromosome number 14 covering single genes (PF14_0508 and PF14_0758) pyridine nucleotide transhydrogenase, putative and EMP1-trafficking protein (PTP3). It has been reported that disruption of the latter gene showed a reduced level of expression of PfEMP-1, which suggests that the protein encoded by PF14_0758 plays a role in trafficking and the display of the virulence protein PfEMP-1 on the host erythrocyte (Maier, et al., 2008). All these genes showing amplifications in the data may be important for the parasite to survive in the host environment.

Chromosomes	Region	Size of variation	Number of genes	Gene symbol	Description
Chr-8	249817-278491	28.6kB	3	MAL8P1139 PF08_0122 PF08_0123	A conserved Plasmodium protein with unknown function, conserved Plasmodium protein with unknown function, U2 snRNA/tRNA pseudouridine synthase, putative
Chr 10	997753-1005203	7.4kB	1	PF10_0232	Chromodomain helicase-DNA binding protein 1 homolog,, putative (CHD1)
Chr10	581505-587163	5.6kB	1	PF10_0143	Transcriptional co-activator ADA2 (ADA2).
Chr10	650889-653243	2.3kB	1	PF10_0159	Glycophorin Binding Protein(GBP)
Chr12	962081-977987	15.90kB	6	PFL1160C PFL1145W PFL1135C PFL1155W PFL1150C PFL1140w	Conserved Plasmodium protein with unknown function, mitochondrial carrier protein, putative, Conserved Plasmodium protein with unknown function, GTP cyclohydrolaseI (GCHI),ribosomal protein L24, putative, 50s integral membrane protein, putative.
Chr12	526681-532619	5.9kB	2	PFL0585W PFL0590C	Polyubiquitin (PfpUB) non-SERCA-type Ca2+ -transporting P-ATPase (ATP4)
Chr13	408857-412219	3.3kB	1	PF13_0048	mRNA-decapping enzyme 2, putative (DCP2
Chr14	2194809-2198003	3.19	1	PF14_0508	pyridine nucleotide transhydrogenase, putative
Chr14	3233161-3236299	3.13	1	PF14_0758	EMP1-trafficking protein (PTP3)

Table3.3: Amplifications other than Variant Surface Antigen genes (VSA) genes in the region of Pf-2

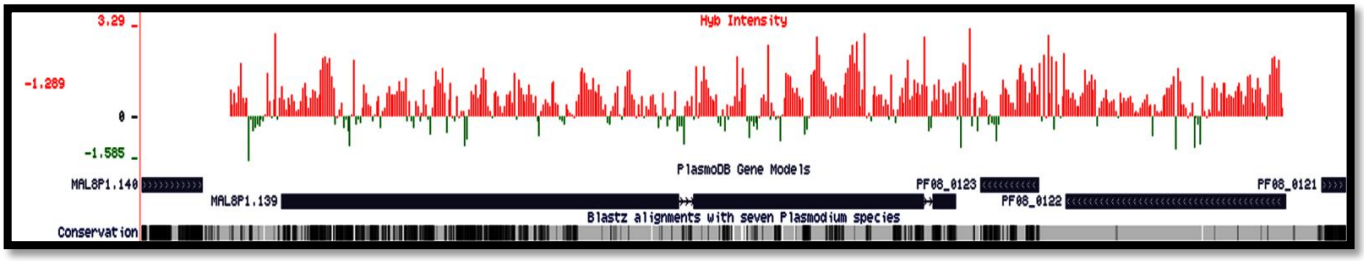


Figure-3.6-Continuous amplification on chromosome-8 A continuous amplification of 28 kb region (249817-278491) of chromosome-8, the X- axis represent the baseline intensity of the probes, the Y-axis represents the hybridized probe intensity, red bars are showing amplifications present in the region and green bars are showing deletions present in the region. The genes present in the region, are near the centromeric location on the chromosome-8 that covers for 3 genes (MAL8P1.139, PF08_0122), conserved *Plasmodium* membrane protein, unknown function, (PF08_0123) tRNA pseudouridine synthase, putative. The white double headed arrows in single exons are representing the 5'-3' orientation of transcription of the genes, the arrows in between the exons are representing introns and strand in which genes are present (Right hand side arrow forward strand, left hand side arrow representing reverse strand). The area without any bars is representing no variations. Conservation of the region in the Plasmodial species is shown in grey and black bars, where the black bars represent the higher conservation.

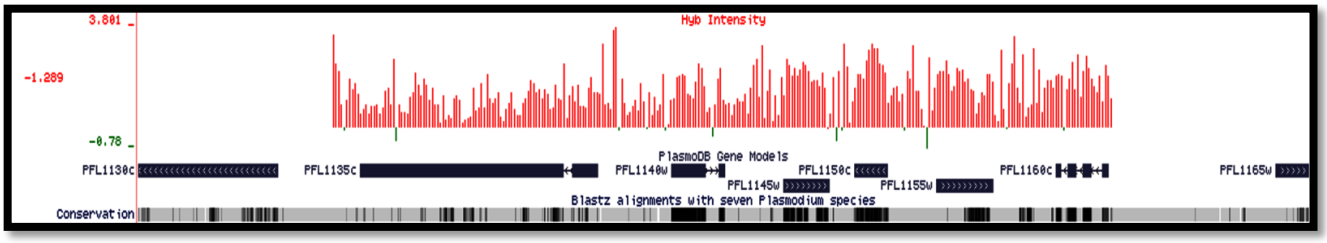


Figure-3.7-Continuous amplification on chromosome-12: Amplification of 15.90 kb is present near close proximity to the centromere region of chromosome 12 (Figure-3), which covers 6 genes, namely (PFL1160C), conserved *Plasmodium* protein, unknown function, (PFL1145W) mitochondrial carrier protein, putative, (PFL1135C) conserved *Plasmodium* protein, unknown function, (PFL1155W) GTP cyclohydrolase I (GCH1), (PFL1150C) ribosomal protein L24, putative, (PFL1140W). The X- axis represent the baseline intensity of the probes, the Y-axis represents the hybridized probe intensity, red bars are representing probe intensity as amplifications present in the region and green bars are showing deletions present in the region. Regions without any bar are showing no variations. Conservation of the region in the Plasmodial species is shown in grey and black bars, where black bars represents the higher conservation.

Conclusion:

The implications of deletions or amplifications in the coding regions of the genes remain unanswerable at this point. At this level of the study, it is not possible to speculate whether the differences that we are seeing in Pf-isolate-2 are instrumental in causing even partly, the severe disease state exhibited by the patient. However, we are reporting some putative and novel variations, like amplification of 28 kb region in chromosome 8 that cover for covers for 3 genes (MAL8P1.139, PF08_0122), conserved *Plasmodium* protein unknown function, (PF08_0123) tRNA pseudouridine synthase, putative. Which are not reported before in the copy number variation studies with clinical isolates causing severe malaria. The data clearly show that there is a difference in the isolate, which has been taken from the patient showing cerebral malaria (Pf-2) than from that with uncomplicated malaria (Pf-1). Differences exist not only in coding regions, but also in the intergenic locations. Thus, there could be implications at the level of gene regulation, which could affect disease conditions. However, these differences could also be inherently natural differences between two parasites isolates without any direct correlation with the disease states, which could be influenced by various other factors. This study is just a preliminary attempt to report the variations or differences in the genomic segments of parasite isolate causing severe malaria with that of the isolate causing uncomplicated malaria. To get a real picture of the genomic variations, these analyses should be done with a large set of severe clinical samples from different parts of India.

Chapter IV

Cloning and protein expression of conserved Plasmodium protein with unknown function (PFD0300W and PFL2320W)

The expression of recombinant proteins in heterologous system is an important and frequently used tool in malaria research. However, the expression of *P. falciparum* protein remains problematic because of high A/T or G/C content and the presence of higher frequency of rare codons that code for amino acids Arginine and lysine repeats in the *P. falciparum* genome. The presence of these rare codons is thought to be the reason of early termination of mRNA translation processes. Therefore, many of *Plasmodium* proteins are expressed in truncated forms or in the form of inclusion bodies. Inspite of modulating factors like change of temperature, pH of media and codon optimization there may be a negligible increase in recombinant protein expression. In this chapter, the results of the trials of protein expression of hub genes PFD0300W (PF3D7_0406000)

and PFL2320W (PF3D7_1248400) that codes for conserved *Plasmodium* protein with unknown function derived from the systems network will be discussed.

Results and Discussion:

4.1-Genome variation analysis of PFL2320W and PFD0300W:

In relation with the previous chapter, we have analysed the presence of genomic variation in the selected hub genes from the whole genome co-expression systems network. CGH data have been cross-checked for these two genes. In comparative genomic hybridization, these two genes have not shown any genomic variations, i.e. absence of log ratio of hybridized probes as shown in **figure 4.1** and **figure 4.2**. For variation data analysis at the level of point mutations, the genes were amplified from genomic DNA, and one clone from field isolates and few amplicons were sent for sequencing. In case of PFL2320W, sequencing of the clone, as well as a few of the amplicons, showed no change from the available 3D7 annotated gene sequence from PlasmoDB database, **figure. 4.3.** Whereas in case of PFD0300W the partial sequence obtained after sequencing of clone and from a few field isolates suggests the absence of any variation in the gene **figure. 4.4**.

Figure 4.1: PFD0300W, genome variation analysis: Absence of log ratio suggests, absence of any genomic variation in the genomic region (321019 - 322545) covering for PFD0300W in the CGH data.

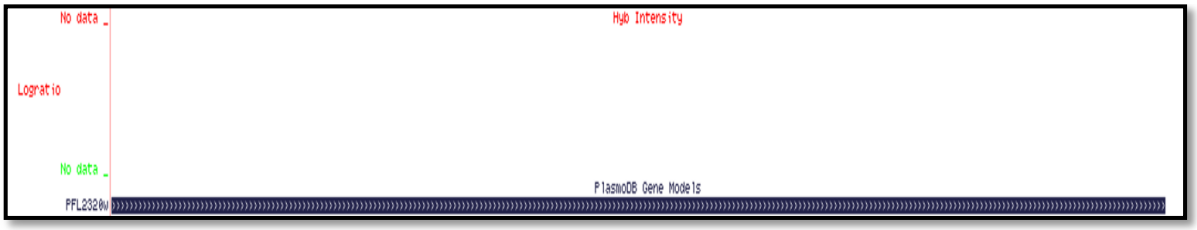
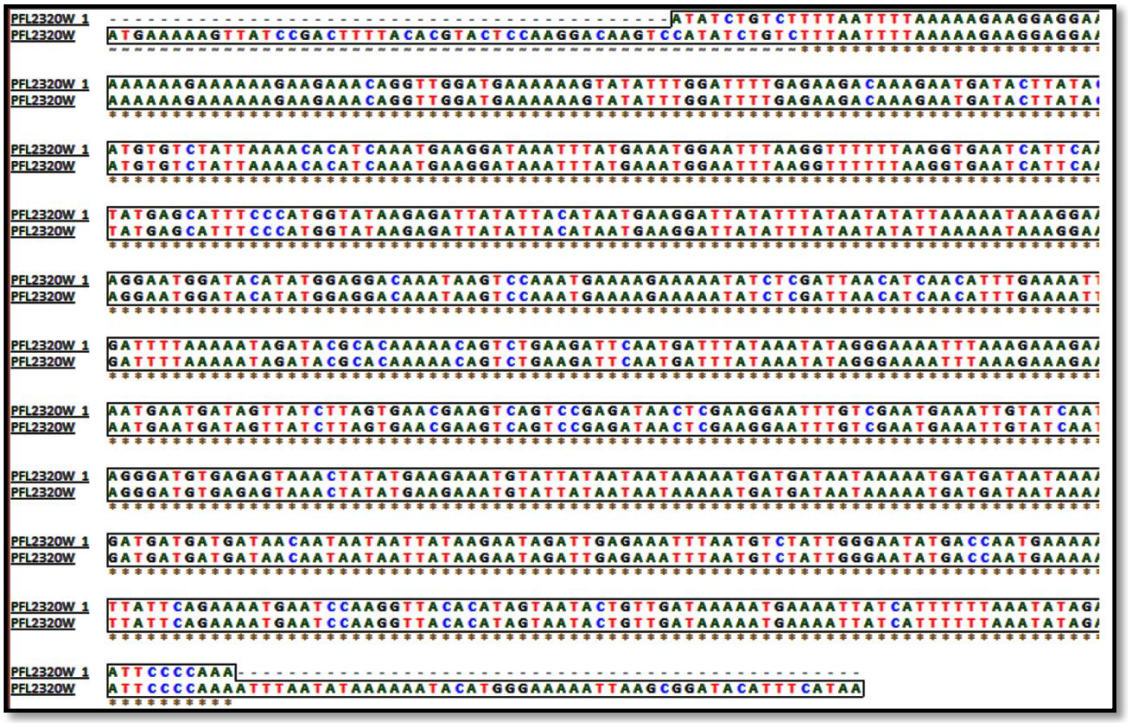
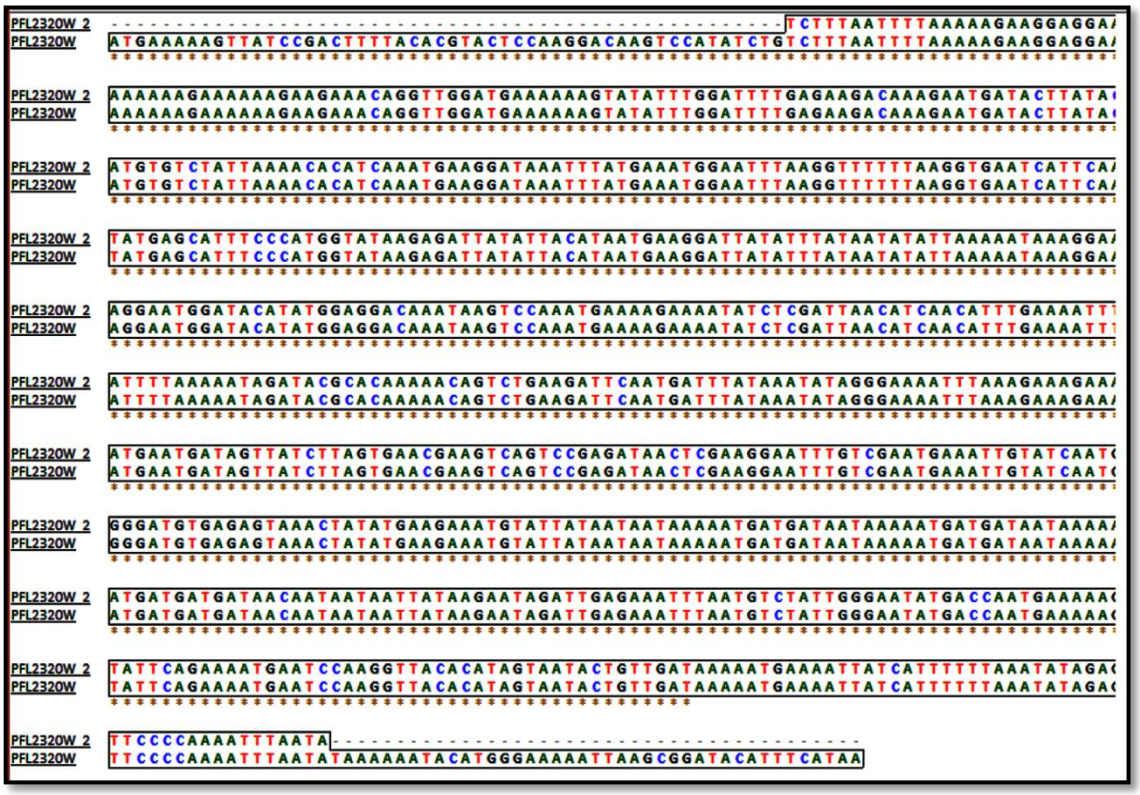


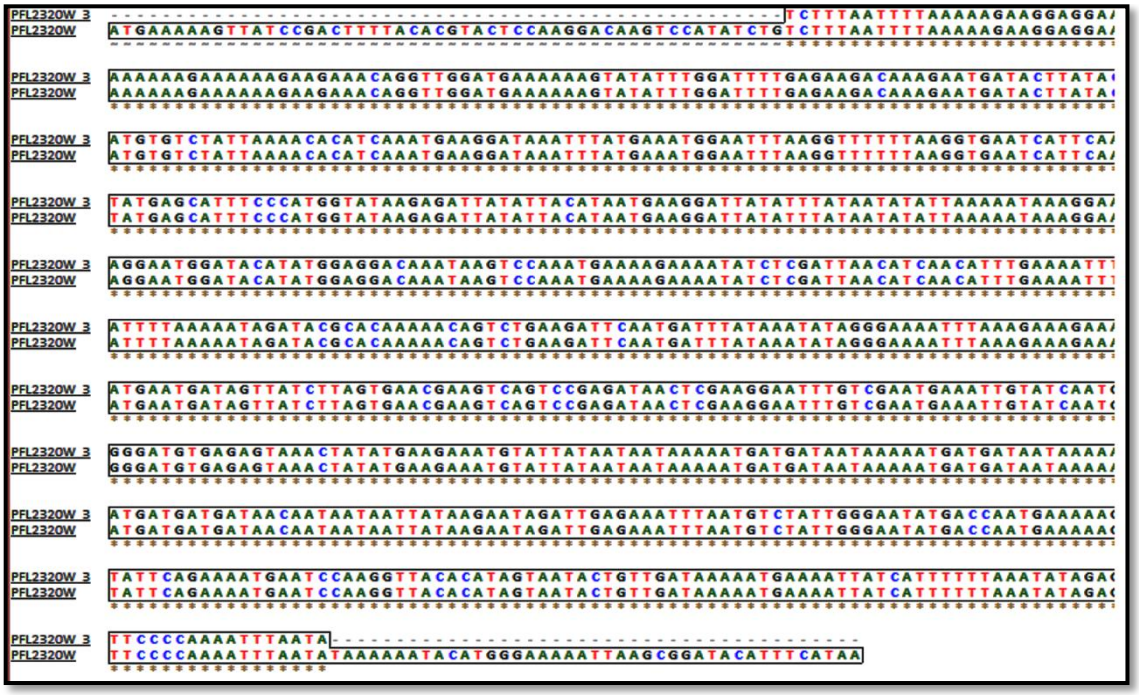
Figure 4.2: PFL2320W genome variation analysis: Absence of log ratio suggests, absence of any genomic variation in the genomic region (1981906 - 1982763) of covering for PFL2320W in the CGH data.



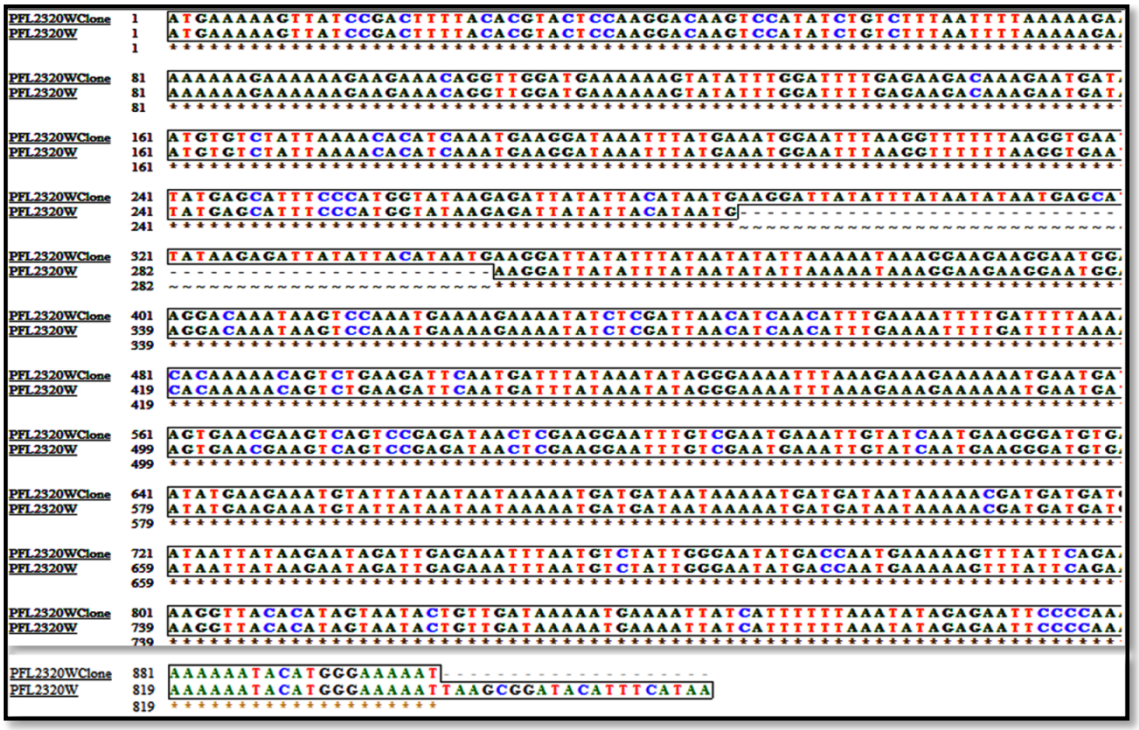
(A) Sample: 1



(B) Sample: 2

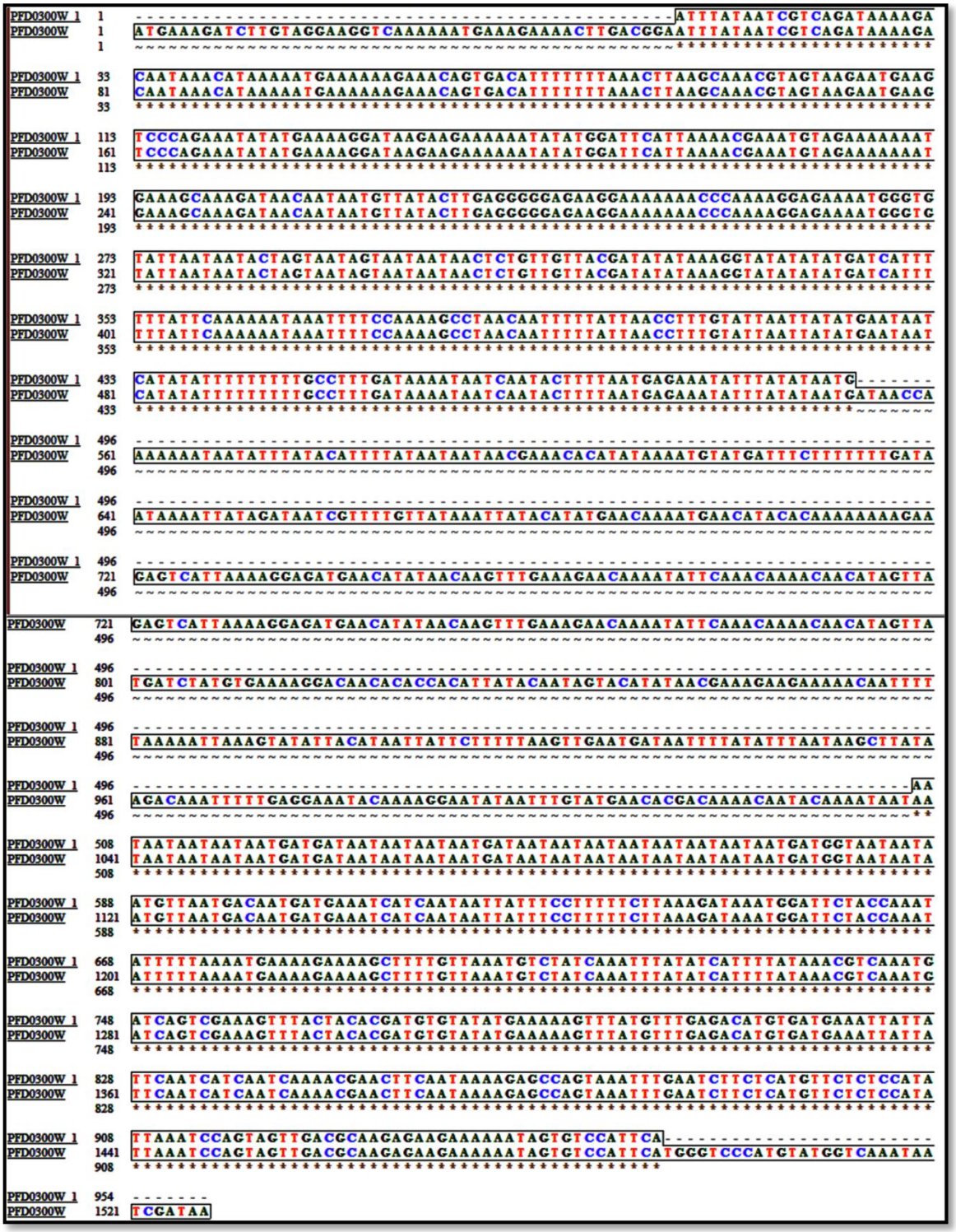


(C) Sample-3

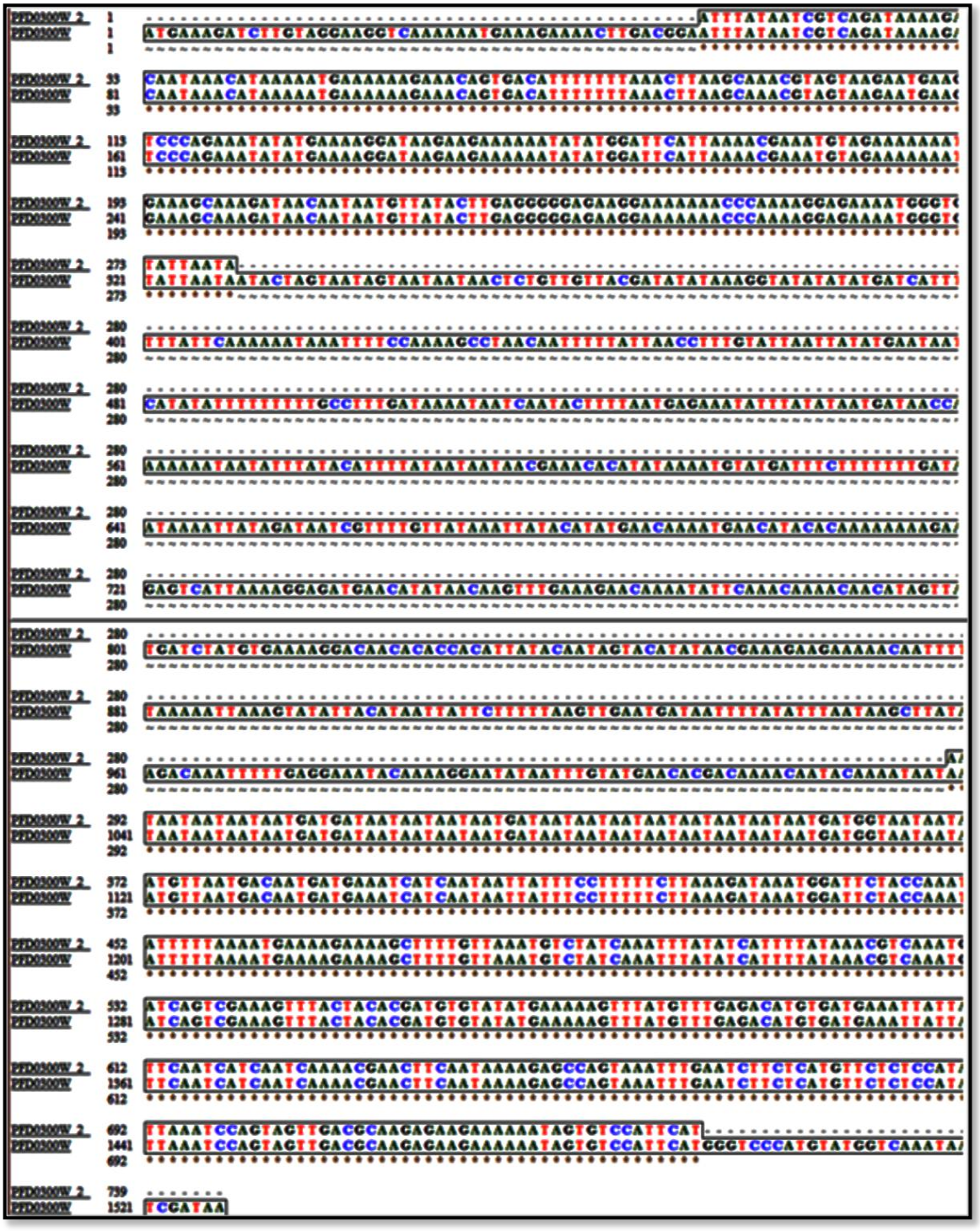


(D) PFL2320W Clone

Figure 4.3: Variation analysis of PFL2320W sequencing result of field isolates and clone A-C) Sequencing result from Sample-1, Sample-2, Sample-3; where PFL2320W_1, PFL2320W_2, PFL2320W_3 represents a sequence of field isolates and PFL2320W represents.



(A) Sample-1



(B) Sample-2

Figure 4.4- Aand B: Sequencing data of PFD0300W Sample-1 and Sample-2; PFD0300W_1 and PFD0300W_2 represents the sequencing result from field isolate, and PFD0300W represents the PlasmoDB sequence.

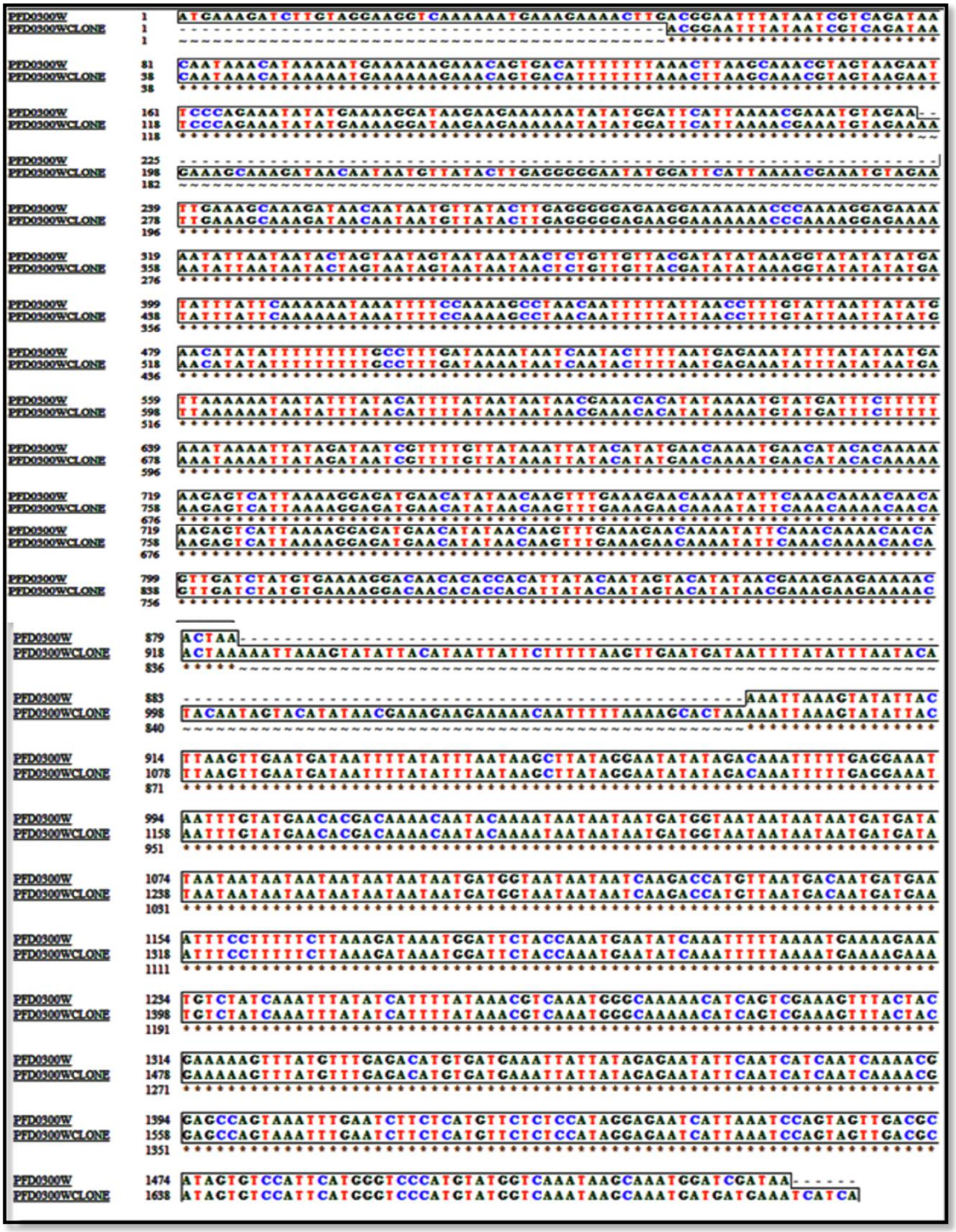


Figure 4.4-C: Sequencing result of PFD0300W clone; PFD0300W represents the sequence of a gene annotated in PlasmoDB, whereas PFD0300W clone represents the sequence obtained from sequencing of PFD0300W clone.

4.2- Amplification of PFD0300W and PFL2320W:

Reaction conditions for amplification of PFD0300W and PFL2320W genes have already been discussed in the materials and method section 2.8.2. Initially, the conventional cloning of amplicons were done in expression vector pRSETA, PFD0300W was cloned in between the XhoI and NcoI restriction enzyme sites **figure 4.5** and PFL2320W was cloned between Xho I and KpnI enzyme sites **figure 4.6** both the clones were further sequenced for the presence of the genes in the expression vector.

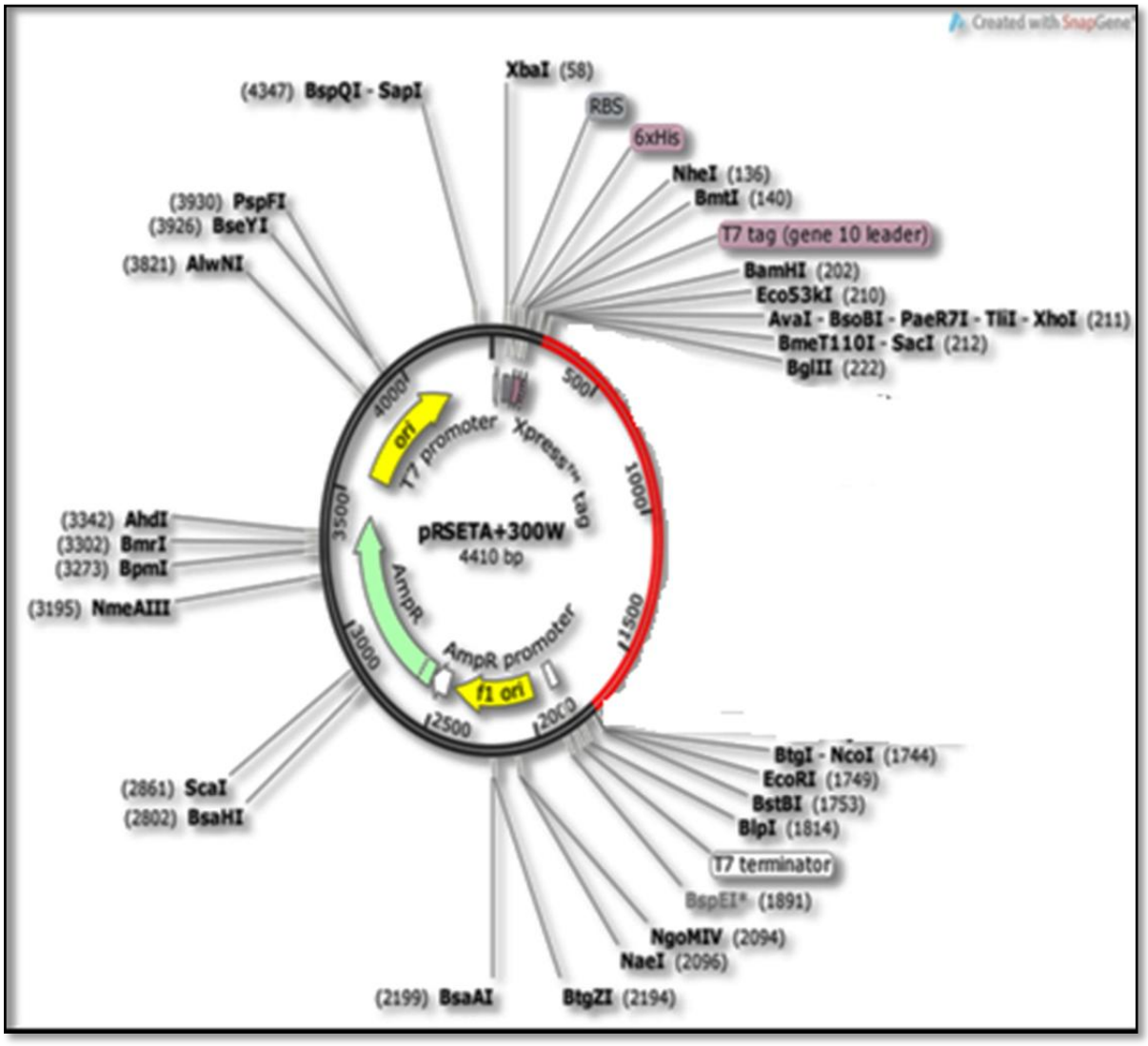


Figure 4.5: pRSETA-PFD0300W clone map- Red part depicts PFD0300W gene and black part represents the backbone of the vector.

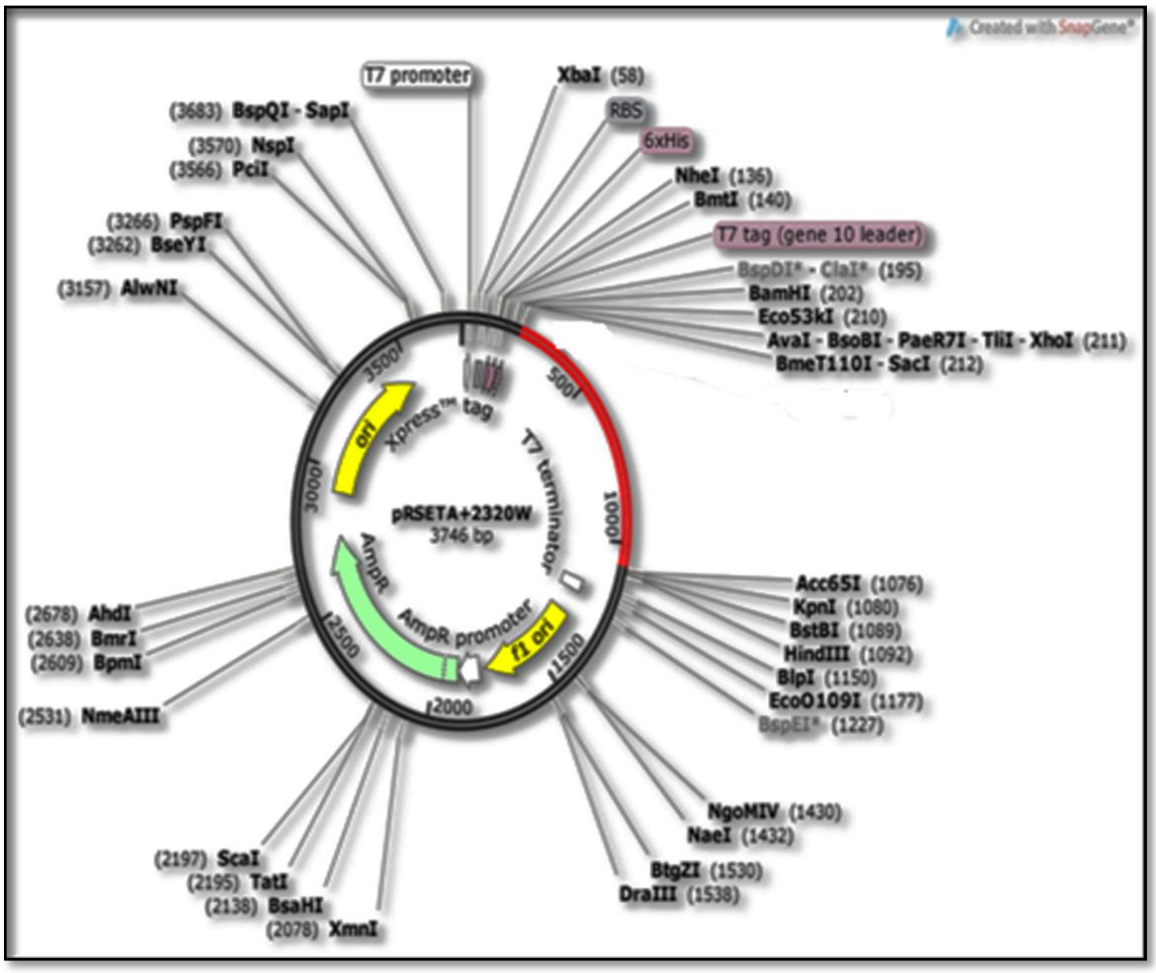


Figure 4.6: **pRSETA-PFL2320W clone map:** Red part depicts PFL2320W gene and black part represents the backbone of the vector.

After successful validation, through sequencing, the recombinant clone was transformed into *E. Coli* host cells for protein expression studies. The different host cells, which were used, were BL21DE3, BL21DE3pLysS, Rosetta and Rosetta pLySs, BL21DE3 host cell along with RIG plasmid. Expression studies were performed at different time points and temperatures along with different media and IPTG concentrations **table 4.1**. Since the pRSETA is an expression vector it has a 6X histidine as a fusion tag, adding which increases the molecular weight of the proteins by 4 kDa over the actual estimated size of the protein band of PFL2320W. The expected size of the protein band of PFL2320W with His-tag is ~37.8 kDa and for PFD0300W, ~60.67 kDa respectively. Different strategies were applied for the protein expression of the two proteins in pRSETA system but without any success. The reason could be the presence of RARE codons in the genes. The Presence of RARE codons was checked using RARE codon caltor, Programmed by

Edmund Ng, available at http://people.mbi.ucla.edu/sumchan/caltor.html **figure: 4.7** and **figure 4.8**.

S.No.	Conditions Used			Host			
	IPTG	Temperature Degree centigrade	Nutrient media	BL21 DE3	BL21DE3+ Rig	BD3pLysS	Rosetta
1	0.5mM	37°C, 28°C, 16°C	LB Broth, Terrific Borth, SOC media	Did not get the successful expression of proteins In any of the host.			
2	0.75mM	37°C, 28°C, 16°C					
3	1mM	37°C, 28°C,16°C					
4	2mM	37°C, 28°C,16°C					

Table 4.1: Protein expression studies at different temperature and conditions

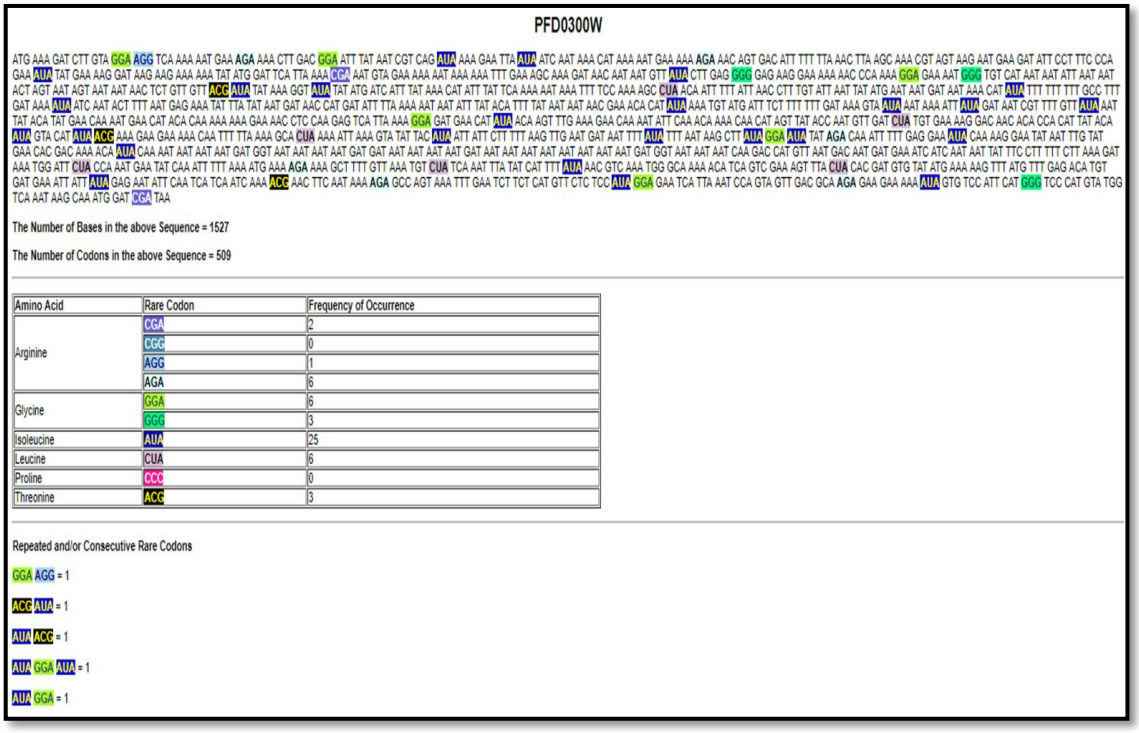
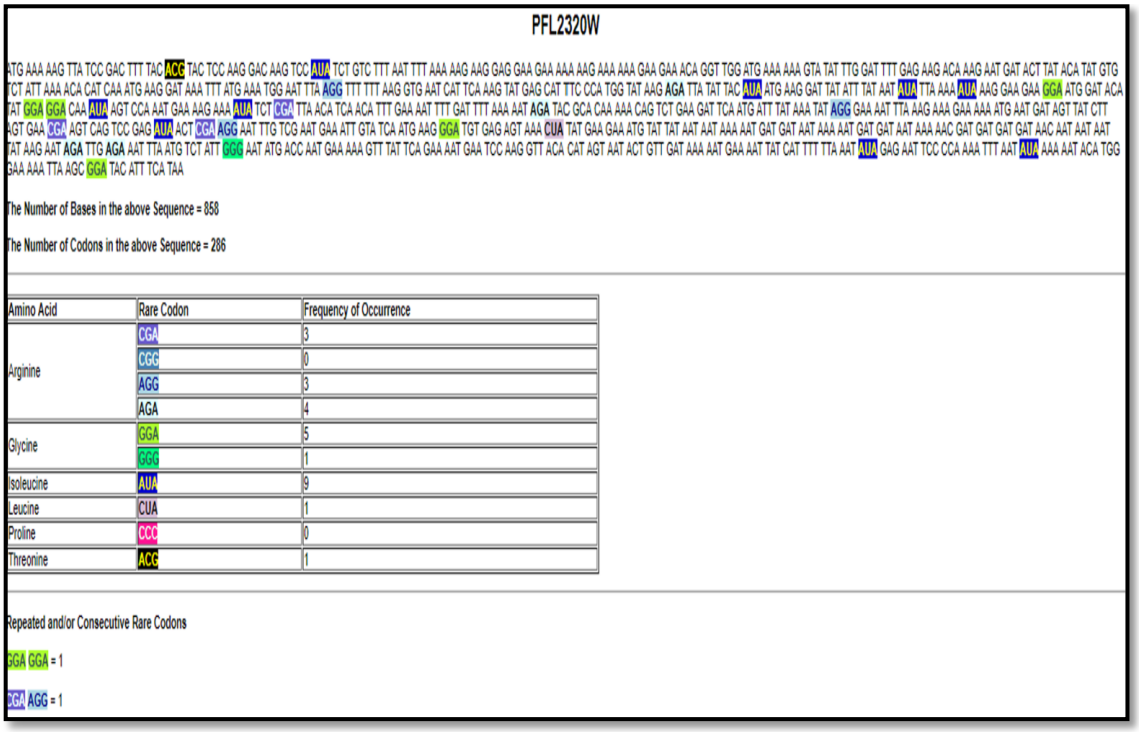


Figure 4.7: Presence of RARE codons in PFD0300W: Different colours are depicting different RARE codons for the amino acids, Arginine, Isoleucine, Leucine, Threonine.The frequency of the RARE codon of Isoleucine is more than other RARE codons.



PFL2320W

ATG AAA AAG TTA TCC GAC TTT TAC ACG TAC TCC AAG GAC AAG TCC AUA TCT GTC TTT AAT TTT AAA AAG AAG GAG GAA GAA AAA AAG AAA AAA GAA GAA ACA GGT TGG ATG AAA AAA GTA TAT TTG GAT TTT GAG AAG ACA AAG AAT GAT ACT TAT ACA TAT GTG TCT ATT AAA ACA CAT CAA ATG AAG GAT AAA TTT ATG AAA TGG AAT TTA AGG TTT TTT AAG GTG AAT CAT TCA AAG TAT GAG CAT TTC CCA TGG TAT AAG AGA TTA TAT TAC AUA ATG AAG GAT TAT ATT TAT AAT AUA TTA AAA AUA AAG GAA GAA GGA ATG GAT ACA TAT GGA GGA CAA AUA AGT CCA AAT GAA AAG AAA AUA TCT CGA TTA ACA TCA ACA TTT GAA AAT TTT GAT TTT AAA AAT AGA TAC GCA CAA AAA CAG TCT GAA GAT TCA ATG ATT TAT AAA TAT AGG GAA AAT TTA AAG AAA GAA AAA ATG AAT GAT AGT TAT CTT AGT GAA CGA AGT CAG TCC GAG AUA ACT CGA AGG AAT TTG TCG AAT GAA ATT GTA TCA ATG AAG GGA TGT GAG AGT AAA CUA TAT GAA GAA ATG TAT TAT AAT AAA AAT GAT GAT AAT AAA AAT GAT GAT AAT AAA AAC GAT GAT GAT AAC AAT AAT AAT TAT AAG AAT AGA TTG AGA AAT TTA ATG TCT ATT GGG AAT ATG ACC AAT GAA AAA GTT TAT TCA GAA AAT GAA TCC AAG GTT ACA CAT AGT AAT ACT GTT GAT AAA AAT GAA AAT TAT CAT TTT TTA AAT AUA GAG AAT TCC CCA AAA TTT AAT AUA AAA AAT ACA TGG GAA AAA TTA AGC GGA TAC ATT TCA TAA

The Number of Bases in the above Sequence = 858

The Number of Codons in the above Sequence = 286

Amino Acid	Rare Codon	Frequency of Occurrence
Arginine	CGA	3
	CGG	0
	AGG	3
	AGA	4
Glycine	GGA	5
	GGG	1
Isoleucine	AUA	9
Leucine	CUA	1
Proline	CCG	0
Threonine	ACG	1

Repeated and/or Consecutive Rare Codons

GGA GGA = 1

CGA AGG = 1

Figure 4.8: Presence of RARE codons in PFL2320W: Different colours are depicting different rare codons for the RARE amino acids, Arginine, Isoleucine, Leucine, Threonine, the frequency of the RARE codon of Isoleucine is more than other RARE codons.

4.3-Codon Optimization of PFL2320W and PFD0300W:

The presence of RARE codons could be the reason that the heterologous gene expression of *Plasmodium* proteins (PFL2320W and PFD0300W) was difficult to obtain. To increase the possibility of protein expression, *E. coli* hosts like Rosetta and RIG plasmid were used, but even after using above-mentioned plasmid and *E. coli* host protein expression was not successful for both the proteins. To overcome this problem gene sequences was outsourced (Invitrogen, Thermo Fisher Scientific) for the optimization of the rare codons according to the codon usage of the *E.coli* host. The codon that were optimized according to host is shown in **figure 4.9 and figure 4.10**. The optimized constructs were obtained in the expression plasmid pET100D-TOPO, with 6X Histidine as a fusion tag **figure 4.11.**

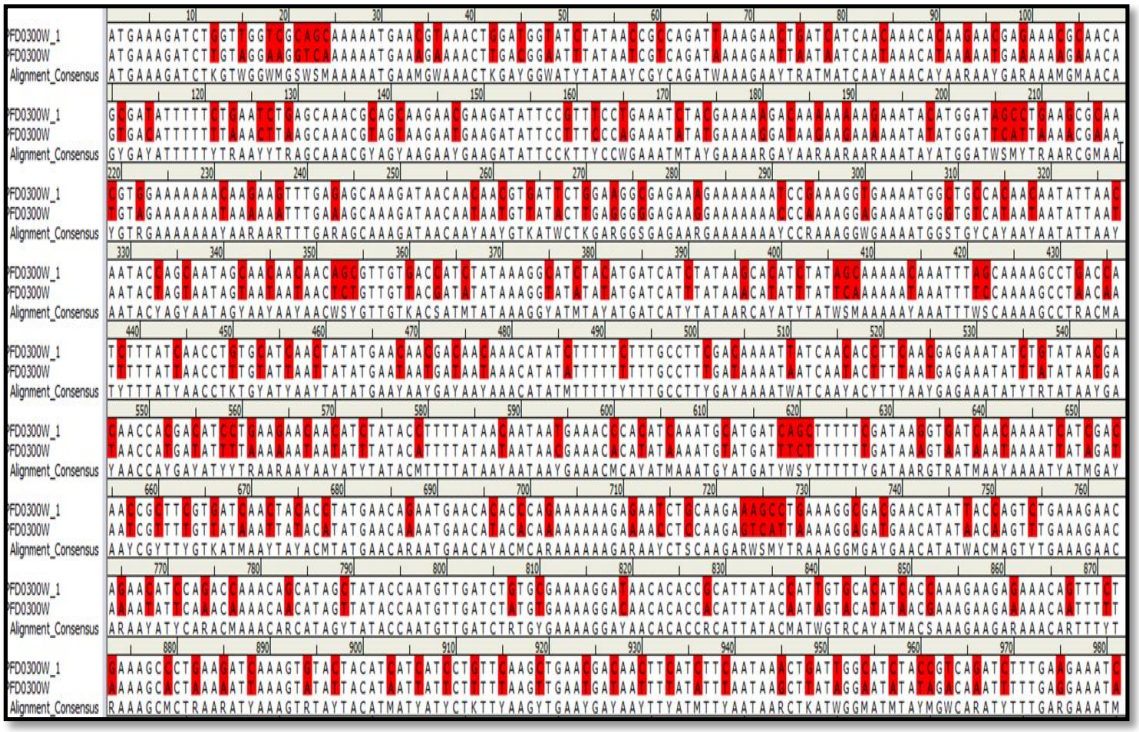


Figure 4.9-Codon optimization of PFD0300W: PFD0300W_1 representing the optimized gene sequence, whereas PFD0300W represents the original gene sequence, the red colour is depicting the change in the gene sequence.

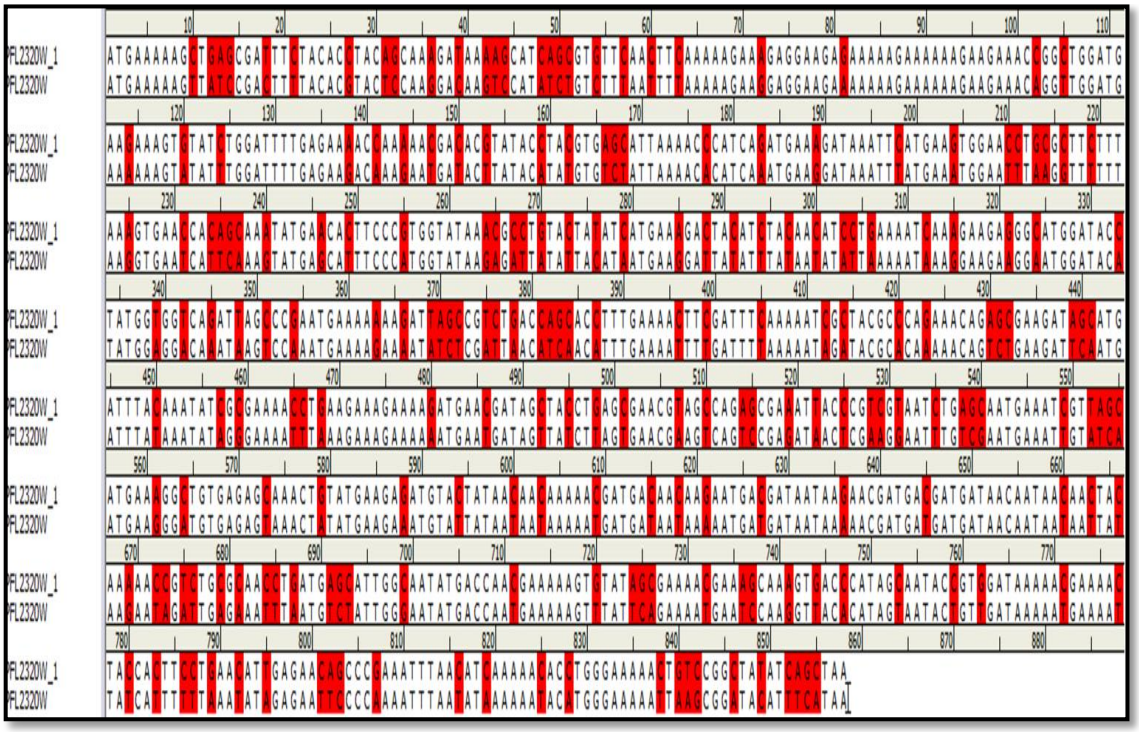


Figure 4.10-Codon optimization of PFL2320W: PFL2320W_1 representing the optimized gene sequence, whereas PFL2320W represents the original gene sequence, the red color is depicting the change in the gene sequence.

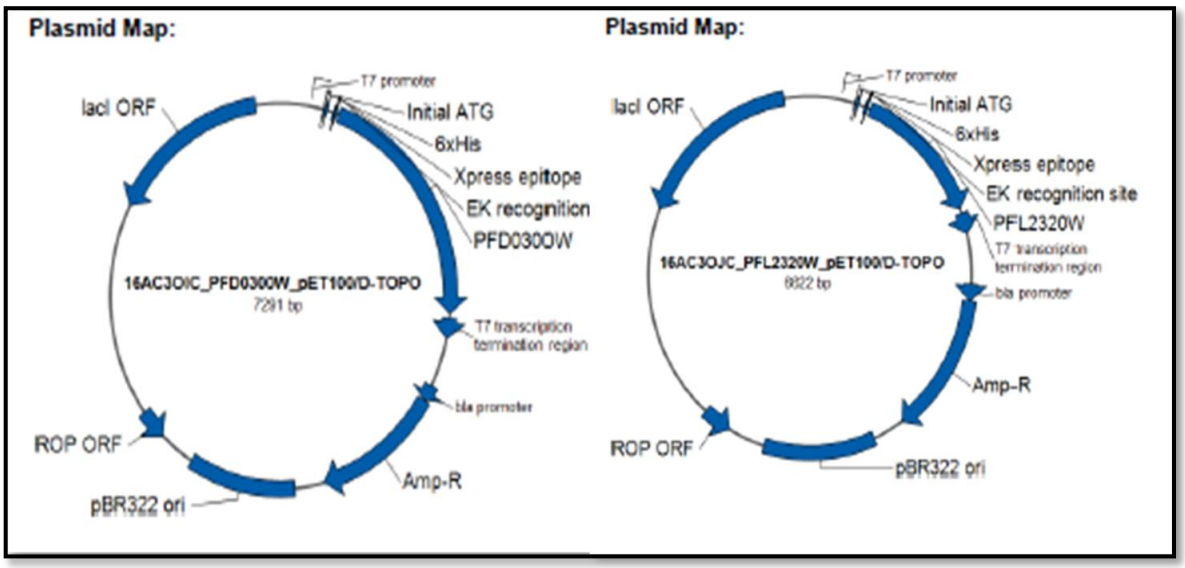


Figure 4.11-Construct vector map: pET100D-TOPO PFD0300W and PFL2320W construct vector map, present in between vector ATG and T7 transcription termination region.

4.4-Protein expression of codon optimized construct:

The previous expression studies with the trials used to express the cloned gene from *P. falciparum* proteins in pRSETA had given us conditions which were used for expression of the codon-optimized constructs in *E. coli* hosts, Different expression protocols, and conditions were used for the protein expression of the constructs as given in **table 4.2**.

Host Cells used	Temp Used	IPTG concentration	Media Used	Buffers used	Resistant Marker used
BD3	37° C, 18°C,20°C	0.5mM, 1mM, 1.5mM,2mM	LB media	Denaturing Buffer and Native Buffer	Ampicillin
BD3 star	37 °C,18°C,20°C	0.5mM, 1mM, 1.5mM,2mM	Lb media, SOC media	Native Buffer	Ampicillin and Carbinicillin
BD3 pLysS	37°C,18°C, 20°C	0.5mM, 1mM, 1.5mM,2mM	LB media	Native Buffer	Ampicillin, Carbinicillin and Chloroamphinicol
Arctic cells	12°C	0.5mM, 1mM, 1.5mM,2mM	LB media	Native Buffer	Gentamycin suphate and Carbinicillin

Table 4.2- Protein expression of construct: Different protein expression protocol for protein expression of construct.

4.4.1-Protein expression of PFL2320W construct using BL21DE3, Bl21D3 pLysS, BD3 star cells:

BD3 star cells, BD3pLysS and BL21DE3 host cells were used for the protein expression. The protocol used for protein expression in all the host cells has already been discussed in materials and methods section 2.10.1 As shown in **table 4.2** different strategies have been employed for the heterologous protein expression in *E. coli* cells, where the downstream processing involved different temperatures, i.e 18°C-37°C, a change of media, and IPTG concentrations. A possible faint expressed protein band at the expected position, i.e. at ~37.8 kDa in an induced fraction of clone from whole cell lysate at time points 4hr, 6hr, and 9hr with 1mM IPTG were seen **figure 4.12**. Further, the protein fractions of clone induced of 6hr, and 9hr time points were collected and sonicated in native buffer, to get the protein from insoluble fraction **figure 4.13**. To check the availability of protein in the insoluble fraction or in inclusion bodies, denaturing buffer with urea was used. However, no protein was detected in supernatant or in the pellet fraction of the lysate unlike that obtained in the native buffer.

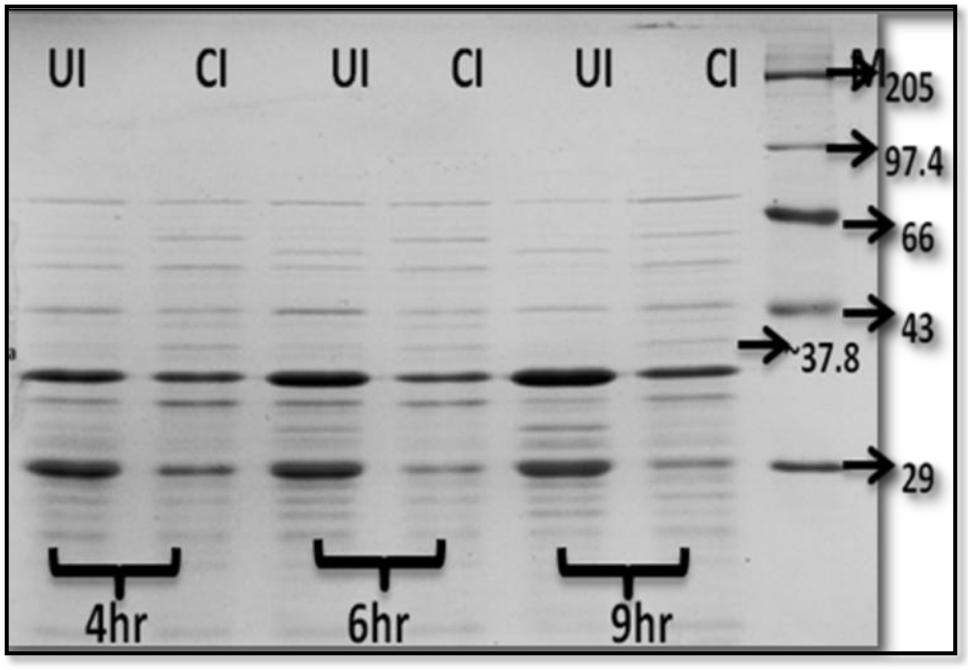


Figure 4.12: Protein expression of construct at 37°C with 1mM IPTG concentration: Lane 1-7: CUI 4hr, CI 4hr, CUI6hr, CI 6hr, CUI 9hr, CI 9hr, PMWH (29-205kDa) the arrow shows the approximate size of PFL2320W protein band i.e ~37.8kDa.

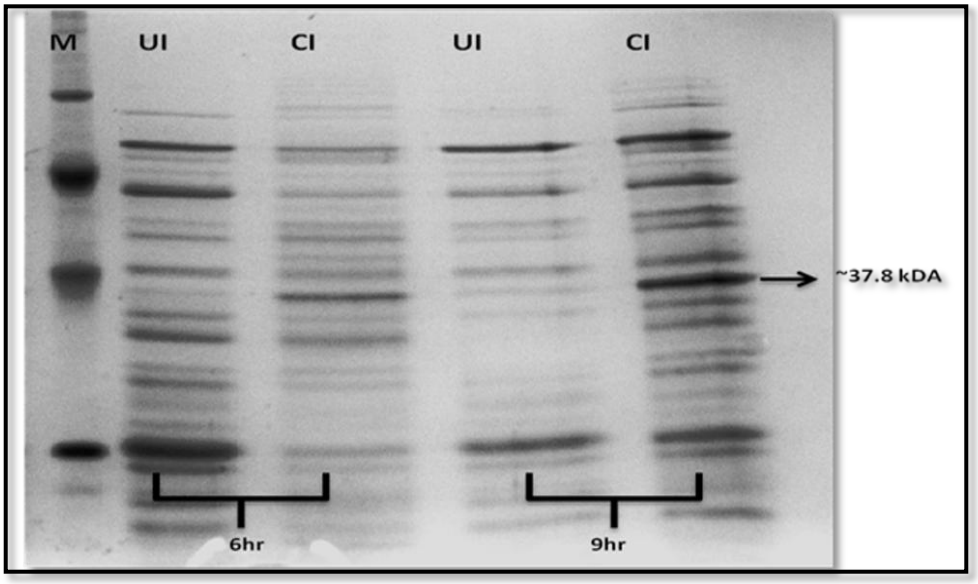


Figure 4.13: Protein expression in Native buffer: Lane 1-5: High molecular weight marker (PMWH-29-205kDa), UI6hr, CI6hr, UI9hr, CI9hr, at 37°C with 1mM IPTG.The arrow is showing approximate size of PFL2320W protein band.

4.4.2-Western blotting of PFL2320W:

Since a band was seen in the native gel western blotting of PFL2320W was performed using RGS 6X histidine primary antibody (QIAGEN). As a control cloned Suf A protein of 21 kDa from *P. vivax* was used. This was a protein which was expressed in the lab with the HIS tag. To check the successful transfer of the protein bands Ponceau-S stain was applied. We could not detect the recombinant protein with the antibody against the His -tag, i.e. ~37.8 kDa protein band in an induced fraction of the clone. In the case of the control, however, we were able to detect the Suf A protein at the desired size **figure 4.14**.

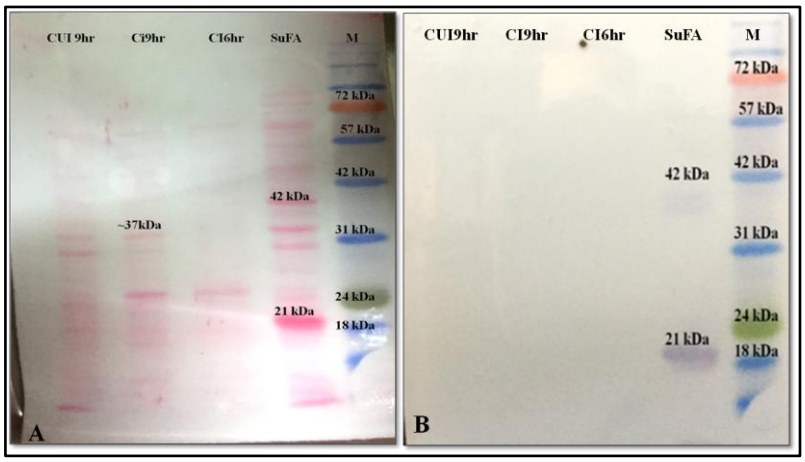


Figure 4.14: A & B Ponceau-S stain of protein expression and Western Blot analysis of PFL2320W: A) Lane 1-5; CUI9hr, CI9hr, CI6hr, SufA, pre-stained marker (18-240kDa). B) Lane 1-5; CUI9hr, CI9hr, CI6hr, SufA, pre-stained protein ladder (18-240kDa).

4.4.3-Protein expression of PFD0300W constructs using BL21DE3, BL21DE3 pLysS and BL21DE3 star cells:

Similar to PFL2320W host cells were used for protein expression, the protocol used for protein expression in all the host cells has already been discussed in materials and methods section 2.10.1. As shown in **table 4.2**, the same strategy for PFD0300W protein expression was employed. However, even at different temperatures and IPTG concentrations, we were not able to see any expressed bands at expected size of ~ 60.67 kDa at any of the time points. Crude lysates, as well as sonicated lysates, were analysed to check the presence of protein in insoluble fraction, but neither in crude extract nor in denaturing or native lysate the expressed protein bands were detected in induced samples at any time point**.** Initially, a trial with BL21DE3 for protein expression of PFD0300W was performed at 37°C, with 1mM IPTG; the crude lysate were checked on 12% polyacrylamide gel at 0hr, 3hr, 9hr and 20hr time points and no bands were obtained at ~60.67 kDa **figure 4.15**. Protein expression of the clone using super broth at 37°C was also attempted as per the protocol of (Yadav and Ockenhouse, 2003), again no expression was obtained at the desired size ~60.67 kDa **figure 4.16**. Reports have been suggested that to prevent the formation of aggregates or inclusion bodies expression at lower temperature is favourable for heterologous expression of protein (San-Miguel, et al., 2013; Palomares et al., 2004). Thus, we attempted different temperature for protein expression of PFD0300W but did not obtain expressed protein bands at the expected size

of ~60.67kDa **figure 4.17**. A trial of protein expression in the native and denatured buffer was also performed with similar results **figure 4.18-A, B.**

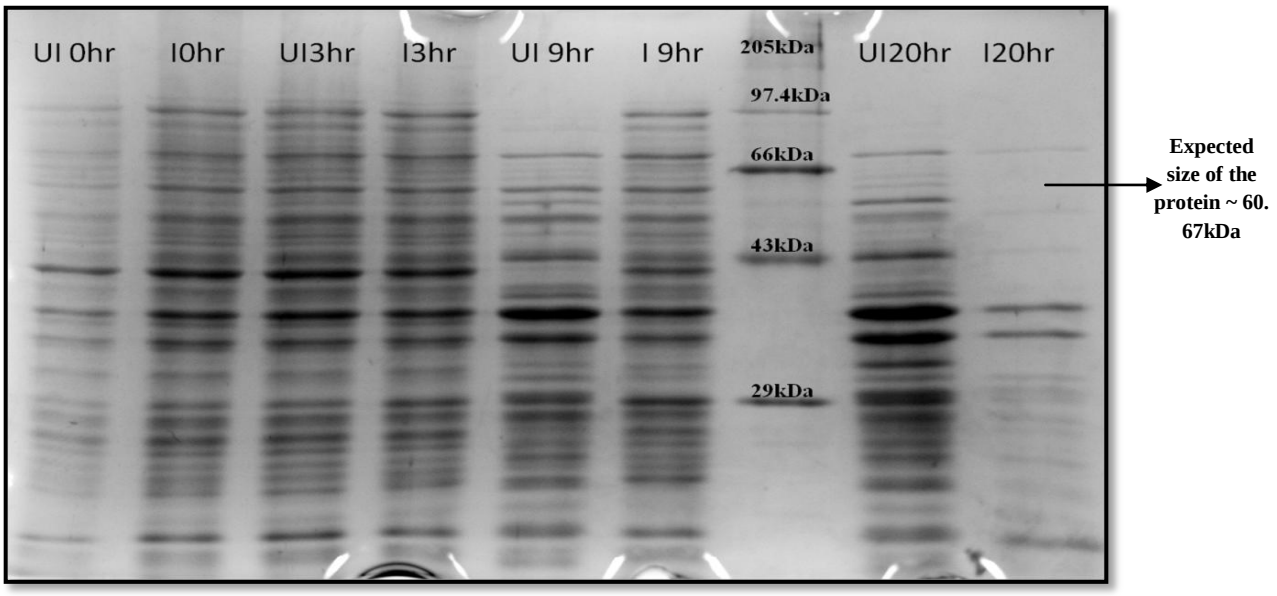


Figure 4.15: PFD0300W crude extract protein expression in BL21DE3 cells at 37°C with 1mM IPTG concentration: Lane 1-9: CUI 0hr, CI 0hr, CUI 3hr,CI 3hr, CUI 9hr,CI 9hr, Marker (PMWL 29kDa-205kDa) CUI 20hr, CI 20hr.

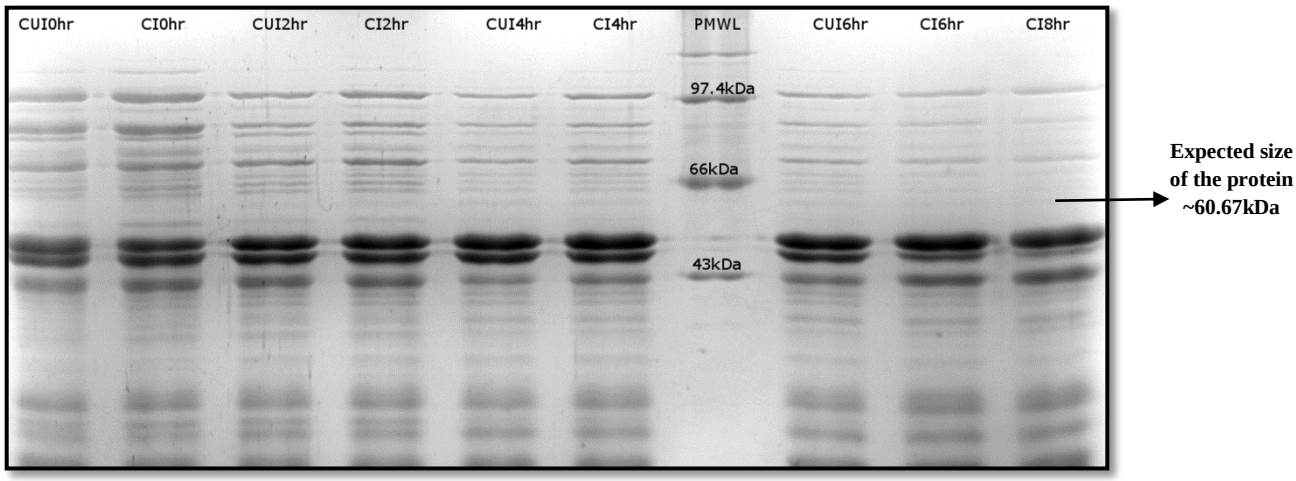


Figure 4.16: Protein Induction of PFD0300W in BL21DE3 host cell using super broth at 37°C with 1mM IPTG: Lane 1-10; crude cell lysate CUI0hr, CI0hr, CUI2hr, CI2hr, CUI4hr, CI4hr, Protein high molecular weight marker PMWL, CUI6hr, CI6hr, CI8hr

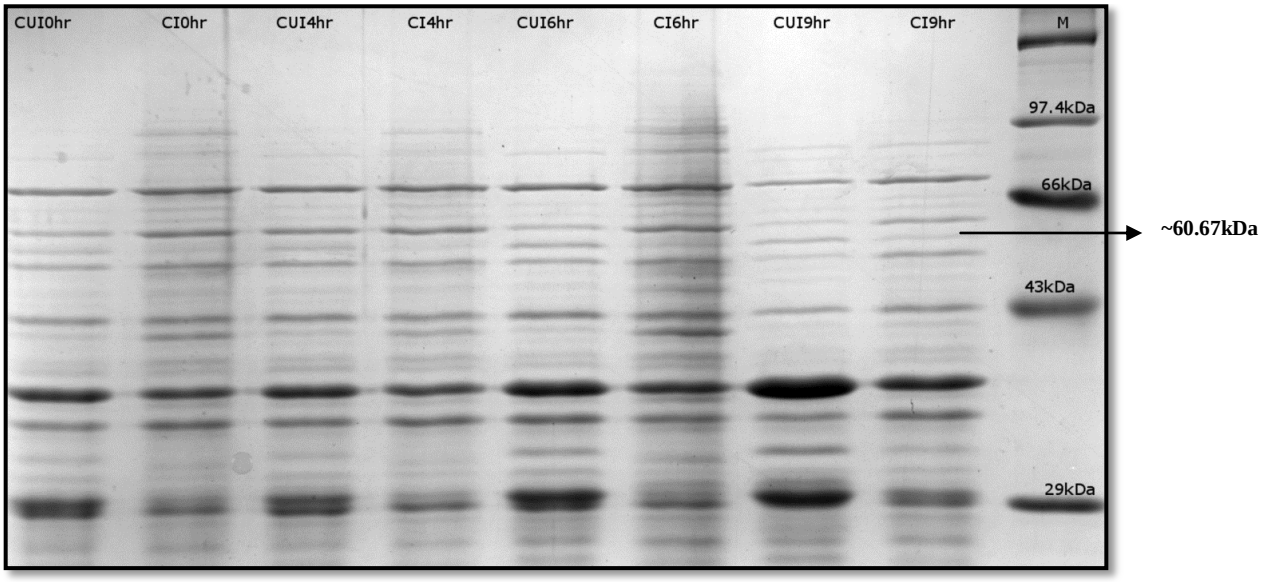


Figure 4.17-Protein Induction of PFD0300W in BL21DE3 host cell at 15°C with 1mM IPTG: Lane 1-9; crude cell lysate CUI0hr, CI0hr, CUI4hr, CI4hr, CUI6hr, CI6hr, CUI9hr, CI9hr, Protein high molecular weight marker PMWH(29-205kDa).

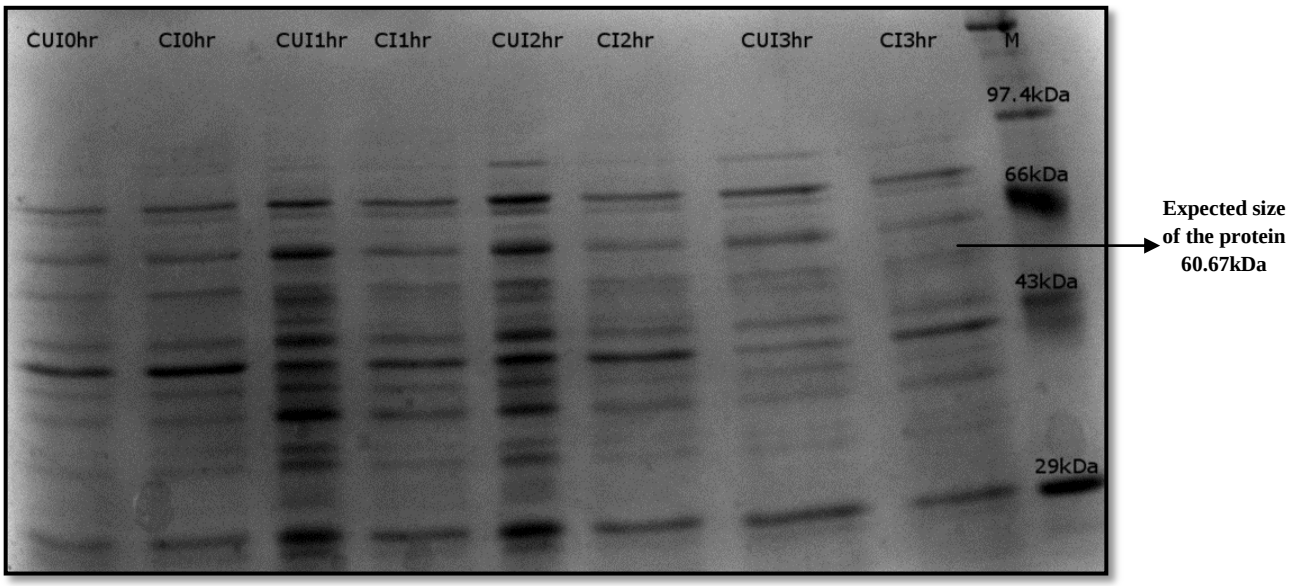
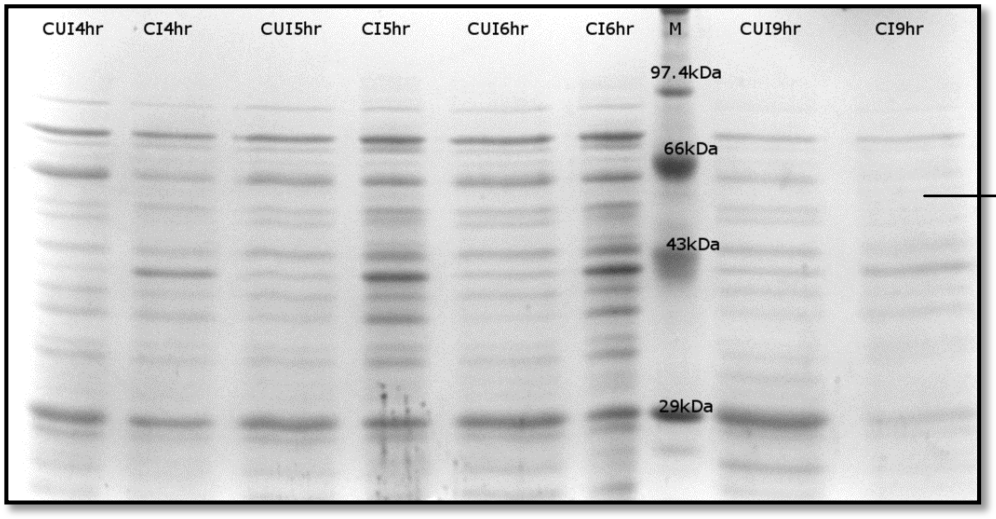


Figure 4.18-A: Protein Induction of PFD0300W in BL21 D3 host cell at 37°Cwith 1mM IPTG in native buffer: Lane 1-9; Sup CUI0hr, Sup CI0hr, Sup CUI1hr, Sup CI1hr, Sup CUIhr, SupCI2hr, Sup CUI3hr, Sup CI3hr, Protein high molecular weight marker PMWL (29-205kDa)



B

Figure 4.18-B-Protein Induction of PFD0300W in BL21DE3 host cell using at 37°C with 1mM IPTG in Native buffer supernatant: Lane 1-9; Sup CUI4hr, Sup CI4hr, Sup CUI5hr, Sup CI5hr, Sup CUI6hr, Sup CI6hr, Sup CUI9hr, Sup CI9hr, Protein high molecular weight marker PMWH (29-205 kDa).

4.5-Protein expression of Construct in Arctic Express (DE3) Competent cells:

The method used for the protein expression in Arctic cells has already been discussed in materials and methods section 2.10.2., However, even with the arctic cells, we have obtained an expression of the cpn 60 chaperonin but could not detect any band at ~60.67 kDa figure **4.19**.

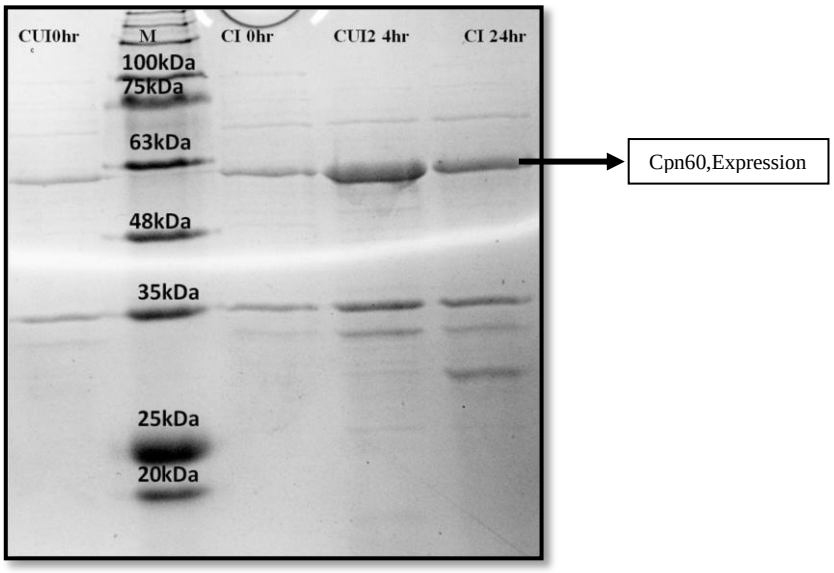


Figure 4.19: PFD0300W expression in Arctic Express competent cells: Lane1-5; CUI0hr, BR-Biochem, Prestained marker (11-245kDa), CI0hr, CUI24hr, CI24hr.

4.6-mRNA transcript analysis of PFL2320W:

Since we were not able to detect the expressed protein under any of the conditions used for the protein expression studies, we decided to check whether, at the transcript level, the mRNA specific for the proteins were detectable. Although it is known that high mRNA levels do not necessarily signifies protein expression. To decipher whether the two constructs mentioned earlier are at all expressing at the mRNA level, RT-PCR using specific primers were performed on the cDNA synthesized from mRNA isolated from the constructs transformed in the host BL21DE3 cells grown at 37°C with 1mM IPTG harvested at different time points **figure 4.20**, **figure 4.21**. RNA isolation and cDNA preparation of total RNA of bacterial cells are given in materials of methods in section 2.4 and 2.5.

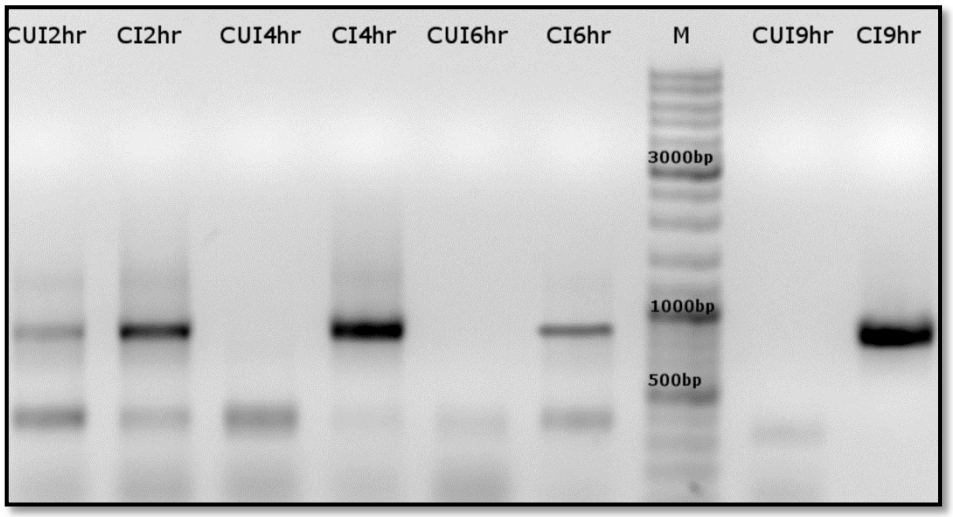


Figure 4.20: cDNA amplification of PFL2320W from DNase treated mRNA samples of clone in BL21DE3 cells at 37°C; Lane1-9; CUI2hr, CI2hr, CUI4hr, CI4hr, CUI6hr, CI6hr, Marker SM0331(100-10,000bp) Fermentas.

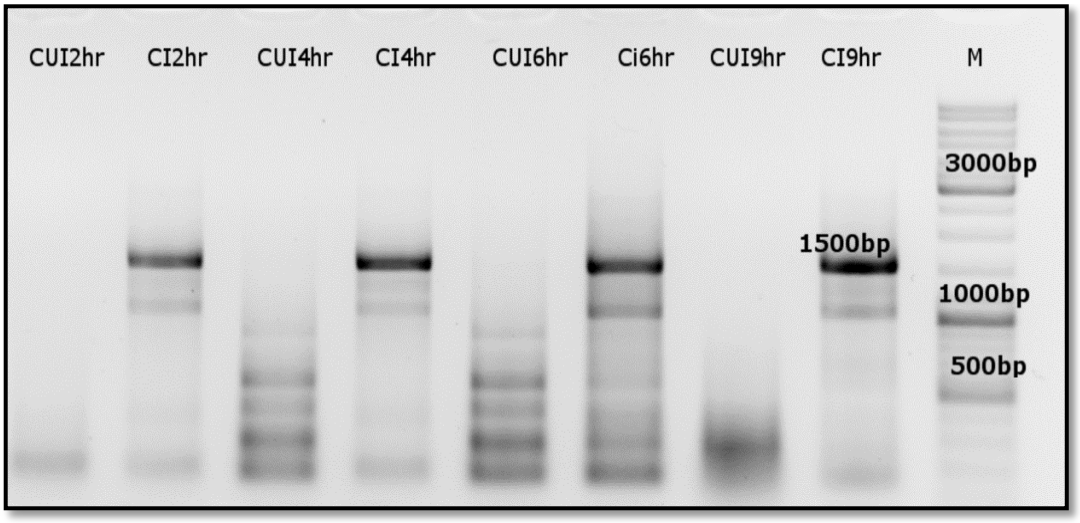


Figure 4.21: cDNA amplification of PFD0300W from DNase treated mRNA samples of clone in BL21DE3 cells at 37°C; Lane1-9; CUI2hr, CI2hr, CUI4hr, CI4hr, CUI6hr, CI6hr, CUI9hr, CI9hr Marker SM0331 (100-10,000bp), Fermentas.

We have detected the transcripts for both PFL2320W and PFD0300W codon-optimized genes. However, repeated protein expression studies did not yield any satisfactory results.

Conclusion:

In conclusion of the chapter, even after trying many protein expression protocols and different conditions we were not able to express both the proteins in a heterologous *E. coli* based system. Initially, we predicted that the reason of no protein expression could as the presence of rare codon. However, even after optimizing the protein for rare codons, the optimized construct did not show expression in any of the *E. coli* hosts. On checking for the mRNA transcript expression in BL21DE3 it has been found that the transcript is present at all the time points suggesting that the mRNA is stable. The reason of mRNA expression and not the protein expression could be the toxicity of the protein to the host cell, mRNA secondary structure that prevents interaction with cellular machinery, post-translational modification of protein or the nature of protein could be intrinsically disordered which will not allow it to express in any particular conformation (Redwan, E. M. 2006). To confirm the physicochemical property of protein and disordered nature of protein, in-silico analysis of both the proteins were performed as discussed in chapter-v.

Chapter V

In-silico analysis and immunolocalization of conserved Plasmodium protein with unknown function (PFD0300W and PFL2320W)

In spite of the decrease in malaria burden, and mortality rates, the threat of malaria due to *P. falciparum* still prevails, this is because of the parasite resistance to the drugs due to random genetic events and selection pressure. *P. falciparum* has developed resistance to nearly all the antimalarials present in current use, although the resistance to antimalarials varies with the geographical locations. The *P. falciparum* genome codes for 5,300 genes out of which about 60 % of the gene codes for hypothetical/ putative proteins (Gardner. et al.,2002). These hypothetical proteins could be putative drug targets. A weighted gene co-expression based systems, network from *in- vivo* samples of patients suffering from complicated malaria (Jaundice and Jaundice along with renal failure) and uncomplicated malaria were generated from which a set of differentially regulated modules were generated and analysed, that provides a set of hub genes from each module which could be putative drug targets. In this chapter, we will be discussing the in-silico analysis of

two conserved *Plasmodium* protein PFL2320W and PFD0300W with unknown function and also their immunolocalization in cultured *P. falciparum* parasites from India.

5.1- Results of in-silico analysis:

The hub genes were extracted from the co-expression based differentially regulated systems network from the uncomplicated and complicated disease manifestation patients. Up-regulated module and a steady-state module were selected for the selection of hub genes. Gene products that were found to be up-regulated in the up-regulated module in all the complicated cases i.e Jaundice (JAU), Jaundice & Renal failure (JRF), and steady-state gene product from the steady-state module used were respectively enriched in the translational biological process and oxidation and reduction process which are important biological processes for the parasite survival machinery. Results of computational analysis of the proteins are presented in the following sequence **figure 5.1.**

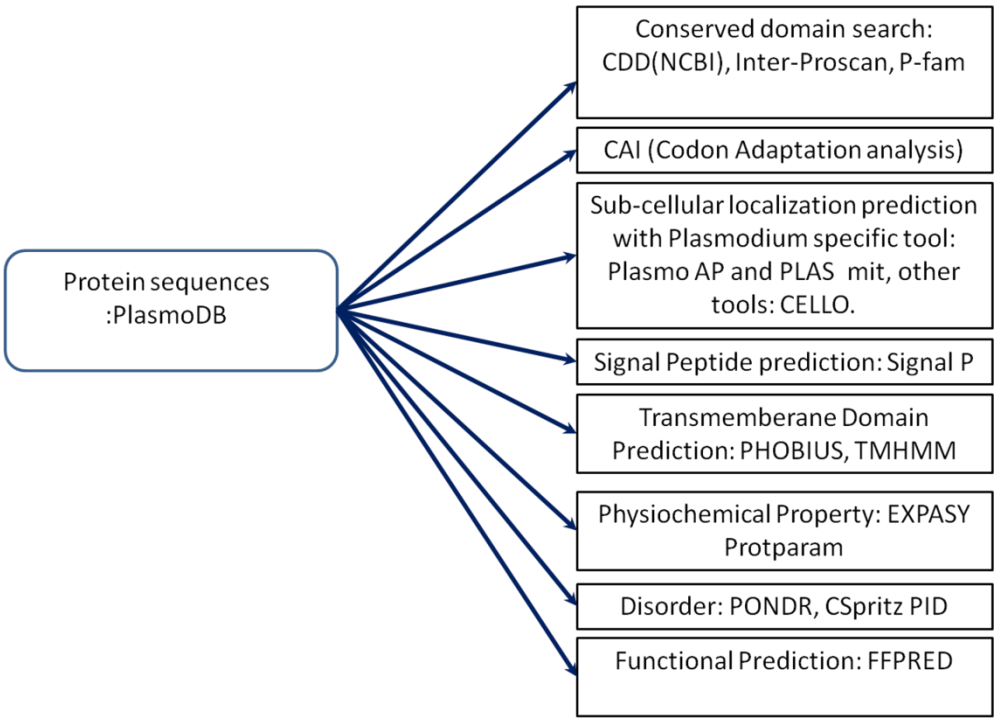


Figure 5.1: Computational workflow used for the analysis of proteins

5.1.1- Conserved Domain Data analysis:

A conserved domain analysis using CDD, available at NCBI, revealed that PFD0300W contains an NBD sugar kinase HSP_70 actin superfamily domain in the ordered region of the protein from the stretch of 164 -218 amino acid sequence, **figure: 5.2**. The e-value was significant 0.08, in the NCBI CDD (using the e-value threshold of 0.100 as a CDD search parameter). We were unable to get any significant conserved domain with other Conserved domain search tool like P-fam and InterProScan. On the other hand, PFL2320W does not contain any conserved domain in the entire sequence of the protein, as a default threshold value of CDD or at e-value of 0.100 **figures 5.3** or with other conserved domain search tools.

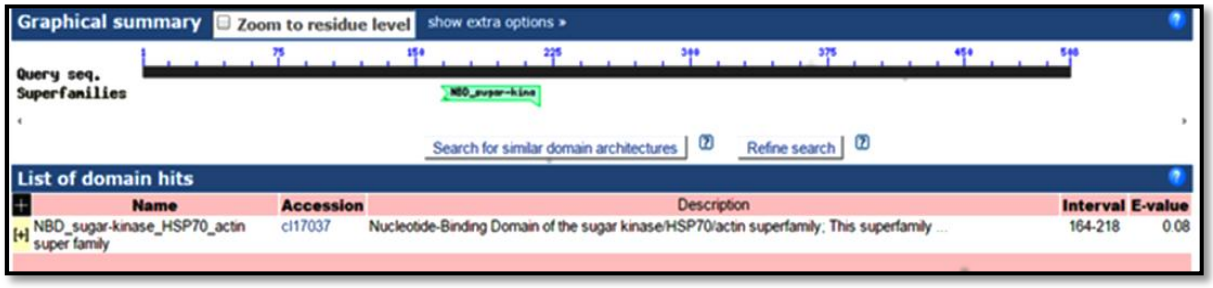


Figure 5.2: Conserved domain of PFD0300W from a stretch of 164-218: The black line represents the protein sequence, the green bar represents NBD_sugar kinase_HSP70_actin superfamily conserved domain in light green colour in PFD0300W sequence.

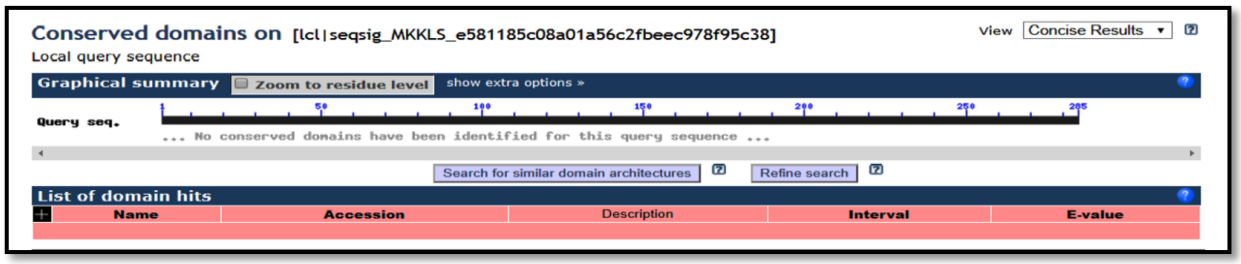


Figure 5.3: Conserved domain of PFL2320W: No conserved domain was found in PFL2320W at default threshold value of e-value of 0.1

5.1.2- Codon Adaptation Index:

Like codon biasness, there could be heterogenicity in the usage of the code within the species. To access the codon usage of the genes, codon adaptation Index of PFD0300W and PFL2320W were assessed and found to be 0.753 and 0.699 in comparison with the housekeeping gene seryl-tRNA synthetase which is 0.809, these CAI index suggests that

PFD0300W is relatively highly expressed and PFL2320W is moderately expressed in *P. falciparum* in comparison to the housekeeping gene seryl-tRNA synthetase and it suggests that it is worthwhile to consider these proteins for further investigation, CAI index was calculated using CAIcal (available at http://genomes.urv.es/CAIcal/).

Name	**Length of the gene**	**CAI**
PFD07_0073 (Seryl tRNA synthetase)	1620bp	0.809
PFD0300W	1527bp	0.753
PFL2320W	858bp	0.699

Table 5.1: Codon Adaptation Index of proteins: CAI index of PFD0300W, PFL2320W and Seryl tRNA synthetase.

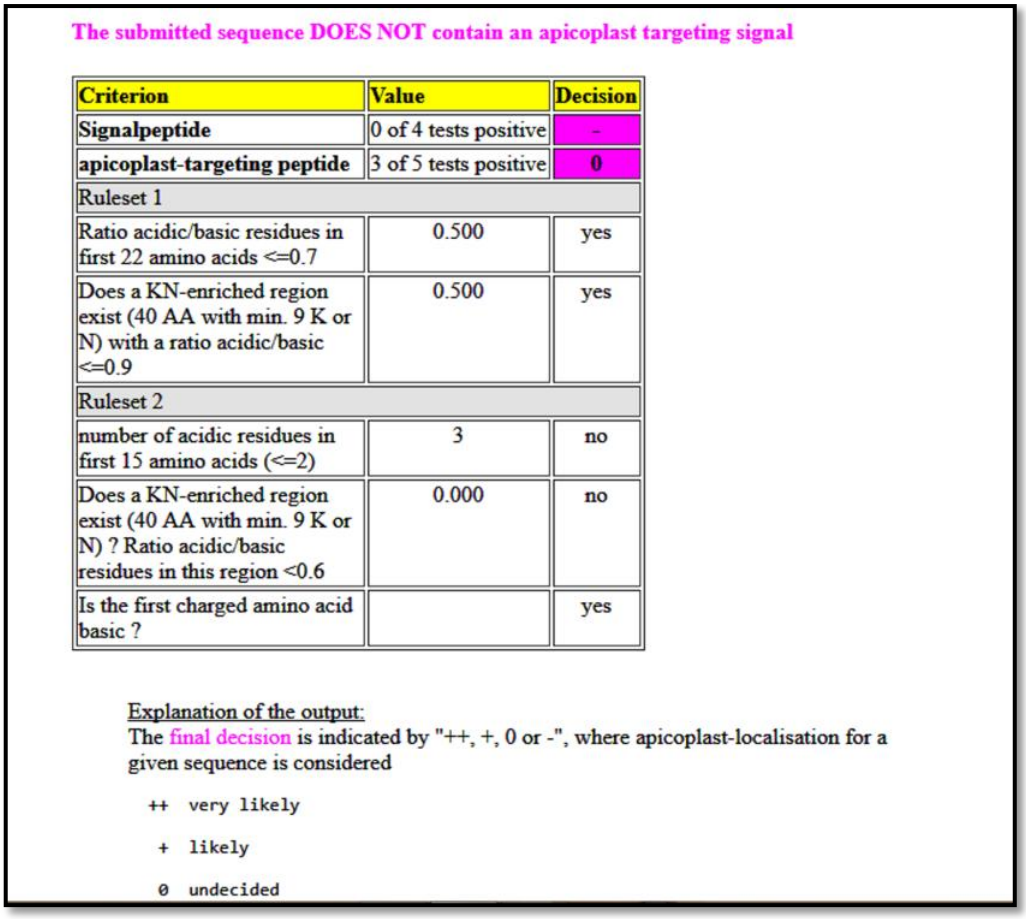
5.1.3-Prediction of sub-cellular localization signal peptide and physicochemical characterization of PFD0300W and PFL2320W:

Since, both the proteins are novel and conserved *Plasmodium* protein with unknown function and have not been explored yet, it is important to know the physicochemical property of the proteins and the sub-cellular localization of the protein.

5.1.3.1: Prediction of Subcellular localization

Targeting of the two genes was predicted using *Plasmodium* specific tools like PlasmoAP and PlasMit used for the prediction of apicoplast targeting and mitochondrial targeting signal respectively, these tools suggests that PFD0300W and PFL2320W not to be localized to apicoplast **figure 5.4** and **figure 5.5** or in mitochondria **figure 5.6** The signal peptide prediction for PFL2320W and PFD0300W suggests absence of signal sequences in both the proteins **figure 5.7**. Other than *Plasmodium*-specific localization tools other tools were also used for cellular localization like CELLO. Prediction of PFD0300W and PFL2320W by CELLO suggests the nuclear localization of the proteins with a reliability score of 3.141 for PFD0300W and 2.942 for PFL2320W **figure 5.8**. The physicochemical properties of both the proteins were accessed by an ExPASY protparam tool (http://web.expasy.org/cgi-bin/protparam/protparam) with the help of which Pi values, aliphatic index, and hydrophobicity of proteins were calculated. The Pi value of PFL2320W is 9.15 with a molecular weight of 34.72 kDa and PFD0300W was

8.76 with molecular weight of 60.67 kDa. The Pi value suggests that both the proteins are basic in nature with more of the basic amino acid residues in the proteins, which might be useful for protein extraction and purification purposes. The aliphatic index of PFL2320W is 48.8 and of PFD0300W is 77.85, higher the aliphatic index higher the thermostability of globular proteins. This suggests that PFD0300W is stable at higher temperature, whereas PFL2320W is less stable at higher temperature. The Grand average of hydropathy (GRAVY) of PFL2320W is -1.358 and PFD0300W is -0.914, which suggests that PFL2320W is more hydrophilic than PFD0300W.



The submitted sequence DOES NOT contain an apicoplast targeting signal

Criterion	Value	Decision
Signalpeptide	0 of 4 tests positive	-
apicoplast-targeting peptide	3 of 5 tests positive	**0**
Ruleset 1		
Ratio acidic/basic residues in first 22 amino acids <=0.7	0.500	yes
Does a KN-enriched region exist (40 AA with min. 9 K or N) with a ratio acidic/basic <=0.9	0.500	yes
Ruleset 2		
number of acidic residues in first 15 amino acids (<=2)	3	no
Does a KN-enriched region exist (40 AA with min. 9 K or N) ? Ratio acidic/basic residues in this region <0.6	0.000	no
Is the first charged amino acid basic ?		yes

Explanation of the output:
The final decision is indicated by "++, +, 0 or -", where apicoplast-localisation for a given sequence is considered

++ very likely

+ likely

0 undecided

Figure 5.4: PlasmoAP analysis of PFD0300W: The PlasmoAP prediction for PFD0300W suggest the absence of apicoplast targeting signal in protein.

The submitted sequence DOES NOT contain an apicoplast targeting signal

Criterion	Value	Decision
Signalpeptide	0 of 4 tests positive	-
apicoplast-targeting peptide	5 of 5 tests positive	++
Ruleset 1		
Ratio acidic/basic residues in first 22 amino acids <=0.7	0.400	yes
Does a KN-enriched region exist (40 AA with min. 9 K or N) with a ratio acidic/basic <=0.9	0.538	yes
Ruleset 2		
number of acidic residues in first 15 amino acids (<=2)	2	yes
Does a KN-enriched region exist (40 AA with min. 9 K or N) ? Ratio acidic/basic residues in this region <0.6	0.538	yes
Is the first charged amino acid basic ?		yes

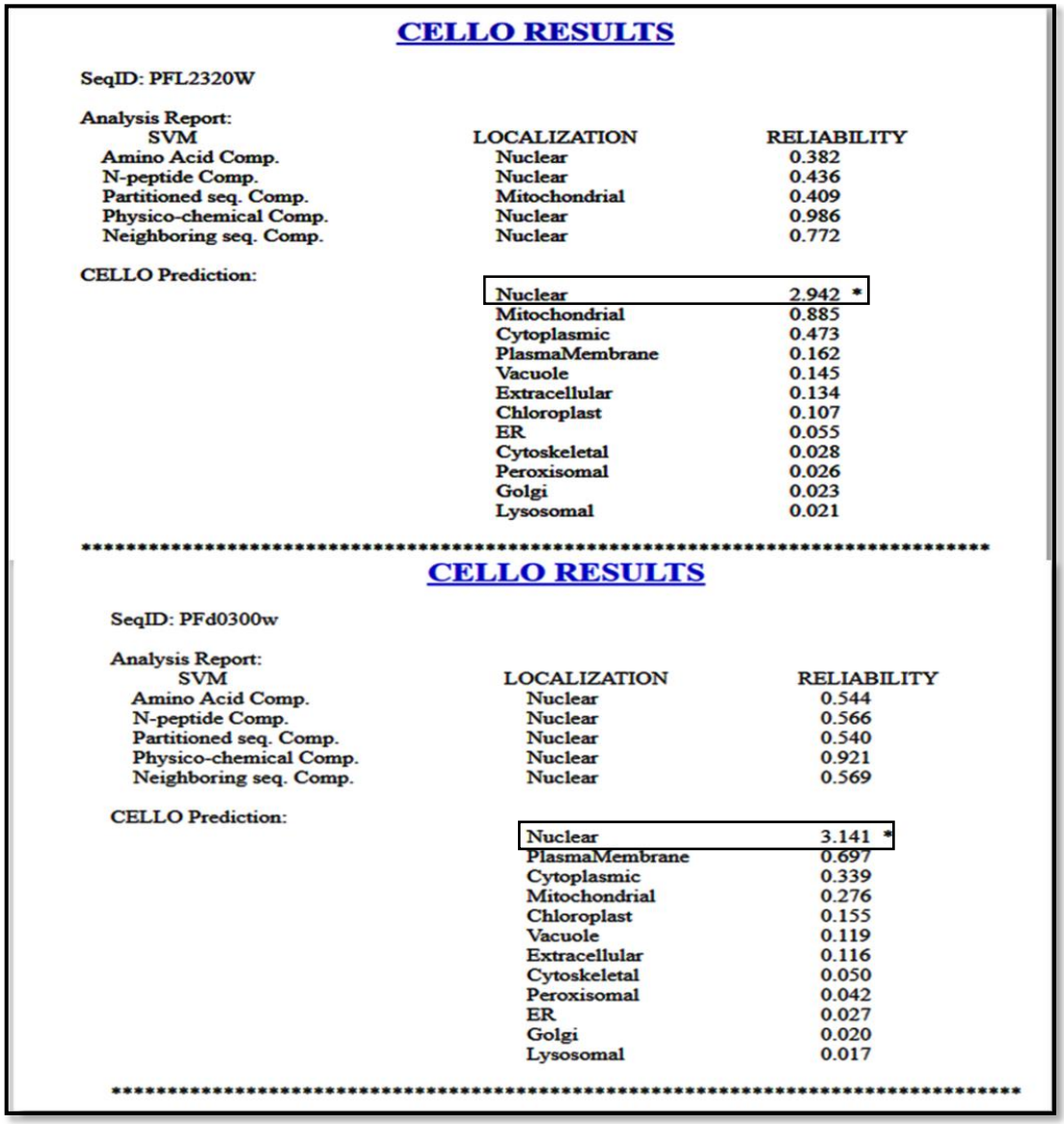
Explanation of the output:
The final decision is indicated by "++, +, 0 or -", where apicoplast-localisation for a given sequence is considered

```
++  very likely
 +  likely
 0  undecided
```

Figure 5.5: PlasmoAP analysis of PFL2320W: The PlasmoAP prediction for PFL2320W suggest the very likely presence of apicoplast targeting peptide but the absence of apicoplast targeting signal peptide in the protein.

Nr	Length	Name	Jury	Strict
0	80	>PFD0300W	non-mito (99%)	n/a
Nr	Length	Name	Jury	Strict
0	284	>PFL2320W	mito (91%)	failed (possibly mito)

Figure 5.6: PlasMit analysis of PFL2320W and PFD0300W: PlasMit analysis suggests the absence of mitochondrial targeting signal in PFD0300W and suggests that PFL2320W could be partially mitochondrial protein.



CELLO RESULTS

SeqID: PFL2320W

Analysis Report:

SVM	LOCALIZATION	RELIABILITY
Amino Acid Comp.	Nuclear	0.382
N-peptide Comp.	Nuclear	0.436
Partitioned seq. Comp.	Mitochondrial	0.409
Physico-chemical Comp.	Nuclear	0.986
Neighboring seq. Comp.	Nuclear	0.772

CELLO Prediction:

Nuclear	2.942 *
Mitochondrial	0.885
Cytoplasmic	0.473
PlasmaMembrane	0.162
Vacuole	0.145
Extracellular	0.134
Chloroplast	0.107
ER	0.055
Cytoskeletal	0.028
Peroxisomal	0.026
Golgi	0.023
Lysosomal	0.021

****

CELLO RESULTS

SeqID: PFd0300w

Analysis Report:

SVM	LOCALIZATION	RELIABILITY
Amino Acid Comp.	Nuclear	0.544
N-peptide Comp.	Nuclear	0.566
Partitioned seq. Comp.	Nuclear	0.540
Physico-chemical Comp.	Nuclear	0.921
Neighboring seq. Comp.	Nuclear	0.569

CELLO Prediction:

Nuclear	3.141 *
PlasmaMembrane	0.697
Cytoplasmic	0.339
Mitochondrial	0.276
Chloroplast	0.155
Vacuole	0.119
Extracellular	0.116
Cytoskeletal	0.050
Peroxisomal	0.042
ER	0.027
Golgi	0.020
Lysosomal	0.017

****

Figure 5.7: CELLO analysis of PFD0300W and PFL2320W: CELLO analysis suggests that both PFL2320W and PFD0300W are nuclear localized.

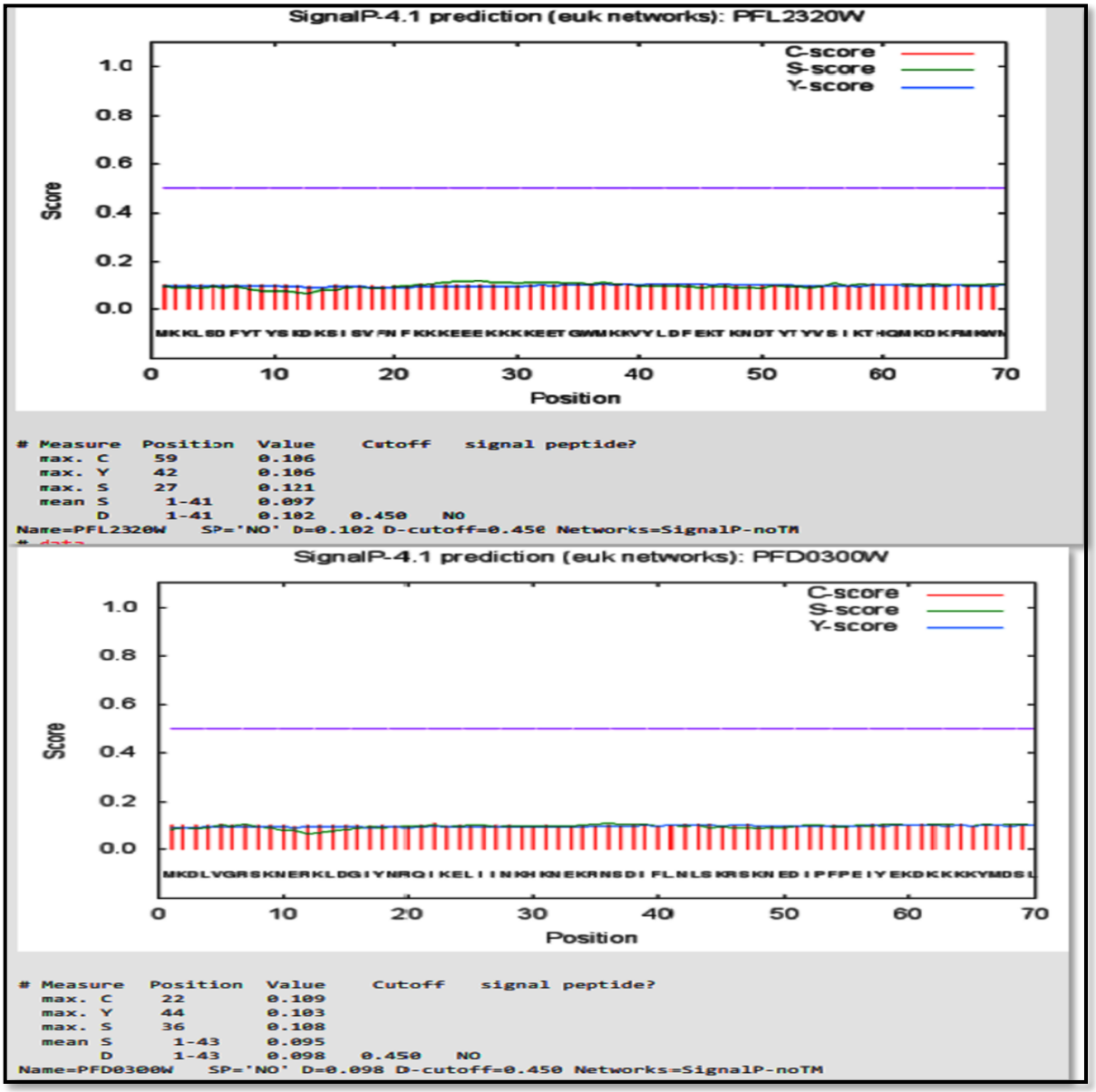


Figure 5.8: Signal peptide prediction of PFL2320W and PFD0300W; Signal P suggests the absence of signal peptide in PFL2320W and PFD0300W.

5.1.4- Transmembrane domain prediction:

Transmembrane domain analysis was performed for these proteins using the well-known Transmembrane prediction tools like PHOBIUS **figure 5.8** and **figure 5.9** TMHMM **figure5. 10** and **figure 5.11**.The analysis by PHOBIUS and TMHMM revealed that the N-terminus of both the proteins PFL2320W and PFD0300W are with zero transmembrane helices i.e. both of these proteins do not contain any of the transmembrane helices or any of the transmembrane domain.

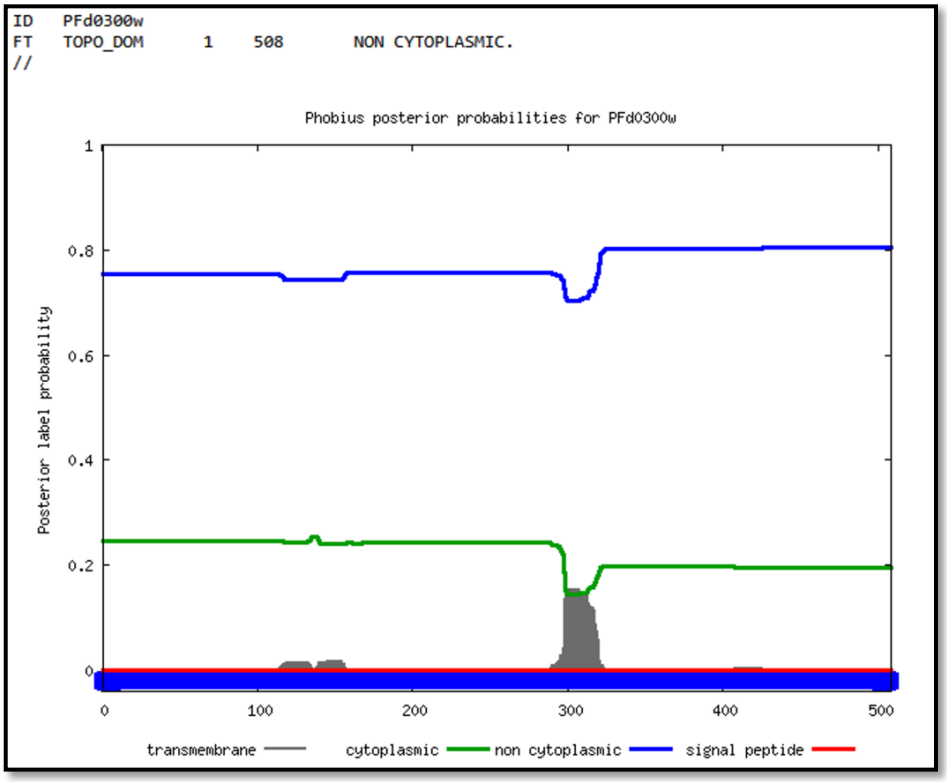


Figure 5.9: PHOBIUS transmembrane prediction of PFD0300W: Transmembrane prediction of PFD0300W by PHOBIUS.

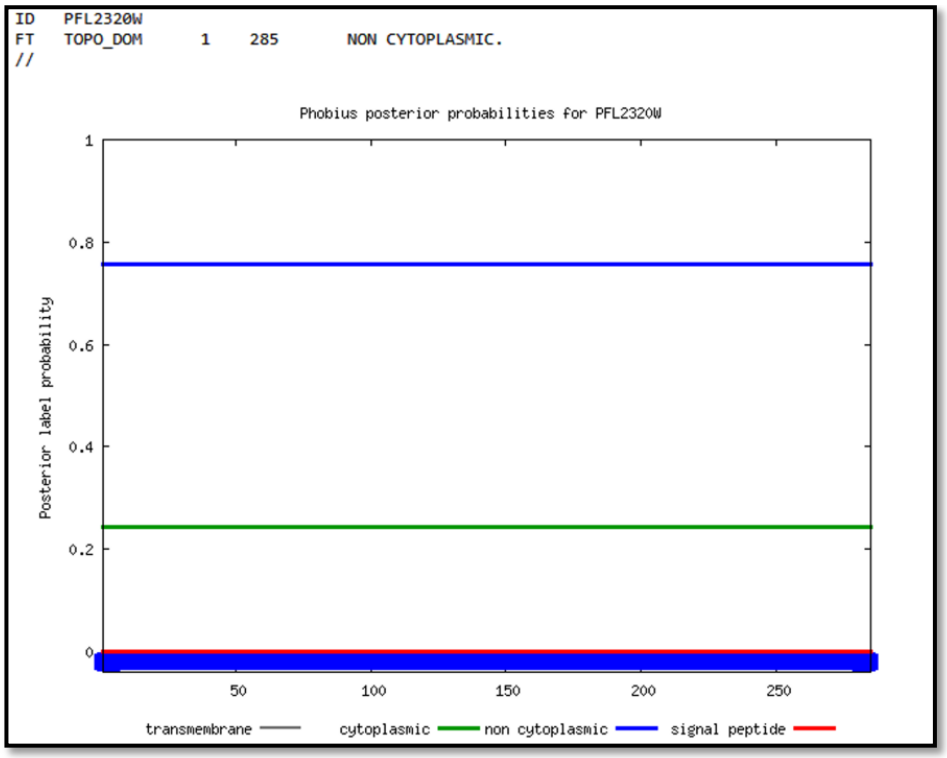


Figure 5.10: PHOBIUS analysis of transmembrane prediction of PFL2320W: Figure depicts the absence of transmembrane helices in PFL2320W and protein is non-cytoplasmic

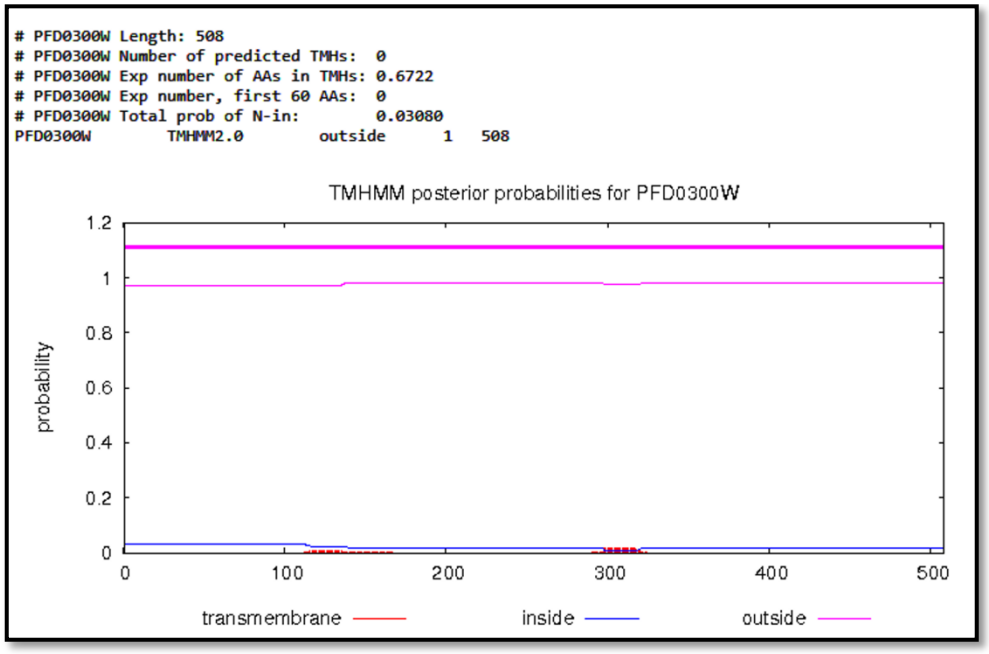


Figure 5.11: TMHMM analysis of PFD0300W: TMHMM analysis of PFD0300W no transmembrane domain has been detected in protein.

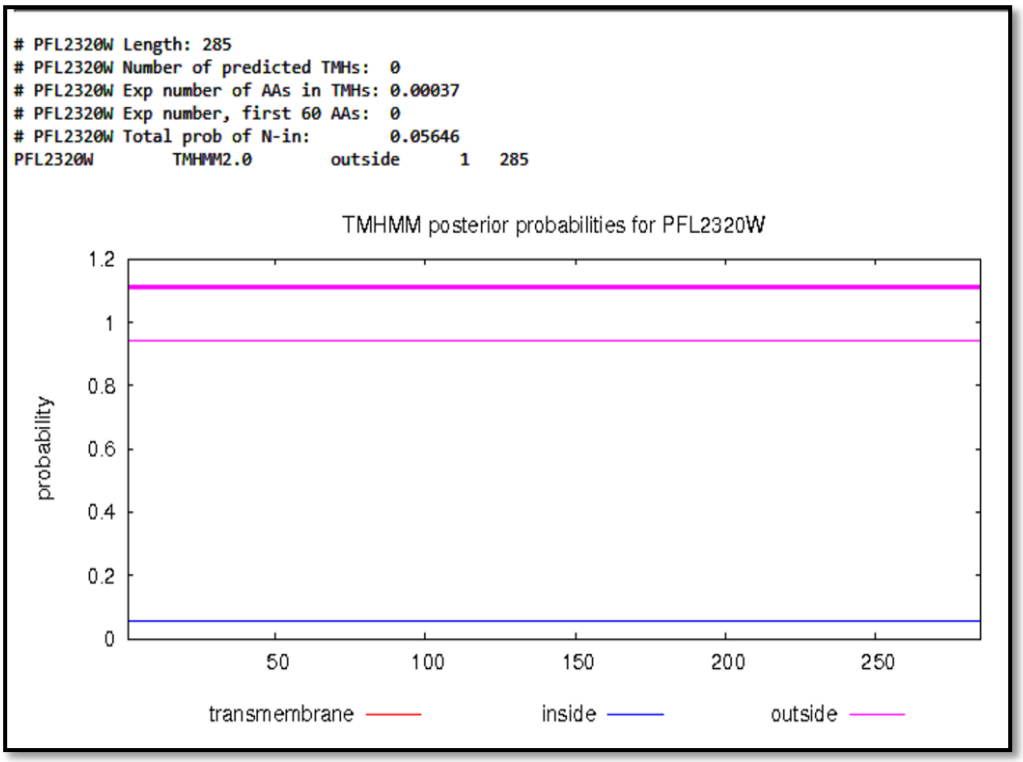


Figure 5.12: TMHMM analysis of PFL2320W: TMHMM analysis of PFL2320W shown that no transmembrane domain is present.

5.1.5- Protein disorder analysis:

Many of the proteins in *P. falciparum* are intrinsically disordered proteins so it was required to check for the disordered nature of these proteins. The result obtained from databases, PONDR VSL-2, VXL-2 tool, and CSpritz PID database available at suggest that percentage of disorderliness of PFD0300W is 46.45% with 3 disordered regions of >30 amino acids and one disordered region >50 amino acid. A total of 10 disordered segments has been found in the protein. Similarly, the disordered protein analysis of PFL2320W by PONDR VSL-2, VXL-2 figure 5.13 and CSpritz PID database suggests the percentage disorderness of PFL2320W is 48.77% with 1 disordered region of >30 amino acids and 1 disordered region of >50 amino acids. This suggests that these proteins could be important for parasite and can take any of the structural conformation or may involve in regulatory pathways of the parasite.

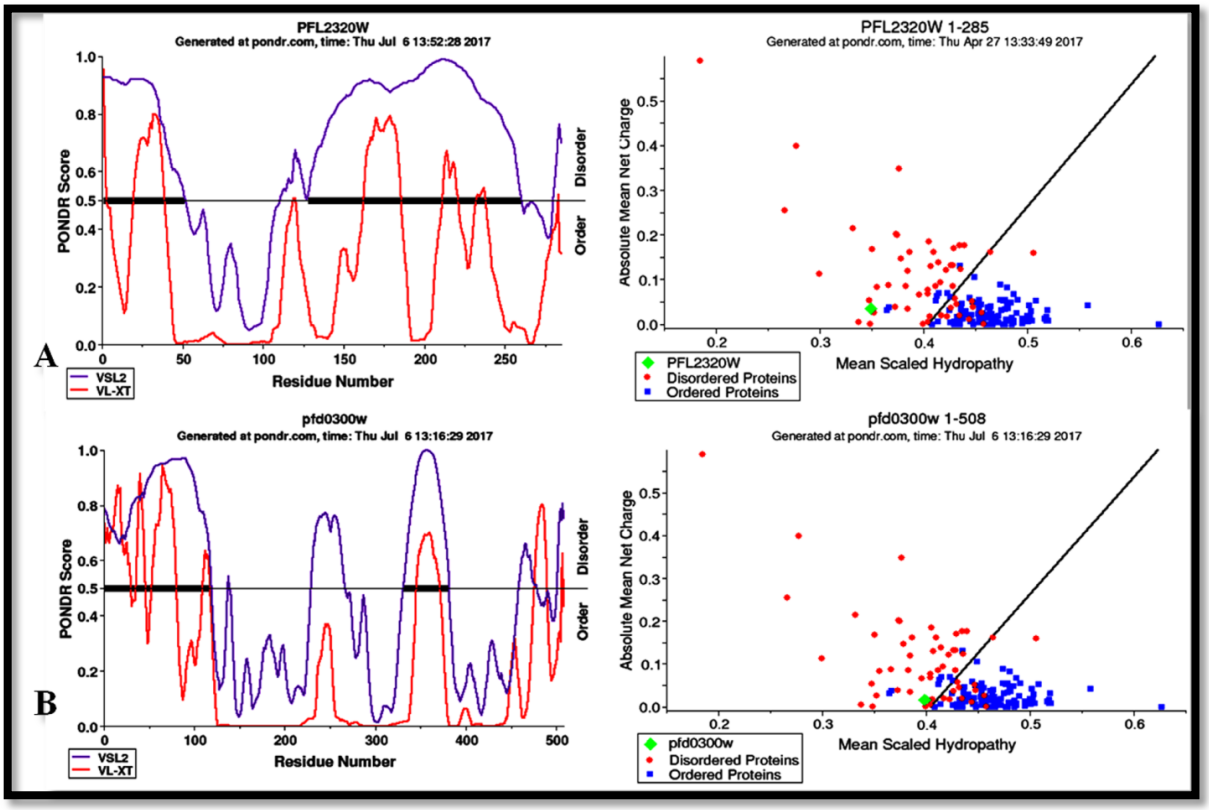


Figure 5.13: A and B: PONDR result of disorderedness of protein: Blue curves are representing long disordered region predicted using the VSL-2 tool and red curves are representing short disordered region using the VXL-2 tool in PFD0300W and PFL2320W.

5.1.6- Structure prediction of PFD0300W and PFL2320W:

Homology modelling of the two proteins for the generation of the tertiary structure was tried using different online tertiary structure prediction tools as explained in materials and methods section 2.19.7. Because of the conserved nature of the proteins within the *Plasmodium* species, we were not able to find any template for the proteins with a percentage identity of more than 25%. Thus, a valid tertiary structure of these proteins was not possible to predict. However, the secondary structure prediction of proteins, these proteins was performed using the PsiPRED tool, which gives the overall prediction of alpha helixes and beta sheets of the protein **figure 5.14**.The secondary structure prediction of PFL2320W suggests the presence of 11 alpha helixes and 5 beta sheets with presence of coiled region in between the alpha helixes and beta sheets. Similarly, the secondary structure prediction of PFD0300W suggests the presence of 20 alpha helixes and 8 beta strands with region of coiled-coil in between the beta strand and alpha helixes.

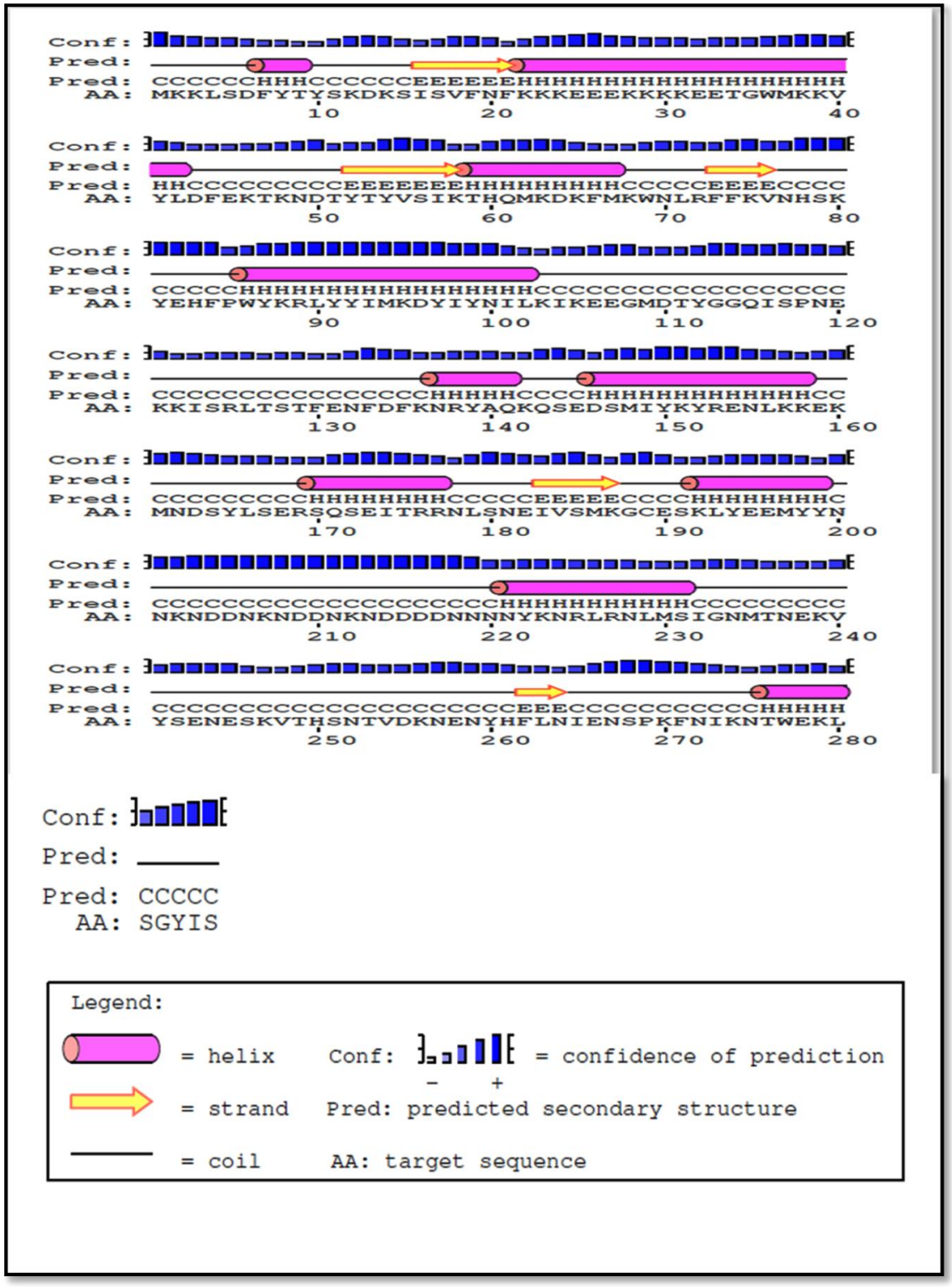


Figure 5.14:Secondary Structure of PFL2320W by PsiPred: The pink cylindrical structure represents alpha helixes whereas yellow arrowhead represents beta strands and the single line represents the coiled coil.

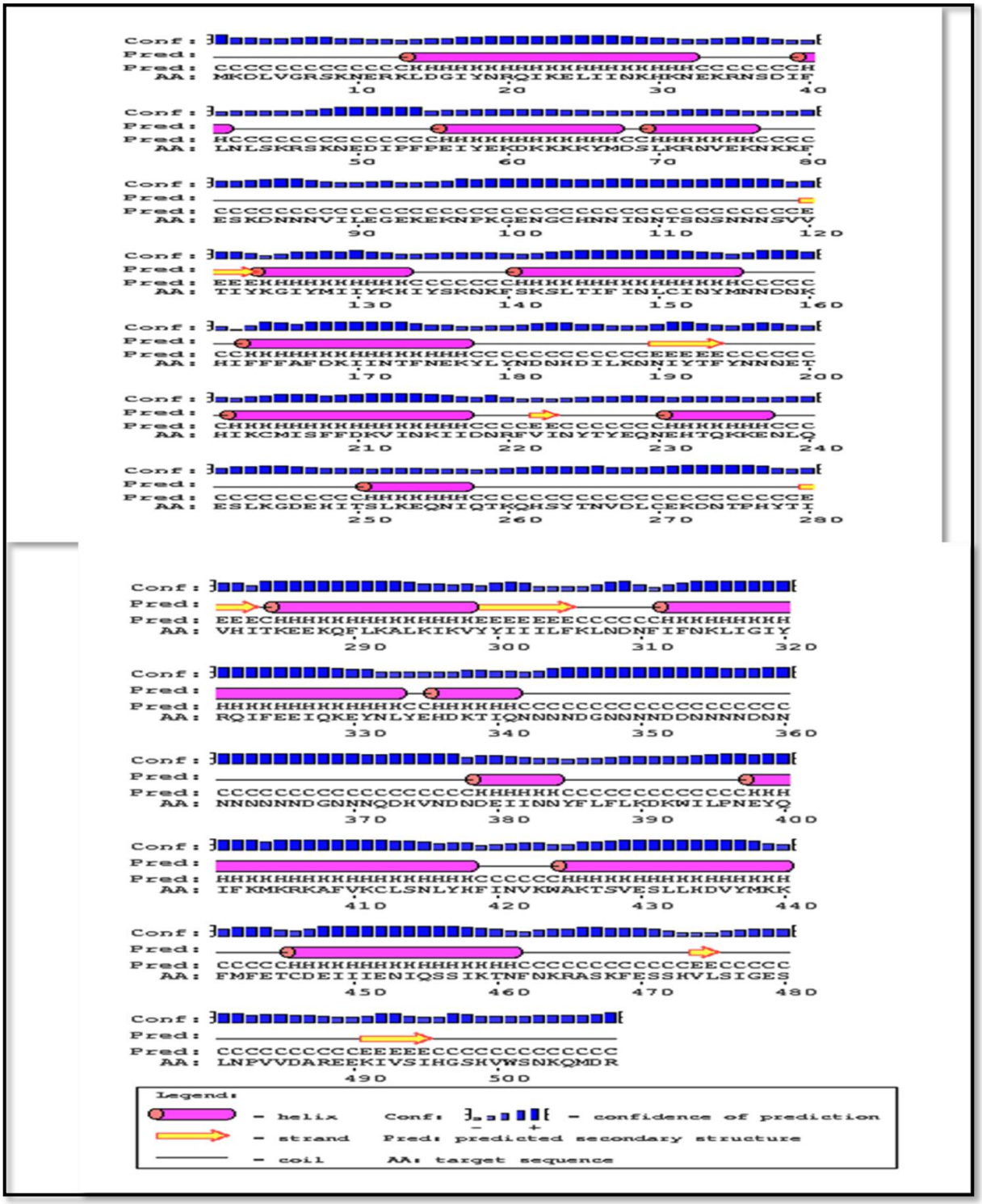


Figure 5.15: Secondary Structure of PFD0300W by PsiPred: The pink cylindrical structure represents alpha helixes whereas yellow arrowhead represents beta strands and the single line represents the coiled coil.

5.1.7-Functional Prediction of PFL2320W and PFD0300W:

Functional prediction of both the proteins was also done using FFPRED which is a tool of PsiPRED, The functional prediction analysis of PFD0300W shows that the gene is enriched in biological processes, like regulation of metabolic process, cellular macromolecule biosynthetic process, regulation of gene expression, regulation of RNA biosynthetic process, regulation of nitrogen compound metabolic process, mRNA metabolic process, RNA processing, regulation of transcription etc. The involvement of PFD0300W shows higher probability for nucleic acid binding, cytoskeletal protein binding, RNA binding, catalytic activity, actin binding, phosphotransferase activity, kinase binding, nucleotide binding etc, Whereas PFL2320W showed involvement in biological process like regulation of metabolic process, regulation of gene expression, cellular macromolecular biosynthetic process, regulation of RNA biosynthetic process, m RNA splicing, via spliceosome, RNA processing, The molecular function in which these proteins are involved are nucleic acid binding, transferase activity, transition metal ion binding etc. The higher probability of involvement of PFL2320W in biological process and molecular function suggest that PFL2320W may involve in oxidation and reduction metabolic process which could be important for parasite survival **figure 5.16** represents the biological process of the protein PFD0300W and PFL2320W, **figure 5.17** represents the molecular function of the proteins PFD0300W and PFL2320W.

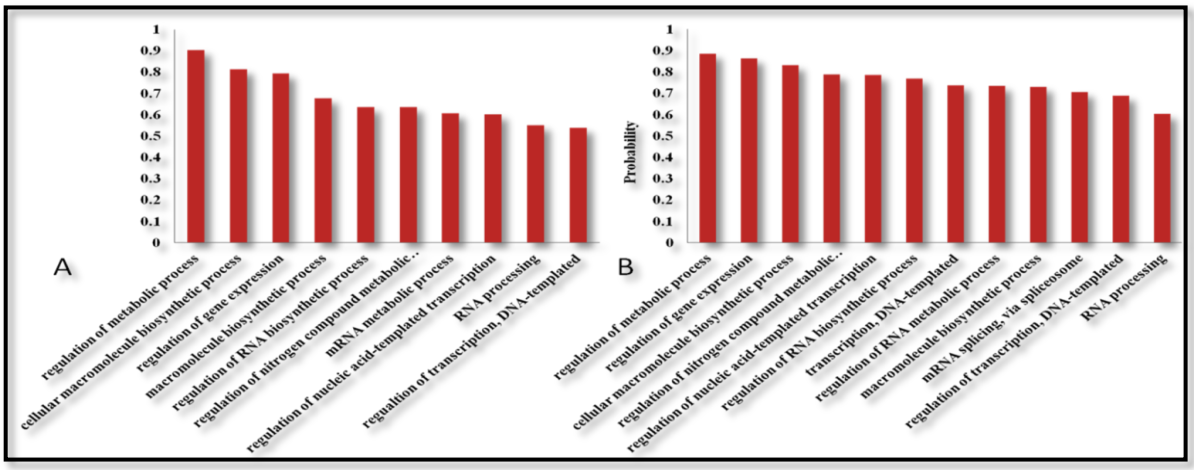


Figure 5.16 FFPRED analysis of PFD0300W and PFL2320W: A) Represent the biological process in which PFD0300W is involved x-axis represents the biological process, the y-axis represents the probability of biological processes in which it is involved. B) Represents the biological process in which PFL2320W is involved, the x-axis represents the biological process and the y-axis represents the probability of occurrence.

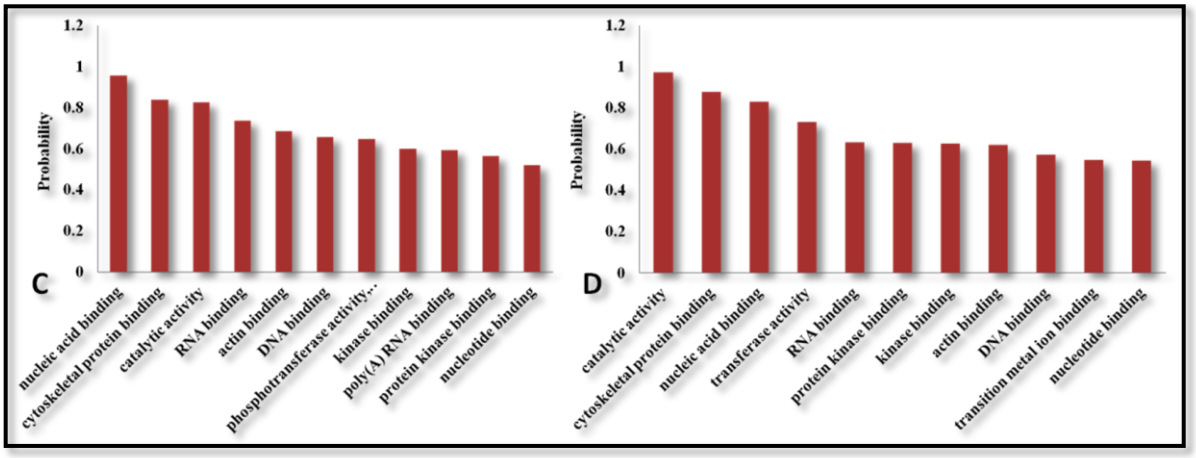


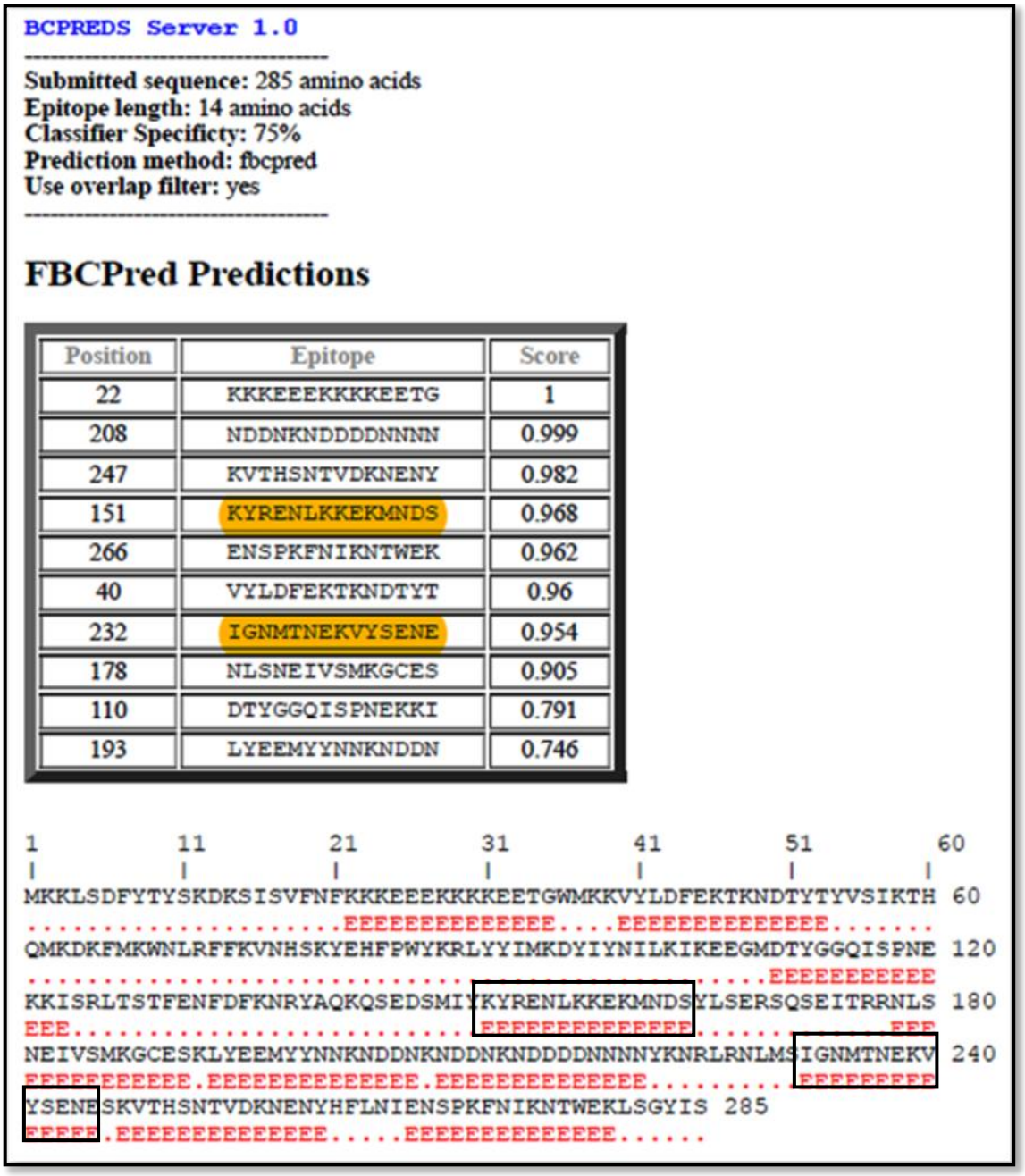
Figure 5.17: The Molecular function of PFD0300W and PFL2320W:C) represents the molecular function of PFD0300W where the x-axis represents the molecular function in which PFD0300W is involved and the y-axis represents the probability, D) represents the molecular function of PFL2320W where the x-axis represents the molecular function and Y-axis represents the probability of involvement of protein.

5.2-Immunolocalization and epitope prediction and immunization:

The conserved *Plasmodium* protein with unknown function and PFD0300W and PFL2320W could not be expressed in an *E. coli* based system.Thus for the immunolocalization process, their sequences were analysed for the presence of B-cell epitopes. Based on these epitopes peptides were synthesized and used to raise antisera, which was then being used for immunolocalization in *P. falciparum* culture isolates. The next section of this chapter describes the B-cell epitope prediction using BCPRED tool, peptide immunization, and Immunolocalization study of both the proteins.

5.2.1- BCPRED analysis for the generation of peptides:

A number of epitopes were predicted from the BCPRED tool, however, only two of the epitopes for each of the proteins were selected predicted with prediction score of either 0.9 or 1.Also, the epitopes were chosen where the position of the epitopes were quite distant from each other, for each individual protein, **figure 5.18** and **figure 5.19**.



BCPREDS Server 1.0

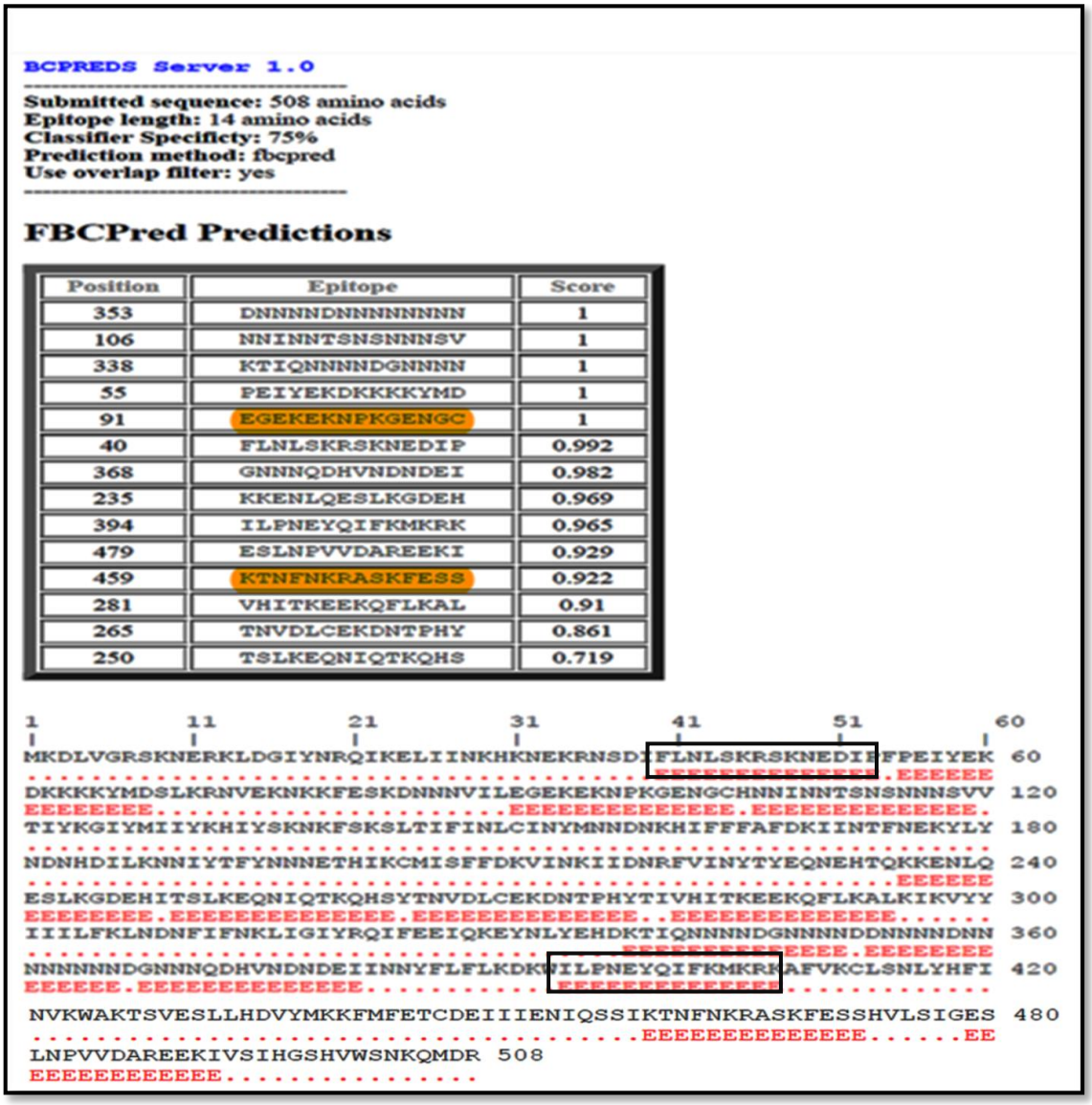
Submitted sequence: 285 amino acids
Epitope length: 14 amino acids
Classifier Specificty: 75%
Prediction method: fbcpred
Use overlap filter: yes

FBCPred Predictions

Position	Epitope	Score
22	KKKEEEKKKKEETG	1
208	NDDNKNDDDDNNNN	0.999
247	KVTHSNTVDKNENY	0.982
151	KYRENLKKEKMNDS	0.968
266	ENSPKFNIKNTWEK	0.962
40	VYLDFEKTKNDTYT	0.96
232	IGNMTNEKVYSENE	0.954
178	NLSNEIVSMKGCES	0.905
110	DTYGGQISPNEKKI	0.791
193	LYEEMYYNNKNDDN	0.746

```
1          11         21         31         41         51         60
|          |          |          |          |          |          |
MKKLSDFYTYSKDKSISVFNFKKKEEEKKKKEETGWMKKVYLDFEKTKNDTYTYVSIKTH 60
......................EEEEEEEEEEEEEE....EEEEEEEEEEEEEE.......
QMKDKFMKWNLRFFKVNHSKYEHFPWYKRLYYIMKDYIYNILKIKEEGMDTYGGQISPNE 120
.............................................EEEEEEEEEEE
KKISRLTSTFENFDFKNRYAQKQSEDSMIYKYRENLKKEKMNDSYLSERSQSEITRRNLS 180
EEE.........................EEEEEEEEEEEEEE.....EEE
NEIVSMKGCESKLYEEMYYNNKNDDNKNDDNKNDDDDNNNNYKNRLRNLMSIGNMTNEKV 240
EEEEEEEEEEE.EEEEEEEEEEEEEE.EEEEEEEEEEEEEE.........EEEEEEEEE
YSENESKVTHSNTVDKNENYHFLNIENSPKFNIKNTWEKLSGYIS 285
EEEEE.EEEEEEEEEEEEEE.....EEEEEEEEEEEEEE......
```

Figure 5.18: Epitope prediction of PFL2320W: Epitope of PFL2320W highlighted epitopes were selected for the antibody generation, position of the epitopes were from 151-165 and another one was from 232-246.



BCPREDS Server 1.0

Submitted sequence: 508 amino acids
Epitope length: 14 amino acids
Classifier Specificty: 75%
Prediction method: fbcpred
Use overlap filter: yes

FBCPred Predictions

Position	Epitope	Score
353	DNNNNDNNNNNNNNN	1
106	NNINNTSNSNNNSV	1
338	KTIQNNNNDGNNNN	1
55	PEIYEKDKKKKYMD	1
91	EGEKEKNPKGENGC	1
40	FLNLSKRSKNEDIP	0.992
368	GNNNQDHVNDNDEI	0.982
235	KKENLQESLKGDEH	0.969
394	ILPNEYQIFKMKRK	0.965
479	ESLNPVVDAREEKI	0.929
459	KTNFNKRASKFESS	0.922
281	VHITKEEKQFLKAL	0.91
265	TNVDLCEKDNTPHY	0.861
250	TSLKEQNIQTKQHS	0.719

Figure 5.19: Epitope prediction of PFD0300W: Selected Epitopes of PFD0300W are highlighted for the antibody generation, the position of the epitopes was from 91-105 and another one was from 459-473.

5. 2.2 -Immunization of PFD0300W and PFL2320W:

After selection and synthesis of the peptides, immunization experiments were initiated in out-bred Swiss Albino mice using Freund's adjuvants and un-conjugated peptides using the protocol described in chapter-2 section 2.12.1, (IAEC/RES/22/04) before proceeding with the immunolocalization studies the antibody titres against the peptides was evaluated using an indirect ELISA format.

5.2.3- ELISA of PFD0300W and PFL2320W:

ELISA was performed to check the antibody titre of the specific peptide-based antibody the method used for the ELISA analysis as discussed in materials and method section 2.13, only final boost of blood sera was taken and antibody titre of those sera were analysed f**igure 5.18**., the antibody dilution of 1:200 was used for the localization studies.

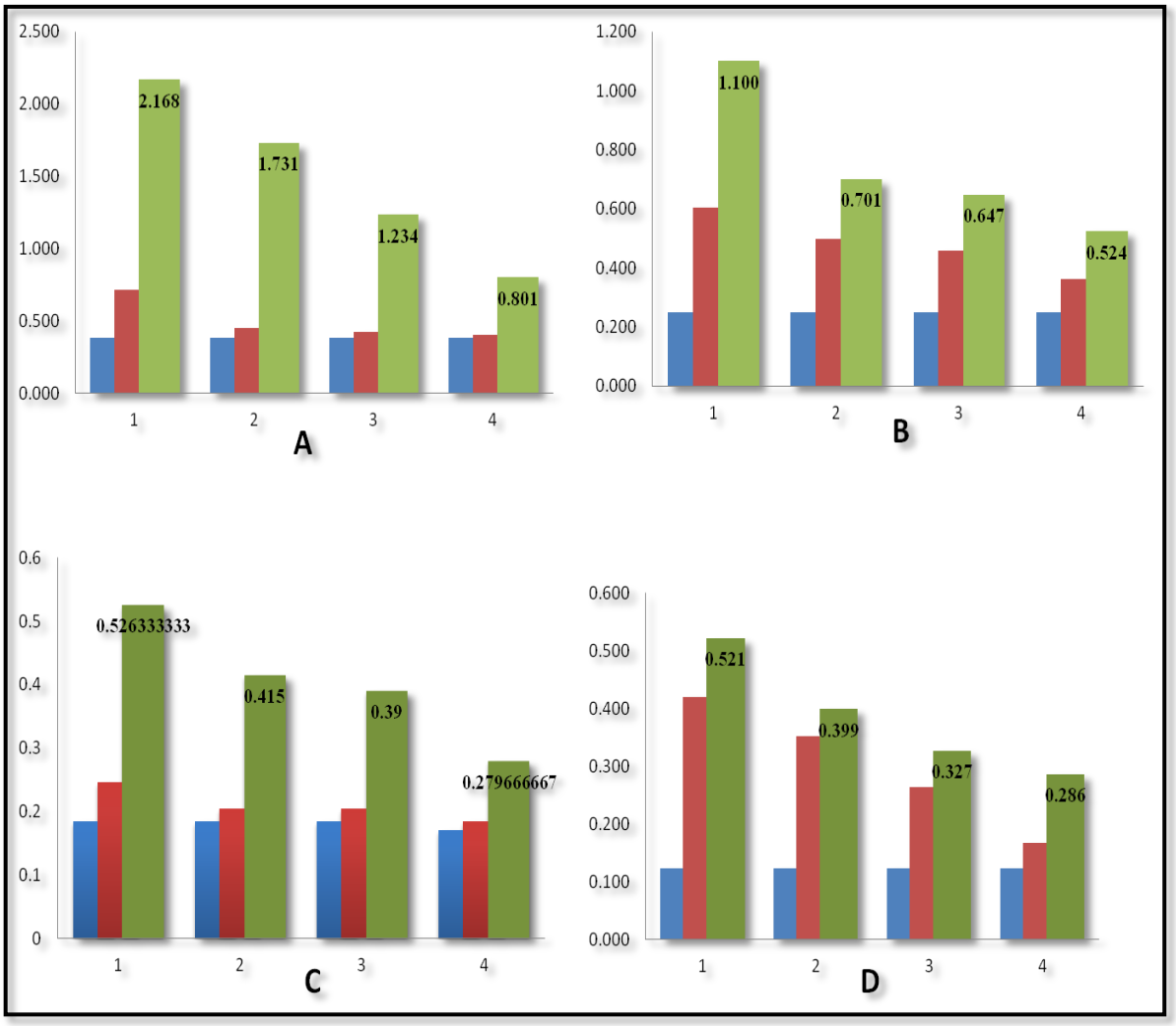


Figure 5.20: ELISA result of E1-300W, E2-300W, E1A-2320W, E3-2320W: A) ELISA result of E1-300w, blue bar represents PBS, maroon bar represents Pre-bleed, green bar represents final sera. B) E2-300w, blue bar represents PBS, the maroon bar represents pre-bleed serum, green bar represents final sera. C) E1A-2320w, blue bar represents PBS, the maroon bar represents pre-bleed, green bar-represent a final boost. D) E3-2320w, blue bar represent PBS, the maroon bar represents pre-bleed sera, green bar represents a final boost.

5.3- Localization study of PFD0300W and PFL2320W:

5.3.1- Localization of PFD0300W:

The co-expression network analysis, of the transcripts of this protein, were seen to be up-regulated in case of severe disease condition i.e., Jaundice and Jaundice + Renal failure. In a gene modification study by Bushell et al in 2017, on the asexual stage-specific gene of *P. berghei* revealed failure of the gene modification of ortholog of this protein, which suggests that this protein is important in the asexual blood stages of the *P. berghei*. The alignment of PFD0300W with orthologous gene PBANKA_1003700 of *P. berghei* ANKA strain shows 53.61% sequence identity **figure 5.21**. Our experiments show the presence of PFD0300W in the cytoplasm of all the blood stages of *Plasmodium* including gametocytes. To see whether this cytoplasm localization is restricted to the endoplasmic reticulum (ER), experiments were performed using ER tracker and the results indicates that the protein is not localized in ER. The nature of the cytoplasmic distribution of the protein did not in any way suggest localization in other organelles like in nucleus or in the mitochondria or Apicoplast, although our *in-silico* analysis data suggested possibilities of nuclear localization of this protein. **figure 5.22-(A-C), 5.22-(D-F)**, localization with E2-300W peptide-based antibody also show the protein to be localized in cytoplasm **figure 5.23-A, 5.23-B**.

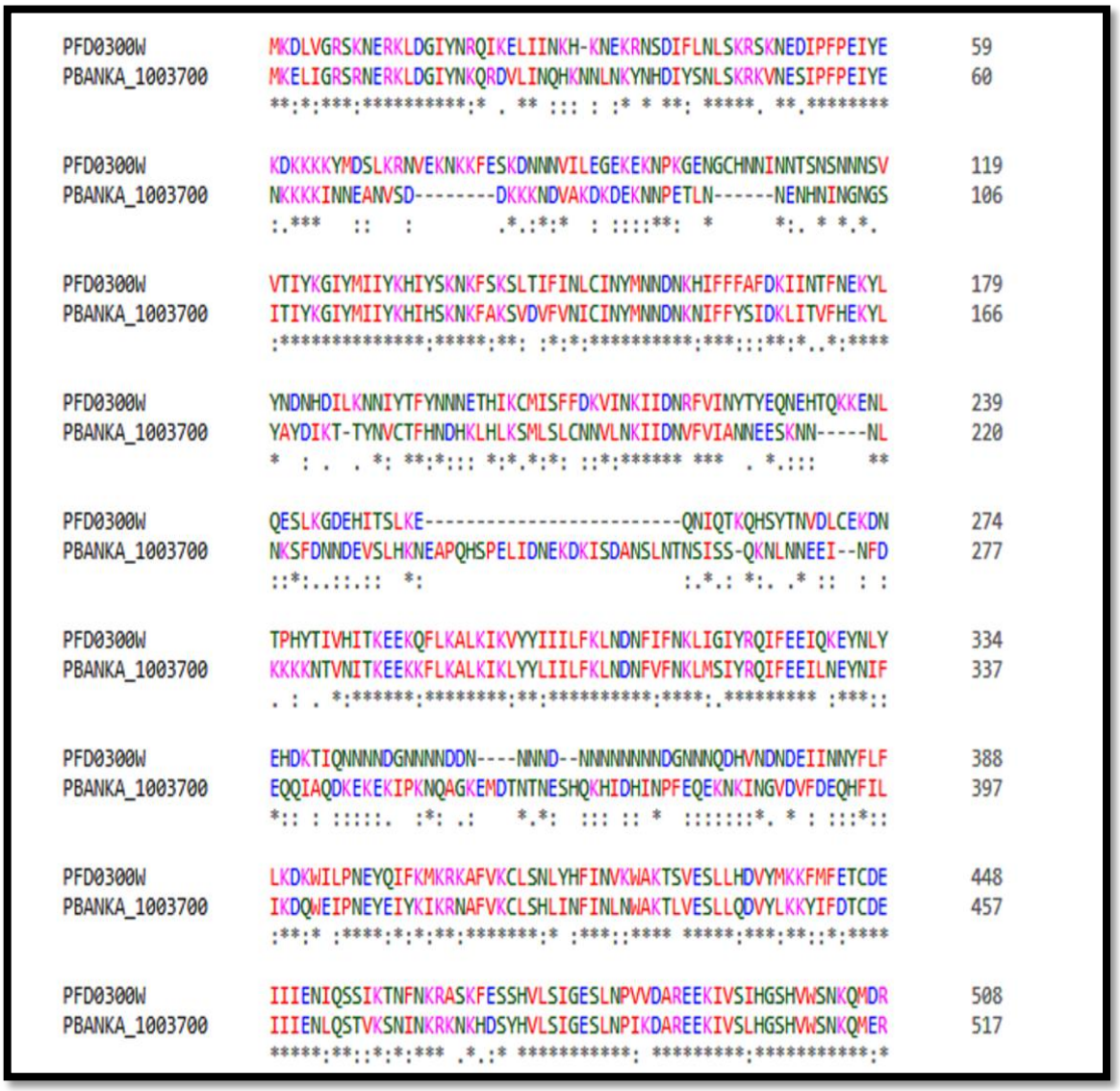


Figure5.21- Alignment of PFD0300W and *P.berghei* ANKA orthologous gene PBANKA_1003700.

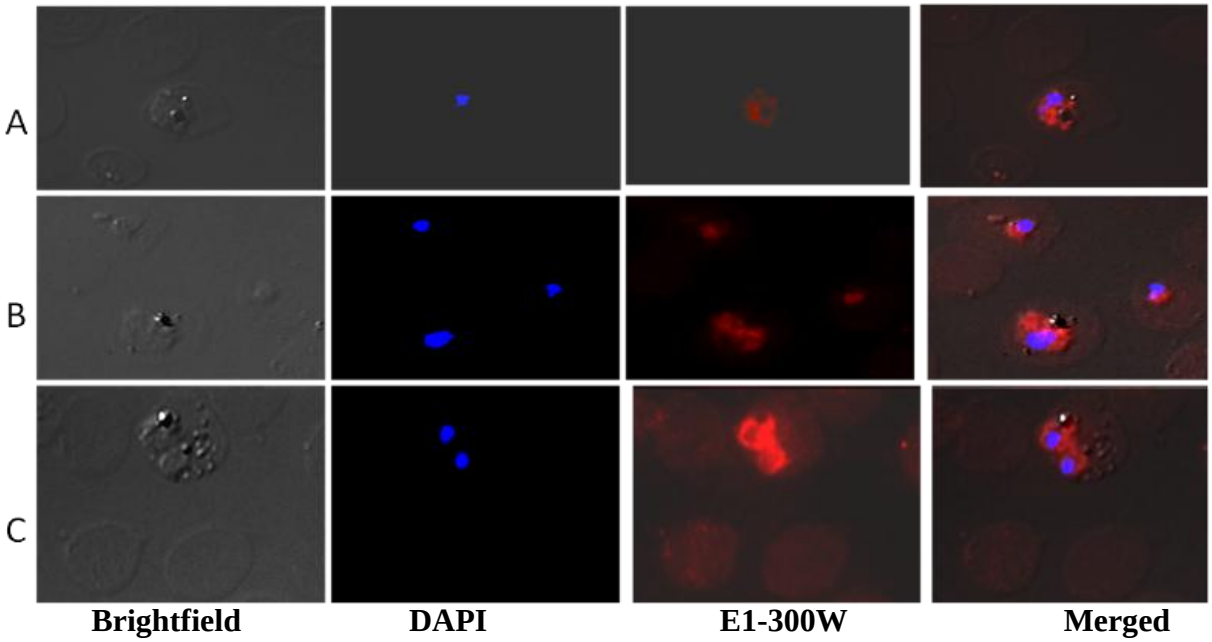


Figure 5.22 (A-C): Localization of E1-300W-Alexa flour- 494 labelled; E1-300W-Alexa flour-494 labelled (red) localized in cytoplasm of asexual stages of lab adapted *P. falciparum* culture of Indian isolate (RKL-9), DAPI is used as a nuclear stain**;** A, B, C represents late trophozoite stage.

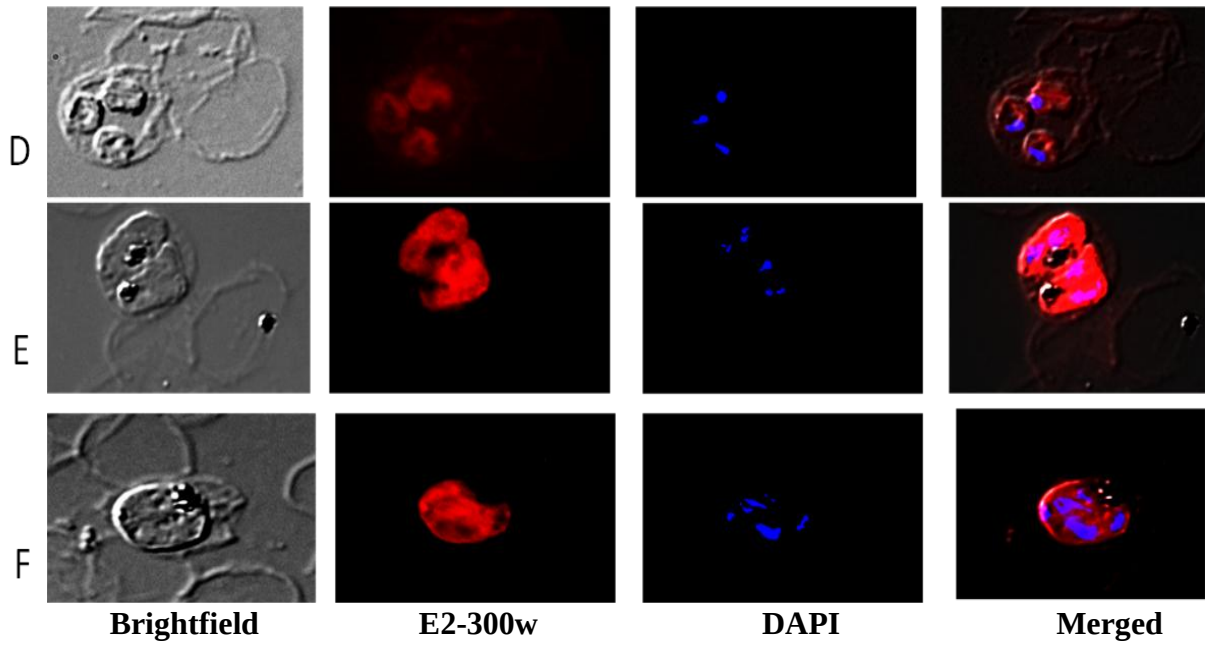


Figure 5.22 (D-F): E2-300W, asexual and sexual stage localization: D) localization of E2-300W Alexa flour-494 (red) labelled in cytoplasm of late trophozoite stage of lab adapted *P. falciparum* culture of Indian isolate (RKL-9) (E) localization of E2-300W Alexa flour-494 (red) labeled in cytoplasm of sexual gametocyte stage III, DAPI was used as a nuclear stain.

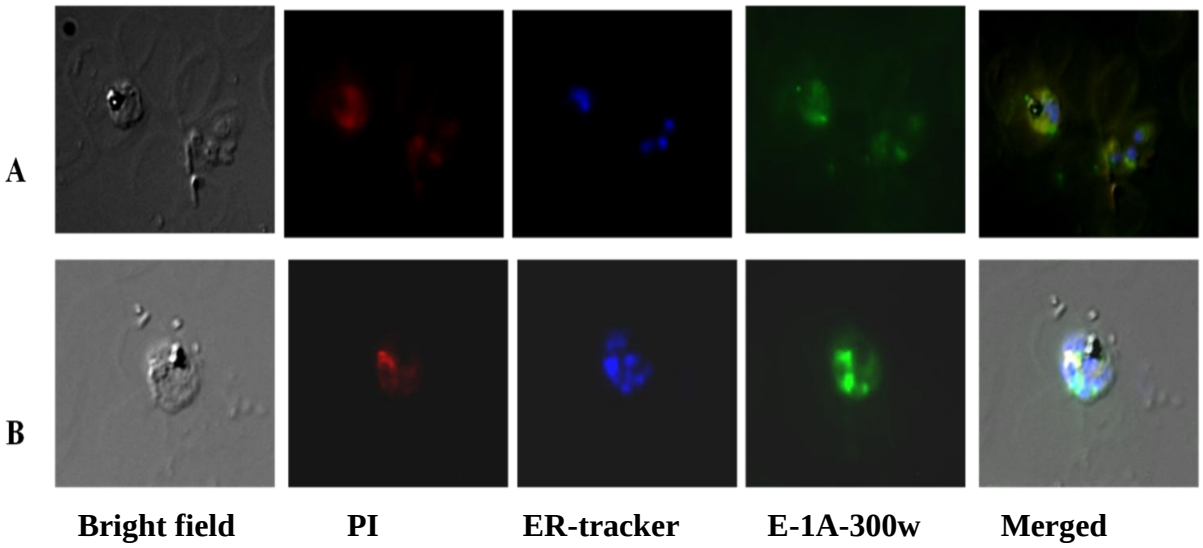


Figure 5.23 (A-B): Localization of E1-300W FITC labelled antibody in cultured Indian strain (RKL-9) along with ER tracker A) Represents the E1-300W localization (green) with ER-tracker (blue) in late trophozoite stage. B) Represents the E1-300W (green), localization with ER-tracker (blue) in schizont.

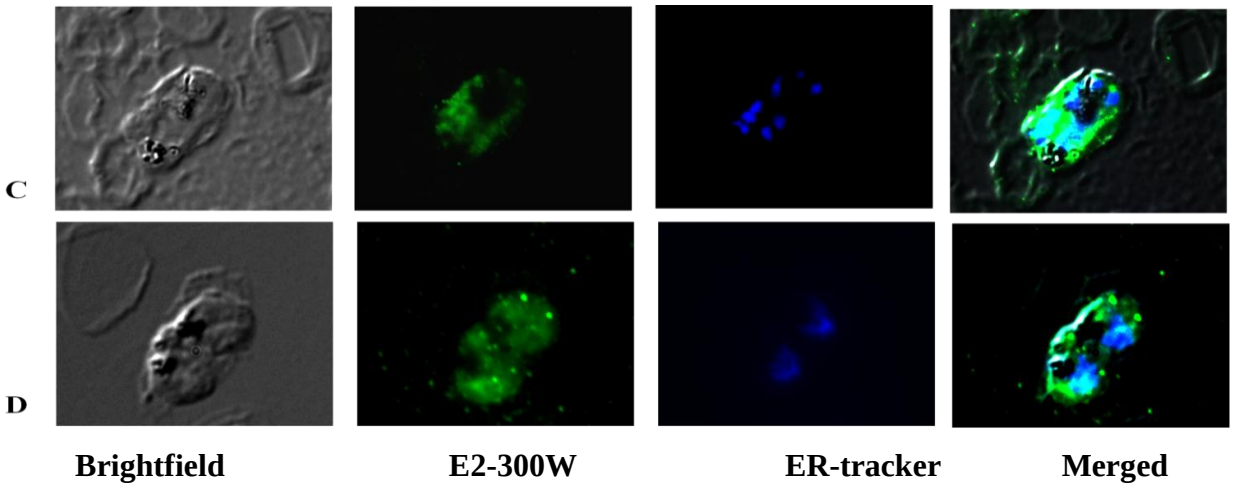


Figure 5.23 (C-D): Localization of E2-300W FITC labelled antibody in cultured Indian strain (RKL-9) along with ER tracker C) E-2- 300W FITC labelled (green) localization with ER- tracker (blue) in late trophozoite stage D) Localization of E2-300W FITC conjugated (green) in late trophozoite stage along with ER- tracker (blue).

5.3.2- Localization of PFL2320W:

The orthologue of PFL2320W in rodent malaria shows the disruption of gene in *P. berghei* in asexual stages which suggests that the protein is not essential for the asexual stages of *P. berghei*, The sequence identity of PFL2320W and *P. berghei*, orthologous protein PBANKA_1461300 is 39.16% **figure 5.24**, this indicates that the proteins are different from each other in amino acid composition and there may be differing levels of requirement in the two *Plasmodium* species. Unless a knockout study is performed in the *P. falciparum* essentiality of the protein cannot be predicted. It has been reported that the protein is found to have an apicoplast-targeting signal, and our in-silico analysis suggest the same, although in PFL2320W according to the PlasmoAP prediction tool apicoplast targeting peptide is present but signal peptide is absent. In our localization studies, we have tried to check for localization of this protein in the apicoplast or mitochondria using apicoplast targeting streptavidin conjugated Qdot and the mitotracker for the mitochondria. Although in some of our slides we did see localization to the apicoplast and no localization in the mitochondria, as we could not satisfactorily repeat the results with the reagents used. We conclude for the protein encoded by conserved *Plasmodium* protein PFL2320W, we have been able to show expression of the protein in slides prepared from *P. falciparum* culture, but are not in a position to conclusively state that the protein is localized in apicoplast **figure 5.25-A and 5.25-B, 5.25-C**.

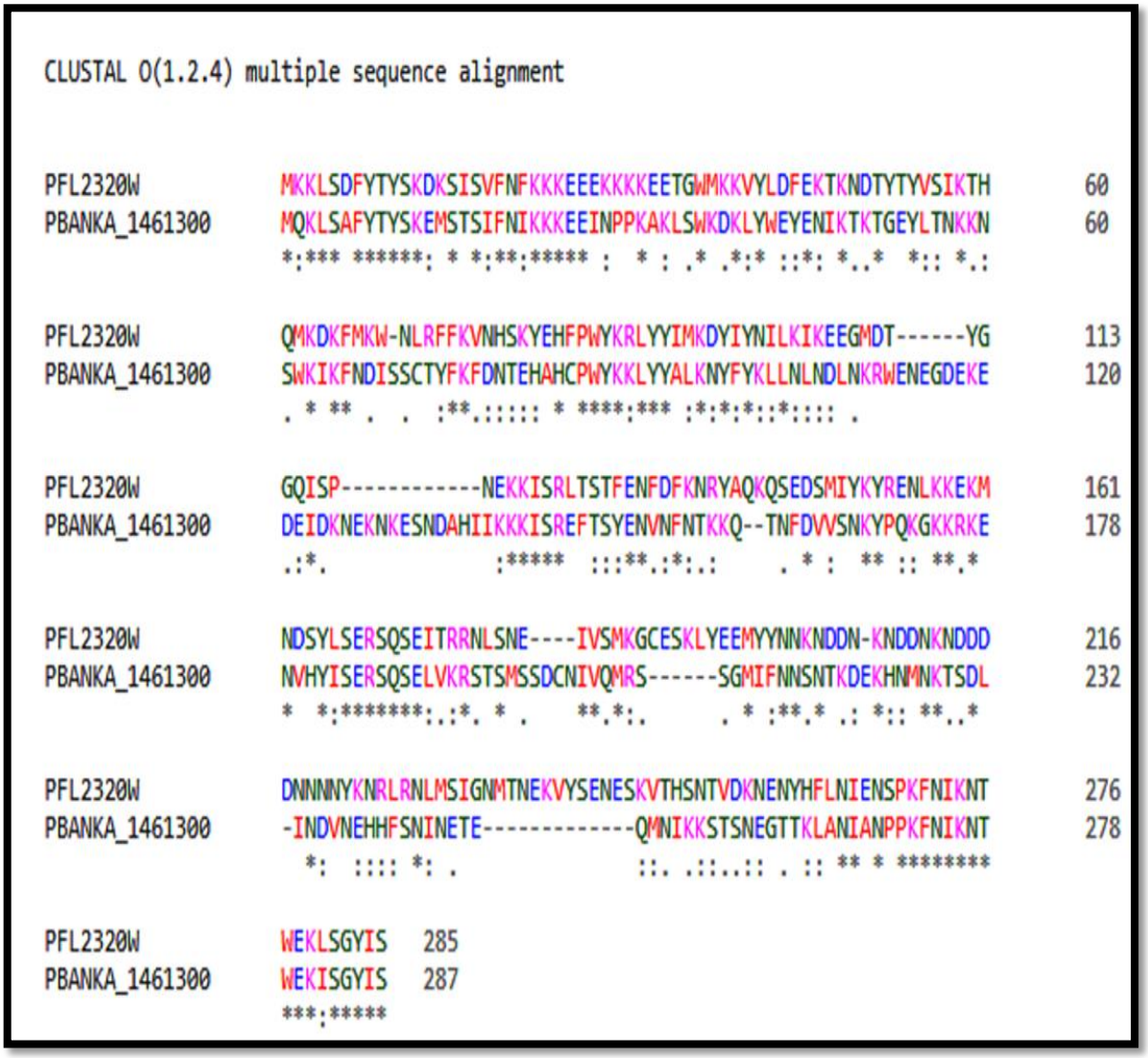


Figure 5.24: Alignment of PFL2320W of *P. falciparum* with orthologous strain of *P. berghei* ANKA strain.

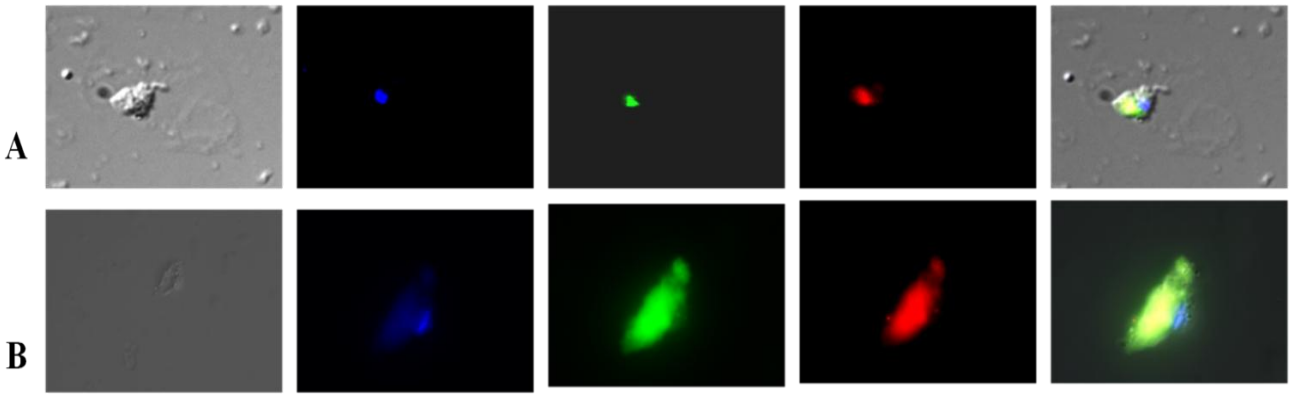


Figure-5.25: **E1A-2320W with an apicoplast tracker (Qdot) in sexual stages of RKL-9, cultured Indian isolate;** A) E1A-2320W-FITC labelled (green) with Qdot (red) in the gametocyte stage V of cultured Indian isolate strain (RKL-9) where gametocyte is present outside the cell. B) E1A-2320W FITC-labelled secondary antibody (green), with apicoplast targeting dye (Qdot) (red) in gametocyte stage III of cultured Indian isolate, RKL-9.

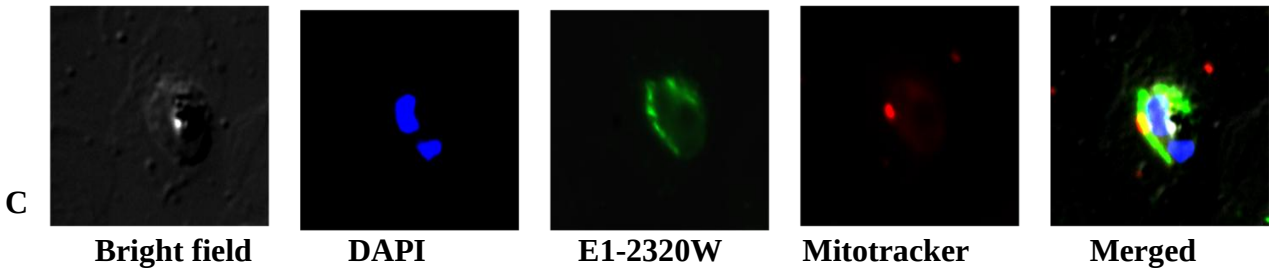


Figure 5.25 C) E1A-2320W FITC labelled (green) with Mito-tracker dye (red) in late trophozoite stage of cultured Indian isolate RKL-9 strain.

Conclusion:

In-silico analysis of PFD0300W suggests the enrichment of protein in biological process like regulation of metabolic process, mRNA processing, regulation of gene expression, and regulation of nucleic acid-templated transcriptional machinery. All these processes are important for the parasite survival. The CDD (Conserved Domain Database) analysis predicted the presence of NBD (Nucleotide Binding Domain) in the ordered segment of the protein, along with this the molecular functional enrichment of the protein suggests the nucleic acid binding, actin binding, DNA binding, kinase binding, nucleotide binding and phosphotransferase activity. The trial of 3D structural analysis of the protein was not successful because of the absence of suitable template; this is because of the conserved nature of the protein.

In-silico analysis of PFL2320W suggests that protein is enriched in biological processes like regulation of metabolic process, cellular macromolecule biosynthetic process, regulation of nitrogen compound, regulation of gene expression, mRNA splicing- via spliceosome, and in RNA processing. We have not found any conserved domain in the protein with any of the conserved domain search tool. The disordered protein analysis suggested that the protein is intrinsically disordered which suggest that protein can play any of the important roles in the parasite which could be essential for the protein. Like PFD0300W the trial of 3D structural analysis of the protein was not successful because of the absence of suitable template; because of the conserved nature of the protein.

The localization of PFD0300W is restricted to the cytoplasm and is not present in the nucleus or in endoplasmic reticulum of the parasite. It is significant that this protein appears to be expressed in all the asexual blood stages and in the gametocytes stages of the parasite.

We have been able to show the expression of the protein PFL2320W at different levels in slides prepared from *P. falciparum* culture, but are not in a position to conclusively state whether the protein is localized in apicoplast. Although in some images, suggest localization to the apicoplast and not in the mitochondria or nucleus.

Chapter VI

Conclusion and Future Directions

6.1. Highlights:

The main objectives of the thesis were to analyze the genomic differences in the parasite isolate causing complicated malaria in comparison with the parasite isolate causing uncomplicated malaria, isolated from a adult malaria patients, and to clone, express and localize the hub genes derived from the Weighted Gene Co-expression network (WGCN). During the thesis work, a custom- based 2x400K tiling array was used with overlapping probes at the space of 56 bp for comparative genomic hybridization analysis between the two parasite isolates. This work for the first time unravels the genomic differences between the parasite isolates causing differing disease manifestations from the Indian subcontinent.

Subsequently, cloning, expression and localization of the two hub genes derived from different modules of a weighted gene co-expression network were attempted. From the up-regulated module PFD0300W was selected and from the steady-state module,

PFL2320W was selected both of these genes encodes for conserved *Plasmodium* protein with unknown function. This part of the study identified that the proteins are intrinsically disordered and does not contain any of the transmembrane domain and signal peptide sequences.

6.2. Specific conclusions:

1. A custom-based 2x400K tiling array was designed for the comparative genomic hybridization analysis of two different *P. falciparum* parasite isolates causing different disease manifestations.
2. Copy number variations have been seen in all the chromosomes of the parasite isolate causing complicated disease manifestation when compared with the uncomplicated malaria.
3. Deletions have been seen mostly in sub-telomeric and telomeric regions of the chromosomes and were found mostly in multigene family genes, encoding for the variant surface antigens. Other than the sub-telomeric and telomeric regions, deletions were also present in the intergenic regions and other parts of the chromosomes.
4. Novel variations like an amplification of 28 kb region in chromosome 8 that covers for 3 genes (MAL8P1.139, PF08_0122), conserved *Plasmodium* membrane protein, unknown function,(PF08_0123) tRNA pseudouridine synthase, putative. Which are not reported before in the copy number variation studies with clinical isolates causing severe malaria.
5. The hub genes PFD0300W and PFL2320W were successfully cloned from field isolates for protein expression. However, no expression has been detected even after using codon optimized constructs. However, RNA transcripts were detected in induced clones at different time points post induction, suggesting stable transcript for the two proteins.
6. CDD (Conserved Domain Database) analysis of PFD0300W suggests the presence of Nucleotide Binding Domain, in the protein. Where as CDD analysis of PFL2320W shows the absence of any of the conserved domains in the protein. The tertiary structure prediction of both the proteins was not possible because of

the lack of suitable templates, both proteins being conserved *Plasmodium* protein with unknown function.

7. The analysis for the prediction of intrinsically disordered nature of the protein suggests that both the proteins are intrinsically disordered with PFD0300W showing 46.45% disorderedness, with 3 disordered region of >30 amino-acids, and one disordered region of >50 amino-acids, whereas PFL2320W showed 48.77% disorderedness with one disordered region of >30 amino-acids, and one disordered region of >50 amino-acids.
8. The Immunolocalization study of PFD0300W suggests the cytoplasmic localization, of the protein, which was detected in all the asexual and the sexual stages of the Indian cultured strain RKL-9. However, the protein was not seen to be localized in the ER or in nucleus.
9. The expression of PFL2320W was seen in both sexual stages and asexual stages; however, the definitive localization of PFL2320W was not achieved with the reagents available.

6.3. Direction for future research (future perspectives)

- The comparative genomic hybridization study of the parasite isolates causing complicated malaria with that of the isolate causing uncomplicated malaria provide us good information regarding genome variation. To get a more accurate picture of the genomic variations and their implications these analyses should be done with a large set of clinical samples from different parts of the India.

- Having both the proteins expressed in a heterologous system would be of great advantage for characterization of these since our attempts with these in the *E. coli* system were not successful, other host systems like *Pichia pastoris*, Chinese Hamster Ovary cells etc., could also be tried.
- There are multiple hub genes present in the network, which encode conserved *Plasmodium* proteins. Similar studies should be performed for at least localization of some more of the same.
- Since, PFD0300W is present in all the parasite stages, it would be interesting to try and identify small molecules which could prevent the functioning of this

protein. Failure of the knockout studies in orthologous protein of PFD0300W in PB_ANKA strain suggests the importance of protein for the parasite and thus this could be looked upon as a putative drug target.

References:

- Aashish, A., & Manigandan, G. (2015). Complicated vivax malaria, an often underestimated condition-Case Report. *Journal of Family & Community Medicine*, *22*(3)
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17)
- Amino, R., Thiberge, S., Martin, B., Celli, S., Shorte, S., Frischknecht, F., & Ménard, R. (2006). Quantitative imaging of *Plasmodium* transmission from mosquito to mammal. *Nature Medicine*, *12*(2).
- Anand, A. C., & Puri, P. (2005). Jaundice in malaria. *Journal of Gastroenterology and Hepatology*, *20*(9).
- Antinori, S., Galimberti, L., Milazzo, L., & Corbellino, M. (2012). Biology of human malaria plasmodia including *Plasmodium knowlesi*. *Mediterranean Journal of Hematology and Infectious Diseases*, *4*(1).
- Aradhya, S., & Cherry, A. M. (2007). Array-based comparative genomic hybridization: clinical contexts for targeted and whole-genome designs. *Genetics in Medicine*, *9*(9).
- Baldwin, M. R., Li, X., Hanada, T., Liu, S. C., & Chishti, A. H. (2015). Merozoite surface protein 1 recognition of host glycophorin A mediates malaria parasite invasion of red blood cells. *Blood*, *125*(17).
- Bannister, L. H., & Sherman, I. W. (2009). *Plasmodium*. *Encyclopedia of Life Sciences (ELS)*. *Chichester: John Wiley & Sons, Ltd*.
- Bartoloni, A., & Zammarchi, L. (2012). Clinical Aspects of Uncomplicated and Severe Malaria. *Mediterranean Journal of Hematology and Infectious Diseases*, *4*(1).

- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., ... & Studholme, D. J. (2004). The Pfam protein families database. *Nucleic Acids Research*, *32* (suppl_1).

- Bengtsson, D., Sowa, K. M., Salanti, A., Jensen, A. T., Joergensen, L., Turner, L., ... & Arnot, D. E. (2008). A method for visualizing surface-exposed and internal PfEMP1 adhesion antigens in *Plasmodium falciparum* infected erythrocytes. *Malaria Journal*, *7*(1).

- Benito-Sanz, S., Belinchon-Martínez, A., Aza-Carmona, M., de la Torre, C., Huber, C., González-Casado, I., ... & Heath, K. E. (2017). Identification of 15 novel partial SHOX deletions and 13 partial duplications, and a review of the literature reveals intron 3 to be a hotspot region. *Journal of Human Genetics*, *62*(2).

- Bernabeu, M., Danziger, S. A., Avril, M., Vaz, M., Babar, P. H., Brazier, A. J., ... & Gomes, E. (2016). Severe adult malaria is associated with specific PfEMP1 adhesion types and high parasite biomass. *Proceedings of the National Academy of Sciences*, *113*(23).

- Bhattacharyya, M., & Chakrabarti, S. (2015). Identification of important interacting proteins (IIPs) in *Plasmodium falciparum* using large-scale interaction network analysis and in-silico knock-out studies. *Malaria Journal*, *14*(1).

- Bushell, E., Gomes, A. R., Sanderson, T., Anar, B., Girling, G., Herd, C., ... & Mather, M. W. (2017). Functional profiling of a *Plasmodium* genome reveals an abundance of essential genes. *Cell*, *170*(2).

- Carret, C. K., Horrocks, P., Konfortov, B., Winzeler, E., Qureshi, M., Newbold, C., & Ivens, A. (2005). Microarray-based comparative genomic analyses of the human malaria parasite *Plasmodium falciparum* using Affymetrix arrays. *Molecular and Biochemical Parasitology*, *144*(2).

- Cheeseman, I. H., Gomez-Escobar, N., Carret, C. K., Ivens, A., Stewart, L. B., Tetteh, K. K., & Conway, D. J. (2009). Gene copy number variation throughout the *Plasmodium falciparum* genome. *BMC Genomics*, *10*(1).

- Chou, K. C., & Shen, H. B. (2007). Euk-mPLoc: a fusion classifier for large-scale eukaryotic protein subcellular location prediction by incorporating multiple sites. *Journal of Proteome Research*, *6*(5).

- Coughlin, C. R., Scharer, G. H., & Shaikh, T. H. (2012). Clinical impact of copy number variation analysis using high-resolution microarray technologies: advantages, limitations and concerns. *Genome Medicine*, *4*(10).

- Cowman, A. F., Berry, D., & Baum, J. (2012). The cellular and molecular basis for malaria parasite invasion of the human red blood cell. *Journal of Cell Biology*, *198*(6).

- Cox, F. E. (2010). History of the discovery of the malaria parasites and their vectors. *Parasites & Vectors*, *3*(1).

- Das, A., Holloway, B., Collins, W. E., Shama, V. P., Ghosh, S. K., Sinha, S., ... & Lal, A. A. (1995). Species-specific 18S rRNA gene amplification for the detection of *P. falciparum* and *P. vivax* malaria parasites. *Molecular and Cellular Probes*, *9*(3).

- de Koning-Ward, T. F., Dixon, M. W., Tilley, L., & Gilson, P. R. (2016). *Plasmodium* species: master renovators of their host cells. *Nature Reviews Microbiology*, *14*(8).

- Dharia, N. V., Sidhu, A. B. S., Cassera, M. B., Westenberger, S. J., Bopp, S. E., Eastman, R. T., ... & Wirth, D. F. (2009). Use of high-density tiling microarrays to identify mutations globally and elucidate mechanisms of drug resistance in *Plasmodium falciparum*. *Genome Biology*, *10*(2).

- Dinkel, H., Michael, S., Weatheritt, R. J., Davey, N. E., Van Roey, K., Altenberg, B., ... & Jödicke, L. (2011). ELM—the database of eukaryotic linear motifs. *Nucleic Acids Research*, *40*(D1).

- Dosztányi, Z., Csizmok, V., Tompa, P., & Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, *21*(16).

- du Manoir, S., Speicher, M. R., Joos, S., Schröck, E., Popp, S., Döhner, H., ... & Cremer, T. (1993). Detection of complete and partial chromosome gains and losses by comparative genomic in situ hybridization. *Human Genetics*, *90*(6).

- Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., & Obradović, Z. (2002). Intrinsic disorder and protein function. *Biochemistry*, *41*(21).

- Feagin, J. E., Harrell, M. I., Lee, J. C., Coe, K. J., Sands, B. H., Cannone, J. J., ... & Gutell, R. R. (2012). The fragmented mitochondrial ribosomal RNAs of *Plasmodium falciparum*. *PLoS One*, *7*(6).

- Foth, B. J., Ralph, S. A., Tonkin, C. J., Struck, N. S., Fraunholz, M., Roos, D. S., ... & McFadden, G. I. (2003). Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*. *Science*, *299*(5607).

- Frech, C., & Chen, N. (2013). Variant surface antigens of malaria parasites: functional and evolutionary insights from comparative gene family classification and analysis. *BMC Genomics*, *14*(1).

- Gardiner, D. L., & Trenholme, K. R. (2015). *Plasmodium falciparum* gametocytes: playing hide and seek. *Annals of Translational Medicine*, *3*(4).

- Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., ... & Paulsen, I. T. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*, *419*(6906).

- Garg, S., Saxena, V., Kanchan, S., Sharma, P., Mahajan, S., Kochar, D., & Das, A. (2009). Novel point mutations in sulfadoxine resistance genes of *Plasmodium falciparum* from India. *Acta Tropica*, *110*(1).

- Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook* ,Humana press.

- Gopal, G. J., & Kumar, A. (2013). Strategies for the production of recombinant protein in Escherichia coli. *The Protein Journal*, *32*(6).

- Guo, X., Wang, Y., Wang, C., & Chen, J. (2015). Identification of several hub-genes associated with periodontitis using integrated microarray analysis. *Molecular Medicine Reports*, *11*(4).

- Hartinger, D., Heinl, S., Schwartz, H. E., Grabherr, R., Schatzmayr, G., Haltrich, D., & Moll, W. D. (2010). Enhancement of solubility in *Escherichia coli* and purification of an aminotransferase from *Sphingopyxis sp.* MTA144 for deamination of hydrolysed fumonisin B 1. *Microbial Cell Factories*, *9*(1).

- He, H., Sun, D., Zeng, Y., Wang, R., Zhu, W., Cao, S., … Deng, H. (2017). A systems genetics approach identified *GPD1L* and its molecular mechanism for obesity in human adipose tissue. *Scientific Reports*, *7*(1799).

- Heinberg, A., Siu, E., Stern, C., Lawrence, E. A., Ferdig, M. T., Deitsch, K. W., & Kirkman, L. A. (2013). Direct evidence for the adaptive role of copy number variation on antifolate susceptibility in *Plasmodium falciparum*. *Molecular Microbiology*, *88*(4).

- Hempelmann, E., & Krafts, K. (2013). Bad air, amulets and mosquitoes: 2,000 years of changing perspectives on malaria. *Malaria Journal*, *12*(1).

- Hoffman, S. L., Bancroft, W. H., Michael, G., James, S. L., Burroughs, E. C. B., Stephenson, J. R., & Morgan, M. J. (1997). Funding for malaria genome sequencing. *Nature*, *387*(6634).

- Howes, R. E., Battle, K. E., Mendis, K. N., Smith, D. L., Cibulskis, R. E., Baird, J. K., & Hay, S. I. (2016). Global epidemiology of *Plasmodium vivax*. *The American Journal of Tropical Medicine and Hygiene*, *95*(6_Suppl).

- Jani, D., Nagarkatti, R., Beatty, W., Angel, R., Slebodnick, C., Andersen, J., ... & Rathore, D. (2008). HDP—a novel heme detoxification protein from the malaria parasite. *PLoS Pathogens*, *4*(4).

- Jethwaney, D., Lepore, T., Hassan, S., Mello, K., Rangarajan, R., Jahnen-Dechent, W., ... & Sultan, A. A. (2005). Fetuin-A, a hepatocyte-specific protein that binds

Plasmodium berghei thrombospondin-related adhesive protein: a potential role in infectivity. *Infection and Immunity*, *73*(9).

- Jiang, H., Yi, M., Mu, J., Zhang, L., Ivens, A., Klimczak, L. J., ... & Su, X. Z. (2008). Detection of genome-wide polymorphisms in the AT-rich *Plasmodium falciparum* genome using a high-density microarray. *BMC Genomics*, *9*(1).

- Käll, L., Krogh, A., & Sonnhammer, E. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Research*, *35*(suppl_2).

- Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-based protein structure modelling using the RaptorX web server. *Nature Protocols*, *7*(8).

- Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., & Pinkel, D. (1992). Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, *258*(5083).

- Kempaiah, P., Dokladny, K., Karim, Z., Raballah, E., Ong'echa, J. M., Moseley, P. L., & Perkins, D. J. (2016). Reduced Hsp70 and Glutamine in Pediatric Severe Malaria Anemia: Role of Hemozoin in Suppressing Hsp70 and NF-κB Activation. *Molecular Medicine*, *22*(570).

- Kidgell, C., Volkman, S. K., Daily, J., Borevitz, J. O., Plouffe, D., Zhou, Y., ... & Mboup, S. (2006). A systematic map of genetic variation in *Plasmodium falciparum*. *PLoS Pathogens*, *2*(6).

- Kochar, D. K., Das, A., Kochar, S. K., Saxena, V., Sirohi, P., Garg, S., ... & Gupta, V. (2009). Severe *Plasmodium vivax* malaria: a report on serial cases from Bikaner in northwestern India. *The American Journal of Tropical Medicine and Hygiene*, *80*(2).

- Kochar, D. K., Saxena, V., Singh, N., Kochar, S. K., Kumar, S. V., & Das, A. (2005). *Plasmodium vivax* malaria. *Emerging Infectious Diseases*, *11*(1).

- Köhler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J. M., ... & Roos, D. S. (1997). A plastid of probable green algal origin in Apicomplexan parasites. *Science*, *275*(5305).

- Kumar, N., Singh, J. P., Pande, V., Mishra, N., Srivastava, B., Kapoor, R., ... & Anvikar, A. R. (2012). Genetic variation in histidine rich proteins among Indian *Plasmodium falciparum* population: possible cause of variable sensitivity of malaria rapid diagnostic tests. *Malaria Journal*, *11*(1).

- Kumar, S. V., Ranjan, S., Saxena, V., Rajesh, V., Roy, S. K., Kochar, D., ... & Das, A. (2005). *Plasmodium falciparum*: genetic diversity of C-terminal region of MSP-1 in isolates from Indian sub-continent. *Experimental Parasitology*, *110*(4).

- Lamikanra, A. A., Merryweather-Clarke, A. T., Tipping, A. J., & Roberts, D. J. (2015). Distinct mechanisms of inadequate erythropoiesis induced by tumor necrosis factor alpha or malarial pigment. *PloS One*, *10*(3).

- Lavstsen, T., Salanti, A., Jensen, A. T., Arnot, D. E., & Theander, T. G. (2003). Sub-grouping of *Plasmodium falciparum* 3D7 var genes based on sequence analysis of coding and non-coding regions. *Malaria Journal*, *2*(1).

- Le Roch, K. G., Chung, D. W., & Ponts, N. (2012). Genomics and integrated systems biology in *Plasmodium falciparum*: a path to malaria control and eradication. *Parasite Immunology*, *34*(2-3).

- LeRoux, M., Lakshmanan, V., & Daily, J. P. (2009). *Plasmodium falciparum* biology: analysis of in vitro versus in vivo growth conditions. *Trends in Parasitology*, *25*(10).

- Lim, L., & McFadden, G. (2010). The evolution, metabolism and functions of the apicoplast. *Philosophical Transactions: Biological Sciences*, *365*(1541).

- Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., & Russell, R. B. (2003) Protein disorder prediction: implications for structural proteomics. *Structure*, *11*(11).

- Liu, W., Li, Y., Learn, G. H., Rudicell, R. S., Robertson, J. D., Keele, B. F., ... & Gonder, M. K. (2010). Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature*, *46 7*(7314).

- Liu, X. S. (2007). Getting started in tiling microarray analysis. *PLoS Computational Biology*, *3*(10).

- Llinas, M., Bozdech, Z., Wong, E. D., Adai, A. T., & DeRisi, J. L. (2006). Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. *Nucleic Acids Research*, *34*(4).

- Maier, A. G., Rug, M., O'Neill, M. T., Brown, M., Chakravorty, S., Szestak, T., ... & Newbold, C. (2008). Exported proteins required for virulence and rigidity of *Plasmodium falciparum*-infected human erythrocytes. *Cell*, *134* (1).

- Malaria, S. (2014). Severe malaria. *Tropical Medicine & International Health*, *19*.

- Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M., ... & Lanczycki, C. J. (2005). CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Research*, *33*(suppl_1).

- Mazhar, F., & Haider, N. (2016). Respiratory manifestation of malaria: An Update. *International Journal of Medical Research & Health Sciences*, *5*(5).

- McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, *16*(4).

- Miller, J. A., Oldham, M. C., & Geschwind, D. H. (2008). A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *Journal of Neuroscience*, *28*(6).

- Mishra, N., Prajapati, S. K., Kaitholia, K., Bharti, R. S., Srivastava, B., Phookan, S., … Valecha, N. (2015). Surveillance of artemisinin resistance in *Plasmodium falciparum* in India using the kelch13 molecular marker. *Antimicrobial Agents and Chemotherapy*, *59*(5).

- Mishra, S. K., Mohapatra, S., & Mohanty, S. (2003). Jaundice in *falciparum* malaria. *The Journal, Indian Academy of Clinical Medicine, 4* (1).

- Mita, T., & Tanabe, K. (2012). Evolution of *Plasmodium falciparum* drug resistance: implications for the development and containment of artemisinin resistance. *Japanese Journal of Infectious Diseases, 65*(6).

- Mobegi, V. A., Duffy, C. W., Amambua-Ngwa, A., Loua, K. M., Laman, E., Nwakanma, D. C., ... & Kwiatkowski, D. P. (2014).Genome-wide analysis of selection on the malaria parasite *Plasmodium falciparum* in West African populations of differing infection endemicity. *Molecular Biology and Evolution*, *31*(6).

- Mockler, T. C., & Ecker, J. R. (2005). Applications of DNA tiling arrays for whole-genome analysis. *Genomics, 85*(1).

- Mohan, Alladi, Surendra K. Sharma, and Srinivas Bollineni (2008). Acute lung injury and acute respiratory distress syndrome in malaria. *Jounal of Vector Borne Diseases* 45(3).

- Mohandas, N., & An, X. (2012). Malaria and human red blood cells. *Medical Microbiology and Immunology*, *201*(4).

- Mok, B. W., Ribacke, U., Sherwood, E., & Wahlgren, M. (2008). A highly conserved segmental duplication in the subtelomeres of *Plasmodium falciparum* chromosomes varies in copy number. *Malaria Journal, 7*(1).

- Nair, S., Miller, B., Barends, M., Jaidee, A., Patel, J., Mayxay, M., … Anderson, T. J. C. (2008). Adaptive copy number evolution in malaria parasites. *PLoS Genetics*, *4*(10).

- Nayak, K. C., Kumar, S., Gupta, B. K., Gupta, A., Prakash, P., & Kochar, D. K. (2014). Clinical and histopathological profile of acute renal failure caused by *falciparum* and *vivax* monoinfection: an observational study from Bikaner, northwest zone of Rajasthan, India. *Journal of Vector Borne Diseases*, *51*(1).

- Pakalapati, D., Garg, S., Middha, S., Acharya, J., Subudhi, A. K., Boopathi, A. P., ... & Das, A. (2013). Development and evaluation of a 28S rRNA gene-based nested PCR assay for *P. falciparum* and *P. vivax*. *Pathogens and global health*, *107*(4).

- Palomares, L. A., Estrada-Moncada, S., & Ramírez, O. T. (2004). Production of recombinant proteins: challenges and solutions. In *Recombinant Gene Expression: Reviews and Protocols* (pp. 15-51). Humana press.

- Pathak, A. . b c, Mårtensson, A. . d g, Gawariker, S. ., Mandliya, J. ., Sharma, A. ., Diwan, V. . f, & Ursing, J. . (2014). Characterization of drug resistance associated genetic polymorphisms among *Plasmodium falciparum* field isolates in Ujjain, Madhya Pradesh, India. *Lipids in Health and Disease*, *13*(1).

- Paul, R. E., Coulson, T. N., Raibaud, A., & Brey, P. T. (2000). Sex determination in malaria parasites. *Science*, *287*(5450).

- Perlmann, P., & Troye-Blomberg, M. (2002). Malaria and the immune system in humans. In *Malaria Immunology* (Vol. 80). Karger Publishers.

- Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10).

- Pinzón, M. A., Pineda, J. C., Rosso, F., Shinchi, M., & Bonilla-Abadía, F. (2013). *Plasmodium vivax* cerebral malaria complicated with venous sinus thrombosis in Colombia. *Asian Pacific Journal of Tropical Medicine*, *6*(5).

- Price, R. N., Douglas, N. M., & Anstey, N. M. (2009). New developments in *Plasmodium vivax malaria*: severe disease and the rise of chloroquine resistance. *Current Opinion in Infectious Diseases*, *22*(5).

- Prudêncio, M., Rodriguez, A., & Mota, M. M. (2006). The silent path to thousands of merozoites: the *Plasmodium* liver stage. *Nature Reviews. Microbiology*, *4*(11).

- Przybytkowski, E., Ferrario, C., & Basik, M. (2011). The use of ultra-dense array CGH analysis for the discovery of micro-copy number alterations and gene fusions in the cancer genome. *BMC Medical Genomics*, *4*(1).

- Puigbò, P., Bravo, I. G., & Garcia-Vallve, S. (2008). CAIcal: a combined set of tools to assess codon usage adaptation. *Biology Direct*, *3*(1).

- Rajesh, V., Singamsetti, V. K., Vidya, S., Gowrishankar, M., Elamaran, M., Tripathi, J., ... & Das, A. (2008). *Plasmodium falciparum*: genetic polymorphism in apical membrane antigen-1 gene from Indian isolates. *Experimental Parasitology*, *119*(1).

- Ramaprasad, A., Pain, A., & Ravasi, T. (2012). Defining the protein interaction network of human malaria parasite *Plasmodium falciparum*. *Genomics*, *99*(2).

- Redwan, E. M. (2006). The optimal gene sequence for optimal protein expression in *Escherichia coli*: principle requirements. *Arab Journal of Biotechnology*, *9*.

- Ribacke, U., Mok, B. W., Wirta, V., Normark, J., Lundeberg, J., Kironde, F., … Wahlgren, M. (2007). Genome wide gene amplifications and deletions in *Plasmodium falciparum*. *Molecular and Biochemical Parasitology*, *155*(1).

- Roy, S., & Reif, A. M. (2013). Evaluation of calling algorithms for array-CGH. *Frontiers in Genetics*, *4*.

- Saha, S., & Raghava, G. P. S. (2006). Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins: Structure, Function, and Bioinformatics*, *65*(1).

- Sambrook, J., & Russell, D. W. (2001). Molecular cloning: a laboratory manual. third.*Cold spring Harbor Laboratory Press, New York*.

- San-Miguel, T., Pérez-Bermúdez, P., & Gavidia, I. (2013). Production of soluble eukaryotic recombinant proteins in *E. coli* is favoured in early log-phase cultures induced at low temperature. *SpringerPlus*, *2* (1).

- Saxena, V., Garg, S., Ranjan, S., Kochar, D., Ranjan, A., & Das, A. (2007). Analysis of elongation factor Tu (tuf A) of apicoplast from Indian *Plasmodium vivax* isolates. *Infection, Genetics and Evolution*, *7* (5).

- Scherf, A., Lopez-Rubio, J. J., & Riviere, L. (2008). Antigenic variation in *Plasmodium falciparum*. *Annu. Rev. Microbiol.*, *13* (62*)*.

- Selzer, R. R., Richmond, T. A., Pofahl, N. J., Green, R. D., Eis, P. S., Nair, P., ... & Stallings, R. L. (2005). Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes, Chromosomes and Cancer*, *44*(3).

- Sharma, D., Lather, M., Mallick, P. K., Adak, T., Dang, A. S., Valecha, N., & Singh, O. P. (2015). Polymorphism in drug resistance genes dihydrofolate reductase and dihydropteroate synthase in *Plasmodium falciparum* in some states of India. *Parasites & Vectors*, *8*(1).

- Smith, J. D., Rowe, J. A., Higgins, M. K., & Lavstsen, T. (2013). Malaria's deadly grip: cytoadhesion of *Plasmodium falciparum*-infected erythrocytes. *Cellular Microbiology*, *15*(12).

- Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research, 33* (suppl_2).

- Straimer, J., Gnädig, N. F., Witkowski, B., Amaratunga, C., Duru, V., Ramadani, A. P., ... & Gregory, P. D. (2015). K13-propeller mutations confer artemisinin resistance in *Plasmodium falciparum* clinical isolates. *Science*, *347*(6220).

- Sturm, A., Amino, R., Van de Sand, C., Regen, T., Retzlaff, S., Rennenberg, A., ... & Heussler, V. T. (2006). Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. S*cience*, *313*(5791).

- Subudhi, A. K., Boopathi, P. A., Pandey, I., Kaur, R., Middha, S., Acharya, J., ... & Das, A. (2015). Disease specific modules and hub genes for intervention strategies: a

co-expression network based approach for *Plasmodium falciparum* clinical isolates. *Infection, Genetics and Evolution, 35*.

- Trampuz, A., Jereb, M., Muzlovic, I., & Prabhu, R. M. (2003). Clinical review: Severe malaria. *Critical Care, 7*(4).

- Tyagi, S., Pande, V., & Das, A. (2014). Mitochondrial genome sequence diversity of Indian *Plasmodium falciparum* isolates. *Memórias do Instituto Oswaldo Cruz*, *109* (4).

- Vanderberg, J. P., & Frevert, U. (2004). Intravital microscopy demonstrating antibody-mediated immobilisation of *Plasmodium berghei* sporozoites injected into skin by mosquitoes. *International Journal for Parasitology*, *34*(9).

- Wassmer, S. C., & Grau, G. E. R. (2017). Severe malaria: what's new on the pathogenesis front?. *International Journal for Parasitology*, *47*(2).

- Wassmer, S. C., Lépolard, C., Traoré, B., Pouvelle, B., Gysin, J., & Grau, G. E. (2004). Platelets reorient *Plasmodium falciparum*–infected erythrocyte cytoadhesion to activated endothelial cells. *The Journal of Infectious Diseases*, *189*(2).

- Weiss, M. M., Hermsen, M. A., Meijer, G. A., Van Grieken, N. C., Baak, J. P., Kuipers, E. J., & Van Diest, P. J. (1999). Comparative genomic hybridisation. *Molecular Pathology*, *52*(5).

- Wozniak, A., Sciot, R., Guillou, L., Pauwels, P., Wasag, B., Stul, M., ... & Debiec-Rychter, M. (2007). Array CGH analysis in primary gastrointestinal stromal tumors: cytogenetic profile correlates with anatomic site and tumor aggressiveness, irrespective of mutational status. *Genes, Chromosomes and Cancer*, *46*(3).

- World malaria report (2017). Geneva: *World Health Organization; 2017*

- Yadava, A., & Ockenhouse, C. F. (2003). Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems. *Infection and Immunity*, *71*(9).

- Yang, Y., Faraggi, E., Zhao, H., & Zhou, Y. (2011). Improving protein fold recognition and template-based modelling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates. *Bioinformatics, 27* (15).

- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics, 13* (1).

- Yeung, S., Pongtavornpinyo, W., Hastings, I. M., Mills, A. J., & White, N. J. (2004). Antimalarial drug resistance, artemisinin-based combination therapy, and the contribution of modeling to elucidating policy choices. *The American Journal of Tropical Medicine and Hygiene*, *71*(2_suppl).

- Yu, C. S., Lin, C. J., & Hwang, J. K. (2004). Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Science*, *13*(5).

- Zdobnov, E. M., & Apweiler, R. (2001). InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, *17*(9).

- Zeeshan, M., Alam, M. T., Vinayak, S., Bora, H., Tyagi, R. K., Alam, M. S., ... & Udhayakumar, V. (2012). Genetic variation in the *Plasmodium falciparum* circumsporozoite protein in India and its relevance to RTS, S malaria vaccine. *PLoS One*, *7*(8).

- Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*(1).

- Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics, 9* (1).

Web sites:

- http://anchor.enzim.hu/
- www.imtech.res.in/raghava/ bcepred/
- http://bioinf.cs.ucl.ac.uk/psipred/
- http://bioinf.cs.ucl.ac.uk/psipred/
- http://cello.life.nctu.edu.tw/
- http://dis.embl.de/
- http://genomes.urv.es/CAIcal/
- http://iupred.enzim.hu/
- http://people.mbi.ucla.edu/sumchan/caltor.html
- http://phobius.sbc.su.se/cgi-bin/predict.pl
- http://plasmodb.org/plasmo/
- http://raptorx.uchicago.edu/
- http://v4-4.plasmodb.org/Help/Tools-Apicoplast-Plasmoap PredictionOfApicoplastTargetingSignals.shtml
- http://web.expasy.org/cgi-bin/protparam/protparam
- http://www.cbs.dtu.dk/services/SignalP/
- http://www.cbs.dtu.dk/services/TMHMM/
- http://www.ch.embnet.org/software/TMPRED_form.html
- http://www.pondr.com/
- http://www.sbg.bio.ic.ac.uk/phyre/html/
- http://www.who.int/malaria/media/artemisinin_resistance_qa/en/
- https://omictools.com/sparks-x-tool
- https://toolkit.tuebingen.mpg.de/#/hhpred
- https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
- https://zhanglab.ccmb.med.umich.edu/I-TASSER/

Appendix-A1:

Reagents and chemicals used in the study

A1.1: Reagents used for parasite genomic DNA isolation:

Reagents used:

Lysis buffer (1X): NaCl=10mM, Tris (pH=8.0)= 50mM, EDTA (pH=8.0) 10mM, SDS=1% SDS Solution

Lysozyme : 50mg/ml (Ameresco)

Proteinase-K : 20mg/ml (Merck Bioscience)

RNase A : 20mg/ml (Merck Bioscience)

Sodium Acetate : 3M (pH 5.2)

Tris-saturated (pH 8.0) Phenol

Chloroform: Isomylalchol (24:1)

Absolute alcohol (99.5%)

Ethanol (70%)

A1.2- Reagents used for RNA isolation using Tri-reagent:

Choloform, 1-Bromo-3-chloropropane,Iso-propanol/2-propanol,75%Ethanol,DEPC treated water (Diethylpyrcarbonate water)

A1.2.1- Reagents used for DNA isolation using TRI-reagent:

Back extraction buffer: 4 M Guanidine Thiocyanate, 50 mM Sodium Citrate NaCl, 1 M Tris base, prepared in ddH2O and sterile filtered. For 25 mL,combine 11.82 g Guanidine Thiocyanate, 0.37 g Sodium Citrate NaCl and 3.03 g Tris base were used.

Iso-propanol

70% ethanol,

Tris-EDTA (TE) buffer (Ph8.0)

A1.3- Reagents used for the Plasmid DNA isolation:

Solution I: 50mM Glucose, 25mM Tris-chloride (pH-8.0), 10mM EDTA (pH8.0)

Solution II: 0.2N NaOH, 1%SDS (freshly prepared)

Solution III: 3M Sodium Acetate (pH 5.2)

A1.4- Reagents used for expression study:

A1.4.1:Media Preparations:

LB media (Himedia) - 25 gm for 1000ml of AQ Mili-Q-water

A1.4.2-Common Components between all the media's used for the protein expression:

Chemicals	SOC media	Terrific Broth	Super Broth
Tryptone	20 gm	12 gm	32gm
Yeast extract	5gm	24 gm	20gm
NaCl	0.5gm	NA	5gm
Glycerol	NA	0.4%	0.8%

A1.4.2.1- SOC media: For 1000 ml:

Dissolve all the above reagents. Then add 10ml of 250mM KCl solution and pH should be adjusted to 7.0 with 5N NaOH, the final volume is adjusted to 1000 ml with autoclaved mili Q water. Finally, autoclave the media. After autoclaving add 20mM filter sterlized glucose, i.e. 20 ml from 1M filter sterlised glucose solution (Sambrook et al, volume-3)

A1.4.2.2: Terrific Broth preparation: for 1000 ml:

For terrific broth preparation same mixtures were used as it was used in SOC media. mixture was dissolved and after autoclaving into the media 100 ml of sterile 0.17M $KH_2PO_{4,+}$0.72M K_2HPO_4 solution was added (This solution is made by dissolving 2.31gm of KH_2PO_4 and 12.54gm of K_2HPO_4 in 90 ml of water after solutes are completely dissolved adjust the volume by adding autoclaved mili-Q water) (Sambrook et al, volume-3).

A1.4.2.3: Super Broth media: For 1000 ml:

100 ml of Monobasic and Dibasic phosphate solution was added to the media as it has been used in terrific broth preparation (Yadava and Ockenhouse, 2003).

A1.4.3:Buffers used for Sonication and purification of the proteins:

For detection of proteins in a soluble and insoluble fraction i.e. in inclusion bodies native and denaturing conditions were used for the sonication of the cells, since proteins were HIS-tagged imidazole was used in the native buffer where protein is purified by using different concentration of imidazole for the efficient binding with the Ni-NTA gravity column (Qiagen) column which was used for purification of the proteins.

A1.4.3.1: Native Buffer composition:

Buffer	Composition	Molarity	pH
Equilibration/	NaH2PO4	50 mM	8.0
Sonication Buffer	NaCl	300 mM	
	Immidazole	10 mM	
Wash Buffer	NaH2PO4	50 mM	8.0
	NaCl	300 mM	
	Immidazole	20 mM	
Elution Buffer-1	NaH2PO4	50 mM	8.0
	NaCl	300 mM	
	Immidazole	100 mM	
Elution buffer-2	NaH2PO4	50 mM	8.0
	NaCl	300 mM	
	Immidazole	250 mM	
Elution Buffer-3	NaH2PO4	50 mM	8.0
	NaCl	300 mM	
	Immidazole	500 mM	

A1.4.3.2: Denaturing buffer composition:

Denaturing purification of protein is a pH based purification, where urea being a strong denaturing agent is used f buffers remains the same, but pH of the buffers varies. As mentioned in the following table:

Buffer	Composition	Molarity	pH
Buffer B or Binding Buffer	Urea	8M	8.0
	NaH_2PO_4	100mM	
	Tris-Cl	100mM	
Buffer C or Wash Buffer	Urea	8M	6.3
	NaH_2PO_4	100mM	
	Tris-Cl	100mM	
Buffer D or elution buffer	Urea	8M	5.9
	NaH_2PO_4	100mM	
	Tris-Cl	100mM	
Buffer E or Elution Buffer	Urea	8M	4.5
	NaH_2PO_4	100mM	
	Tris-Cl	100mM	

A1.5: Reagents for SDS Polyacrylamide Gel Electrophoresis (PAGE):

A1.5.1:10% SDS:

10 gm of SDS was weighed and dissolved in 40 ml of water and kept at 37ºincubator for 3 hrs until the clear solution is observed, once the clear solution was observed volume was makeup till 100 ml in a measuring cylinder with the AQMQ water.

Note: Since SDS is detergent it will cause frothing so do not stir the mixture

A1.5.2: Acrylamide and N N'-methylene- bis- acrylamide:

A stock solution was prepared by adding 29% (w/v) acrylamide and 1% (w/v) N,N' bis acrylamide prepared with warm water (to assist dissolution of bisacrylamide). The solution was stored in a dark bottle at room temperature.

A1.5.3: 10% Ammonium per-sulphate (APS) for 2 ml:

Ammonium persulphate provides free radicals that help in the polymerization of acrylamide and bis acrylamide. 0.2 gm of APS was weighed and dissolved in 2 ml of deionized water and stored at 4ºC.

A1.5.4: 1XSDS gel loading buffer for 100 ml:

Chemicals	100ml
Tris (pH6.8)0.5M	14.2ml
BPB 80mM	6mg
10%SDS	20ml
βmercaptoethanol(5%)	5ml
60% glycerol	15ml
Water	Upto 100ml

A1.5.5: 5X TGS (Tris Glycine SDS buffer) recipe for 500 ml:

Chemicals	**500 ml**
Tris base	7.55gm
Glycine	47gm
SDS(10%)	25ml
Water	Q.S

A1.5.6:Coomassie Brilliant blue solution R250 and G250:

Typically R-250 and G250 is chemical forms of Coomassie dye. The R-250 is a red tinted form were as G-250 is green tinted form, in acidic conditions, these dyes bind to proteins through basic amino acids (primarily arginine, histidine and lysine).

A1.5.7: Preparation of R-250 (0.1% w/v) staining solution for 400 ml:

Dissolve 400mg of CBB R-250 in 80 ml of deionized water, keep it for some time at room temperature until it dissolves completely.Filter the solution with Whatman filter paper after complete filtration add 120 ml of methanol in the filtrate. Finally, add 20% of glacial acetic acid in the solution, i.e (40ml of GAA+160 ml of de-ionized water).

A1.5.8:The destaining solution for CBB staining:

A1.5.8.1: The destaining solution I:

Reagents	Destaining solution I	Destaining solution II
Methanol	40%	5%
GAA	7%	7%
Water de -ionised	QS	QS

A1.6: Western Blotting for the confirmation of protein expression:

Western blotting was done for both the proteins to confirm the expression of the proteins,

A1.6.1:Reagents and Materials required for western blotting for 1000ml:

Reagents	Concentration	Amount in grams
10x TBS,pH=7.5	Tris=500mM Nacl=150mM	60.5gm 87.5gm
Transfer buffer	Tris=45mM Glycine=39mM Methanol=20%	5.45gm 2.927gm 200ml
Blocking buffer	5% skim milk/3%BSA	5gm in 100ml of 1XTBS with 0.02%tween 20

A1.7: <u>Reagents used for ELISA</u>:

Reagents Used	Quantity
Peptide solution	150ng/ml
10x Coating buffer (pH9.4);	Na_2CO_3:1.59gm $NaHCO_3$:2.93gm
Wash Buffer	1x PBS+0. 2% Tween 20
Blocking Buffer:	5%Skim milk powder in 1XPBS
Primary antibody or serum dilution	1:50,1:100,1:200,1:400 dilute with blocking buffer
Secondary Antibody	Goat anti-mouse IgG, HRP conjugated antibody (1:1000) dilution should be made in blocking buffer
Substrate	1XTMB/H_2O_2 Specific for ELISA.
Stop solution	2N H_2SO_4

A1.8: <u>Reagents used for immunolocalization study:</u>

Reagents	
Reagents	**Quantity**
10X PBS (500ml)	NaCl: 40 gm, KCl : 1gm, Na_2HPO_4:7.2gm, KH_2PO_4: 0.12gm for 1000ml
0.2% Saponin	5ml from 2%saponin in 50ml of 1XPBS
Wash Buffer	1x PBS+0. 2% Tween 20
3%Blocking Buffer:	1.5gm of BSA in 50 ml of 1XPBS and filter sterlised with 0.22µm filter.
Primary antibody or serum dilution	1:200 antibody in 3%BSA solution
Secondary Antibody	Goat anti-mouse IgG, HRP conjugated antibody (1:1000) dilution should be made in blocking buffer

A1.8.1:Secondary antibody used for tracking oraganelles:

Secondary antibody	Oraganelles	Stock concentration	Working concentration
DAPI solution **PI solution**	Nucleus	1mg/ml(Solvent used for dissolving =Di methyl formamide)	100ng/ml
Q-Dot(invitrogen)	Apicoplast		1:1000 dilution in 6% BSA
ER-Tracker (molecular probes)	Endoplasmic Reticulam	1mM	0.5µM
Mito-Tracker-Red (Invitrogen)	Mitochondria	1mM	0.5µM

A1.9: 4% Paraformaldehyde preparation according to Sambrook et.al.Volume-2:

4 gm of paraformaldehyde was dissolved in 50 ml of autoclaved mili-Q water and with 1 ml of 1N NaOH was added. A magnetic stirrer was set at temperature 65ºC and the mixture was allowed to dissolve with gentle stirring.10 ml of 10x PBS (Phosphate buffer saline) was added to the mixture and it was allowed to cool down at room temperature (25-30º). After cooling pH was adjusted to 7.4 using 1N HCl solution. After the proper pH was achieved solution was filtered using 0.45-µm membrane filter to remove particulate matter and stored at -20º in small aliquots.

Approvals taken from ethical committees for the experiments performed in the thesis are as follows:

Title	**Approval No**	**Date Approved**	**Date Expired**
Immunolocalization of *P. falciparum* proteins using culture and patients slide material	IBSC/Pilani/2017-01/09	07.10.2017	06.10.2020
Cloning and expression of Hub genes obtained from a co-expression network of *P.falciparum* in plasmid based *E.coli* systems	IBSC/Pilani/2017-01/08	07.10.2017	06.10.2020
CGH array analysis for un-complicated and complicated *P.falciparum* samples	IBSC/Pilani/2017-01/11	07.10.2017	06.10.2020
Transcriptome Profiling, network analysis and characterization of hypothetical or putative proteins of *Plasmodium falciparum.*	IAEC/RES/22/04	08/10/2016	07/10/18

Any parasitised blood from humans was collected earlier under approval from the SP medical college and associated group of hospitals ethics committee on informed consent by trained clinicians. Approval number: No.F.(ACAD)SPMC/2003/2395),

List of Publications

List of Publications:

1. Amit Kumar Subudhi, P.A. Boopathi, **Isha Pandey**, Ramandeep Kohli, Sheetal Middha, Jyoti Acharya, Sanjay K. Kochar, Dhanpat K. Kochar, Ashis Das Disease specific modules and hub genes for intervention strategies: a co-expression network based approach for *Plasmodium falciparum* clinical isolates**,** Infect. Genet. Evol., 35 (2015), pp. 96–108

2. Amit Kumar Subudhi, P.A. Boopathi, **Isha Pandey**, Ramandeep Kohli, Rohan Karwa, Sheetal Middha, Jyoti Acharya, Sanjay K. Kochar, Dhanpat K. Kochar, Ashis Das, Plasmodium falciparum complicated malaria: Modulation and connectivity between exportome and variant surface antigen gene families, Molecular and Biochemical Parasitology, Volume 201, Issue 1, May 2015, Pages 31-46.

3. **Isha Pandey**, Ramandeep Kaur, Amit Kumar Subudhi, P.A.Boopathi, Raja C Mugasimangalam, Sudha N.Rao, Mohammed Aiyaz, Sheetal Middha, Jyoti Acharya, Sanjay Kochar, Dhanpat Kochar, **Ashis Das**. **A Tiling array based approach to predict Copy Number Variations between *Plasmodium falciparum* clinical isolates from the Bikaner region, (Manuscript under review).**

4. **Isha Pandey**, Agham P. singh, Ishan Wadi, C.R. Pillai, Ashis Das. ***In-silico* analysis and Immunolocalization study of conserved *Plasmodium* protein with unknown function from *in-vivo* systems network (Manuscript under review)**

Abstracts and Posters

Posters presented in National and International conferences:

- **Isha Pandey**, Ramandeep Kaur, Amit Kumar Subudhi, P.A.Boopathi, Raja C Mugasimangalam, Sudha N.Rao, Mohammed Aiyaz, Sheetal Middha, Jyoti Acharya, Sanjay Kochar, Dhanpat Kochar, Ashis Das. **A Tiling array based approach to predict Copy Number Variations between *Plasmodium falciparum* clinical isolates from Indian Subcontinent,**Presented at international conference "BITS Gene and Genome Regulation" 18th -20th Feb 2016, Department of Biological Sciences, Birla Institute of Technology and Science, Pilani, Pilani Campus, Rajasthan, India.

- **Isha Pandey**, Ramandeep Kaur, Amit Kumar Subudhi, P.A.Boopathi, Sheetal Middha, Jyoti Acharya, Sanjay Kochar, Dhanpat Kochar, Ashis Das. **Comparative Genome Analysis of *Plasmodium falciparum* isolates from India.** Presented at 12th International Conference on Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases, held on December 11 – December 13, 2014 in Royal River Hotel, Bangkok, Thailand.

- **Isha Pandey**, Ramandeep Kaur, Amit Kumar Subudhi, P.A.Boopathi, Sheetal Middha, Jyoti Acharya, Sanjay Kochar, Dhanpat Kochar, Ashis Das. **Glimpse of cytokine gene expression from *P. falciparum* infected human host**. Molecular Parasitology & Systems Biology Lab, Department of Biological Sciences, Birla Institute of Technology and Science, Pilani, Rajasthan. Presented at the 24th National Congress of Parasitology jointly organised by Regional Medical Research Centre for Tribal's & Indian Society of Parasitology at Regional Medical Research Centre for Tribal's, Jabalpur from 27-29 April 2013.

  .

Workshops Attended:

- Attended; National workshop on Computational Systems Biology and Dose response modelling, MGDC, Bhartidasan University, Tiruchirapalli.(March 2-4,2012.

Biography of Prof. Ashis Kumar Das

Dr. Ashis Kumar Das joined BITS, Pilani, Pilani campus as Assistant Professor in the year 1998 and is currently a Professor in the Department of Biological Sciences, BITS Pilani, Pilani campus. He has been Head of the Department of Biological Sciences (Group leader) and Dean, Research and Counsultancy Division (R&C) BITS, Pilani. He has completed his schooling from Calcutta boys school, Calcutta, Bachelors degree from University of Calcutta, Calcutta, India, Master's from University of Calcutta, Calcutta, India, and his Ph.D degree from Jawaharlal Nehru University (National Institute of Immunology), New Delhi, India. He has worked as Post Doctoral Fellow in Department of Molecular microbiology and Immunology, SHPH, John Hopkins University, Baltimore, USA and has been a WHO fellow at the malaria Branch, Centres for Disease Control and Prevention, Atlanta, Georgia, USA. His current research interests include the study of molecular biology and immunology of the human malaria parasite. And also remains enthusiastically involved in formulating and evolving diagnostic procedure for the human malaria parasites based on 18S and 28S r RNA. His interest lies in the systems biology-based approach to attempt and enhance the understanding of parasite biology, from the patients suffering with severe malaria (cerebral malaria, severe malarial anaemia, acute renal failure, hepatic dysfunction etc). His research interest also involves cloning, characterization and immunolocalization of hub genes derived from the systems network. He has published research articles in the journals of international repute and has presented his research at national and international conferences and, as invited speaker. He has attracted funding from funding agencies like University Grant Commission (UGC), Council of Scientific Research and Industrial Research (CSIR) and Department of Biotechnology (DBT), and also from private organizations like Dabur, and filed 3 Indian patents.

Biography of Isha Pandey

Isha Pandey has completed her B.Sc (Biotechnology with C.B.Z) from HNB Garhwal University, Dehradun, Uttarakhand and M.Sc Biotechnology from Amity University, Noida, Uttar Pradesh. After completion of masters, subsequently she joined as a research scholar in Department of Biological Sciences BITS Pilani, Pilani Campus, Rajasthan, and enrolled as a Ph.D student under the supervision of Prof. Ashis Kumar Das. During her Ph.D tenure, she got a chance to present a poster at the 12th International Conference on Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases, held on December 11-December 13, 2014 in Royal River Hotel, Bangkok Thailand. Her current research interest involves the understanding of the transcriptome and genome biology of the malaria parasite and the mechanism by which malaria parasite invades the erythrocytes.

The complete research work presented in this thesis can be summarized in the form of following sections:

Chapter-1-Introduction and review of literature: Malaria in humans is caused by six species of *Plasmodium* of which *Plasmodium falciparum* is considered to be the most virulent species and responsible for much of the mortality and morbidity. In spite of a decrease in malaria cases and malaria mortality rate since 2010, there are still over 216 million new malaria cases worldwide in 2016 and tragically about 429,000 malaria deaths, worldwide. The challenges to eradicate malaria is difficult because of the complex life cycle of the parasite which is coupled with its inherent property to acclimatize itself to the host environment. Additional parasite factors that may contribute to its pathogenesis are virulence factors which may interact directly with the host by using multiple host cell invasion pathways.The other reasons of the pathogenicity and complicacy of the disease could be the genomic variations, like insertions, deletions, single point mutations, allelic variations, etc. Although, genome variation studies have been done and have given a tremendous amount of information about the regions showing variations in different culture isolates, our understanding about the genomic variations in the isolates causing complicated manifestation is still illusive.

In order to study genomic variations, different, next generation techniques have been designed, among them array Comparative Genomic Hybridization is one of the robust techniques to study the whole genome chromosomal aberrations. CGH a new chromosome analysis technique was first discoverd by Kallioniemi et al in 1992. Later on, array CGH was developed with arrayed oligonucleotide probes on glass slides. Oligonucleotides in a tiling array covers the whole genome or contigs of the genome in an unbiased fashion. A tiling array usually spans the entire genome of an organism. The probes can be non-overlapping or partially overlapping and tiled to cover the entire genome end to end or spaced at regular intervals. The array CGH tiling array provides various opportunities to determine the comparative analysis between the treated and the control samples. A Custom designed high density overlapping 25 mer tilling microarray was used to detect the SNPs and small insertions and deletions (indels) and copy number variations to distinguish the three *Plasmodium falciparum* laboratory strains HB3, Dd2 and 3D7, where 3D7 was used as

a reference strain and Dd2 was compared with the 3D7 strain, Deletion of 6 subtelomeric gene has been detected in chromosome num-2 of Dd2 strain in comparison with the 3D7 strain an amplification event has also been found in Dd2 on chromosome-5 surrounding *pfmdr-1* gene and also on chromosome-12 near *pfgch-1* gene.In fosmidomycin resistant mutants of Dd2, an amplification event has been found in chromosome-14 compared to the parental Dd2 strain.

With the complete genome sequence of *Plasmodium falciparum* in 2002, it was predicted that it would provide insights into parasite genome biology and provide information leading to discovery of new drug and vaccine candidates. As stated before, the genome of *Plasmodium falciparum* consists of approximately 60% of the genes that codes for hypothetical proteins and many of these do not show any sequence homology with other organisms, especially the human host. One of the objectives of functional genomics and systems biology is to provide indicators of antimalarial targets, in this context transcriptomics studies have led to the discovery of differential expression of different genes in different stages of the parasite life-cycle.System level functionality of the genes is extensively explored nowadays with the use of gene co-expression networks. The network construction is fairly straightforward, where genes are represented as nodes. These nodes are connected if the genes are significantly co-expressed in the diseased samples. The question remains, whether the connection between the nodes is meaningful or not? To this end, a general framework of soft thresholding has been designed by Zhang and Horvath in 2005, where each connection has been given a weight, known as weighted gene co-expression networks (WGCN). This network is more biologically relevant than unweighted gene co-expression networks. In WGCN different modules are generated in which one gene is highly connected with the other genes in that subnetwork or module, this HUB gene could be highly related to the external gene information.

A study from our lab on construction of co-expression based systems network using microarray co-expression data from field samples revealed the vital role of hub genes and have shown the enrichment of these genes into several crucial biological processes that are important for parasite survival and pathogenicity, In this study ,some of the hub genes identified were up or stably expressed genes, e.g. PF14_0150 was found to be up-regulated in the network whereas two other PF08_0085 and PF07_0083 were found to be stably

expressed genes.This data suggests that whether the genes are differentially expressed or otherwise, being hub genes the proteins coded by these genes could be essential for the parasite metabolic activities, and thus for survival. In addition to annotated genes, some of the unknown function proteins or hypothetical proteins were also reported as hub genes and could also play a major role in parasite survival. Characterization of these hub genes could lead to the discovery of novel drug target or molecules for other intervention/prevention strategies.

The primary objective of this thesis was to dissect and report the genomic variation to identify and report large scale amplifications and deletion at the genome level between *P. falciparum* parasite isolates causing complicated and uncomplicated disease manifestations.

The Second objective of the thesis involved identification and characterization of key molecules from a co-expression based systems network generated based on parasite RNA from complicated and uncomplicated patient samples. Further correlation of the array CGH data with the regions containing the selected genes and analysis of minor genetic variations like SNPs in the field isolates for the selected molecules was performed.Subsequently cloning and expression of selected molecules was attempted and, In-silico based analysis of selected molecules and Peptide-based immunolocalization studies were done.

Chapter-2-Materials and Methods: To achieve the objectives mentioned above various techniques was used based on the exhaustive literature survey. The methodology adopted for comparative genome hybridization analysis and reagents and kits used for the characterization and localization of the proteins are detailed in this chapter. The chapter starts with detailing of the methodology used for the comparative genome hybridization analysis, the results of this analysis are discussed at length in chapter-III. Following this detailing, the next section of this chapter includes a selection of the hub genes from co-expression based systems network for the study. The rest of the chapter is divided into two parts: The first part includes experimental details and procedures used for the cloning, expression and localization studies of the proteins which includes blood sample collection from malaria infected patients, isolation of parasite DNA/RNA, confirmation of the infection by PCR, amplification of selected hub genes, their cloning and expression trials. The results of the experiments performed are discussed at length in chapter-IV. The second

part of the chapter includes *in-silico* based approach for the analysis of the selected proteins using various bio-informatics tools. This includes the analysis of conserved domain in proteins, and disordered nature of protein, functional prediction of proteins, secondary structure prediction and physicochemical nature of the selected proteins with the help of various freely available online software and peptide based immunolocaization studies are discussed in chapter -V.

Chapter-3-Comparative genomic hybridization of *Plasmodium falciparum* clinical isolates:This chapter focuses on the results obtained from the Comparative Genomic Hybridization (CGH) data analysis of Pf-isolate-2 (complicated) in comparison with the Pf-Isolate-1 (uncomplicated) and details about the variations (loss/gain) present in different regions of the genome of Pf-isolate-2. A detailed description of significant regions that are showing aberrations of more than 1kb, like, sub-telomeric regions and amplifications and deletions other than the subtelomeric regions are reported in this chapter. Regions showing variation comprises a total of 637 unique genes in the genome of Pf-isolate-2. The loss/gains varied in size from 100 nucleotides to 47,000 bp and included intra and intergenic regions. Out of a total of 637 genes, amplifications was seen in 511 genes and deletions in 176 genes. There are 27 genes in the regions of Chromosome-4, 6,7,8,10,11,12,13 in Pf-2 in which there is the presence of both amplifications/gains and deletions/loss. These regions are rich in Variant Surface Antigen family genes. The deletion and amplifications covered smaller regions and did not cover more than two contiguous genes. Most of the variation was seen in the sub-telomeric and telomeric regions. However regions covering for genes showing novel and putative variations are also reported. The differences reported in Pf-2 could also be inherently natural differences between two parasites isolates without any direct correlation with the disease states, which could be influenced by various other factors. This study is just a preliminary attempt to report the variations or differences in the genomic segments of parasite isolate causing severe malaria with that of the isolate causing uncomplicated malaria. To get a real picture of the genomic variations, these analyses should be done with a large set of severe clinical samples from different parts of India.

Chapter-4-Cloning and protein expression of conserved *Plasmodium* protein with unknown function (PFD0300W and PFL2320W): In this chapter, the results of the trials of protein expression of hub genes PFD0300W (PF3D7_0406000) and PFL2320W (PF3D7_1248400) that codes for conserved *Plasmodium* protein with unknown function derived from the systems network are discussed. The genome variation in the selected gene in relation with the aCGH data is discussed. For variation data analysis at the level of point mutations the genes were amplified from genomic DNA, and one clone from field isolates and few amplicons were sent for sequencing. In case of PFL2320W, sequencing of the clone as well as a few of the amplicons showed no change from the available 3D7 annotated gene sequence from PlasmoDB database. In case of PFD0300W the partial sequence obtained after sequencing of clone and from a few field isolates suggests the absence of any variation in the gene. Initially, the conventional cloning of amplicons were done in expression vector pRSETA, further sequenced for the presence of the genes in the expression vector. For expression studies different *E.coli* host cells were used like BL21DE3, BL21DE3pLysS, Rosetta and Rosetta pLySs, BL21DE3 host cell along with RIG plasmid. Expression studies were performed at different time points and temperatures along with different media and IPTG concentrations. Different strategies were applied for the protein expression of the two proteins in pRSETA system but without any success. The reason could be the presence of RARE codons in the genes. To overcome the problem gene sequences were outsourced (Invitrogen, Thermo Fisher Scientific) for the optimization of the rare codons according to the codon usage of the *E.coli* host. The optimized constructs were obtained in the expression plasmid pET100D-TOPO, with 6X Histidine as a fusion tag. Even after using different host cells and different expression conditions the proteins were not being able to expressed. We checked for the mRNA transcript expression in BL21DE3 and found that the transcript is present at all the time points for the genes, suggesting that the mRNA is stable. The reason of mRNA expression and not the protein expression could be the toxicity of the protein to the host cell, mRNA secondary structure that prevents interaction with cellular machinery, post-translational modification of protein or the nature of protein could be intrinsically disordered which will not allow it to express in any particular conformation.

Chapter 5-In-silico analysis and immunolocalization of conserved *Plasmodium* protein with unknown function (PFD0300W and PFL2320W):In this chapter, the in-silico analysis of two conserved *Plasmodium* protein PFL2320W and PFD0300W with unknown function and also their immunolocalization in cultured *P. falciparum* parasites from India is discussed. In-silico analysis was done using different online available tools. In order to check for the presence of conserved domains in both the proteins, CDD, InterPro scan and P-fam was used. In PFD0300W conserved domain analysis using CDD, available at NCBI, revealed that PFD0300W contains an NBD sugar kinase HSP_70 actin superfamily domain, where as in PFL2320W no conserved domain was detected.The in-silico analysis for both the proteins for signal peptide prediction, transmembrane domain prediction suggests the absence of transmembrane domain and signal peptides in both the proteins. The disorder prediction analysis suggests the presence of 46.45% disorderness in PFD0300W and 48.77% disorderness in PFL2320W, homology modeling for both the proteins was not able to achieved due to absence of any suitable template for either of the proteins. Peptide based subcellular localization was done in chloroquine resistant lab adapted *P. falciparum* culture of Indian isolate (RKL-9) which is a field isolate collected from Rourkela (Odisha) that exhibits higher gametocytaemia i.e >2% using antisera raised against predicted peptides in swiss albino mice, which suggests that protein (PFD0300W) is localized in cytoplasm and not in endoplasmic reticulum or nucleus of the parasite. Further, the protein is localized in most of the satges i.e asexual and sexual stages of the parasite. For sub-cellular localization of PFL2320W apicoplast targeting marker Q-Dot and Mito-tracker were used, along with the anti peptide antisera. Although the presence of the protein for PFL2320W was detected, accurate localization for the same was not reproducible.

Chapter-6-Conclusion: In conclusion, by analysing array CGH data, here we have seen some novel and putative variations in addition to the already reported variations in the *P. falciparum* isolate causing complicated manifestation in comparison with the *P. falciparum* parasite isolate causing uncomplicated disease manifestation.The in-silico analysis of the conserved *Plasmodium* proteins with the unknown function from *P. falciparum* showed that both the proteins are intrinsically disordered and one of the proteins (PFD0300W) is localized in cytoplasm in most of the asexual stages of the parasite. Whereas the other

protein PFL2320W expression was detected in asexual and sexual stages of the parasite, however, accurate localization was not reproducibly obtained.