# Genetic and Structural Characterization of Serine Repeat Antigen (SERA) in the Human Malarial Parasite *Plasmodium* from Indian Isolates

**THESIS**

Submitted in Partial Fulfillment

of the Requirements for the Degree of

**DOCTOR OF PHILOSOPHY**

By

**C.N.RAHUL**

ID. No: 2010PHXF801H

Under the supervision of

**Prof. Vidya Rajesh**


**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**HYDERABAD CAMPUS**

**2015**

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI - HYDERABAD CAMPUS**

## CERTIFICATE

This is to certify that the thesis entitled ***"Genetic and Structural Characterization of Serine Repeat Antigen (SERA) in the Human Malarial Parasite Plasmodium from Indian Isolates"*** submitted by Mr. **C.N.Rahul, ID.No. 2010PHXF801H** for the award of Ph. D. degree of the Institute embodies the original work done by him under my supervision.

Signature in full of the supervisor: _____

Name in capital block letters: **VIDYA RAJESH**

Designation: **Associate Professor,**
**Dept. of Biological Sciences**

Date: _____

*Dedicated To Almighty*

# Acknowledgement


- *First of all, I would like to thank the almighty for giving me the opportunity to do the research work with good health and prosperity.*
- *The following PhD thesis encompasses my individual work although I could have never reached the heights or explored the depths without the help and guidance of lot of people.*
- *My deep sense of gratitude to* ***Prof. B. N. Jain,*** *Ex-Vice Chancellor (BITS) and* ***Prof. V.S.Rao,*** *Acting Vice Chancellor & Director (BITS-Pilani, Hyderabad Campus), for permitting me to carry out my research work in the campus.*
- *I wish to express my sincere gratitude and heartfelt thanks to my supervisor,* ***Vidya Rajesh, Associate Professor,*** *Department of Biological Sciences****,*** *BITS-Pilani, Hyderabad campus, for instilling in me the qualities of being a good researcher and biologist. Her excellent guidance, constant encouragement and suggestions have been the major driving force throughout my research work.*
- *I am grateful to all the former FIC's, Research & Consultancy Division* ***Prof. M. Moorthy Muthukrishnan, Prof. P. Yogeeswari*** *and present HOD****,*** *Biological Sciences Department,* ***Prof. SumanKapur*** *of BITS, Pilani, Hyderabad campus for their valuable support. I acknowledge* ***Prof. Ramakrhishna Vadrevu*** *and* ***Dr. Pranav Narayan****, Biological Scieince Department, BITS, Pilani, Hyderabad campus for their valuable suggestions in the capacity of my Doctoral Advisory Committee members.*
- *I wish to express my thanks and take immense pleasure in acknowledging all the faculty of Department of Biological Sciences, BITS-Pilani, Hyderabad Campus for their support and encouragement to carry out this research work.*
- *My special gratitude and sincere thanks to* ***Prof. A. K. Das,*** *Department of Biological Sciences, BITS, Pilani and* ***Prof.N.Rajesh,*** *Department of Chemistry, BITS-Pilani , Hyderabad Campus for their valuable advice and suggestions during my entire research work. I extend my acknowledgement to faculty of Chemistry, Pharmacy and Physics, BITS-Pilani, Hyderabad Campus for motivating me.*
- *I place my deep sense of gratitude and loving thanks to my family members for their blessings, encouragement and kind advice to complete this work.*

- *I deeply acknowledge **DST (Department of Science & Technology), New Delhi, India** for funding my Initial project fellowship for the first year of my research work and **UGC (University Grant Commission**), **New Delhi, India** for funding me with a **UGC-CSIR fellowship** for the remaining years of my research work. I would also like to acknowledge the commercial sequencing service centres, Bioserve India Pvt.Ltd, Hyderabad, India and **Invitrogen,** BioServices India Pvt. Ltd, Gurgaon, India for providing reliable and accurarte DNA sequencing results.*
- *My humble and deep thanks to **Mr. Shiva Krishna**, former Project Assistant BITS-Pilani, Hyderabad Campus and **Mr.Atul Pawar**, former ME Biotechnology student, BITS-Pilani, Hyderabad Campus who were associated with my research work and also co-authored the publications gained from this research work.*
- *This research work would not be complete without the constant support of blood samples obtained from our collaborators. **Dr. M.Meera,** Sir Ronald Ross Institute of Tropical and Communicable Diseases, Fever Hospital, Nallakunta, Hyderabad, India for collection of samples across Andhra Pradesh, **Dr. Sandhya Phadke,** Ehrlich laboratories, Royapetta, Chennai, India for providing samples from Tamil Nadu and **Dr. Nalini Rastogi** and **Dr. Aditya K.Pandey** for collecting samples from Orissa and Allahabad respectively.*
- *I would like to acknowledge my deep regards to my fellow research students **Madhu Poornima** and **Manasi** and lab technicians for their encouragement and maintaining a healthy, positive and cordial environment all through the years of my research work in this campus.*

***I humbly dedicate this thesis to God***

*Date:* *C.N.Rahul*

# Abstract

The burden of malaria caused by *Plasmodium* is huge in India and widespread across the globe. Genetic diversity of stage specific antigens in malaria parasite is a major obstacle hindering effective vaccine development. Understanding the extent and dynamics of genetic polymorphism in vaccine antigens is important to guide rational vaccine design. In this context, exploring genetic polymorphism in highly expressed blood stage antigens of malaria parasite would help in evaluating and validating them as potential targets. SERAs (Serine Repeat Antigens) are promising candidates in asexual blood stage of malaria, both, as a vaccine and drug target. SERAs belong to a multigene family representing several homologues that are expressed mainly in blood stage and have been reported in several *Plasmodium* species. The PfSERA5 in *P. falciparum* is the most well characterized member till date and is reported to be predominantly expressed in the late blood stage of parasitic life cycle. This antigen is well explored in terms of localization, processing, immunological potential and its indispensable role in parasite exit (egress) from infected RBCs. Genetic diversity of PfSERA5 in field isolates is reported from several geographic locations across the world. However, in comparison to *P.falciparum*, the research focus on *P.vivax* SERA is negligible worldwide. Moreover, there are no reports on genetic diversity characterization of SERA in the human malaria parasite *Plasmodium* from India.

Vaccine research on *P.vivax* malaria is of high priority, as its disease burden is huge and affects large populations annually. Also, recent reports from India substantiate this benign form can cause severe illness and death. Hence, we investigated, the genetic polymorphism of two highly transcribed PvSERA genes PvSERA4 and PvSERA5 in Indian field isolates. Our results of sequence polymorphism in C-terminal region of PvSERA

demonstrate extensive diversity in PvSERA5 with major deletions, insertions and SNPs and signifying this region to be under positive selection. On the other hand, PvSERA4 C-terminal region showed high sequence conservation in Indian field isolates. We further, evaluated the genetic diversity of these genes in the central and N-terminal regions. The central domain of both PvSERA4 and PvSERA5 from Indian field isolates show high sequence conservation. Evaluation of N-terminal region in PvSERA5 shows high genetic diversity with non-synonymous substitutions, while the N-terminal region of PvSERA4 shows deletions and insertions in the glutamine rich tetrameric repeat units, contributing to its diversity. These reports representing genetic diversity in terminal regions of both PvSERA genes are first of its kind from India and worldwide, leading to identification of novel haplotypes. This study will help in elucidating the allelic variants and aid in understanding the antigenic variations and immune mediated selection mechanisms contributing to the functional significance of these two *vivax* proteins.

In highly transcribed member PfSERA5, the central region is structurally well characterized to have a unique putative papain like cysteine protease motif. This is implicated to be a useful drug target for intervention in *P.falciparum* malaria as it is established to play an indispensable role in parasite exit from blood stage. Our report of high sequence conservation of the PvSERA4 central region coupled with its high expression indicates a possibility of its critical role in *P.vivax* parasite biology. As there is no information on the structural characterization of *P.vivax* SERA members, we deduced the 3D model of this central region of PvSERA4 by homology modelling using PfSERA5 as structural template. Further, we performed MD simulation to acquire refined protein structure of this central domain. The 3D model showed the presence of an unusual antiparallel Beta hairpin motif between catalytic residues similar to haemoglobin binding motif of *Plasmodial* hemoglobinases. Apart from PvSERA4, on similar lines we also

modelled central regions of other highly expressed PvSERA members like PvSERA2 and PvSERA5, as studies substantiate importance of targeting other SERA members based on evidence of functionally redundant roles in blood stage. Our PvSERA structural models together with report of its genetic conservation in multiple field isolates recommends it to be reliable drug target for *P.vivax* malaria.

The characterization of sequence diversity of blood stage *Plasmodium falciparum* SERA5 (PfSERA5) is lacking in a malaria endemic country like India, although it has been recently reported from nine countries worldwide. In PfSERA5, the N-terminal (47KDa) protein contributes to its vaccine potential and shows repeat polymorphism mainly of two types Octamer Repeat (OR) and Poly Ser-Repeat (SR). We report existence of unique repeat polymorphism and novel haplotypes for both OR and SR regions within the N-terminal region. Several isolates from India were identical to low-frequency African haplotypes. Unique finding of our study was an Indian isolate showing a deletion in the perfectly conserved 14 mer sequence within OR region. Indian haplotypes showed broad diversity with representation of all three allelic clusters reported worldwide. Evaluation of these haplotypes helps to understand variant specific immunity and aids in malaria vaccine research.

PfSERA5 central region is indicated to play a crucial indispensable role in parasite egress although its exact functional role remains an enigma. This central region consists of a prodomain and a putative enzymatic domain. The structure of the enzyme domain is known but the prodomain structure has not been elucidated so far. Using *insilico* approach we attempted to predict the structure of the prodomain. Our predicted prodomain structure showed structural similarity to calcium binding proteins. Taking clue from this observation, we performed Molecular Dynamics (MD) simulation on the putative enzyme domain both in presence and absence of calcium and our MD results substantiated conformational

differences in presence of calcium. These results lend support to the importance of calcium in parasite egress and also indicated calcium dependent activity and stability of SERA5. These computational results gives impetus to gain future insights by designing calcium dependent inhibition studies on SERA5.

Our results on genetic diversity characterization and *in-silico* analysis of these highly expressed SERA genes from both *Plasmodial* species – *P.falciparum* and *P.vivax* is first of its kind from India. This study gives further scope for evaluating and validating the potential of *Plasmodial* SERA genes as drug and vaccine development targets against malaria.

**Keywords:** *Plasmodium vivax*; *Plasmodium falciparum*; PvSERA5; PvSERA4; PfSERA5; Genetic diversity; Repeats ; Insertions; Deletions; SNPs; Haplotypes; Positive selection; India; Homology modelling; Antigenicity index;

# Table of contents

# Abbreviations & Symbols

| | |
|---|---|
| SERA | Serine Repeat Antigen |
| Pf | *Plasmodium falciparum* |
| Pv | *Plasmodium vivax* |
| Mab | Monoclonal Antibody |
| RBC | Red Blood Cell |
| Ab | Antibody |
| $^0$C | Degree Celsius |
| DNA | Deoxy ribonucleic acid |
| RNA | Ribonucleic acid |
| SNP | Single Nucleotide Polymorphism |
| EDTA | Ethylene diamine tetra acetic acid |
| EtOH | Ethanol |
| SDS | Sodium dodecyl sulphate |
| dNTPs | deoxy ribonucleotide triphosphate |
| Mins | Minutes |
| M | Molar |
| ng | nano gram |
| µl | Micro litre |
| rpm | Revolutions per minute |
| ml | milli litre |
| G | Gram |
| KDa | Kilo Dalton |
| MSA | Multiple Sequence Alignment |
| *Sal I* | Salvador I (*P. vivax* standard strain) |
| 3D7 | *P.falciparum* standard strain |
| WHO | World Health Organization |

**Amino Acids**

| | | |
|---|---|---|
| Alanine | Ala | A |
| Arginine | Arg | R |
| Aspargine | Asn | N |
| Aspartic acid | Asp | D |
| Cysteine | Cys | C |
| Glutamine | Gln | Q |
| Glutamic acid | Glu | E |
| Glycine | Gly | G |
| Histidine | His | H |
| Isoleucine | Ile | I |
| Leucine | Leu | L |
| Lysine | Lys | K |
| Methionine | Met | M |
| Phenylalanine | Phe | F |
| Proline | Pro | P |
| Serine | Ser | S |
| Threonine | Thr | T |
| Tryptophan | Trp | W |
| Tyrosine | Tyr | Y |
| Valine | Val | V |

**Nucleotide bases**

| | |
|---|---|
| Adenine | A |
| Guanine | G |
| Cytosine | C |
| Thymine | T |

# List of Figures

# List of Tables

# CHAPTER 1

# INTRODUCTION

# Introduction

## Malaria

Malaria is a widespread disease caused by the parasite *Plasmodium* and is transmitted to humans by the female anopheles mosquito vector. It remains a major threat among parasitic infections and affects diverse populations across the world. Four main *Plasmodial* species, viz. *P. falciparum, P.vivax, P.malaria* and *P.ovale* are known to cause malaria in humans. Of these, *P. falciparum* & *P. vivax* are the most common human malaria parasites (WHO, 2014). *P. falciparum* is known to be associated with the life threatening form of malaria. Recently, *P. knowlesi*, which usually causes malaria among monkeys is added as one of the species affecting humans. (Jongwutiwes, Putaporntip, Iwasaki, Sata, & Kanbara, 2004).

## Origin & Brief History

Malaria is one of the world's oldest known diseases, and its history predates humans by millennia (Carter & Mendis, 2002; W. Liu et al., 2014). The earliest written report of malaria is by a Chinese emperor Huang Di Neijing (Yellow Emperor, 2697–2590 BCE) who referred to these repeated paroxysmal fevers associated with enlarged spleens and had a tendency of gaining epidemic occurrence (F. E. G. Cox, 2002). The name malaria is derived from 'mal'aria' (bad air in Medieval Italian). Early research, has indicated that for about 1000 years herbal remedies have been used to treat malaria (Willcox & Bodeker, 2004). In 1820 ($19^{th}$ century) Pierre Joseph Pelletier and Joseph bienaimé caventou separated the antimalarial drugs quinine and chinconine as alkaloids from the tree bark (Haas, 1994). Later, French physician Charles Louis Alphonse Laveran observed pigmented parasites inside red blood cells, at a military hospital in Constantine, Algeria (F. E. Cox, 2010). In 1885, Camillo Golgi observed that all parasites present in the blood divided almost

simultaneously at regular intervals and that, the divisions coincided with attacks of fever (Santamaria, 1994). Proof of malaria transmission by mosquito was discovered by Sir Ronald Ross in 1897, an army surgeon working in Secunderabad, India (Ross, 1897).

**Prevalence of Malaria**

According to World Malaria report fact sheet 2014, latest estimates indicate that around 198 million cases of malaria have been reported in 2013 with 584,000 deaths (WHO, 2014) . The incidence of malaria is widespread across the globe with sub Saharan Africa being the most affected region (Greenwood & Mutabingwa, 2002; Price et al., 2007).

In the South-East Asian region, WHO estimates that around 1.32 billion are prone to the risk of malaria. Indian population holds the highest risk contributing about 76% of total malaria cases from South East Asia (Kondrashin & Rooney, 1992; Kumar, Valecha, Jain, & Dash, 2007; WHO, 2012).

**Malaria epidemiology in India**

Malaria in India is a major challenge and imposes great socioeconomic burden on the human population. Malaria epidemiology in India is complex because of its geo-ecological diversity, multi-ethnicity and wide distribution of its six known anopheline vectors involved in transmitting malaria (Das et al., 2012; Kumar et al., 2007).

According to the historical background of malarial incidence in India, the highest incidence rate reported so far was in the year 1950 with 75 million cases and 0.8 million deaths (www.nvbdcpgov.in). This led to the launch of NMCP (National Malaria Control Programme) in the year 1953 which was successful in reducing the number of malaria cases to <50,000 and bringing down the mortality rate to zero by the year 1961. Since then, the numbers of reported malaria cases have declined gradually in the country. In 2009, the number of confirmed cases declined steeply to 1.6 million with ~ 1100 deaths as reported

by WHO (Das et al., 2012). But, evidence from several other epidemiological survey studies have indicated that the scale of burden caused by malaria has been greatly underestimated (Dhingra et al., 2010; Hay, Gething, & Snow, 2010) and that the true burden of malaria in India is still uncertain.

Four principle factors have been attributed to malaria in India which are detailed in Table 1.1.

**Table 1.1 Factors contributing to malaria incidence in India**

| Principle parameter and Contributing cause | Effects | References |
|---|---|---|
| **Unique topography with vast geographic and climate diversity** (from tropical monsoon in the south of the country to temperate in North) | **Influences distribution of parasites**<br>Both *P.vivax* and *P.falciparum* are predominant although the two infecting species vary in proportion across the country.<br><br>Eg: *P. vivax* is predominant in Tamil Nadu, *P.falciparum* is predominant in Orissa while in west (example:Gujarat) mixed infections of both species prevalent. | (Das et al., 2012) |
| **Genetic diversity in the malaria parasite.** High genetic diversity within parasite genes involved in immunity and drug resistant genes. | **Influence parasite gene function**<br>Several studies on parasite genetic diversity in India, which include mainly functionally significant immune and drug resistant gene shown to be highly polymorphic<br>Eg: *Plasmodium* Merozoite Surface protein (MSP) gene and Chloro Quine resistance *P. falciparum* chloroquine resistance transporter (PfCRT) gene. It has been reasoned that this is due to unique eco-epidemiology of India in comparison to other malaria endemic regions across the globe. | (V. Singh, Mishra, Awasthi, Dash, & Das, 2009) |
| **Enigmatic malaria vector diversity**<br>Malaria vectors in India vary in vectorial capacity. The potential of the mosquito to serve as a vector depends on the ability to support sporogony, mosquito abundance, and contact with humans, which are all influenced by climatic and ecological factors. | **Influence distribution of vectors.**<br>Among six species of anopheles responsible for the malaria burden, *An.culicifacies* accounts 60-65% of malaria in India. These are distributed into number of sibling species which are although morphologically indistinguishable have diverse vector capacity and ecological distribution. | (Adak, Kaur, & Singh, 1999) |
| **Indians exhibit high susceptibility to malaria and are evolving malaria resistance** mainly caused due to different geographic distribution of human diseases like thalassemia, G6PD deficiency and duffy negative blood group. | This resulted in evolution of genetic variants to protect against malaria in India. | (Gaikwad, Ashma, Kumar, Trivedi, & Kashyap, 2005) |

**The National Malaria control Programmes in India**

The National Vector Borne Disease Control Programme (**NVBDCP**) is the agency involved in prevention and control of all vector borne diseases in India. This is one of the technical departments of the Directorate General of Health Services under the ministry of health. It is responsible for framing technical guidelines, policies and monitoring operations through regular reports on malaria control. It has deployed major strategies to achieve its objectives of well-informed and self-sustained health care systems in India. Presently, their program activities are in accordance to their Millennium development goal to achieve halting and reversing incidence of malaria and other vector borne diseases by 2015. Their important initiatives through National Rural Health Mission (**NRHM)** are :

1) Provision of village based Accredited Social and Health Activities (ASHAs)

2) Trained personnel in malaria diagnosis using rapid diagnostic tests (**RTD**) and

3) antimalarial drug administration (Source: NVBDCP).

The NVBDCP achieves evaluation of its programmes through National Institute of Malaria Research **(NIMR)**. It's one of the permanent institutes of Indian Council of Medical Research **(ICMR).** NIMR regularly evaluates new insecticides and diagnostic kits, conducts field trials and monitors resistance to insecticides among vectors and drug therapy among parasites (www.mrcindia.org).

***Plasmodium* Life Cycle**

Malaria parasite exhibits a complex lifecycle involving two different hosts. A part of the lifecycle is completed in the insect vector female anopheles mosquito (primary host) and remaining is completed in the humans (secondary host). The infection begins, when sporozoites in the saliva of mosquitoes are injected into human blood by a feeding mosquito. These sporozoites are carried by the blood stream to the liver, where they invade

hepatocytes. Intracellular parasites replicate by asexual reproduction in hepatocytes to form merozoites, known as exo-erythrocytic cycle. The resultant newly formed merozoites are released into the blood stream, which further invades the red blood cells (RBCs). In the RBC, the merozoites develop into ring, trophozoite and schizont stages and produce further merozoites. Among the newly released merozoites some infect the other erythrocytes and few occur as sexual forms (male and female gametocytes) which are then taken up by the mosquito, and the life cycle continues (Carpenter, Pearson, Mitchell, & Oaks Jr, 1991).

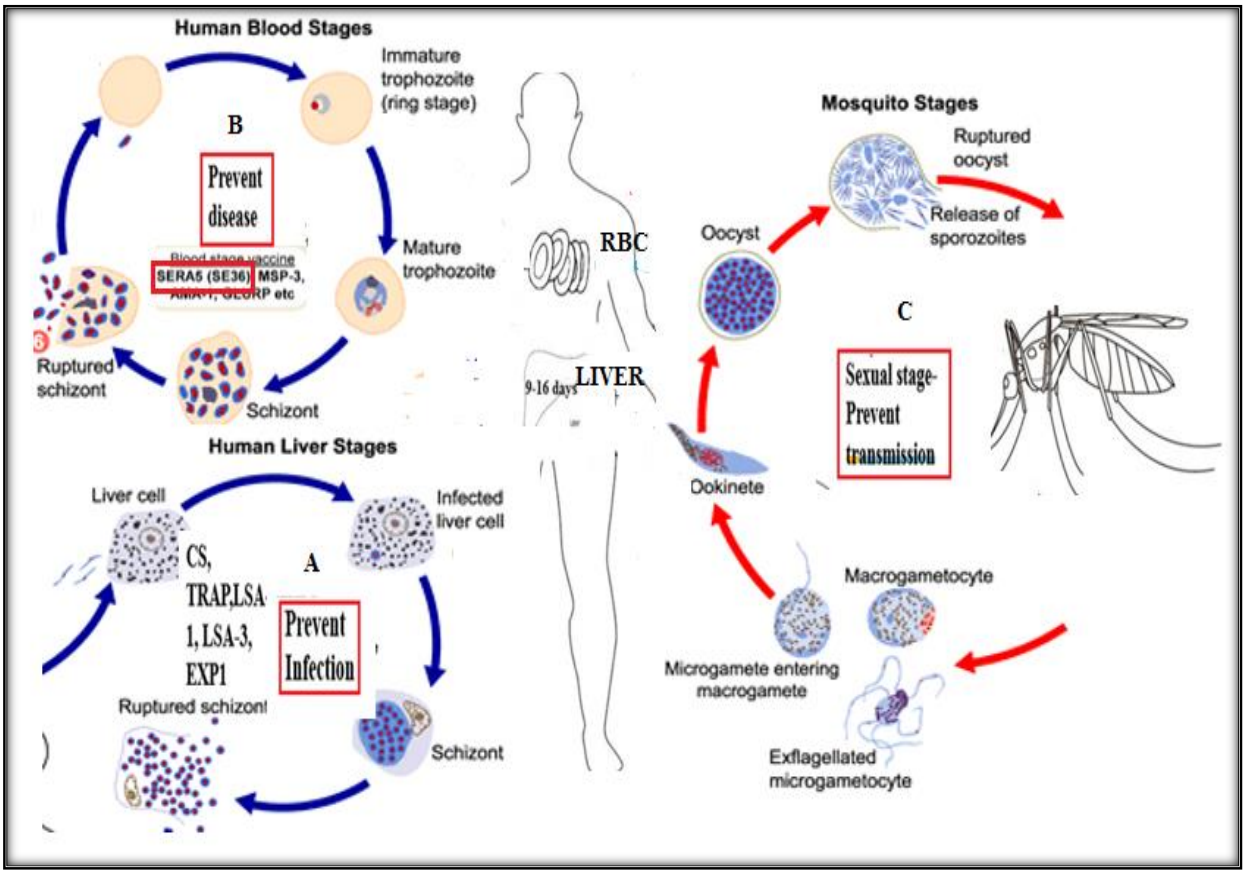



**Figure 1.1. Schematic representation of Life cycle of malaria parasite (Figure modified from : Parasite Image Library, National Centre for Infectious Disease, Division of Parasitic Disease, CDC, Atlanta. http://www.cdc.gov/dpdx/). The malaria parasite life cycle involves two hosts. During a blood meal, a malaria-infected female Anopheles mosquito inoculates sporozoites into the human host. A) Exo-erythrocytic schizogony- Sporozoites infect liver cells and mature**

**into schizonts, which rupture and release merozoites. B) Erythrocytic schizogony- After this initial replication in the liver, the parasites undergo asexual multiplication in the erythrocytes Merozoites infect red blood cells. The ring stage trophozoites mature into schizonts, which rupture releasing merozoites. Some parasites differentiate into sexual erythrocytic stages (gametocytes). The gametocytes, male (microgametocytes) and female (macrogametocytes), are ingested by an Anopheles mosquito during a blood meal. C) Sporogonic cycle- The parasites' multiplication in the mosquito. While in the mosquito's stomach, the microgametes penetrate the macrogametes generating zygotes. The zygotes in turn become motile and elongated (ookinetes) (10) which invade the midgut wall of the mosquito where they develop into oocysts. The oocysts grow, rupture, and release sporozoites, which make their way to the mosquito's salivary glands. Inoculation of the sporozoites into a new human host perpetuates the malaria life cycle. The boxes marked in red indicate the vaccine strategies to prevent transmission, infection and disease and targets identified in liver and blood stages, showing SERA is a blood stage member.**

### The Malaria research progress, problem, and challenges

Mankind has stepped into the second decade of the 21th century and malaria still remains a major threat to worldwide population causing a huge global impact. Over the years, there has been a reason of hope to develop a vaccine against this disease. This hope has been strengthened by various research (Carvalho Goto, Daniel-Ribeiro & Goto, 2002; Doolan, Dobano, & Baird, 2009) which showed that.

a. Clinical immunity is acquired naturally in populations living in intense transmission areas.
b. Protective Immunity against blood stage infection could be achieved by antibodies transferred passively from immune to non-immune individuals.

c. Immunization of primates, humans and mice with radiation-attenuated sporozoites or recombinant antigens could induce partial or sterile (complete protection) anti-parasite immunity.

However, there are multiple challenges and constraints which act as impediments in the development of malaria vaccine:

a) Natural immunity takes many years to be acquired and occurs only when there is heavy or continuous contact with parasite.
b) Extensive polymorphism of *Plasmodial* antigens leading to different antigenic properties (Mercereau-Puijalon, Fandeur, Guillotte, & Bonnefoy, 1991).
c) Host genetic restriction of malaria antigen recognition mediated by Major Histocompatibility Complex (MHC) (Schuler et al., 2001).
d) Lack of both *in vitro/in vivo* surrogate tests to justify reliable protection.
e) Non-availability of adjuvants that can significantly strengthen the response against antigens.

Based on the above mentioned challenges and constraints it has been concluded that

1) Acquisition of immunity to malaria remains a poorly understood phenomena.
2) Reliable correlates of protection are not well defined.

Malaria vaccine need to be inducing immunity more than nature for robust and sustained protection which cannot be achieved by traditional vaccine design. Hence only a multivalent vaccine would be most effective (Zepp, 2010). Developing such a highly effective vaccine would need incorporation of two or more distinct antigens, or two or more alleles of same antigen from several lifecycle stages.

## Targeting diverse vaccine candidates in malaria life cycle

Over the years, despite of all the above mentioned constraints, development of vaccine is considered to be an important approach for prevention of malaria. Programmes involved in malaria vaccine development have mainly focussed on six target stages of parasitic life cycle i.e. 1) Sporozoite 2) Liver stage 3) Merozoite (Surface antigens) 4) Blood 5) Toxins (parasite-derived molecules) 6) Sexual Stages.

Among the stages, earlier studies show that protection in the liver stage could be achieved when irradiated sporozoites were used to immunize mice and humans (Belnoue et al., 2004) proving the feasibility of vaccination in the liver stage. However, it was constrained by the requirement of repeated inoculations with hundreds of radiation attenuated sporozoites for sustained protection. Results in natural infection have indicated that much lower immune responses are attained in liver stage than those by irradiated sporozoites. Hence, it is important to identify better options than the liver stage as a means of effective vaccination strategy.

Targeting the blood stage has received the most attention (Ocana-Morgner, Mota, & Rodriguez, 2003) as malarial symptoms manifest in this stage. Naturally acquired immunity is also mainly directed against antigens of this stage. Also, vaccination strategies against the blood stage are believed to be most effective in eliminating malaria as sterile immunity could be achieved contributing towards reducing the severity and complications of disease (Wipasa, Elliott, Xu, & Good, 2002).

## Understanding Blood Stage Immunity

Blood stage antigens are important targets in malaria life cycle and their efficacy is proven in various animal models. Also, to a partial extent efficacy of these in humans limit parasite growth and minimise clinical disease providing reasonable cause to be optimistic about

development of an effective blood stage antigen (Crewther, Matthew, Flegg, & Anders, 1996; Genton et al., 2002; Ling, Ogun, & Holder, 1994; Perera, Handunnetti, Holm, Longacre, & Mendis, 1998; Richards & Beeson, 2009).

Most of the leading malaria vaccine antigens currently considered for clinical trials are all blood stage antigens located on the merozoite surface or apical organelle and play an important functional role in erythrocyte invasion. (Richards & Beeson, 2009).

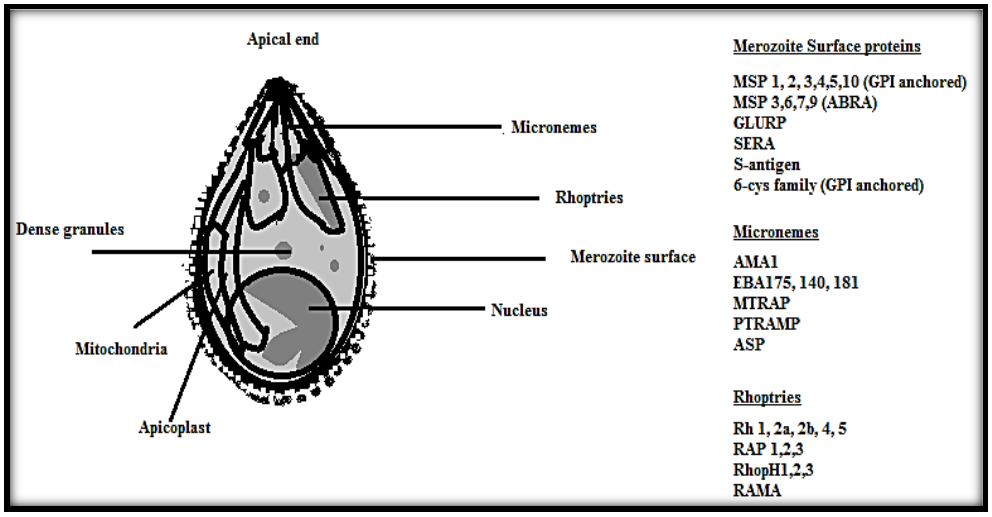



**Figure 1.2. Structure and major antigens of *P.falciparum* merozoite**

**Erythrocyte Invasion** is one of the crucial and well characterized steps in the erythrocytic cycle. The apical end of merozoite has specific organelles involved in erythrocyte invasion including the paired rhoptries and micronemes which are thought to expel the proteins to bind to erythrocyte receptors. Merozoite ligands and vaccine candidate antigens are present in the organelles and on the surface of the merozoites. Surface proteins may be GPI anchored or associated with GPI anchored proteins by molecular interaction (Gilson & Crabb, 2009).

The list of known proteins of merozoite surface and organelles is given in **Figure 1.2** (Richards & Beeson, 2009).

**Vaccine development in Malaria**

To control and eradicate malaria, a long lasting and broadly efficacious malaria vaccine would be the most sustainable approach. In the last decade, towards this aim, establishment of PATH Malaria Vaccine Initiative (MVI) (Barry & Arnott, 2014) has been done to assist and accelerate malaria vaccine development. This initiative has greatly contributed in the evaluation and identification of promising malaria vaccine candidates. The first malaria vaccine technology roadmap was launched by WHO in 2006 to further focus on and unite global vaccine initiatives. It aimed at achieving the development and licencing of first generation malaria vaccine by 2015 that has protective efficacy of more than 50% against severe disease and death, and lasts more than one year (World Health Organization, 2006)

Natural immunity in malaria is directed against the blood stage antigens and in principle are more important than pre-erythrocytic and transmission blocking antigens. However, this protection is very slow and takes several years to develop even in individuals living in endemic areas. Earlier studies in mice indicated that B cells play an important role in protection, where they established that mice lacking B cells developed chronic infection and passive transfer of antibodies could confer protection in mice (Wipasa et al., 2002). Also, the degree of protective immunity in humans, mice and monkeys correlated with level of antibodies and Ab isotypes. Ab response of cytophilic IgG subclass IgG1 and IgG3 are considered most important for protection against *P.falciparum* malaria. All these evidences were clearly established in *P. falciparum* malaria. Additionally, host genetic factors and age are also partial contributors for developing immunity (Aucan et al., 2001). Monoclonal Abs against the merozoite antigens were more protective in blocking erythrocyte invasion than polyclonal ones. These studies suggest the importance of identifying epitopes that induce protective Abs (Ramasamy, Ramasamy, & Yasawardena, 2001). Moreover, variant specific antibodies agglutinate infected RBCs providing variant specific immunity (Marsh &

Howard, 1986) . This was evident in studies conducted in both, children with convalescent stage infection and adults from endemic areas wherein their serum having antibodies had specificity to agglutinate infected RBC and parasitic strain respectively. This has given scope for studying antigenic diversity as an important parameter for guiding malaria vaccine design and deployment (Wipasa et al., 2002).

**Antigenic diversity: A major hurdle in development of Malaria vaccine**

Development and deployment of a multivalent, multi-component malaria vaccine is suggested for malaria with an aim to achieve high anti-parasite immunity (Conway, 1997). Ideally, this would require incorporation of different allelic forms of multiple malarial antigens from worldwide isolates (Targett, 1995). The existence of genetic variability and high frequency of genetic variants circulating in the population along with the fact that the parasite continues to evolve through mutations and sexual recombination under environments of selection pressure as a strategy for immune evasion, makes these, a greater challenge in vaccine design. This would also affect the vaccine efficacy of malaria depending on the alleles chosen. Similar observations are also made with respect to other pathogens like bacteria and rapidly evolving viruses. Hence, it is important to conduct molecular epidemiological studies of leading vaccine candidates from different geographical regions to identify the antigenic diversity, most relevant to vaccine escape and cross-protection (Takala & Plowe, 2009) .

In this context, several malaria antigens are identified as leading vaccine candidates and are under consideration for development in the malaria vaccine. Among these, those that are under clinical trials (Barry & Arnott, 2014) include liver stage CSP (Circumsporozoite Protein), LSA-1 (Liver Stage Antigen-1), Thrombospondin-related antigen (TRAP), Merozoite surface proteins (MSP1, MSP2, MSP3), Serine Repeat Antigens

(SERA) and AMA-1 for *P. falciparum*. Most promising ones among these include AMA-1 and MSP-1. For *Plasmodium vivax,* only two vaccine candidates PvCSP and PvDBP (duffy binding protein) are under clinical trials. MVI has stressed the need to invest more resources in identifying more promising vaccine candidates for progression into clinical trials (Barry & Arnott, 2014) .

**Serine Repeat Antigen (SERA)**

Of the ones listed above, Serine Repeat Antigen (SERA) is recognised by the WHO as a promising malaria vaccine candidate of highest interest amongst *P.falciparum* malaria vaccine antigens (Palacpac, Arisue, Tougan, Ishii, & Horii, 2011). SERA represents a multigene family that encodes a protein with a central protease domain having a putative papain like- cysteine protease motif (Aoki et al., 2002). All SERA members can be classified into two groups Ser-type SERAs (Serine in the active site) and Cys-type SERAs (Cysteine in the active site), based on the amino acid in the active site and their shared homology in the central protease region (Arisue et al., 2011).

***P. falciparum* SERA**

In *P. falciparum*, SERA has 9 homologues and earlier reports suggest that SERA5 (PfSERA5) is the most predominantly expressed antigen in the late asexual blood stage followed by SERA 4, 6 and 8 (Aoki et al., 2002). Majority of the work on SERA is carried with PfSERA5 and it is the well characterized and holds high priority for vaccine development due to following reasons:

1) PfSERA5 is important in the parasite development in the blood stage. The N-terminal domain of SERA5 is shown to raise parasite inhibitory antibodies naturally, making it a potential target for immunological relevance and vaccine design (Aoki et al., 2002).

2) The central region of PfSERA5 has structural features similar to putative cysteine protease motif. This motif is implicated to play a key role in the multistep process of parasite egress (newly formed merozoites are released from the host RBC) in late blood stage of the parasite development (Hodder et al., 2003).

Hence, PfSERA5 is a useful target for both vaccine and drug development (Lee & Fidock, 2008; McCoubrie et al., 2007).

The PfSERA5 gene has a four exon three intron gene structure with first exon coding only for signal peptide (Tanabe et al., 2012). The genetic loci of PfSERA5 is located on chromosome 2 of the *P.falciparum* genome. The standard reference sequence of *P.falciparum* 3D7 strain is available in the primary database NCBI- PUBMED with accession number AE001362 and the gene sequence is available in secondary database PlasmoDB with accession number XM_001349550 (Gardner et al., 2002). The graphical representation of PfSERA5 gene and protein structure is shown in **Figure 1.3**

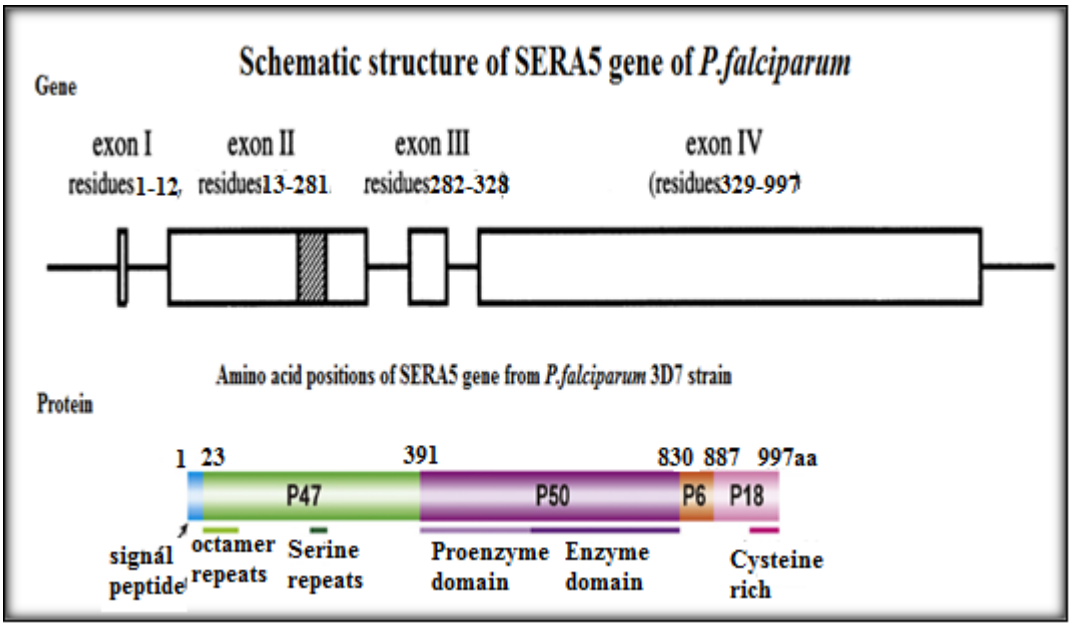



**Figure 1.3 Schematic representation of PfSERA5 gene (Morimatsu, Morikawa, Tanabe, Bzik, & Horii, 1997) and protein product (Tanabe et al., 2012).**

The 120 KDa encoded SERA protein is known as P126/SERP and is secreted in late trophozoite and schizont stage. It undergoes proteolytic processing into N-terminal 47 KDa (P47), central 56KDa protease that is further converted to a 50 KDa (P50) and 6KDa fragment (P6) and C-terminal 18KDa fragment (P18) as represented in **Figure 1.4**. Based on Western Blot and indirect immunofluorescence studies it is understood that all these processing events occur in infected RBCs just prior to schizont rupture. These results further confirm that the P50 is released into the culture medium, while the P47 and P18 are the only components associated with the merozoite surface held together by di-sulphide bridge. Also, parasite inhibitory antibodies were found only against P47 domain of PfSERA5 (Li, Mitamura, Fox, Bzik, & Horii, 2002).

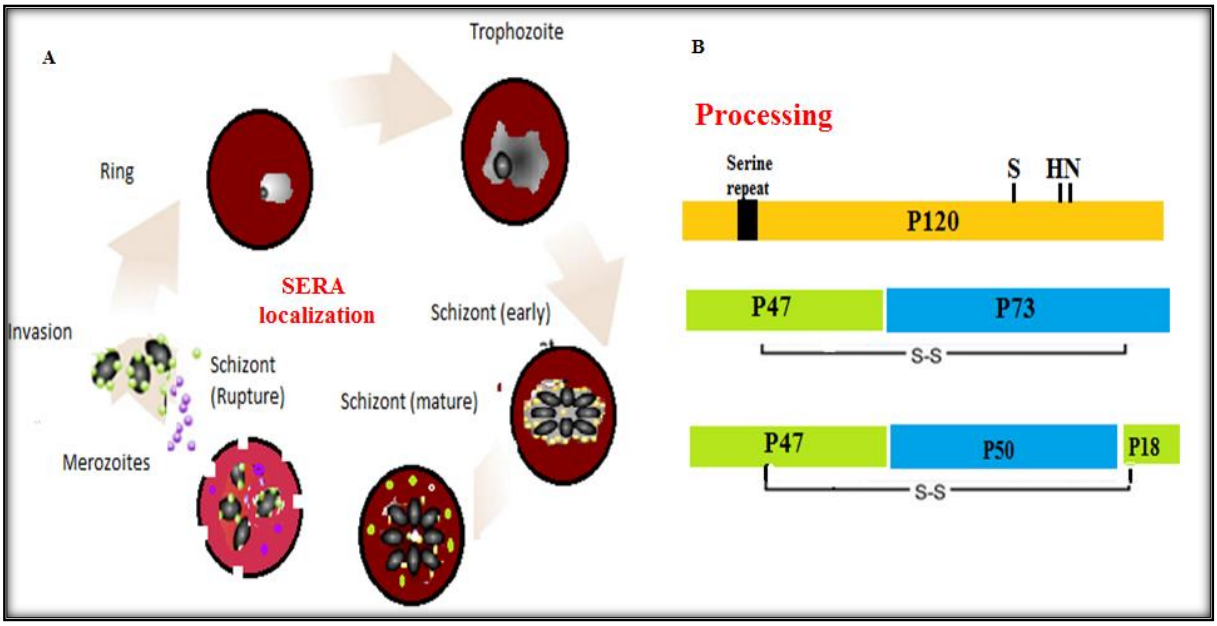



**Figure 1.4 SERA5 localization and processing in *P.falciparum*. A) Localization in blood stage is colour coded to correspond to B) showing PfSERA5 processing events. Figure adopted from (Palacpac et al., 2011).**

The N-terminal domain-P47 of SERA5 is well characterized to be highly immunogenic in natural environment. Antigenic diversity is reported to be predominant in this region in comparison to other domains P50 and P18 that are relatively well conserved (Safitri et al., 2003) . Several studies show that recombinant proteins of this region could induce protective immunity in animal models like aotus and squirrel monkeys (Aoki et al., 2002). Also, studies show that P47 reacts to naturally acquired antibodies obtained from serum of individuals from malaria endemic regions of Uganda and Brazillian amazon (Banic et al., 1998; B. A. Okech et al., 2001) . Moreover, these studies show the ability of serum antibodies against P47 to inhibit parasite growth *in vitro*. Results of epitope mapping show a major epitope mapped within the conserved region of P47 to be a target of parasite inhibitory antibodies. Further, results of fine epitope mapping indicated an epitope in the repeat region to be partly recognized by a monoclonal antibody (mAb) 43E5 (B A Fox, Xing-Li, Suzue, Horii, & Bzik, 1997; Barbara A Fox, Horii, & Bzik, 2002) . The P47 region is majorly encoded by exon II which shows mainly two types of repeats which are separated by a conserved stretch of 100 amino acids. The repeat regions include Octamer Repeat region (OR) with repeat types of eight amino acids and Serine Repeats (SR) with stretch of poly-serine residues (Q. Liu et al., 2000).

Till date, structural insight is available only for the central domain (P50) of PfSERA5 which is well covered and studied. This domain shows homology to archetypical cysteine protease L-cathepsin and is believed to have a cysteine protease structural motif (Hodder et al., 2009). The exact role of PfSERA5 central domain is enigmatic due to the recent evidence of its speculative non catalytic role and the presence of serine in the active site catalytic triad instead of cysteine. Hence, it is currently reviewed to be an important molecular determinant in parasite egress playing an indispensable, non-catalytic role in asexual blood stage (Hodder et al., 2009; Stallmach et al., 2015).

Current studies on antigenic diversity of PfSERA5 indicated that, there is limited repertoire of PfSERA5 alleles that could contribute to variant specific immunity against *P.falciparum* malaria thus effectively serving as a potential candidate malaria vaccine (Tanabe et al., 2012).

### *P.vivax* Serine Repeat Antigen

In *P.vivax,* 12 PvSERA homologues have been identified, of which PvSERA4 and PvSERA5 are the highly expressed members in the blood stage (Palacpac et al., 2006). In contrast to studies in *P.falciparum* SERA, very limited information is available about the genetic characterization of PvSERA members (Kiefer et al., 1996). Research in *P. vivax* malaria lagged disproportionately in comparison to *P. falciparum* (Palacpac et al., 2006). Recent studies support evidence for anti-PvSERA4 antibodies against regions in PvSERA4 which are similar to N-terminal P47-PfSERA5 indicating its immunogenic potential and scope for its consideration for vaccine (Yildiz Zeyrek et al., 2011). Hence, focus on *P.vivax* is a research priority particularly considering the huge burden of malaria caused by *P.vivax* in India leading to search for new candidate genes to escalate malaria vaccine development.

### Gaps in existing research

As it is clear, understanding genetic diversity within malaria parasite is important to evaluate the history of its association with the disease. Also, it may reduce the efficacy of therapeutic drugs and vaccines (Neafsey et al., 2012). Antigenic diversity studies on PvSERA genes have been limited worldwide and the comprehensive analysis done in PfSERA5 reports genetic diversity from different geographic locations (Safitri et al., 2003; Tanabe et al., 2012). However, genetic characterization of SERA genes from a large malaria endemic

country like India is lacking. Hence, considering the importance of these blood stage antigens it is important to study their antigenic diversity from Indian population.

**Scope and objective of work**

The research work presented in this thesis are the first reports on antigenic diversity of SERA in the Indian population. The study gains significance as attempts to eradicate the disease by means of vector control and parasiticidal drugs have met with difficulties owing to growing resistance of anopheles mosquitoes to insecticides and of the parasites to available antimalarial drugs. In the context of SERA being a high priority vaccine candidate our study aims to investigate the genetic diversity of SERA genes in both the highly prevalent species of *Plasmodium – P.vivax* and *P. falciparum* in the Indian context. This genetic polymorphism report from Indian field isolates adds to the existing repertoire of antigenic variations for SERA.

The following are the objectives proposed for the study:

1. To investigate polymorphisms of the Serine Repeat Antigen (SERA) gene of *Plasmodium* species in field isolates from different parts of India.
2. To compare the polymorphism, if found, with reported data from non-Indian strains.
3. Classification and identification of alleles and haplotypes found in different geographic locations in India.
4. Evaluation of data using bio-informatics tools to investigate the impact of polymorphism on the structure and epitopes of SERA.

# CHAPTER 2

# METHODOLOGY

# Methodology

*Plasmodial* antigens are important candidates for vaccine design in malaria and have great research focus worldwide due to the persisting challenges in combating the disease. It is interesting to note that there is global cataloguing process for representing antigenic diversity of leading malaria vaccine antigens of *Plasmodium* sp. (Barry & Arnott, 2014). India, being an important study centre of malaria complexity (Das et al., 2012), many studies have focused on understanding the genetic diversity of malaria vaccine antigens like AMA, CSP and MSP from various geographic regions (Farooq, Malla, & Dubey, 2009; Rajesh et al., 2007; Ranjit & Sharma, 1999). However, there are no reports on SERA from India. Recent reports of severe malaria due to *P.vivax* from India, has propelled *P.vivax* research, in both India and worldwide (Kochar et al., 2009; Price, Douglas, & Anstey, 2009). Diversity studies on *P.vivax* SERA are also relatively very low worldwide. Research on *P.vivax* is neglected mainly due to lack of methods to *in vitro* culture (Valencia, Rodriguez, Acero, Ocampo, & Arevalo-Herrera, 2011). Use of molecular biology tools like Polymerase Chain Reaction (PCR) has aided to overcome culturing limitation allowing direct amplification of parasitic DNA from patient blood and is considered to be an advancement in genotyping of malaria samples (Brito & Ferreira, 2011; Greenwood, 2002). Use of PCR combined with direct sequencing of gene sequence of candidate antigens for malaria is a common approach. However, direct sequencing of PCR products does not reveal mixed strain infections and could over represent genetic diversity. This problem can be mitigated by selection of sequence chromatograms showing non-overlapping peaks for further analysis judging them to be single genotype infection .Cloning of PCR product is suggested to be better choice to improve calculation of genetic diversity but this approach is suggested to be time consuming (Lopez et al., 2012). A general flowchart representation of molecular

approaches to sequence diversity characterization of malaria genes is shown in **Figure 2.1.** A similar PCR based molecular approach was carried out from Indian field isolates to characterize the SERA antigen of the human malaria parasite of both the *Plasmodium* species from this part of the world.

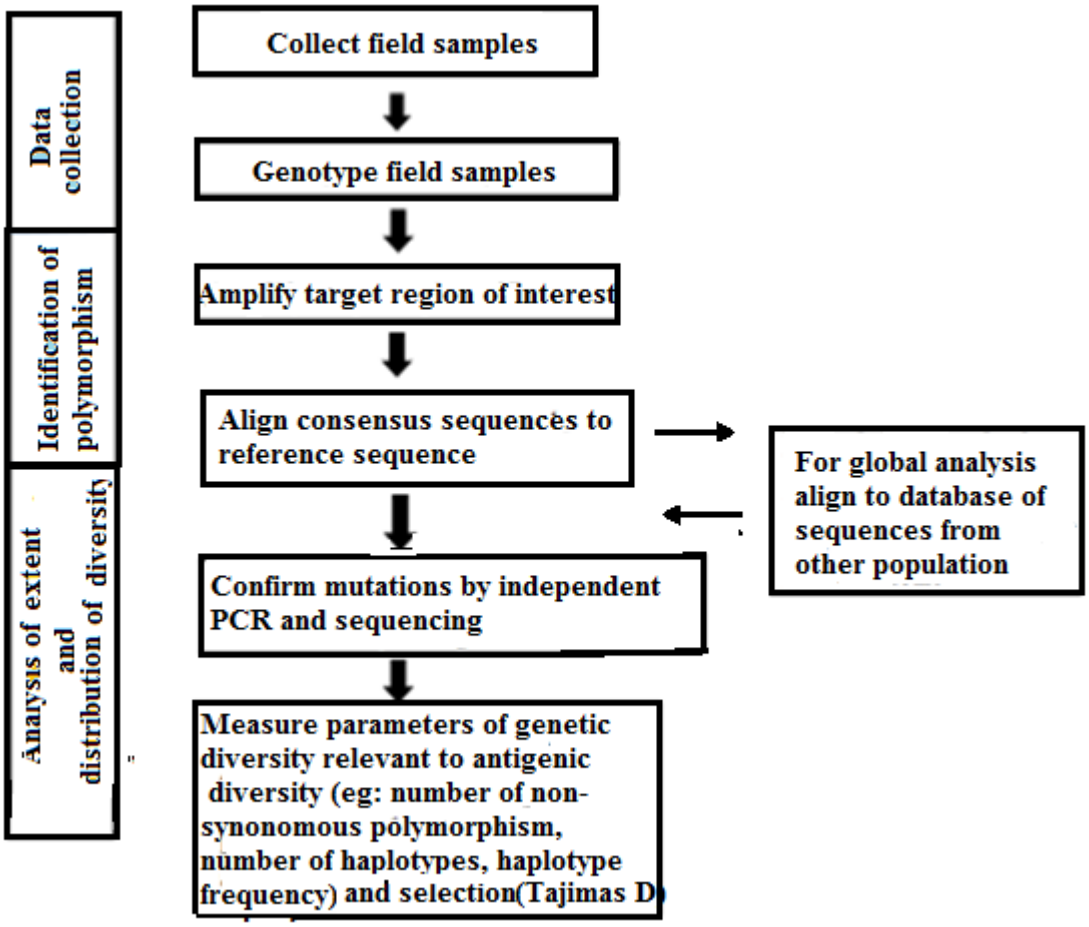



**Figure 2.1 Flowchart representation of molecular approaches to genetic diversity characterization of malaria antigens these are the steps followed in our study. The figure is adopted from (Barry & Arnott, 2014).**

### Collaboration and Blood Collection

Collection of malaria positive blood is an important source to study genetic diversity of SERA genes. Parasitic DNA available in infected blood, is the source for amplification of specific SERA genes by PCR which allows further analysis.

Study sites from different parts of India, mainly Tamil Nadu (Chennai) and Andhra Pradesh (Hyderabad), Orissa (Bhuvaneshwar) and Uttar Pradesh (Allahabad) were chosen for blood sample collection by establishing collaboration with reputed hospitals and labs like

1. Sir Ronald Ross Institute of Tropical and Communicable Diseases, Hyderabad, Telangana, India.
2. Ehrlich laboratories, Chennai, Tamil Nadu, India.
3. Secunderabad Diagnostic and Research Centre (SDRC), Hyderabad, Telangana, India.
4. Clinicians at Allahabad and Orissa also supported us with few samples from their region.

Blood samples were collected by trained technicians from patients who visit hospitals and labs for diagnostic confirmation and treatment of malaria. Care was taken to collect samples after taking the patients consent in the duly signed patient consent form (Appendix I). All samples were given a unique reference ID at these hospitals and stored at - $20^0$C before being transported to BITS for further analysis.

Approval was taken from the Institutional Human Ethical Committee of BITS-Pilani under IHEC approval No. IHEC-22/09-10; Dated 21-04-2010 under the title 'Polymorphism Studies of the Malarial Vaccine Candidate Serine Repeat Antigen (SERA) in Indian Isolates'.

**Parasite infected blood DNA extraction**

1) Lysis buffer mix (10mM Nacl, 50mM Tris, 10mM EDTA, 1% SDS).
2) Proteinase K
3) Phenol- Tris saturated
4) Choloroform
5) Iso-amyl alcohol
6) 3M Sodium acetate
7) 70% and 100% ethanol (EtOH)
8) 1X TE buffer

The blood DNA isolation protocol was adopted from standard DNA extraction procedure (Sambrook & Russell, 2001). 500µl of blood was added to an equal volume of lysis buffer mix. Incubation of the lysate was done at $37^0$C for 45 minutes. 5 µl of Proteinase K stock-20mg/mL was added and incubated for 90 minutes at $55^0$C. DNA was purified once with phenol/chloroform/isoamyl alcohol (25:24:1) and then with chloroform/isoamyl alcohol (24:1). The DNA was precipitated with 3M sodium acetate (1/10), 2 volumes of EtOH (-20 $^0$ C). After incubation in -$20^0$C for 16 h the DNA was centrifuged (15 mins-10,000 rpm) and the pellet was suspended in 1ml cold ethanol (70% v/v). Before use centrifuged pellet was suspended in 30µl of TE buffer and 3 µl of it was loaded on to agarose gel electrophoresis to check for DNA bands (**Figure 2.2**).

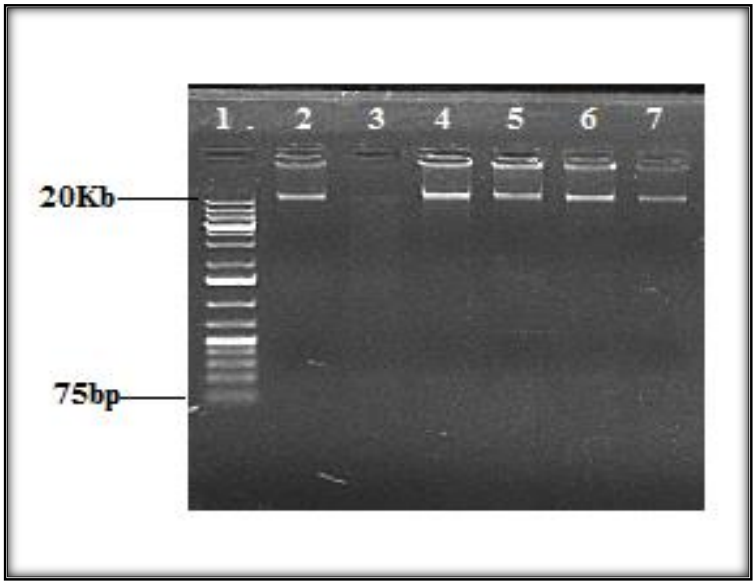



**Figure 2.2 DNA Isolation from malaria positive blood. Lane number 1 indicates DNA ladder 1Kb plus (Thermofisher.Scientific.Cat.No-1333). Lane 2-7 indicates isolated blood DNA.**

### PCR based diagnostic confirmation of malaria infection

A diagnostic PCR test based on the 18s rRNA was performed using the parasite DNA extracted from infected blood samples to differentiate and confirm detection of *P.falciparum* and *P.vivax* infection (Das et al., 1995). The PCR reaction utilized the following primers- one genus specific 5' primer and two species –specific 3'primer in the same reaction cocktail (**Table 2.1**). The PCR results yield different size fragments for *P.falciparum* (~1.4 Kb) and *P.vivax* (~500 bp) confirming species specific malaria infection.

**Table 2.1 Primer details for diagnostic PCR**

| Name of the primer | PRIMER |
|---|---|
| A1- Genus specific | 5'ATCAGCTTTTGATGTTAGGGTATTT3' |
| A289-*P.vivax* specific | 5'TAACAAGGACTTCCAAGC3' |
| A291-*P.falciparum* specific | 5'GCTCAAAGATACAAATATAAGC3' |

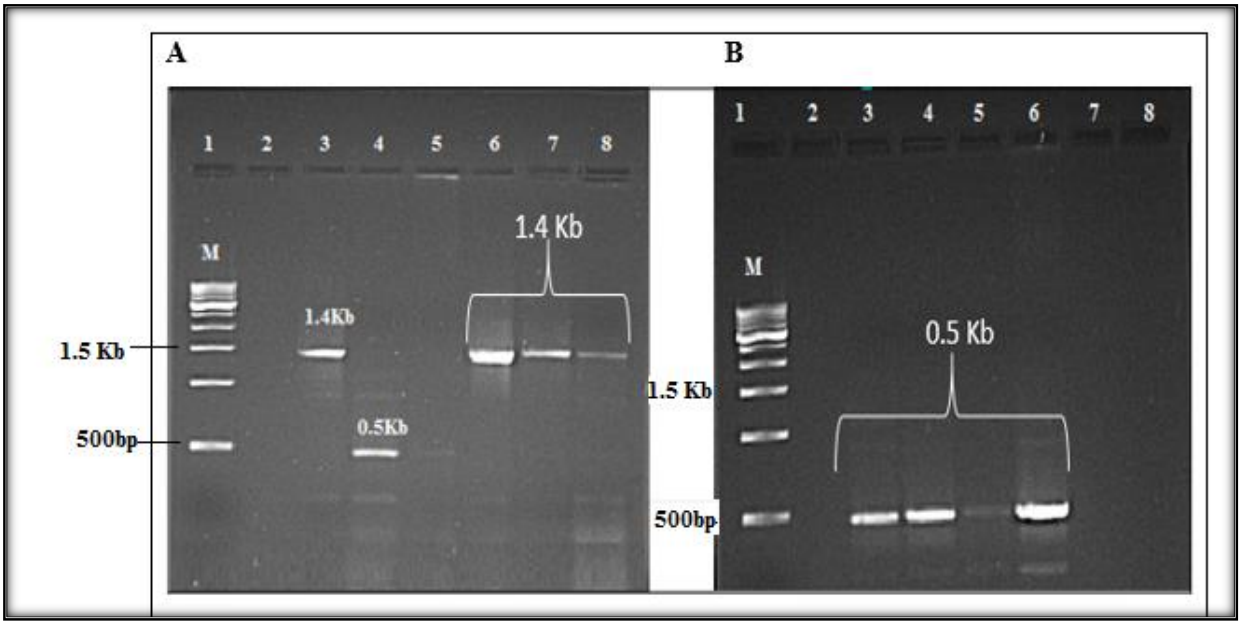



**Figure 2.3 Diagnostic PCR showing differences between *P.falciparum* and *P.vivax* confirming malaria infection. Lane 1 in both A and B indicate 500 bp DNA ladder (GeNei-Merck Millipore # Cat No-105670). DNA bands in image A-lanes 3, 6-8 correspond to 1.4 Kb band confirming *P.falciparum* infection. DNA bands in image B-lanes 3, 4 and 6 correspond to 500 bp confirming *P.vivax* infection.**

### SERA genes from *Plasmodium* species considered for study

Considering the important role of highly expressed PfSERA5 gene in the blood stage we prioritized the amplification and genetic diversity study of this well characterized antigen in *P.falciparum*. On similar lines, studies in *P.vivax* also show high expression of PvSERA4 and PvSERA5 of *P.vivax* among multigene family of 12 homologues (Palacpac et al., 2006) . In comparison to *P.falciparum*, studies on characterization of PvSERA genes is very limited worldwide and there is no reports of SERA genes in Indian isolates. Accordingly we prioritized to study sequence diversity of both PvSERA4 and PvSERA5 in *P.vivax*. Gene details and strategy of amplification of these highly expressed genes are given in the following **Table 2.2** and **Figure 2.4.**

## Table 2.2 Gene details, gene structure and strategy of amplification

| Gene details | Gene structure | | Amplification strategy |
|---|---|---|---|
| PfSERA5 (*P.falciparum*) | **Accession number** | **AE001362** | Amplified fragments of 1.2 Kb (N-terminal region) and 1.8 Kb (Central and C-terminal region) were obtained. This was done to overcome technical constraints associated with large gene size using PCR |
| | **Gene size** | **3435bp(3.4Kb)** | |
| | **CDS** | **2999bp (2.9Kb)** | |
| | **Processing** | **N-terminal -P47 ( encoded by Exon II) and Central-P50 and P18 (encoded by Exon IV)** | |
| PvSERA4 (*P.vivax*) | **Accession number** | **PVX_003845** | On similar lines as done for PfSERA5. Amplified fragments of 1.6 Kb (N-terminal region) and 2Kb (Central and C-terminal region) were obtained. |
| | **Gene size** | **4203bp (4.2Kb)** | |
| | **CDS** | **3698bp (3.6Kb)** | |
| | **Processing** | **Its processing events are unknown, studies and analysis done on highly expressed PvSERA4 are based on analogy with PfSERA5.** | |
| PvSERA5 (*P.vivax*) | **Accession number** | **PVX_003830** | Amplified fragments of 1.2 Kb (N-terminal) and 1.4 Kb (Central and C-terminal) were obtained. |
| | **Gene size** | **3681bp (3.6Kb)** | |
| | **CDS** | **3264bp** | |
| | **Processing** | **Like PvSERA4 its processing events has not yet been explored.** | |

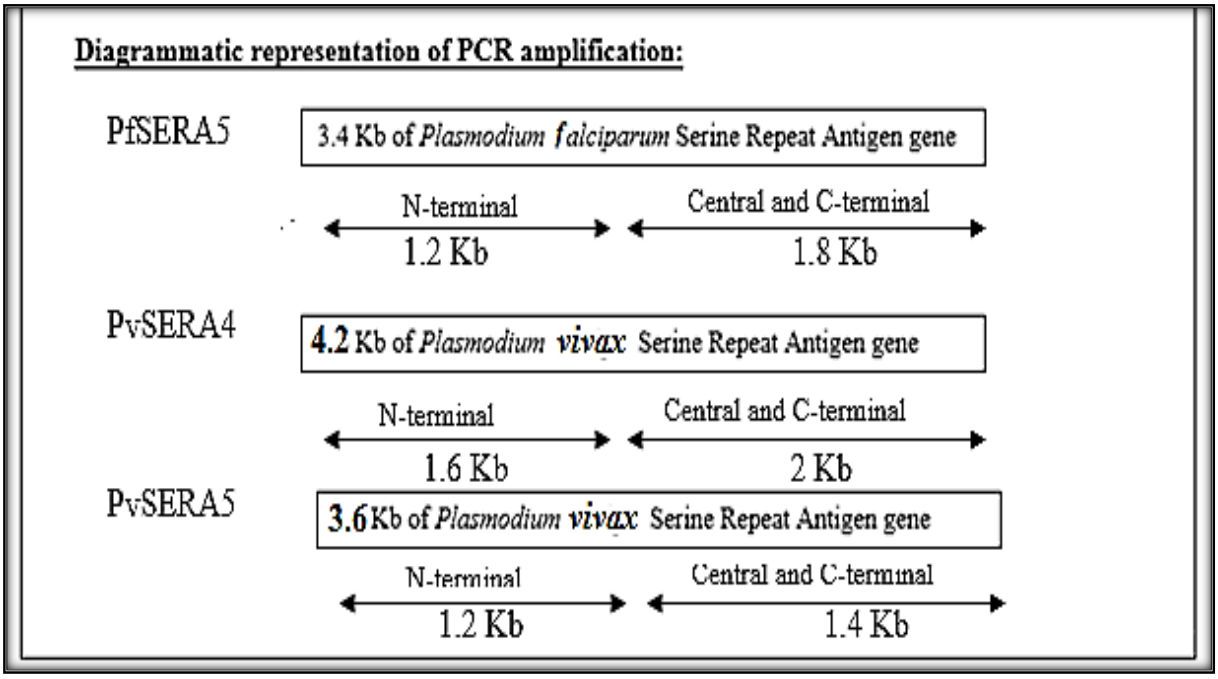
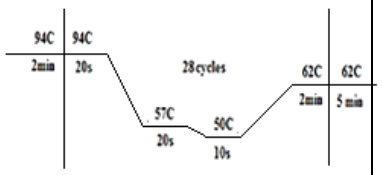
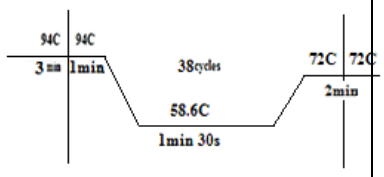
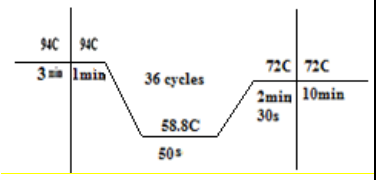
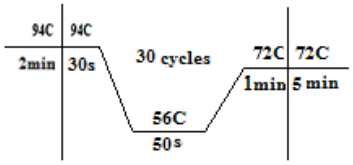
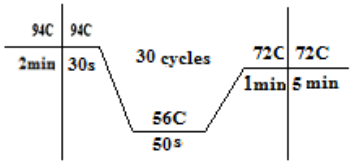



**Figure 2.4 Diagrammatic representation of fragmented gene amplification for all three genes PfSERA5, PvSERA4, PvSERA5.**

### PCR conditions and primer details

Specific customized primers were designed for PCR amplification of SERA genes by using primer design software available online. Designing primers is critical for PCR amplification procedure. Using in house Bioinformatics tools and Gene-runner software the primers were designed for SERA amplification which are given below for PfSERA5 gene, PvSERA4 and PvSERA5 (**Table 2.3**).

**Table 2.3 PCR conditions and primer details for PfSERA5, PvSERA4 and PvSERA5 genes**

| Gene Name and Domain region | Primer Name | Primer Sequence | Thermocyclic profile |
|---|---|---|---|
| PfSERA5 gene is AT rich-72.8% Amplification of P47 region encoded by Exon-II. | Forward primer: CNRSERA5P F1A | 5'GCGGTCGACATGAAGT CATATATTTCCTTG3' |   |
| | Reverse primer: CNRSERA5P F2 | 5'GCGACGCGTTACTAAA AGAGCACATTGAAA3' | |
| Amplification of P50 and P18 encoded by Exon IV. | Forward primer: CNRSERA5P F3: | 5'GCGGTCGACTGCTTTCA ATGTGCTCTTTTAG 3' |   |
| | Reverse primer: CNRSERA5P F4 : | 5'GCGACGCGTTGGAGAG TTATGCCCTATTT 3' | |
| PvSERA4 C-terminal fragment & Central fragment encoding protease | Forward primer: CNRSERA4P V3 | 5'GCGGGGATCCATCTTTG CCAACCTAACGA3' |   |
| | Reverse primer: CNRSERA4P V4 | 5'GCGAAGCTTCTCTCTCC TACGGCAATGG 3' | |
| N-terminal fragment | Forward primer: SERA4PV1 | 5'GCGGGATCCAGCTACT TTTAATTCTACAG3' |   |
| | Reverse primer: 4SERAPV9R | 5'AATGTAATCAGCAAAA TGGGA3' | |
| PvSERA5 C-terminal and Central regions | Forward primer: CNRSERA5P V6: | 5'GCGCGGGAAGAAGGTG CAAAG3' |   |
| | Reverse primer: CNRSERA5P V4 | 5'GCGCCCGTCACACTCTT CCTAC3'. | |
| N-terminal region | Forward primer: SERA5PV1 | 5'GCGCCTAGGATGAAGT CTCCTTCCC3' | |
| | Reverse primer: 5SERAPV10R | 5'GATTCTCTTAGCGGACA ACT3' | |

Representative gel image of amplification achieved for SERA genes of *P.falciparum* PfSERA5 and *P.vivax* PvSERA4 and PvSERA5 in the following **Figures 2.5, Figure 2.6** and **Figure 2.7.**

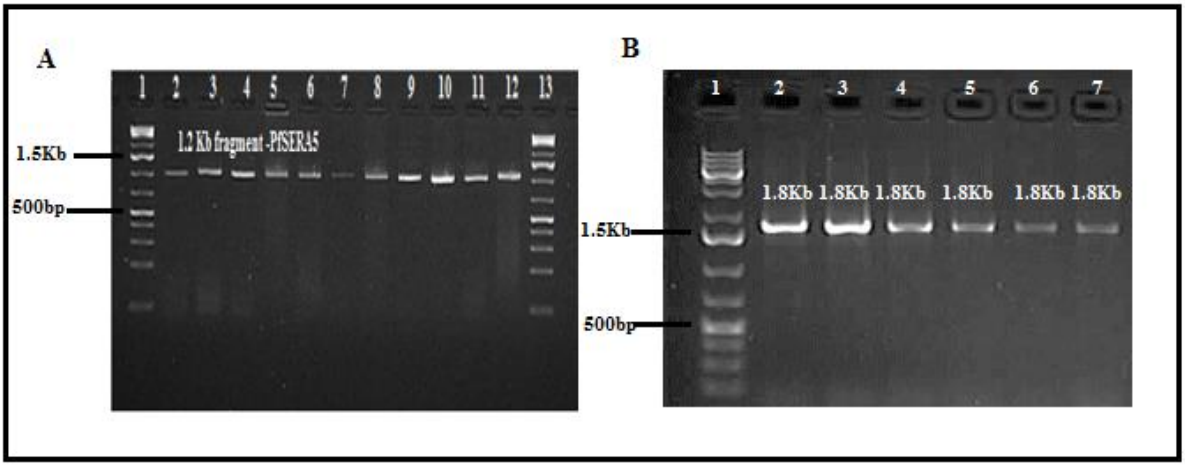



**Figure 2.5 Gene amplification of PfSERA5 A) 1.2 Kb N-terminal fragment encompassing P47 domain B) 1.8Kb Central (P56) and C-terminal region (P18). A) Lane number 1 and 13 are loaded with gene ruler 1Kb plus DNA ladder from Thermo Fisher Scientific # Cat.No- SM1333. B) Lane 1 is loaded with DNA ladder 1Kb plus from ThermoFisher Scientific # Cat.No- SM1333.**

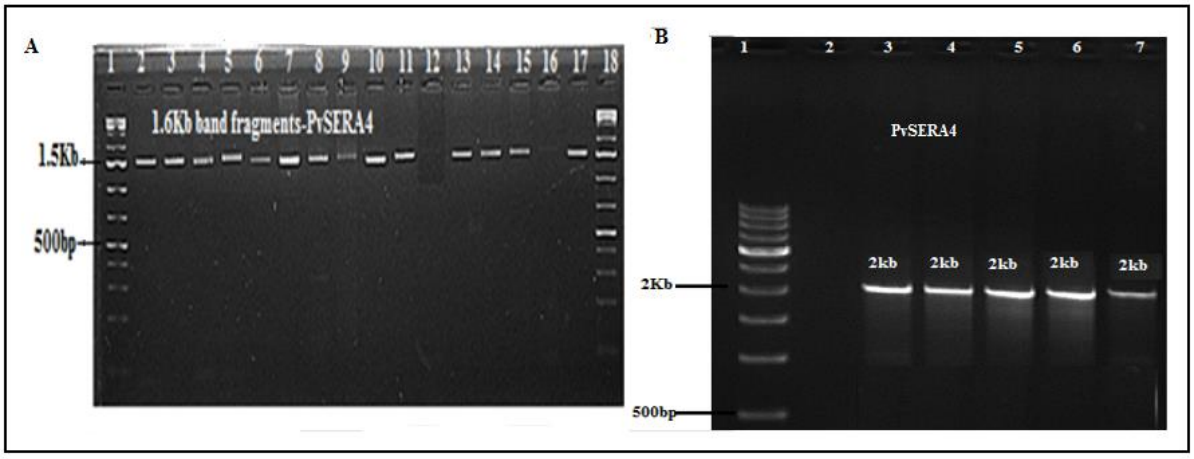



**Figure 2.6 Gene amplification of PvSERA4 A) N-terminal fragment length of 1.6 Kb and B) Central and C-terminal fragments of 2Kb. A) Lane number 1 and 18 show loaded DNA ladder gene ruler 1Kb plus DNA ladder from Thermo Fisher Scientific # Cat.No- SM1333. B) Lane 1 shows DNA ladder.**

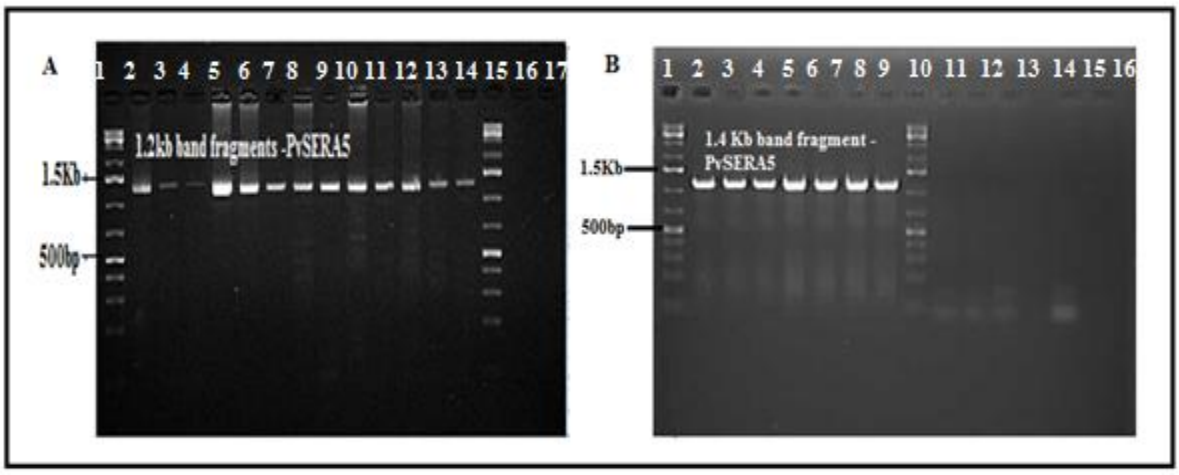



**Figure 2.7 Gene amplification of PvSERA5 A) N-terminal region with fragment length of 1.2 Kb and B) Central and C-terminal region with fragment length of 1.4 Kb. A) Lane number 1 and 15 show loaded DNA ladder from Thermo Fisher Scientific # Cat.No- SM1333.. B) Lane 1 and 10 show loaded DNA ladder from Thermo Fisher Scientific # Cat.No- SM1333.**

### DNA Sequencing

The gel-purified Amplicons were sequenced by commercial sequencing at:

1) Bioserve Pvt. Ltd. Hyderabad, India.
2) ABI Life Technologies-Invitrogen, Gurgaon, India.

About 50–100 ng of the purified DNA was used for sequencing PCR using BigDye® Terminator v3.1 Cycle Sequencing Kits [Applied Biosystems]. Cycling conditions for sequencing PCR were as follows: denaturation at 96 °C for 10 s, annealing at 55 °C for 10 s and extension at 60 °C for 4 min. After 35 cycles, the templates were purified by Ethanol/EDTA precipitation method and sequenced on ABI 3730xls Genetic Analyzer (Applied Bio systems)**.**

Sequences were read multiple times per sample using both the forward and reverse primers. The polyclonality of infection was assessed based on the peaks observed in the chromatograms. Only chromatograms lacking overlapping peaks were judged to be of single infection type and were chosen for further analysis and results. The remaining samples showed overlapping peaks and were non conclusive for single infection and were hence,

eliminated. This was achieved by designing sequencing primers for thorough DNA sequencing further selected as mentioned in the below **Figure 2.8**

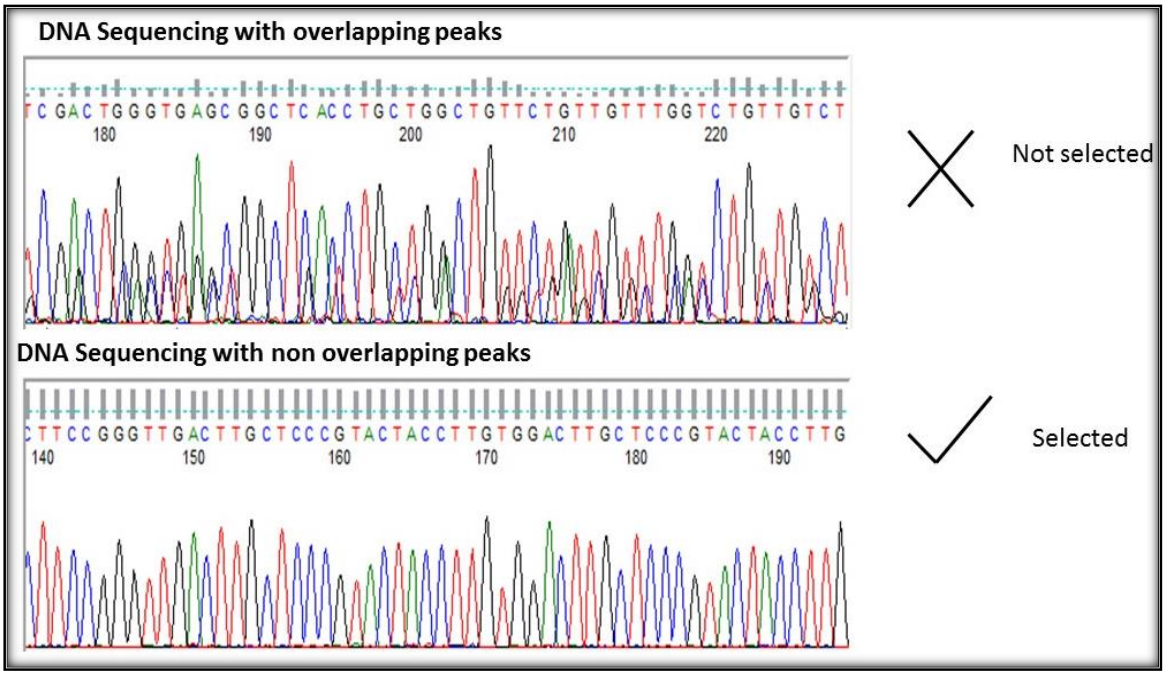



**Figure 2.8 Selection of DNA sequence chromatogram with non-overlapping peaks.**

DNA sequencing results was viewed using available online software-Finch TV version 1.4.0 developed by Geospiza.Inc.

Multiple Sequence Alignment (MSA) was done to represent sequence results using **Clustal W** program available in **BioEdit** software a widely used bio-tool for handling DNA sequences (Hall, 1999). This resulted in identification of mutations by comparing each set of sequences for an individual isolate with that of reference sequence. The reporting of genetic polymorphism not only helps in distinguishing those distinct and unique to particular geographic location but also aids in determining the extent of genetic diversity in SERA genes.

**Statistical and Computational tools used for analysis of sequence results**

Single Nucleotide Polymorphisms (SNP's) are one of the usual form of genetic polymorphism expected in malarial proteins. SNPs in the coding region are of two types, **Synonymous** and **Nonsynonymous** SNPs. In this regard we have employed several measures of genetic diversity in our study (Gunasekera et al., 2007). These are represented in the following **Table 2.4**

**Table 2.4 Statistical analysis of DNA polymorphism**

| Statistical parameter | Definition | Significance | Reference |
|---|---|---|---|
| **Nucleotide diversity ($\pi$)** | An estimate of average number of substitutions between any two sequences. Assuming the sample is random (all sites-synonymous and non-synonomous sites were considered in this calculation and its standard deviation. | Useful parameter to compare level of genetic diversity across nucleotide distance. | (Gunasekera et al., 2007) |
| **The number of segregating sites (S)** | The number of polymorphic sites that are dependent on sample size and length of nucleotide sequence | | |
| **H (number of haplotypes**) | The term refers to set of SNPs on a single chromatid of a chromosome pair that is statistically associated. | | |
| **dN, dS, dN/dS** | **dN**- Number of nonsynonymous mutations per nonsynonymous sites using the Nei-gojobori method with the jukes-cantor correction and standard error.<br>**dS**- Number of synonymous mutations per synonymous sites using the Nei-gojobori method with the jukes-cantor correction and standard error. | **dN/dS**- measure is widely used as indicator of natural selection.<br>**dN/dS>1**- suggest positive selection.<br>**dN/dS<1**-suggest purifying or negative selection | (Mugal, Wolf, & Kaj, 2014) |
| **Tajima's D statistics** | Test to detect departure from neutrality and is based on comparison of estimates of number of segregating sites (S) and mean pair wise differences between sequences ($\pi$) | Recommended to be better statistics particularly when recombination levels are unknown. | |

| | | | |
|---|---|---|---|
| **Recombination analysis and linkage disequilibrium** | Minimum number of recombination events occurring along a sequence is estimated using recombination parameter C In population genetics, linkage disequilibrium (LD) is the non-random association of alleles at two or more loci that descend from single, ancestral chromosome. D' and $R^2$ are indices for LD | Gives insight in to level of recombination as LD values are strongly affected by recombination, since it will break down existing haplotypes and generate new ones thus decreasing LD | (Ramírez-Soriano, Ramos-Onsins, Rozas, Calafell, & Navarro, 2008) |

All the parameters discussed above were analysed using the software tool **DnaSP-Version 5** (Librado & Rozas, 2009)**.**

Phylogenetic tree analysis helps in representing the relationships of haplotypes particular to a geographic region with those of other regions resulting in clustering of alleles and patterns of haplotype distribution, leading to identification of distinct and common haplotypes across various geographic regions. The **Neigbhour–joining** method was used to generate phylogenetic tree using the bio-software tool **MEGA-Ver5.0** (Tamura et al., 2011)**.**

**Computational tools used for structural Analysis**

The structure of the central protease like domain (**P50**) of highly expressed PfSERA5 gene with detailed structural characterization has been carried out in *P.falciparum*. However, virtually no information is available about the structural characterization of SERA genes from *P.vivax*. With an aim to achieve structural characterization of PvSERA, and to elaborate our understanding on unknown functional characteristics of PfSERA5, computational tools were used for structure prediction, modelling and molecular dynamics experiments. The refined structural models obtained would be useful for drug targeting and to design interventions against the blood stage of malaria parasite.

**Structure prediction and modelling tools**

a) **Ab-initio model using I-TASSER online server -** I-TASSER is Iterative Threading Assembly Refinement which is an advanced computer algorithm for protein structure and function prediction. It predicts the spatial conformation of a protein from its primary structure i.e from its sequence of amino acids. This online available computational methods relies on modelling all energetics involved in protein folding and finding the structure with lowest free energy. The generated I-TASSER models are assessed for quality and accuracy based on C-score. C-score is a confidence score that estimates the quality of predicted models by I-TASSER. (Xu & Zhang, 2011). Hence, this was one of the important tools used in our study.

b) **Secondary structure prediction using NPS@ (Network Protein Sequence @nalysis) server -** Secondary structure prediction is an additional useful step that helps in validation of the final 3D structure. Predictions of all 3 state predictions, i.e. helices, strands and coils were done based on consensus from a collection of prediction methods available in this server (Combet, Blanchet, Geourjon, & Deléage, 2000)**. PSIPRED** structure prediction method was also used for reiterating secondary structure prediction (McGuffin, Bryson, & Jones, 2000).

c) **Refinement using ModRefiner and Modloop online servers-** The obtained ITASSER models was subjected to refinement using ModRefiner and Modloop online servers for atomic-level, high-resolution protein structures and for their loop refinements respectively. Modloop relies on the loop modeling routine in MODELLER that predicts the loop conformations by satisfaction of spatial restraints (A Fiser, Do, & Sali, 2000; Andras Fiser & Sali, 2003).

d) **Stereochemical quality and validation of 3D Models using The Structure Analysis and Verification Servers**

- **PROCHECK** (Laskowski, MacArthur, Moss, & Thornton, 1993) checks the stereochemical quality of a protein structure by analysing residue-by-residue geometry and overall structure geometry.
- **ERRAT** (Colovos & Yeates, 1993) analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a 9-residue sliding window, calculated by a comparison with statistics from highly refined structures.

e) **DALI** (Holm & Rosenström, 2010) network service, which validates the structure by giving appropriate matches after comparison with 3D protein structures in protein data bank (PDB).

f) **ProSA web**-ProSA is a tool widely used to check 3D models of protein structures for potential errors. Its range of application includes error recognition in experimentally determined structures, theoretical models and protein engineering (Wiederstein & Sippl, 2007).

g) **Q-site finder** is a tool used to identify the location of binding sites. It uses interaction between binding sites and a simple vander wall probe to locate energetically favourable binding sites (Laurie & Jackson, 2005).

h) **RAMPAGE -** Online web tool for Ramchandran plot assessment developed by Cambridge university (Lovell et al., 2003).

i) **Comparative Modelling of protein structures: Homology modelling** (Baker & Sali, 2001) is used to achieve structural modelling of protein sequences whose structure is not known using templates with whom there is sequence identity and whose empirical 3D structures are available. This was achieved using bio-software Modeller-Ver 9.0

(Sali & Blundell, 1993) for homology modelling. In Modeller the following steps of homology modelling are:

1. Retrieving the sequence
2. Template selection
3. Sequence alignment
4. Building the model
5. Evaluation of the model.

**Molecular Dynamic (MD) simulation:** A computational method used to evaluate behaviour of molecular system which is time dependent and gives detailed information on the fluctuations and conformational changes of protein. It involves use of a computer programme which does mainly energy minimization and dynamic simulation of protein in solution. This approach was used to achieve a) Refinement of modelled protein structure and b) To gain insight in to functional aspects of protein structure based on evaluation of its stability and conformational changes achieved during simulation.

MD simulation of the protein structure was done using the following steps:

a) The GROMACS 4.5.4 package was used to perform Molecular Dynamics simulations (Hess, Kutzner, van der Spoel, & Lindahl, 2008).
b) Appropriate choice of force field GROMOS 96 53 a6 was done and a standard SPC water model was chosen (Cino, Wong-ekkabut, Karttunen, & Choy, 2011).
c) Energy minimization step was done to discard high energy inter molecular interactions.
d) The overall quality and atomic charges were optimized to avoid steric clashes.
e) Whole system was simulated at 300 K over 100ps using NVT ensemble using modified berendsen thermostat (Berendsen, Postma, van Gunsteren, & Hermans, 1981).

f) Subsequently at NPT using Parrinello–Rahman pressure coupling.

g) Finally, a 10 ns MD simulations was carried out with time step of 2 ps.

h) The trajectories were visualized using PYMOL(DeLano, 2002)**.**

i) Results were analysed using RMSD (root mean square deviation) and RMSF (root mean square fluctuation) plots using R-software.

j) Clustering of trajectories was done, the average ensemble model of the largest cluster was further energy minimized to achieve final MD refined 3D model (Nurisso, Daina, & Walker, 2012).

These methods selected and executed, together ensured comprehensive and detailed analysis with utmost precision and accuracy.

# CHAPTER 3

# CHARACTERIZATION OF GENETIC POLYMORPHISM OF HIGHLY EXPRESSED PvSERA GENES OF *P.vivax* FROM INDIAN ISOLATES

# Characterization of genetic polymorphism of highly expressed PvSERA genes of *P.vivax* from Indian isolates

## Background

The development of a malaria vaccine is greatly hindered by extensive diversity in antigen coding genes of malaria parasite. Clinical immunity against malaria happens only after years of repeated exposure and this is mainly contributed by diversity which exists usually in the malaria antigens. Hence, it is important to understand the extent and dynamics of genetic diversity in vaccine antigens to guide a rational vaccine design and in interpreting vaccine efficacy trials in malaria endemic regions (Takala & Plowe, 2009).

*P.vivax* is the most widely distributed human malaria parasite worldwide, hence vaccine development against the infection caused by it, is of high priority. There is also a growing concern in *P.vivax* malaria, about failure of treatment with first-line *P. vivax* antimalarial agents, such as chloroquine as shown by a number of clinical studies. Hence, studies on genetic diversity in *P.vivax* are important. Such an evaluation would lead to understanding of their significant influence on gene flow and rate of new mutations leading to either drug resistance or escape from vaccine induced responses. Genetic diversity studies on several candidate antigen genes of *P.vivax* are limited as research on *P.vivax* has lagged disproportionately in comparison to *P. falciparum*. Genetic diversity studies are also important to seek information on development and deployment of control strategies against *P.vivax* malaria (de Souza-Neiras, de Melo, & Machado, 2007).

Vaccine targets against *P.vivax* malaria are currently very limited in number and search for new candidate genes is highlighted in several recent reviews (Barry & Arnott, 2014; WHO, 2015). Studies on *P. falciparum* SERA indicate that, highly expressed SERA

genes are important in erythrocytic development of the parasite. Among the *P. falciparum* SERA multi-gene family, the highly expressed PfSERA5 is the most well characterized member and is implicated to play an important role in the blood stage (Miller et al., 2002). On similar lines, in PvSERA multigene family of 12 homologues, PvSERA4 and PvSERA5 are the highly expressed members in the blood stage and are proposed to play similar important roles as understood in *P.falciparum* (Palacpac et al., 2006).

The locus comprising SERA related genes were primarily identified in *P.vivax* by molecular cloning and sequence analysis. This locus revealed a ~25Kb genomic DNA of *P.vivax* bearing a cluster of five repeated SERA- like genes VSERA1 to VSERA5 (Kiefer et al., 1996). However, later studies by advanced *in silico* approach revealed six more PvSERA homologues genes downstream to earlier identified five VSERA genes (Palacpac et al., 2006). Additionally, studies show that the exon 4 of PvSERA homologues encodes a stretch of 200 amino acids near the C-terminal region within Sal-I strain which was found to be highly divergent (Kiefer et al., 1996). The sequence differences observed at the terminal regions (N-terminal and C-terminal) across these PvSERA genes necessitate investigation of antigenic diversity of these regions in field isolates also, as they would help in identifying regions under selection pressure which would impact their role as vaccine candidates or as a biomarker for drug resistance (Ferreira, Nunes, & Wunderlich, 2004).

The only reports on field isolates of highly expressed PvSERA genes are from Thailand where the antigenic diversity of C-terminal was compared with Sal-I reference strain (Palacpac et al., 2006). Also, sero-immunological studies conducted in South Eastern Turkey where malaria is exclusively caused by *P.vivax* indicate age dependent antibody responses against N-terminal of PvSERA4 establishing its immunological implication (Yildiz Zeyrek et al., 2011). PvSERA4 could be a potential vaccine candidate antigen similar to *P. falciparum* PfSERA5 which is shown to be refractory to genetic deletion and

important for growth of parasite in erythrocytic stage supporting further investigation on PvSERA4 (McCoubrie et al., 2007). All these evidences encourage studies on PvSERA genes, as very limited information is available with respect to their genetic and functional characterization.

The malaria burden and endemicity caused by *P.vivax* is very high in India as shown in **Figure 3.1**. Most of the states in India have *P.vivax* incidence *of* more than 60% and some states like Tamil Nadu, Karnataka and Uttar Pradesh show more than 90% incidence (Joshi et al., 2008). Keeping in view that no genetic diversity reports are available from any part of the world, our first representation on genetic diversity characterization of PvSERA genes from malaria endemic country like India gains significance. Our study on these highly expressed genes would aid to further establish the vaccine and drug potential of these antigens.

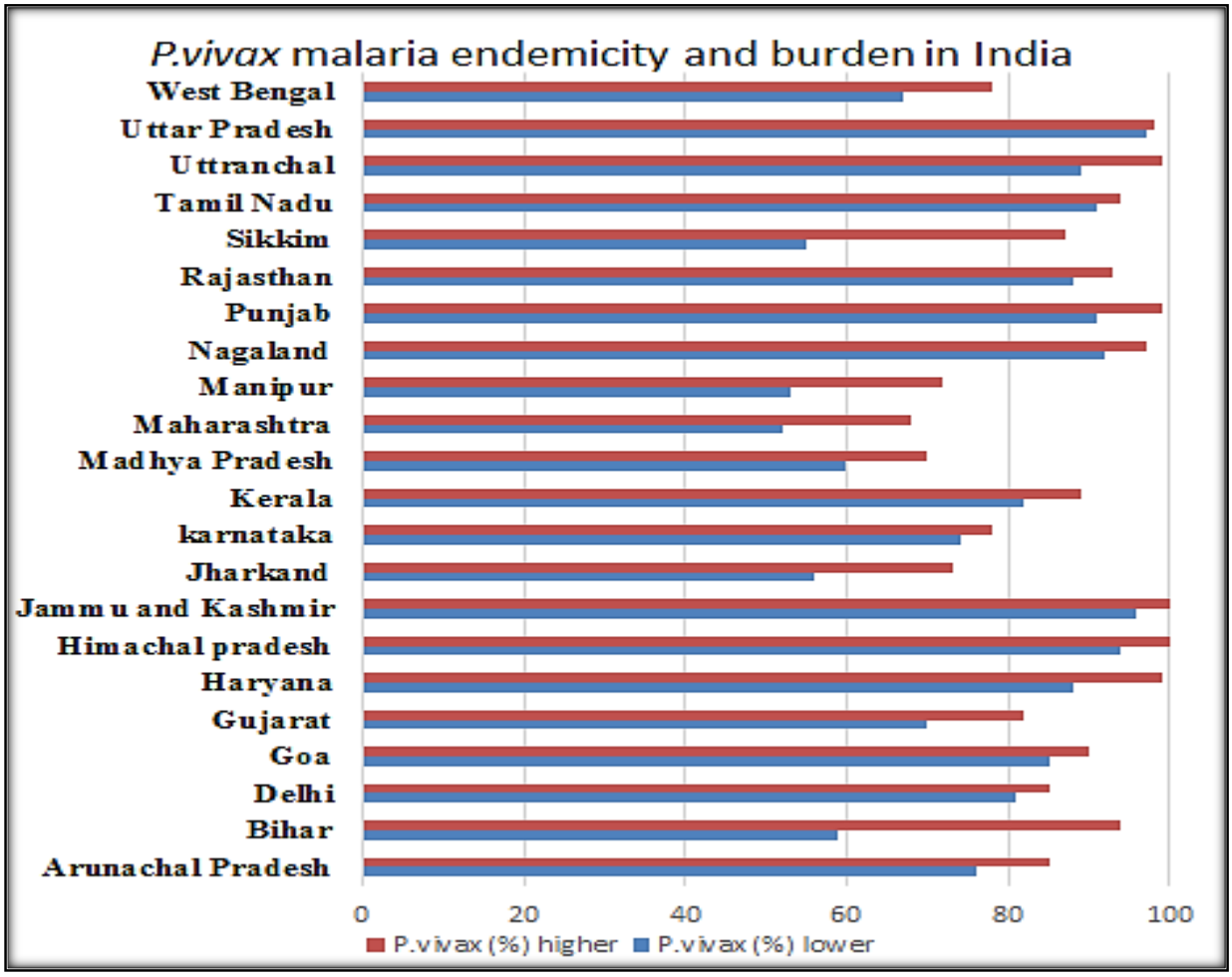



**Figure 3.1 Trend of predominance of *P.vivax* cases in most of the states from India (Joshi et al., 2008).**

Hence, the specific objective of this study is to identify and report the extent of genetic polymorphism in the highly expressed PvSERA4 and PvSERA5 genes of asexual blood stage of *P.vivax* malaria from Indian field isolates. The genetic polymorphism of the exons encoding the N and C-terminal regions of both these genes along with the central region were evaluated to completely characterize the full length genes of the highly expressed members for their sequence variations.

## Results and Discussion

### *A) Genetic polymorphism in the PvSERA4 gene*

Sequence results obtained were analysed by multiple sequence alignment (MSA) done at both nucleotide and protein level using Bio-Edit software (Hall, 1999). Sal-I strain reference sequence was used as standard for comparison and to report sequence results (**Table 3.1**).

**Table 3.1 Genetic regions considered in the study of PvSERA4 gene**

| **Genetic regions representing sequence results considered in this study for PvSERA4 gene** | | **Number of field isolates sequences represented** | **Compared with Standard strain** |
|---|---|---|---|
| C-terminal region | 636 bp (corresponds to amino acid residues 1021–1231aa). | 18 | Gene sequence PvSERA4 and accession number of AAB41486 reported for Sal-I strain from GenBank (Kiefer etal., 1995) |
| Central region | 651bp (corresponds to amino acid residues 684-901 aa). | 22 | |
| N-terminal region | 1,113 base pairs (corresponds to 371 amino acid residues). | 13 | |

### 1) *Genetic polymorphism in the C-terminal region of PvSERA4 gene*

The sequencing results represented in this study includes 18 sequences of C-terminal region PvSERA4. Based on the sequence analysis it is understood that the PvSERA4 C-terminal is completely conserved in Indian field isolates (**Figure 3.2**). Almost, all the amino acid positions show absolute conservation in-comparison to Sal-I reference standard and also between the isolates of two endemic regions. As the sequence results of PvSERA4 from Indian isolates indicated high genetic conservation, further statistical evaluation of its genetic polymorphism were shown to be insignificant.

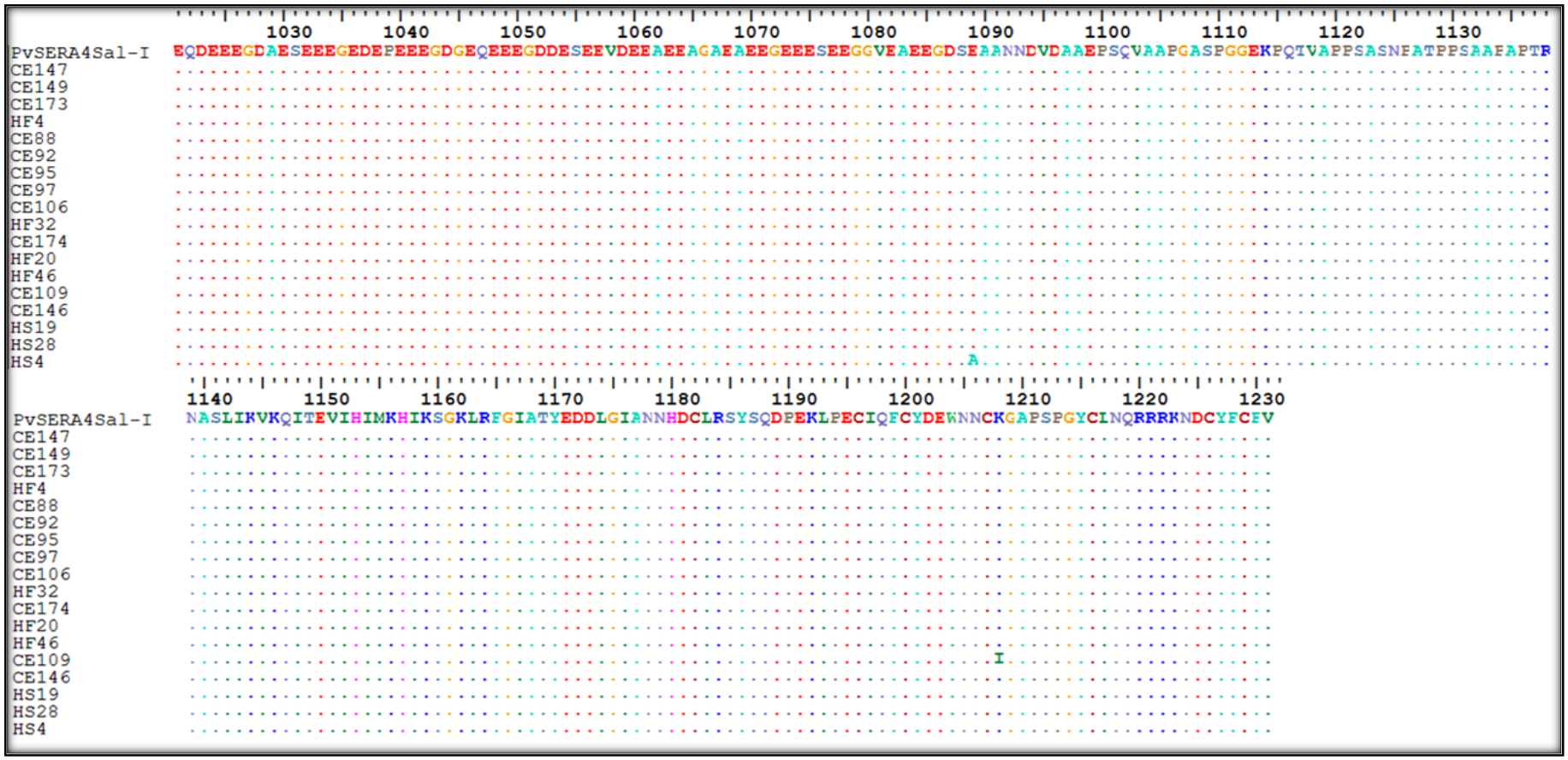



**Figure 3.2 Multiple sequence alignment (MSA) of translated protein sequences of 18 Indian isolates against Sal-I strain in the C-terminal region of the PvSERA4 gene. The MSA was generated using the ClustalW application in Bio Edit software highlighting colour code for each amino acid against Sal-I strain with dots '.' Implying sequence conservation.**

The complete genetic conservation of PvSERA4 sequences in the C-terminal region is shown to be a major observation in this study. This implicates this region could be under functional constraints and potentially important for *P.vivax* parasitic life cycle analogous to similar region in PfSERA enabling it to be a good target for intervention (Aoki et al., 2002; Arisue et al., 2011). However, further results which aid in understanding functional significance of PvSERA4 is required.

**2) *Genetic polymorphism in the central region of PvSERA4 gene***

The growing concern for emerging multidrug-resistant *P. vivax* strains demands the identification of potential new chemotherapeutics to target the blood stage of the parasite (Baird, 2004). Highly expressed PvSERAs are conjectured to be reliable as new selective inhibitor based antimalarial therapeutics for *P.vivax*. The central protease region of PvSERA homologues participates as key molecule in the molecular and biochemical pathway of egress of parasites from blood stage and shares considerable sequence identity and homology among themselves, serving as important drug targets of intervention in the blood stage (McCoubrie et al., 2007; Miller et al., 2002). Sequence results of central region of PvSERA4 for 22 field isolates from two different endemic regions of India were compared with the standard *P. vivax* Sal- I strain sequence available in gene bank AAB41486 (Kiefer et al., 1996). Sequence analysis revealed that PvSERA4 protease domain shows high level of conservation with the standard strain as well as between the isolates of the two endemic regions studied (**Figure 3.3**).

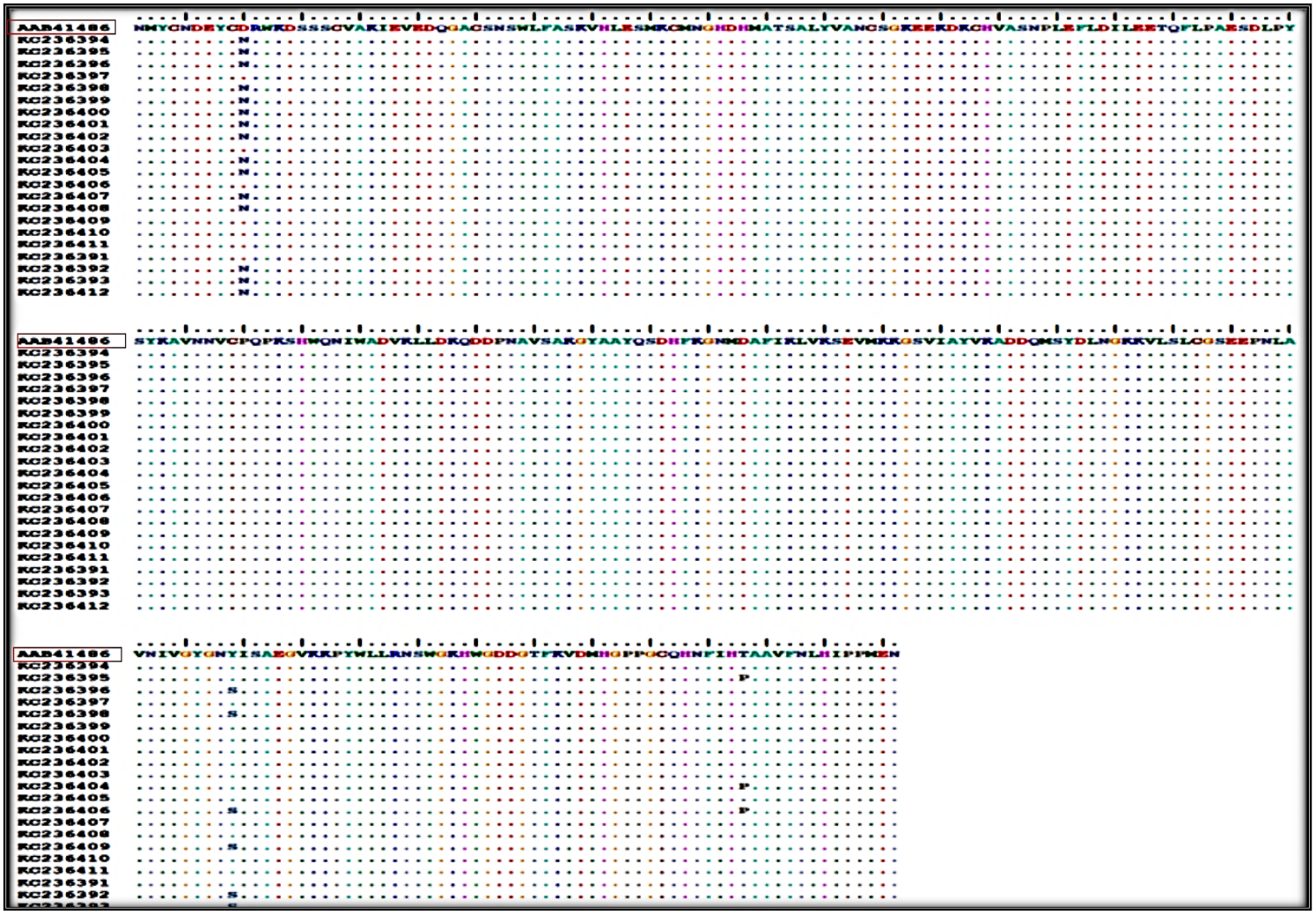


**Figure 3.3 Sequence polymorphism of PvSERA4 protease domain in *P. vivax* Indian isolates. The figure above represents sequence alignment of protease domain of PvSERA4 from 22 Indian field isolates compared against standard sequence from Salvador strain (Sal-I) (Accession No-AAB41486) highlighted in red box.**

Sequence analysis showed that only SNPs were observed in isolates at 3 sites with remaining nucleotide sites showing 99% identity with Salvador-I (AAB41486). Statistical analysis of the genetic polymorphism by computing Tajimas D gives a significant negative value less than zero in most of the nucleotide positions which support the hypothesis of purifying or negative selection operating in PvSERA4 protease domain (**Figure 3.4**).

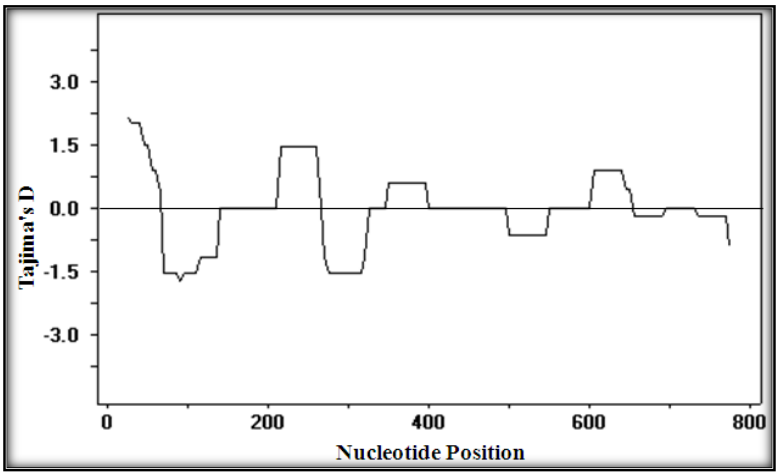



**Figure 3.4 Sliding window analysis of natural selection (Tajima's D values) for PvSERA4 protease domain. Window size= 50 bp, step size =5 bp.**

Also, codon-based Z-test of selection between sequences also revealed evidence for negative or purifying selection (p value less than 5% level of significance). This negative selection indicates that deleterious mutations cannot take over the population and such low genetic diversity identified in PvSERA4 domain reveal their vital role in parasitic life cycle and hence increasing their prospect of suitable drug targeting.

As this region is evident to be under functional constraint, its protein sequence are observed to be genetically conserved across PvSERA homologues in Sal-I Strain. However, the investigation of this genetic conservation in field isolates from various malaria endemic regions

would compound its use as further novel drug targets for *P.vivax* malaria. In this context our investigation of the genetic polymorphism status of highly expressed PvSERA genes pertaining to this central proteolytic region from Indian field isolates is very useful and first of its kind.

### 3) *Genetic polymorphism in the N-terminal region of PvSERA4 gene*

The sequence results represented in this study include 13 sequences in the N-terminal of PvSERA4. We observe presence of different combinations of tetramer repeats with glutamine (Q) as the first amino acid concentrated in the 5'end of N-terminal region, in our chosen standard sequence Sal-I (PVX_003845) of PvSERA4 itself. This repeat region is overall rich in glutamine (Q), Aspartic acid (D) and glutamic acid (E) and Threonine (T). Such repeats are unique to Sal-I strain and were not observed in other reported strains (Neafsey et al., 2012). All the PvSERA4 Indian isolates studied show sequence identity to Sal-I strain and have the same unique pattern of repeats as seen in N-terminal region of Sal-I PVSERA4. Phylogenetic analysis and percentage identity graph also confirm this identity with Sal-I (**Figure 3.5**). The significance of these repeats for *P. vivax* infection has not been studied till date and as far as our knowledge goes its contribution to the disease etiology remains obscure and needs to be further investigated. However, earlier studies on other transmembrane proteins with similar low complexity regions of tandem repeats have been shown to contain variable immunodominant epitopes providing clinical relevance (Brocchieri, 2001). These evidences implicate further exploration of PvSERA4 N-terminal region for any switch between antigenic phenotypes as this strategy of rapid diversification may aid in immune evasion (Reeder & Brown, 1996).

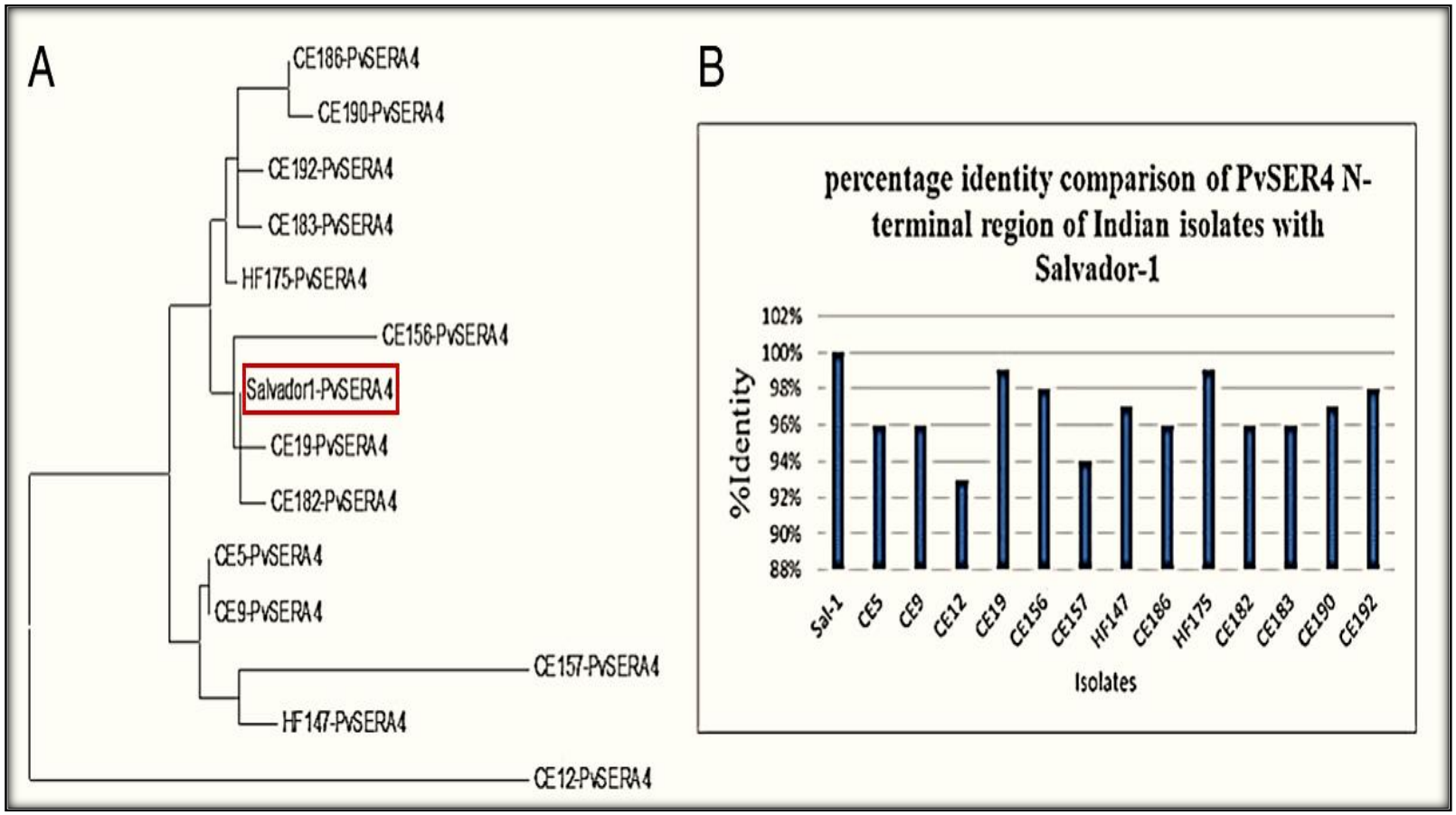



**Figure 3.5 Phylogeny and percentage identity graph for N-terminal region of PvSERA4 gene. A) Phylogenetic comparison of PvSERA4 sequences of Indian field isolates with Salvador-1 strain. B) Graph representing percentage identities of samples with Salvador-1 strain.**

The Indian isolates also show deletions and insertions in the repeat region leading to repeat number variations as depicted in **Figures 3.6** and **3.7**. In addition, nucleotide changes resulting in amino acid substitutions are also seen within these tetrameric repeats, which have been documented in **Table 3.2**. These changes result in 12 novel haplotypes found in this present study. These observations are very significant findings of our study and could provide better leads for further elucidation of the role of N-terminal PvSERA4.

However, in contrast to 5' end, the 3'end of N-terminal region shows high level of conservation as can be observed in **Figure 3.6**.

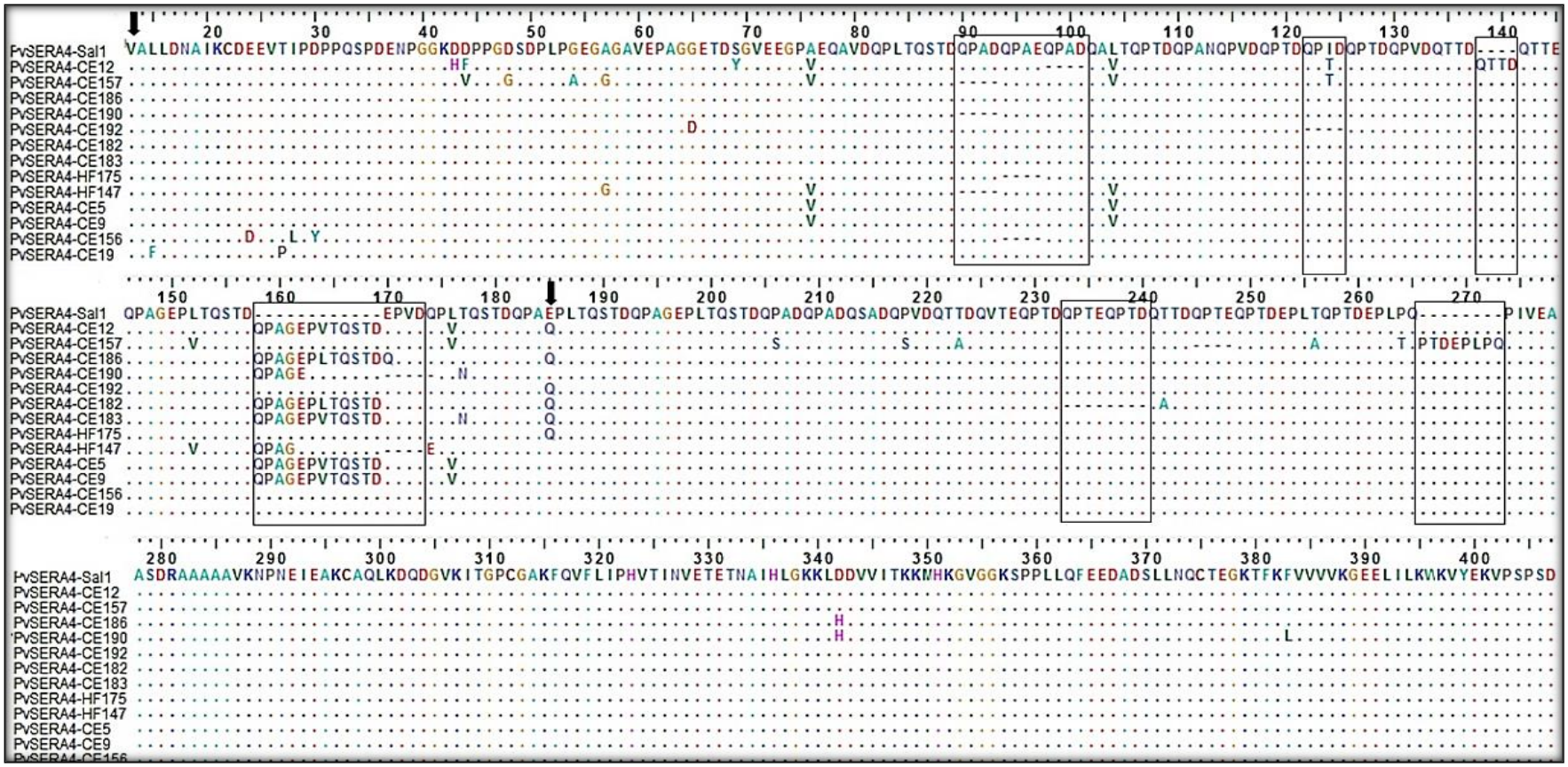
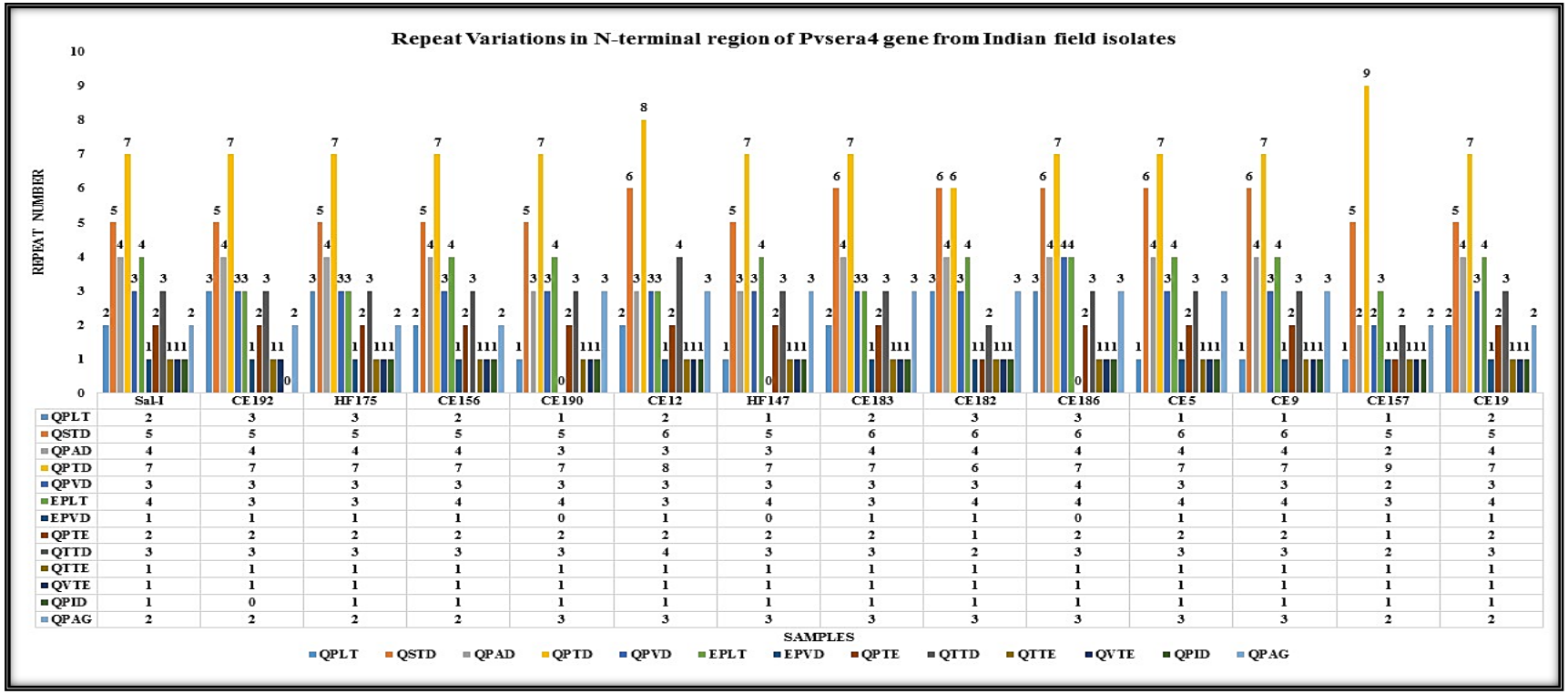



**Figure 3.6 Multiple sequence alignment of PvSERA-4 N-terminal region with '.' representing identity and '-' representing gaps. Arrows represent start codon sites encoded by Exon II and Exon III. Boxes represent deletions and insertions in Indian field isolates.**




| | Sal-I | CE192 | HF175 | CE156 | CE190 | CE12 | HF147 | CE183 | CE182 | CE186 | CE5 | CE9 | CE157 | CE19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| QPLT | 2 | 3 | 3 | 2 | 1 | 2 | 1 | 2 | 3 | 3 | 1 | 1 | 1 | 2 |
| QSTD | 5 | 5 | 5 | 5 | 5 | 6 | 5 | 6 | 6 | 6 | 6 | 6 | 5 | 5 |
| QPAD | 4 | 4 | 4 | 4 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 2 | 4 |
| QPTD | 7 | 7 | 7 | 7 | 7 | 8 | 7 | 7 | 6 | 7 | 7 | 7 | 9 | 7 |
| QPVD | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 3 | 2 | 3 |
| EPLT | 4 | 3 | 3 | 4 | 4 | 3 | 4 | 3 | 4 | 4 | 4 | 4 | 3 | 4 |
| EPVD | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| QPTE | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 |
| QTTD | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 3 | 2 | 3 | 3 | 3 | 2 | 3 |
| QTTE | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| QVTE | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| QPID | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| QPAG | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 |

**Figure 3.7 Graph for repeat number variations in PVSERA4 found in Indian field isolates. On the X axis is the sample ID of field isolates and Y axis includes number of repeats. The repeat types are mainly tetrameric repeats rich in Glutamine (Q) - QPLT, QSTD, QPAD, QPTD, QPVD, QPTE, QTTD, QTTE, QPID, QPAG with exceptions in repeat types EPLT and EPVD.**

**Table 3.2 Variations within N-terminal tetrameric repeats of PvSERA4**

| Repeat variant | Samples |
|---|---|
| QP(L/V)(T/N) | CE12, CE157,CE190,CE183 |
| EP(L/V/A)T/N | HF147, CE157 |
| (E/Q)PVD | CE186 |
| Q(T/A)TD | CE182 |

Statistical analysis of genetic polymorphism in the N-terminal region of PvSERA4 represent high genetic diversity from Indian field isolates which is clearly evident from the statistical values obtained for their haplotype diversity (HD)±SE (0.989±0.031) and nucleotide diversity values per site ($\pi$) for PvSERA4 (Pi =0.01003). Also, Tajima's D analysis indicates a positive score for PvSERA4. However, the calculated Tajima's scores were not significant for the present data set with ($P > 0.1$).

Our present study is the first report on genetic diversity of N-terminal regions of *P. vivax* SERA genes from field isolates. The N-terminal PvSERA4 diversity analysis from our study sheds light on the tetrameric repeats of Sal-I strain and the similarity and differences of these repeats as seen in Indian Isolates.

***B) Genetic polymorphism in the PvSERA5 gene***

The sequence results obtained for each genetic region considered for PvSERA5 were analysed using Bioedit software to obtain multiple sequence alignment by comparison with reference PvSERA5 sequence of Sal-I strain (**Table 3.3**).

**Table 3.3 Genetic regions considered in the study of PvSERA5 gene**

| **Genetic regions representing sequence results considered in this study for PvSERA5 gene** | | **Number of field isolates sequences represented** | **Compared with Standard strain** |
|---|---|---|---|
| C-terminal region | 687 bp (corresponds to amino acid residues 823–1051). | 18 | Gene sequence PvSERA5 and accession number of PVX_003830 reported for Sal-I strain from GenBank (Kiefer etal., 1995) |
| Central region | 774bp (corresponds to amino acid residue 517-775) | 18 | |
| N-terminal region | 711 base pairs (corresponds to 237 aa ) | 19 | |

**1)** ***Genetic polymorphism in the C-terminal region of PvSERA5***

The sequencing results represented here includes 18 sequences for C-terminal region of PvSERA5 gene which is represented as multiple sequence alignment (MSA) in the **Figure 3.8**. The C-terminal region of PvSERA5 sequences from Indian isolates upon sequence alignment with standard Sal-I sequence showed high level genetic polymorphism. They represent several Single Nucleotide Polymorphism (SNPs), most of them in the form of **parsimony informative sites**. A site is parsimony-informative if it contains at least two types of nucleotides (or amino acids), and at least two of them occur with a minimum frequency of two. Insertions and deletions also occur at specific locations distributed along the C-terminal region of the protein. Major deletion includes a 40 amino acid deletion (GLGAAAGPTGQEGVGVAAPGGRGTGGEAIAAAAVVVVGGG) observed between amino acid position 866 and 905 in 11 isolates (CE-6,7,25A,65,71,85,93,106,HS- 23,24,28) and partial deletions in six Indian isolates (CE- 23,25,50,67A,69,HS-25). The sample CE-88 isolate did not exhibit the deletion and was similar to the *P.vivax* Sal I (**Figure 3.8**).

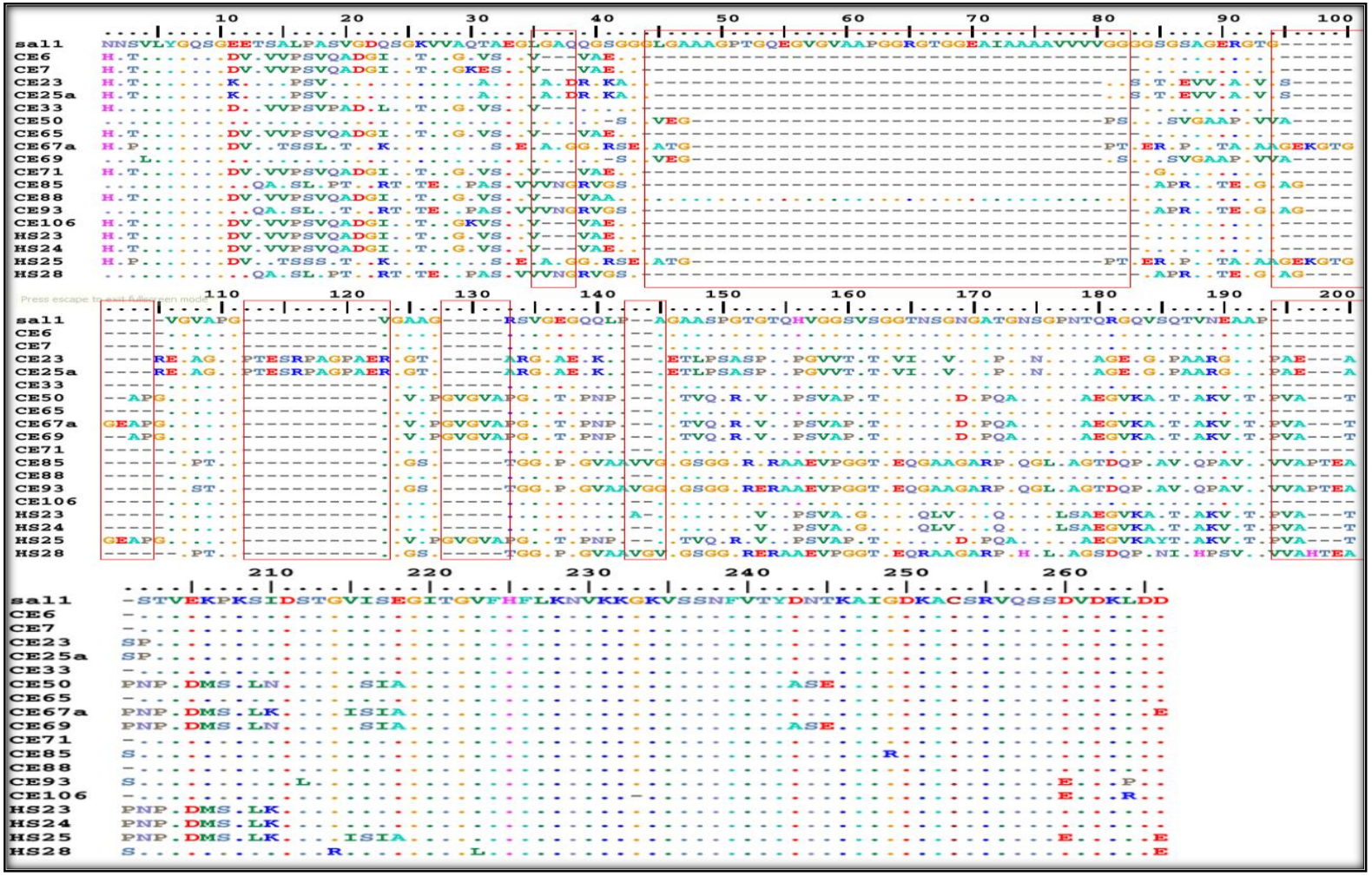



**Figure 3.8 Multiple sequence alignment in the C-terminal region of PvSERA5 from Indian isolates showing translated protein sequences for 18 Indian isolates against Sal I strain. Insertions and deletions are marked in red colour boxes.**

Similarly, unique insertions were also found in our Isolates. Two major insertions of 11 amino acids "GEKGTGGEAPG" and "PTESRPAGPAE'' were seen between amino acids 916-917 and 922-923 of PvSERA 5 protein sequence respectively. Few minor insertions of 1-5 amino acids were observed in these positions as well as at other locations of C-terminal region as represented in **Table 3.4**

**Table 3.4: Insertions observed in PvSERA5 sequences from Indian isolates**

| S NO | INSERTION CODON POSITION | AMINO ACID LENGTH | INSERTED AMINO ACID SEQ | SAMPLE ID |
|---|---|---|---|---|
| 1 | 916-917 | 11 | GEKGTGGEAPG | CE67A, HS25 |
| 2 | 916-917 | 4 | AAPG | CE50,CE69 |
| 3 | 916-917 | 2 | SR | CE23,CE25A |
| 4 | 916-917 | 2 | AG | CE85,CE93,HS28 |
| 5 | 922-923 | 11 | PTESRPAGPAE | CE23,CE25A |
| 6 | 927-928 | 5 | GVGVA | CE50,CE67A,CE69,HS-25 |
| 7 | 937-938 | 2 | VV | CE85 |
| 8 | 937-938 | 2 | VG | CE93,HS28 |
| 9 | 937-938 | 1 | A | HS23 |
| 10 | 986-987 | 8 | VVAPTEAS | CE85,CE93 |
| 11 | 986-987 | 8 | VVAHTEAS | HS28 |
| 12 | 986-987 | 5 | PVATP | CE50,CE67A,CE69,HS25,HS23,HS24 |
| 13 | 986-987 | 5 | PAEAS | CE23,CE25A |

Statistical parameters like Haplotype diversity (HD), average number of pairwise differences (K), and the recombination parameters (R and Rm) were computed using DnaSP Ver- 5 (Librado & Rozas, 2009; Tajima, 1993). Also, $\pi$ diversity was calculated on a slide window of 100 bases, with a step size of 25bp in order to estimate stepwise diversity across C-terminal region (Librado & Rozas, 2009). Test of selection operating in this C-terminal region was done by computing Tajimas D using DnaSP Ver-5 (Librado & Rozas, 2009) and fisher's exact test using MEGA-version 5.0 (Tamura et al., 2011).

At the nucleotide level, analysis revealed that 252 sites show polymorphic mutations. A total of 286 sites were monomorphic (invariant). Out of the 252 variants 19 were found to be

singleton and 233 sites were **parsimony informative sites**. These nucleotide variations led to 131 amino acid changes in the C-terminal region of PvSERA5. Of these 131 amino acid changes, 54 (41.22%) showed di-morphic changes, 50 (38.16%) were trimorphic, 20 (15.26%) tetra-morphic and 7(5.34%) penta-morphic.

***Haplotypes and nucleotide diversity analysis***

Among the 18 PvSERA5 sequences considered in this study, 15 PvSERA sequences were found to be novel haplotypes. The novelty of these haplotypes were further confirmed by using BLAST searches against all the PvSERA5 haplotypes earlier reported from Thailand (Palacpac et al., 2006).

Nucleotide diversity analysis was done by computing parameters available using Dnasp Ver-5 software (Tajima, 1993). Haplotype diversity (Hd) was computed and found to be high (0.977$\pm$ 0.227). The average pairwise nucleotide differences K within Indian isolates 92.69.

The $\pi$-diversity was also computed and overall diversity was found out to be 0.17229$\pm$0.022 SD. Also, estimation of $\pi$-diversity was done by sliding window method using step-size of 25bp which show high diversity between nucleotide position 400-450bp (**Figure 3.9 A**).

Data obtained for the above stated statistical parameters were also estimated for the earlier reported haplotypes from Thailand isolates and the comparative values were distinct as mentioned in the **Table 3.5**

*Test of Neutrality and recombination analysis*

**Tajima's D** test was computed using DnaSP- Ver 5.0 (Librado & Rozas, 2009). Positive Tajimas D value 0.7084 was obtained for the samples considered in this dataset. This indicated this region to be under positive selection. Also, **fisher's exact test** was conducted to reconfirm operating positive selection at 95% Confidence Interval.

**Table 3.5 Measures of DNA sequence polymorphism for C-terminal region of PvSERA5 among Indian and Thailand population Study**

| Study area | No of isolates | Singleton variable sites | Parsimony informative sites | Total number of Mutations | K | H | Hd±SD | π±SD | Tajima's D |
|---|---|---|---|---|---|---|---|---|---|
| INDIA | 18 | 19 | 233 | 322 | 92.69 | 15 | 0.977±0.027 | 0.17229±0.02257 | 0.70842 |
| THAILAND | 8 | 33 | 138 | 205 | 84.61111 | 8 | 0.972±0.064 | 0.14689±0.01442 | 1.78971 |

**K; Average number of pairwise nucleotide differences, H; Number of haplotypes, Hd; Haplotype diversity; π; Observed average pairwise nucleotide diversity, D; Tajima's D test statistics.**

Recombination analysis was done by computing minimum number of recombination events between adjacent polymorphic sites using DnaSP-Ver5 which was found to be 23. Effect of recombination on reported PvSERA alleles were shown graphically using Linkage Disequilibrium (LD) plotted against nucleotide distance. The decline in levels of LD ($R^2$ indices) with increasing nucleotide distance was observed (**Figure 3.9 - B**).

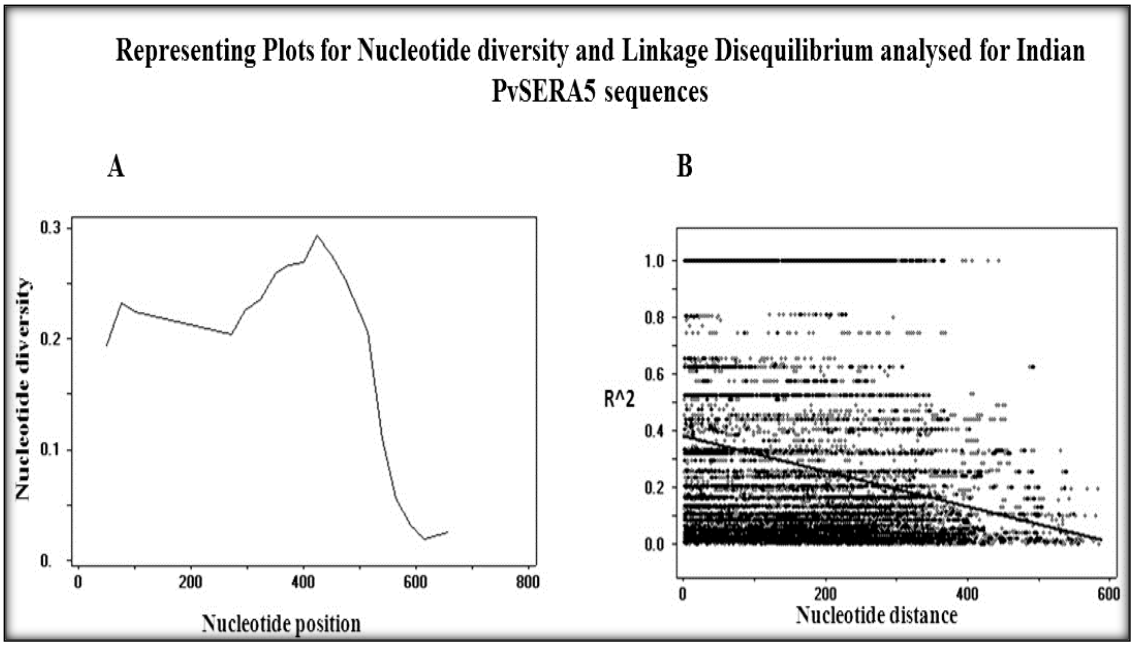



**Figure 3.9 Nucleotide diversity and Linkage Disequilibrium pot for PvSERA5 C-terminal region. A) Plot representing nucleotide diversity versus nucleotide position. B) The linkage disequilibrium (LD) plots showing non-random association between nucleotide variants at different polymorphic sites. The R2 values are plotted against the nucleotide distances with two-tailed Fisher's exact test of significance using DnaSP. The value of LD index (range from -1 to +1).**

## Phylogenetic Analysis

Phylogenetic analysis was done using neighbour joining method (Saitou & Nei, 1987) using MEGA-Version 5 by comparing Indian isolates across Sal-I PvSERA5 sequence and earlier reported Thailand sequences. The analysis indicate that the haplotypes from India are distinct from Thailand haplotypes (**Figure 3.10**). The generated genealogy indicate that the selection is operating on the haplotypes identified from India. This kind of selection is expected to be frequency based which is understood to be positive and mediated by immune selection.

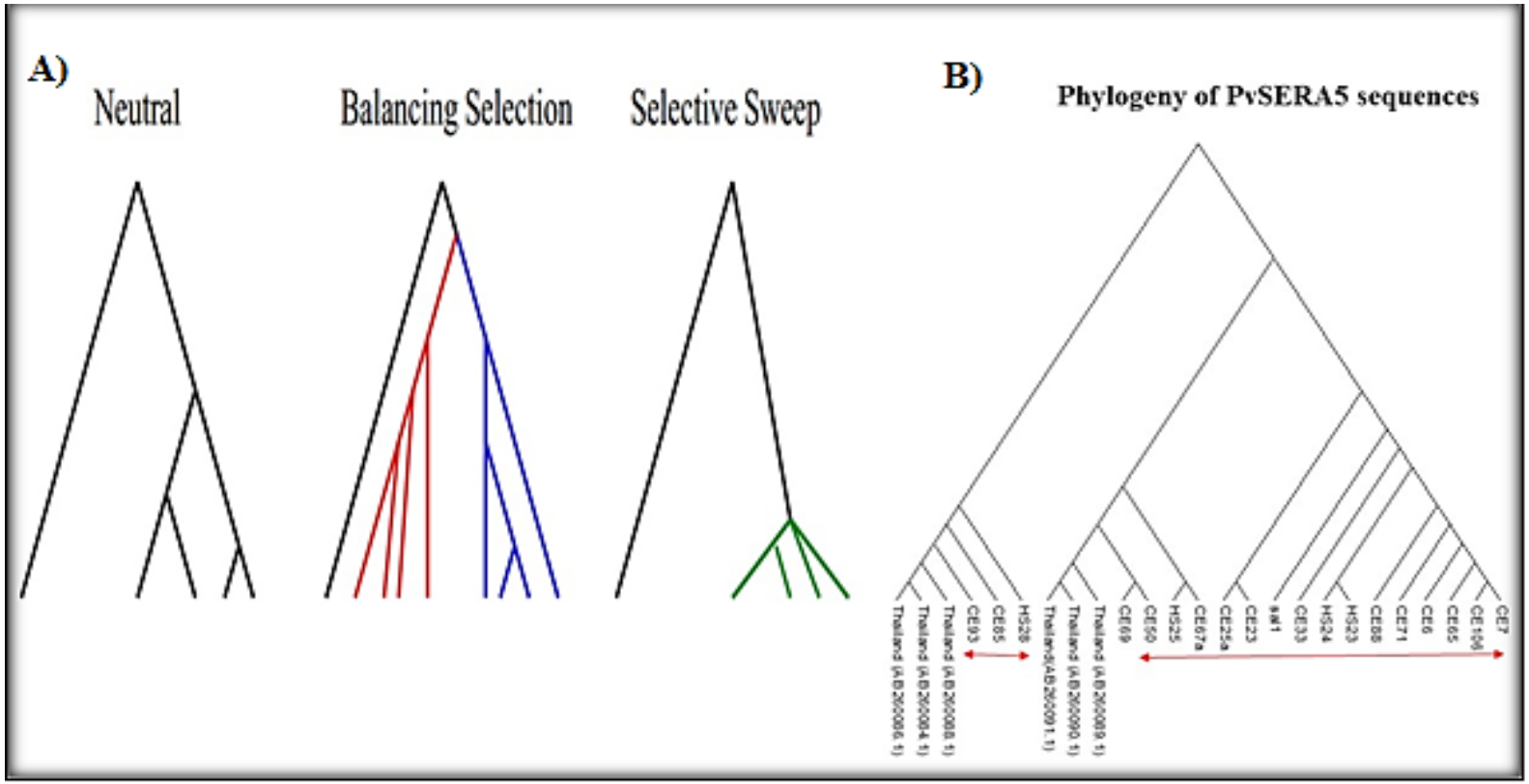



**Figure 3.10 Phylogeny for C-terminal region of PvSERA5 sequences from Indian isolates. A) The different types of phylogenies obtained under natural selection. B) Phylogenetic relationship between PvSERA5 Indian haplotypes (marked arrow Red) across Sal-I strain and earlier reported Thailand sequences.**

Identification of potential vaccine targets for *P.vivax* requires population genetic studies that can aid to identify signatures of balancing selection within surface antigen genes. This enables identification of domains targeted by strong host immune pressure and thus enabling identification of potential vaccine candidates. However, this requires detailed understanding of *P.vivax* genetic diversity, population structure and transmission dynamics (Arnott, Barry, & Reeder, 2012). In the direction contributing to this understanding, we here in this study aims to analyse the extent of genetic polymorphism and to estimate the potential selection mechanisms operating in PvSERA5 gene. This study is first of its kind in Indian population.

The critical analysis of PvSERA5 C-terminal region shows varying degree of deletions, insertions and SNPs which range from monomorphic to pentamorphic which are manifested in the form of novel haplotypes. The haplotypes reported in our study are distinct from earlier reported haplotypes from Thailand for this C-terminal region. This is evident from the phylogenetic studies. Also, analysis of basic population genetic parameters like HD, $\pi$-diversity and K indicate high genetic diversity existing in these PvSERA5 sequences from these endemic regions of Indian isolates. This variety of sequence polymorphism attained is understood to be due to antigenic diversity mechanisms like nucleotide replacement and recombination which create allelic diversity (Ferreira et al., 2004; Ferreira, Zilversmit, & Wunderlich, 2007).

Recombination analysis by computing Rm and LD plots support evidence for maintenance high allelic diversity (Hughes, 1992). The obtained results show high Rm values and strong decline in LD values across nucleotide positions which support high recombination rate as a possible reason for existence of diverse PvSERA5 alleles across Indian isolates.

Role of selection operating in the C-terminal region of PvSERA5 is expected to be statistically significant from Tajima's D and fisher's exact test of neutrality. Also, the pattern

of phylogeny obtained for PvSERA5 sequences from Indian isolates support positive selection operating in this region. This enables a balanced selection of multiple alleles maintained at particular frequency in the population based on the host immune pressure (Conway, 1997).

**2) *Genetic polymorphism in the central region of PvSERA5***

The central region of PvSERA5 gene which encodes a protease like domain which was observed to be genetically conserved similar to PvSERA4 central region in the 18 field isolates from India.

**3) *Genetic polymorphism in the N-terminal region of PvSERA5***

The N-terminal region of PvSERA5 from 19 Indian field isolates were analysed using Sal-I strain (PVX_003830) as standard which is represented as multiple sequence alignment (**Figure 3.11**) performed using Clustal W programme in the BioEdit software (Hall, 1999). Variations found in our field isolates were also compared with recently reported strains from http://www.broadinstitute.org/annotation/genome/plasmodium_vivax/MultiHome.html) - Board Institute *Plasmodium vivax* database to identify novel mutations (Neafsey et al., 2012).

Among the field isolates evaluated only 3 (HF14, CE178, CE182) showed similarity to Sal-I. They cluster with Sal-I with 98% identity. The remaining samples were distant from Sal-I with 70-75% identity. These remaining samples when compared with other reported strains, clustered more closely to with the Brazilian (PVBG_05106) and Mauritian (PVMG_02698) strains (**Figure 3.12**).


| Codon Site | AA change wrt to Sal-1 | Codon Site | AA change wrt to Sal-1 |
|---|---|---|---|
| 30 | Q-->R/P | 85 | S-->G |
| 49 | A-->G | 86 | G-->V |
| 54 | G-->R | 88 | S-->N |
| 56 | S-->P | 95 | V-->P |
| 60 | T-->P/S | 96 | V-->G |
| 62 | V-->N | 97 | P-->L |
| 64 | Q-->N | 100 | G-->S |
| 65 | S-->P | 104 | N-->G |
| 68 | Q-->P | 105 | P-->S |
| 70 | G-->R/P | 110 | S-->P |
| 71 | G-->S/V | 111 | S-->N |
| 72 | G-->E | 113 | G-->S |
| 73 | D-->G | 119 | G-->V |
| 77 | Q-->H | 120 | A-->V |
| 78 | G-->I | 124 | Q-->L |
| 79 | G-->V | 128 | Q-->P |
| 82 | D-->A/G | 129 | A-->T |
| 83 | R-->S | 133 | A-->V |
| 84 | A-->P | 134 | G-->A |
| | | 139 | Q-->E |

**Codon sites mentioned in the above box are variations unique to Indian field Isolates**

**Figure 3.11 Multiple Sequence Alignment file of PvSERA5 N-terminal region: Samples vs Strain comparison. The SNPs represented in the highlighted box are unique to sequences of Indian field isolates and are not observed in strains.**

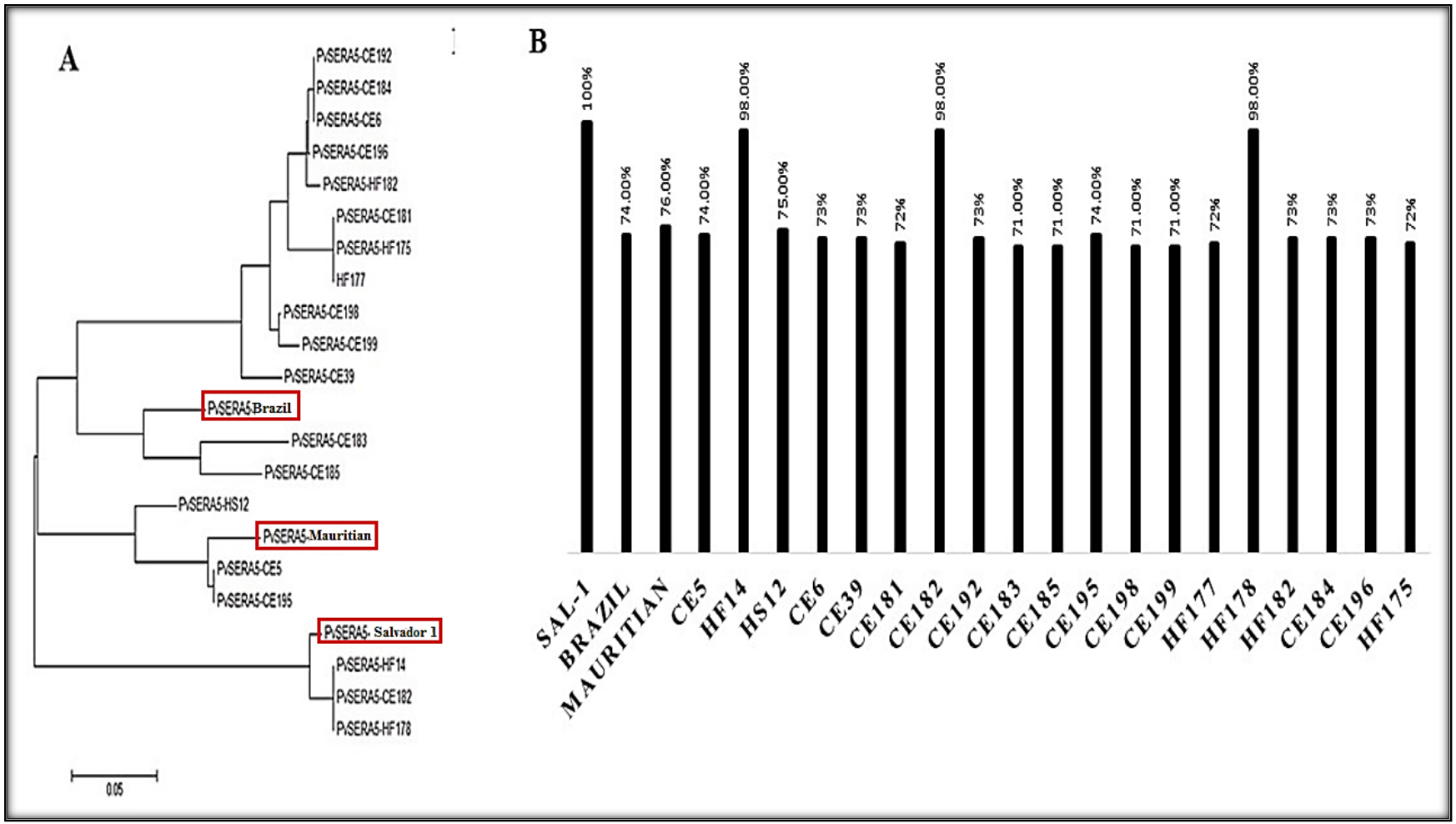



**Figure 3.12 A) Phylogeny of PvSERA5 samples indicating high genetic diversity among Indian field isolates. The samples indicated in the above figure indicate clustering of haplotypes among strains. B) Percentage identity graph for comparison of PvSERA5 sequences of Indian field isolates, Brazil, and Mauritian strains withSal-I.**

***Statistical Analysis of Polymorphism Data***

Statistical analysis of basic population genetic parameters like number of haplotypes (h), haplotype diversity (HD), Nucleotide diversity ($\pi$), Tajima's D test were also computed for the sequence data obtained for PvSERA5 N-terminal region using DnaSP Version 5 (Librado & Rozas, 2009; Tamura et al., 2011). PvSERA5 represent high genetic diversity in field isolates which is clearly evident from the statistical values obtained for their respective haplotype diversity (HD) ± SE (0.965 ± 0.024). Nucleotide diversity values per site PvSERA5 show greater nucleotide diversity (Pi = 0.08753) in comparison to PvSERA4 (Pi = 0.01003). Also, Tajima's D analysis indicates a positive score for PvSERA5. However, the calculated Tajima's scores were not significant for the present data set ($P > 0.1$).

Research highlights of N-terminal diversity of PvSERA5 include the following:

- The samples showed 38 SNPs unique to Indian field isolates. Polymorphism resulted in identification of 16 novel haplotypes from Indian field isolates.
- At the 89th amino acid position in the earlier Brazilian and Mauritian strains there is an insertion of amino acid Glycine (G). The same insertion is also found in 14 of the Indian Isolates. A novel insertion of Valine (V) at this 89th position in two isolates CE183 and CE185 is reported here for the first time.
- Also, we observed that diversity seen in N-terminal region of PvSERA5 was predominantly localized to the 5' region and the 3' end showed conservation in all Isolates.

The present data that indicates that the PvSERA5 from Indian Isolates show novel variations and phylogenetically have haplotype distribution spread across worldwide strains. Our sequence results point toward a need to further understand the PvSERA5 diversity worldwide and emerging genetic recombination patterns if any. This may assist in design of vaccines or to elucidate selection events associated to drug resistance. All these observations on *P. vivax* offer new insight on these highly expressed blood stage members and their role as potential vaccine candidates.

## Conclusion

Our results of genetic diversity studies of both PvSERA4 and PvSERA5 genes from Indian isolates is first study from Indian isolates. Also, it contributes to the antigenic repertoire of very limited worldwide haplotypes of these genes reported from Thailand. Our reports, on the extent of genetic diversity in both PvSERA genes and expected positive selection operating in PvSERA5 would further aid in investigating such studies in various geographic locations to better understand recombination mechanisms and patterns of genetic diversity in these genes. This would further aid to establish these genes as potential candidates for *P.vivax* malaria.

**Summary of PvSERA sequences submitted to the NCBI database (First Submission from Indian Isolates) is detailed in Appendix II**

# CHAPTER 4

# STRUCTURAL CHARACTERIZATION OF CENTRAL PROTEASE DOMAIN OF PvSERA MEMBERS BY USING COMPUTATIONAL APPROACH

# Structural characterization of central protease domain of PvSERA members by using computational approach

## Background

Key proteins of the malaria parasite involved in disease etiology have been major targets for immune based intervention as vaccine candidates as well as drug targets for therapeutics. Understanding the genetic variability of these key proteins is important (Takala & Plowe, 2009).

1) To represent the diversity and structure of local parasite population.
2) Understanding their functional evolution.
3) To evaluate and choose proper vaccine candidates or drug target enzymes.

*Plasmodial* proteases like PvSERA4 are conjectured to be good drug targets because of their high expression profiles in the blood stage and their likely critical role in parasite biology (Palacpac et al., 2006).

*In silico* approach and genetic polymorphism studies will aid to validate this malaria enzyme as probable drug target. The genetic characterization of the PvSERA4 enzyme domain done from field isolates from two endemic regions of India, Tamil Nadu and Andhra Pradesh showed genetic conservation. Only three SNPs resulting in amino acid changes at positions 671 N→D, 870 Y→S and 914 T→P were found. Overall genetic conservation was also confirmed by evaluation of population genetic parameters like Tajima's D and Z-test which indicated negative selection or purifying selection. It became clear through these results that PvSERA4 enzyme played a vital role in the parasite life cycle, had low genetic diversity and deleterious

mutations were unlikely to take over the population, highlighting the prospects of it being a suitable drug target.

In *P.falciparum*, PfSERA5 is a highly expressed member in late trophozoite and schizont stages of parasite and its interference was found to impair blood stage development (Hodder et al., 2009). Key findings indicate their strong involvement in parasite exit and hence serve as target of intervention that would promote parasite clearance in blood stage (Arastu-Kapur et al., 2008; Blackman, 2008; Yeoh et al., 2007) . Similarly, Serine repeat antigen of *P.vivax* are important stage specific protease and PvSERA4 is the highly expressed members in blood stage (Palacpac et al., 2006). Earlier studies have suggested that interventions targeting highly expressed SERA as in PfSERA5 should be designed to also target other SERA members, as findings indicate that these Ser-type SERA members to be monophyletic and could compensate redundant function if expressed strongly relative to PfSERA5. Moreover it is clearly evident that targeting other Ser-type SERA members is also important to design an comprehensive intervention for blood stage malaria (McCoubrie et al., 2007)**(Figure 4.1)**. In the present study, we have modelled the 3D structure of central protease domains of highly expressed PvSERA4, PvSERA5 and PvSERA2 genes of *P.vivax* using PfSERA5 (PDB id – 2wbf) protease domain as template for homology modelling. This attempt to structurally elucidate the central domain of PvSERA protein by *insilico* approach is first of its kind and will aid in rational drug design.

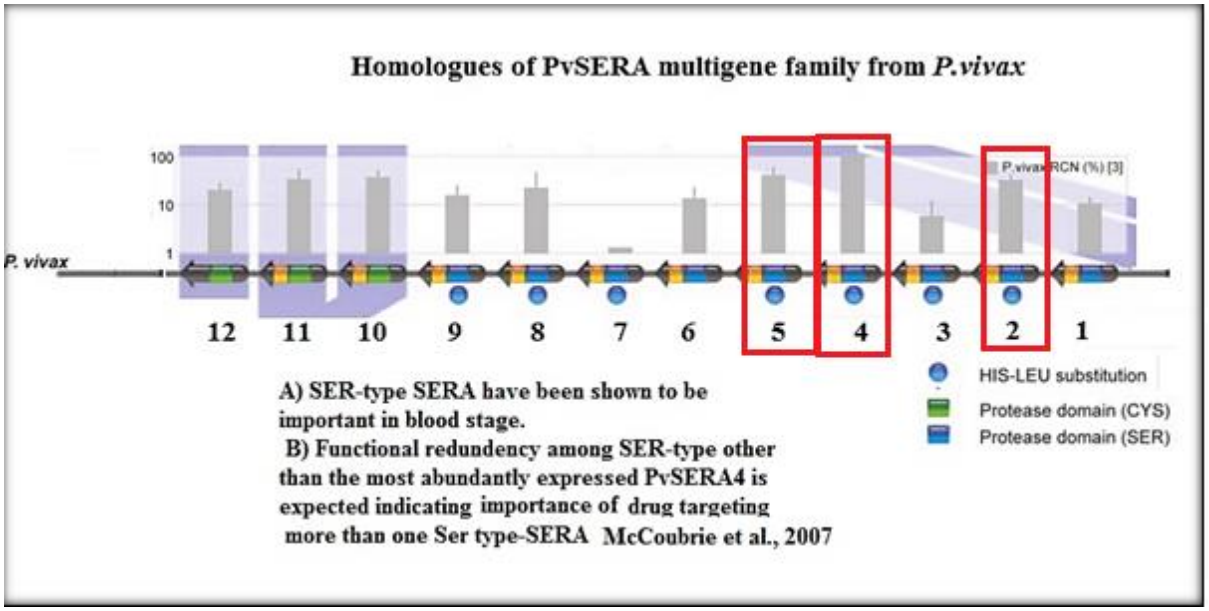



**Figure 4.1 Graphical representation of expression rates of 12 PvSERA genes in the blood stage (McCoubrie et al., 2007). The central protease domain of the SERA genes are coloured Blue – Serine type SERA and Green-Cysteine-type SERA. Blue circles below Serine type SERAs have a specific His-Leu substitution in catalytic residue. Red boxes indicate relatively highly expressed members PvSERA genes in blood stage.**

Structural characterization was done using computational tools like **Modeller 9v9** (Sali & Blundell, 1993)**, SAVS, PROCHECK** (Laskowski et al., 1993)**, ERRAT**(Colovos & Yeates, 1993)**, ModRefiner and Modloop online servers**(A Fiser et al., 2000)**, ProSA web**(Wiederstein & Sippl, 2007)**., Verify3D** (Bowie, Luthy, & Eisenberg, 1991; Luthy, Bowie, & Eisenberg, 1992)**, Q-site finder**(Laurie & Jackson, 2005) **and Molecular dynamic simulations by GROMACS 4.5.4 package** (Hess et al., 2008)**. The details of these methods have already been discussed in Chapter 2.** Finally, a 10ns MD simulation was carried out with time step of 2ps.

## Results and Discussion

Chemotherapy of malaria generally involves killing of parasite by blocking vital proteins involved in metabolic pathways that differ significantly from human host (Phillips, 2001; Philip J Rosenthal, 1998; Tripathi, Mishra, Dwivedi, Tewari, & Verma, 2005). Conserved *Plasmodial* proteins, critical for parasite development and survival can act as potential targets for inhibitor based therapies. Therefore genetic diversity studies from *Plasmodium* infected field isolates would be extremely valuable in identifying protein encoding genes that are of low genetic diversities and structural characterization of such genetically conserved vital proteins would serve as lead targets for rational drug design.

### *Structural Characterization of PvSERA4 protease domain*

PvSERA4 remains to be characterized in terms of its protein structure. Structural insight into this domain would aid in understanding the substrate specificities of PvSERA4 and for designing improved structure-based inhibitors against this enzyme. PvSERA4 central domain sequence shows 55% identity with template PDB id-2wbf the crystal structure of PfSERA5 enzyme domain (**Figure 4.2**). This was used to build the homology model of protease domain of PvSERA4.

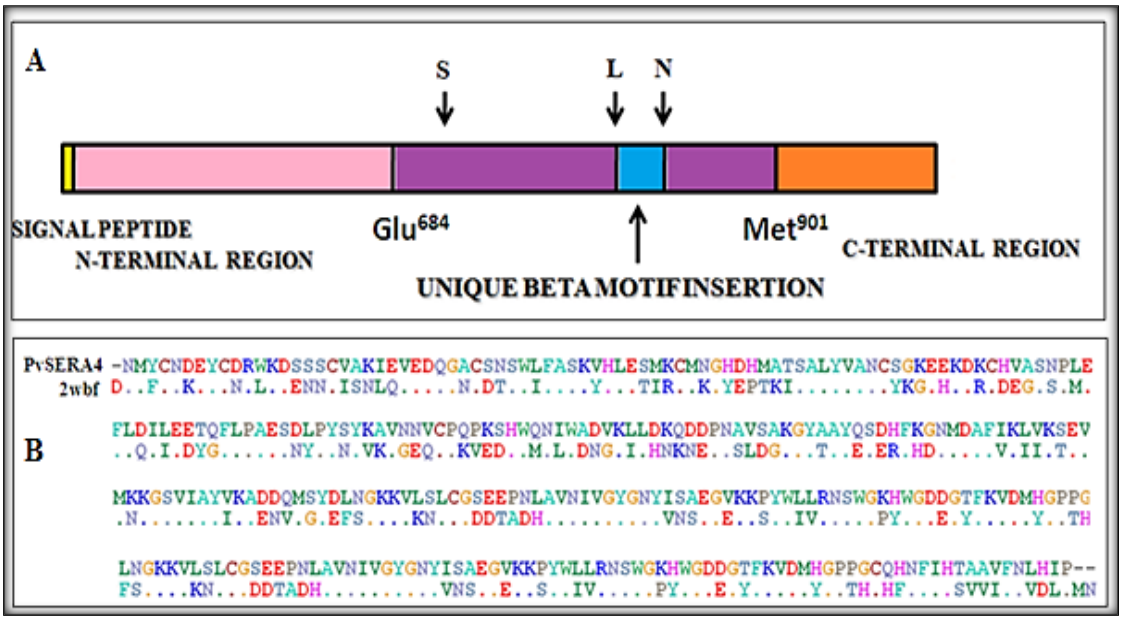



**Figure 4.2 A) Schematic representation of domain organization of PvSERA4 showing it central protease region with catalytic residues Serine, Leucine and Aspargine highlighted with arrows and an unique-beta motif sequence between catalytic residue Leucine and Aspargine. B) Pair-wise sequence alignment of PvSERA4 Sal-I sequence with the PfSERA5 template sequence bearing PDB id-2wbf.**

The stereo-chemical quality of the predicted refined model structure was evaluated through RAMPAGE analysis. This analysis revealed that 98% of the residues were positioned in the most favourable and allowed regions of the Ramachandran plot for enzyme domain, thus demonstrating good quality of the model generated. Reliability of the model was further checked by ERRAT that gave score of 97.09% (**Table 4.1**). ProSA-web "Z" score value of the 3D model lies in the range of native confirmations of crystal structures computed using NMR method. The Z-score for PvSERA4 modelled protein enzyme domain is -7.29, and Z-score for template is -8.72. The stereo-chemical quality and z-score from ProSA-web indicate that model is reliable and 3D model is shown **Figure 4.3**.

**Table 4.1 Stereo-chemical analysis of the modelled protease domain**

| Structural Analysis and Verification | |
|---|---|
| **RAMPAGE analysis** | **98%** residues were positioned in most favourable and allowed regions of the Ramachandran plot |
| **ERRAT** | 97.07%. |

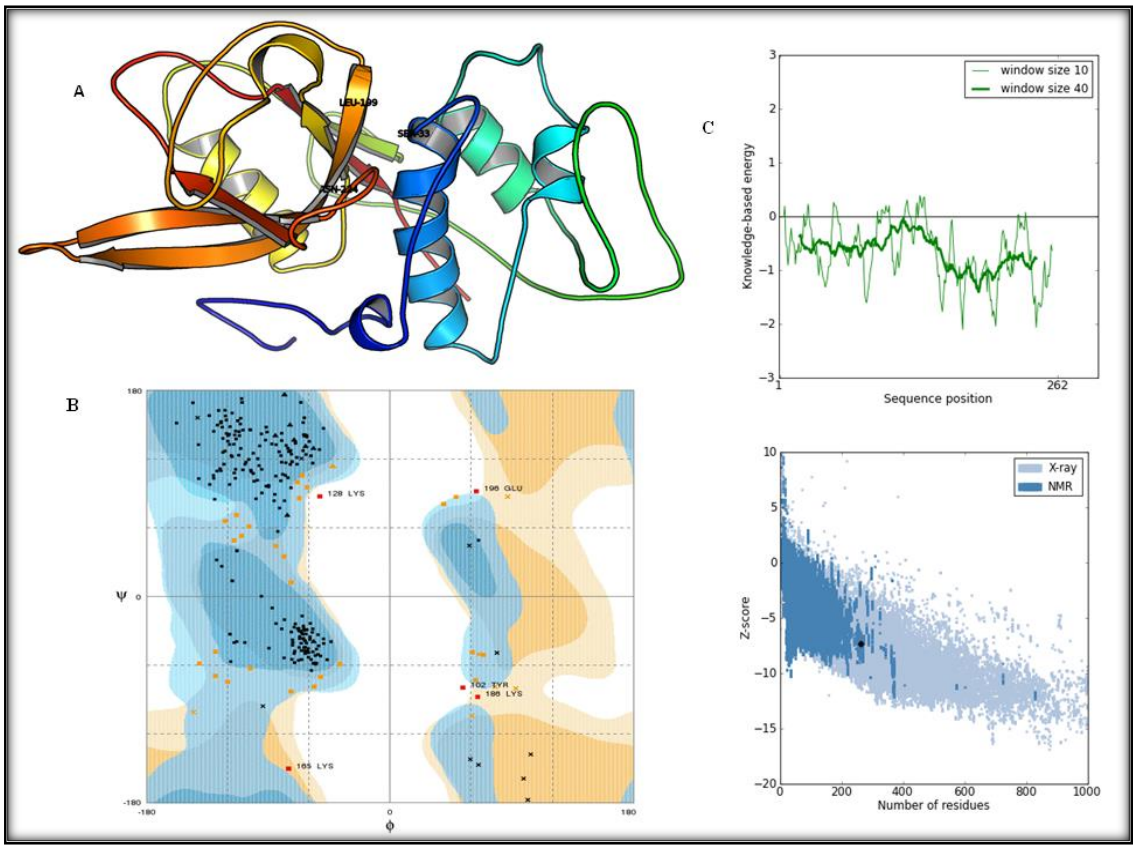



**Figure 4.3 A) Modeled three-dimensional structure of PvSERA4 by homology modeling using comparative modeling approach. Active site residues are highlighted. B) Ramachandran plot values of the 3D Model showing number of residues in favored, allowed and outlier region through RAMPAGE evaluation server. C) ProSA-web Z-scores of all protein chains in PDB determined by X-ray crystallography (light blue) or NMR spectroscopy (dark blue) with respect to their length. The Overall model quality Z- score of the homology model of *Plasmodium vivax* PvSERA4 are highlighted as large black coloured dot and local quality model energy profile.**

***MD Simulation of PvSERA4 protease domain***

The protein model backbone RMSD was analysed for trajectory from the starting structure as a function of time. The backbone RMSD remains constant approximately after 4ns which indicates stability of the modelled protein (**Figure 4.4 A**). The RMSF plot of the model shows that loops are the most dynamic regions and the secondary structures are the most stable (**Figure 4.4 B**). Our results of RMSD and simulation time prove that model is a stable confirmation.

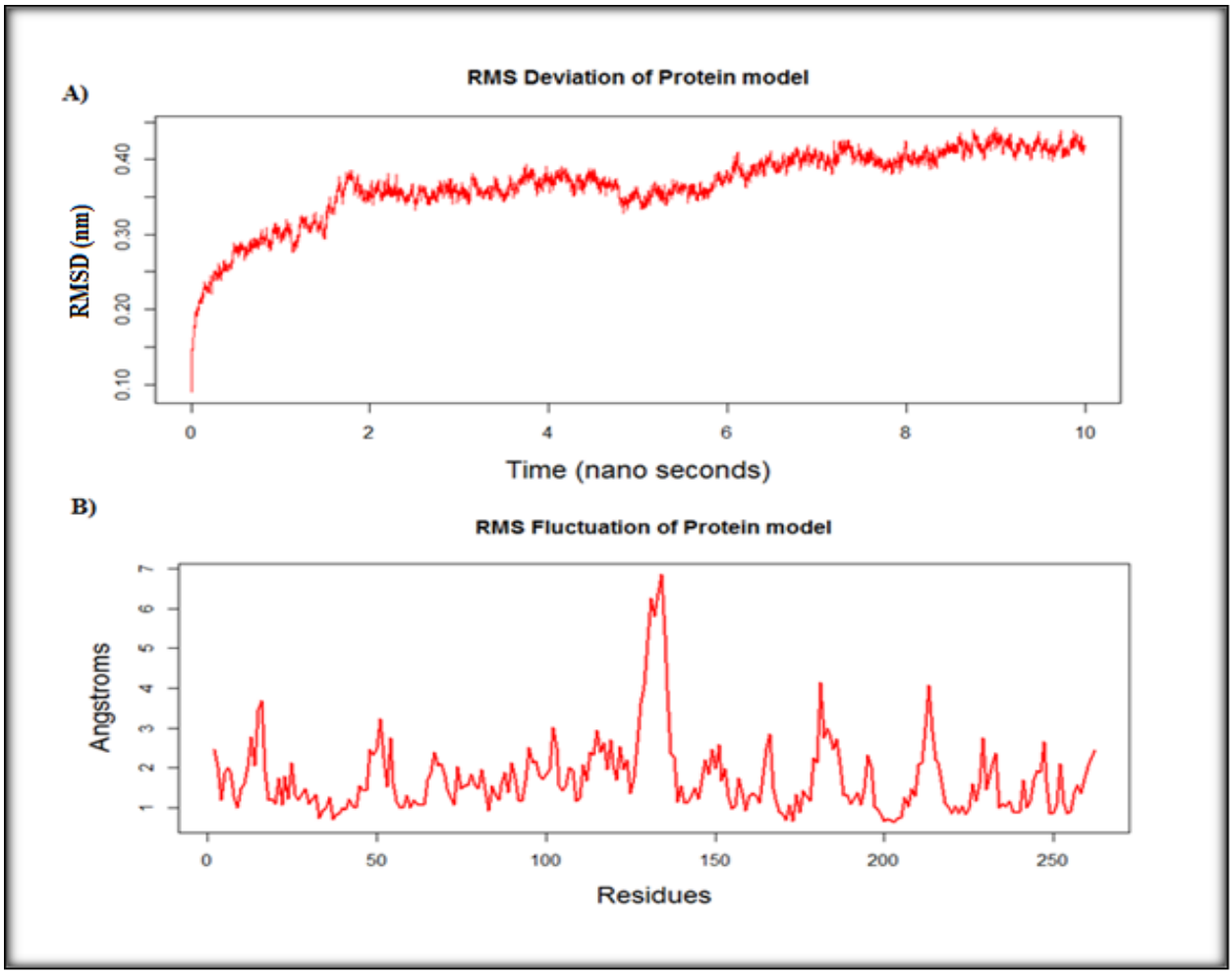



**Figure 4.4 A) RMSD of the backbone atoms of the modelled protein over a time period of 10 ns. B) RMS Fluctuations (a measure of the average atomic mobility) of backbone atoms (N, Cα and C atoms) during the molecular dynamics simulations.**

***Structural analysis***

Based on the generated refined structure, a comprehensive analysis of the sequence and structural features was carried out to gain an understanding of functional aspects of PvSERA4. We observe an unusual beta hairpin motif insertion involving two antiparallel beta strands between active site residues $Leu^{860}$ and $Asn^{885}$ in the enzyme domain of PvSERA4. Similar motifs have earlier been reported in *Plasmodial* hemoglobinases like falcipain, vivipains and are shown to be important in substrate recognition. A sequence comparison of beta hairpin motif across *Plasmodial* proteins including SERAs show high level conservation except for few amino acid changes, indicating its structural and functional significance. Additionally, another characteristic observation made in this sequence comparison is a five amino acid deletion consistently found in all *Plasmodial* SERA members, which could have implications in substrate specificity and recognition (**Figure 4.5 A**).

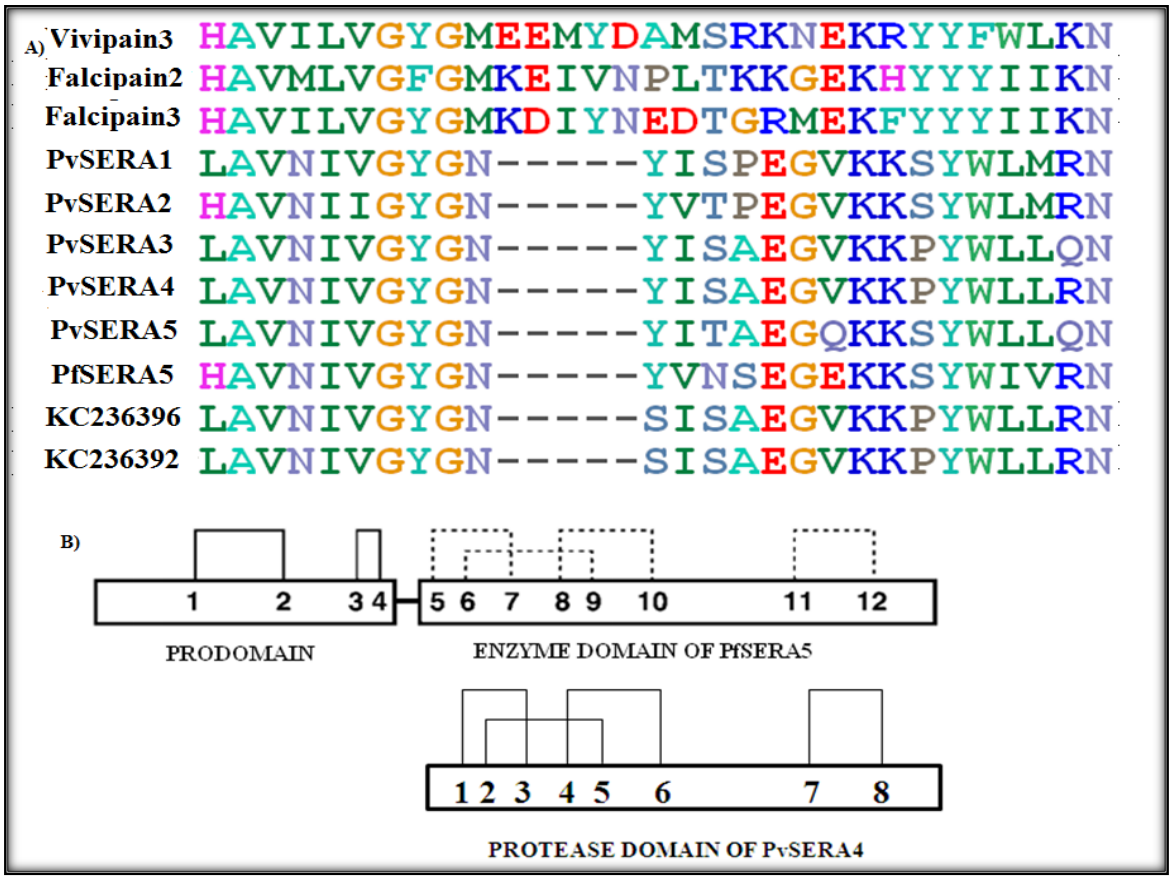



**Figure 4.5 (A) Multiple sequence alignment of unusual Beta motif from different *Plasmodial* and non *Plasmodial* proteases using Bio edit sequence alignment editor (PvSERA1-AAB41488, PvSERA2-AAB41485, PvSERA3-AAB41487, PvSERA4-AAB41486, PvSERA5-AAB41489 and PfSERA5 (NC_000910),KC236396, KC236392 are from Tamil Nadu ,Andhra Pradesh field isolates respectively from the present study. (B) Schematic representation of the arrangement of eight conserved cysteines and Array of disulfide bonds in the enzyme domain of PfSERA5 a well characterized SERA member (serine protease) from *Plasmodium falciparum* and protease domain PvSERA4 from *Plasmodium vivax*.**

Our field isolates showed high level sequence conservation in the cysteine residues in the protease domain of PvSERA4 indicating essential role in maintaining the stable confirmation of the protein. The pattern of disulphide bond formation that has a crucial role in stabilizing the structure and folding of a protein was similar for the predicted model and

template 2wbf. Schematic representation of the arrangement of di-sulphide bridges (**Figure 4.5 B**).

***Prediction of binding sites***

The substrate binding sites of our homology template PfSERA5 (PDB 2wbf) was inferred based on structural comparison with archetypical human enzyme cathepsin L (PDB 1mhw) and is well defined in literature (Hodder et al., 2009). Binding site prediction for our PvSERA4 model was carried out using Q-site finder (Laurie & Jackson, 2005). A structural comparison of these predicted sites of our model with that of binding sites located in the crystal structure (PDB 2wbf) shows the binding site to be in similar conformation as that of template as shown in **Figure 4.6**. The binding site can be probably used for the design of potential inhibitors.

***3D homology searches***

Comparison of the generated 3D model with known protein structures using DALI server gave hits with rmsd values $2.7A^0$-$3A^0$ which were considered to be structural homologs as depicted in **Figure 4.7** and **Table 4.2**. Because of structural similarity, the structural alignment can be used to propose enzyme-inhibitor interactions and help obtain information on ligands for further studies on drug targeting. All these combined structural results of PvSERA4 model validate the significance of this enzyme as a potential target for combating *P.vivax* malaria.

**Table 4.2 Structural homologs of PvSERA4 protease domain identified using Dali server.**

| PDB Id | RMSD (A° ) | Sequence identity (%) | Z-Score | Molecule Description |
|---|---|---|---|---|
| 2wbf-X | 2.8 | 54 | 25.2 | Serine Repeat Antigen 5 :serine protease from *Plasmodium falciparum* |
| 1iwd-A | 2.7 | 18 | 17.8 | Ervatamin B :plant cysteine protease |
| 2b1n-A | 3.0 | 20 | 17.4 | SPE31: cysteine protease from seeds of *Pachyhizus erosus* |
| 1yal-A | 2.7 | 17 | 17.4 | Chymopapain: cysteine protease from *Carica papaya L* |
| 2act-A | 2.8 | 17 | 17.4 | Actinidain Precursor: Proteolytic enzyme from the fruit of *Actinidia chinensis* |
| 1ewp-A | 2.8 | 21 | 17.3 | Cruzain: cysteine protease from *Trypanosoma Cruzi* |
| 2p7u-A | 2.8 | 21 | 17.3 | Cysteine Protease: Rodesian cysteine protease from *T.brucei* rhodesiense |
| 1yvb-A | 2.7 | 23 | 16.6 | Falcipain 2: cysteine protease from *Plasmodium falciparum* |

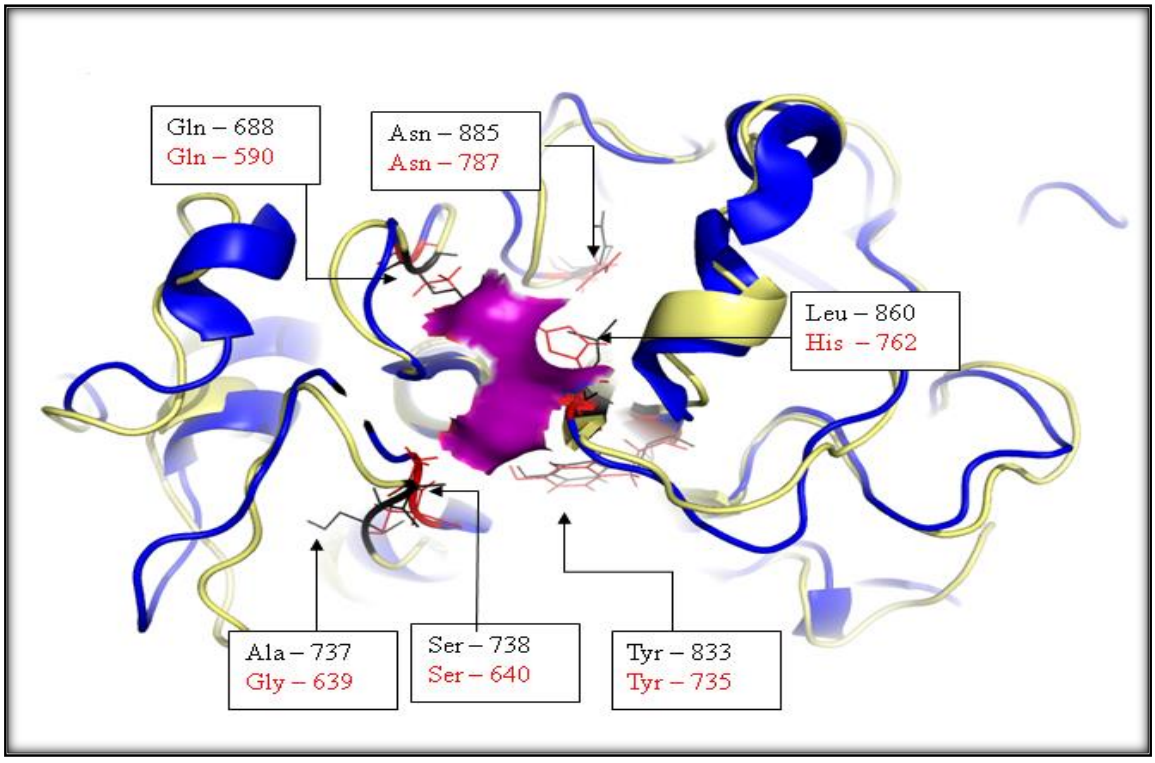



**Figure 4.6 Structural comparison of binding sites in modelled PvSERA4 and PfSERA5. Q-site finder predicted binding site represented in surface mode (purple).PfSERA5 (PDB: 2wbf-**

**colored in blue) binding sites include the following residues (red), Q-590, G-639, S-640, Y-735, H-762, and N-787. Corresponding residues (Black) in PvSERA4 (colored in pale yellow) are Q-688, A-737, S-738, Y-833, L-860, and N-885.**

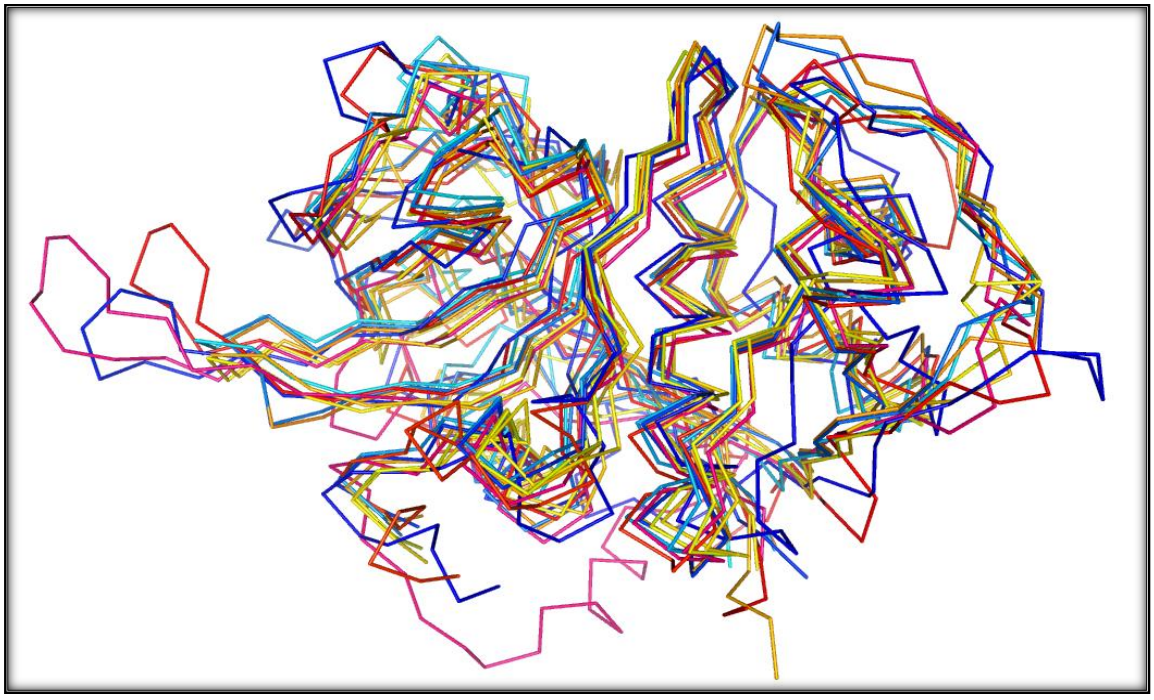


**Figure 4.7 Superimposition of PvSERA4 (dark blue) and similar structures. The backbones are shown as ribbons, PDB id, 2wbf-x (red): serine-repeat antigen protein; PDB id, 1iwd-a (cyan): ervatamin b; PDB id, 2b1n-a (dark yellow): spe31; PDB id, 2act-a (light yellow): actinidain precursor; PDB id, 1ewp-a (light blue): cruzain; PDB id, 2p7u-a (orange): cysteine protease, PDB id, 1yvb-a (pink): falcipain 2.**

### *Homology modelling of other PvSERA members*

Structural modelling of PvSERA 2 and 5 were also done by comparative modelling using PfSERA5 (2wbf) as their percentage sequence identities were significant to build a homology model. The protein sequence alignment and percentage identities of other PvSERA members apart from PvSERA4 is shown in **Figure 4.8.**

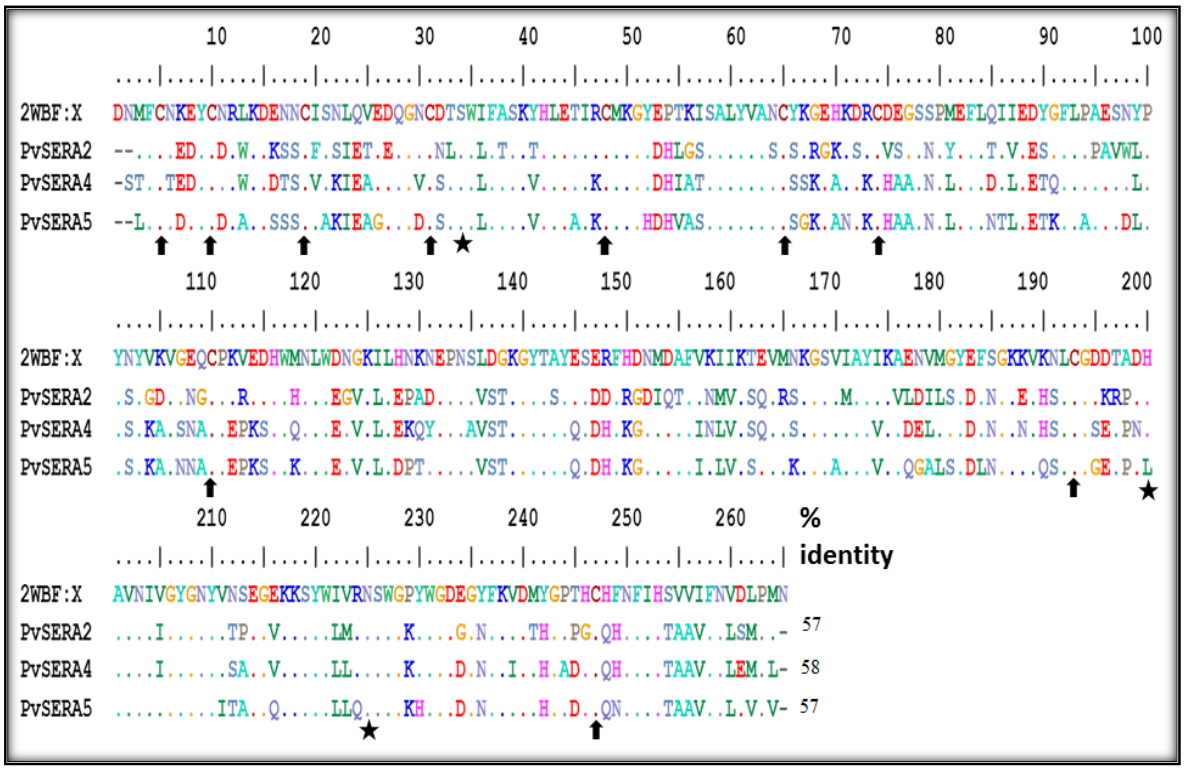



**Figure 4.8 Protein sequence alignment of PvSERA members –PvSERA2, 4 and 5 with template PDB id 2wbf of PfSERA5 from *Plasmodium falciparum*. Black arrow in figure represents conservation of cysteine residues participating in di-suphide bridges. Star mark represents sites of active site residues.**

Results of stereo-chemical analysis by ERRAT, Procheck and verify 3D indicate the models to be good quality models to be used as templates for drug targeting and design **Table**

**4.3** and **Figure 4.9**. Also, observed structural conservation across these PvSERA members would aid in developing common strategies of drug targeting and design.

**Table 4.3 Stereo chemical analysis of modelled protease domain of PvSERA 2, 4 and 5 members.**

| **Structural Analysis and Verification** | | | |
|---|---|---|---|
| **PvSERA** | **ERRAT** | **PROCHECK** | **VERIFY 3D** |
| PvSERA2 | 91.7 | 92.0% core 7.6% allowed 0.4% generously allowed 0.0% disallowed | 100% of the residues had an average 3D-1D score>0.2 |
| PvSERA4 | 93.7 | 88.7% core 9.2% allowed 2.1% generously allowed 0.0% disallowed | 83.71 of the residues had an average 3D-1D score >0.2 |
| PvSERA5 | 91.6 | 89.3% core 9.4% allowed 0.9% generously allowed 0.4% disallowed | 96.20% of the residues had an average 3D-1D score >0.2 |

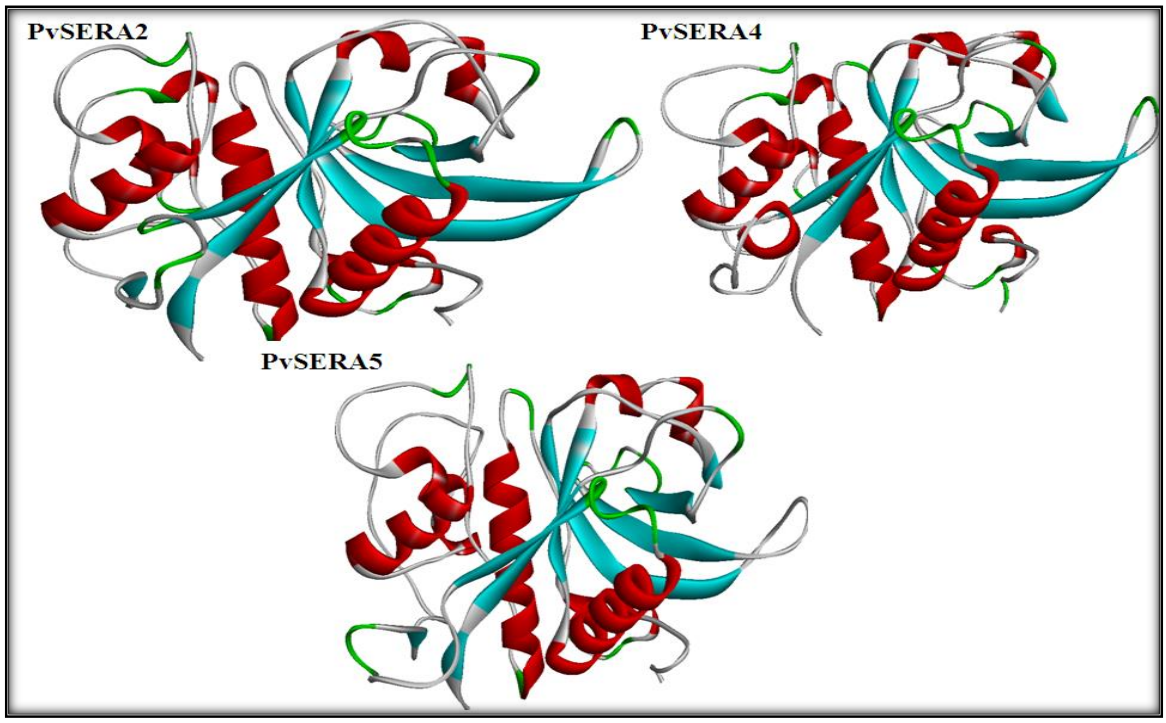



**Figure 4.9 Graphical representation of modelled PvSERA structures using Discovery studio visualizer.**

## Conclusion

This study highlights the significance of genetic conservation in the protease domain of PvSERA4 in multiple field isolates and recommends it as a reliable drug target for *P.vivax* malaria. We have also attempted to present homology models of the central proteolytic domains in other highly transcribed PvSERA genes. The protein models presented in this study would serve as lead targets. However, additional biochemical and functional characterization are required to define their biological role in *P.vivax* and our structural information from this study is significant to provide future insight into rational drug design of PvSERA as antimalarial targets.

# CHAPTER 5

# CHARACTERIZATION OF SEQUENCE DIVERSITY IN *Plasmodium falciparum* SERA5 (PFSERA5) IN INDIAN FIELD ISOLATES

# Characterization of sequence diversity of *Plasmodium falciparum* SERA5 (PfSERA5) in Indian field isolates

## Introduction

The design of antimalarial vaccines and use of antimalarial drugs is hampered extensively by antigenic diversity. *P. falciparum*, the agent for malignant malaria, is of a great clinical and epidemiological importance for elucidating the processes that maintain genetic polymorphism (Escalante, Lal, & Ayala, 1998) and it has generated considerable interest and debate in population genetics. The difficulty in controlling this devastating disease is due to the high levels of genetic diversity in *P. falciparum* which allows rapid evolution and dissemination of advantageous traits such as drug resistance and antigenic variability in the parasite (Schultz et al., 2010).

Extensive polymorphism in surface antigens is one of the major factor for late development of immunity to malaria (Ferreira et al., 2004). Genes encoding such antigens are shown to be under strong immune mediated selection favouring rare alleles (Street, 1997). This kind of selection is referred to as balancing selection which indicate that immune response to each allelic form is highly effective against the parasite in the asexual blood stage (Roy, Ferreira, & Hartl, 2008). Hence, it is important to report and identify genetic polymorphism in malaria antigens.

*P. falciparum* Serine Repeat Antigen 5 (PfSERA5) is a potential blood stage target. It is highly expressed as late blood stage 120 kDa antigen and is processed into a 47 kDa (P47) N-terminal, 50 kDa (P50) central and 18 kDa (P18) C-terminal fragments.

Studies show that only the N-terminal region harbours genetic diversity while the P50 and P18 regions are conserved (B A Fox et al., 1997). The representation of genetic organization of the PfSERA5 showing repeat regions in exon II is shown in **Figure 5.1**. The N-terminal region of PfSERA5 is recognized by human serum from individuals living in endemic areas (Horii et al., 2010; B. A. Okech et al., 2001; Pratt-Riccio et al., 2008). The antibodies against P47 have been found in individuals from Uganda and Amazon wherein they have significantly found to lower parasitemia levels (Banic et al., 1998; B. A. Okech et al., 2001). Also, the SE36 a recombinant derivative of the P47 which lacked the serine repeat region was found to correlate to lower parasitemia in individuals from Solomon Islands (Horii et al., 2010). High titre of IgG anti-SE36 were shown to be associated to protection against severe malaria in infants (B. Okech et al., 2006). All the human and laboratory animals trials showed that P47 region is a promising vaccine candidate (Inselburg et al., 1991; Inselburg, Bathurst, Kansopon, Barr, & Rossan, 1993; Sugiyama, 1996).

In the P47 N-terminal region, two distinctive types of repeat polymorphism are seen a) the Octamer Repeat (OR) region and b) Serine Repeat (SR) region (Morimatsu et al., 1997). Sequence diversity studies reported earlier were limited to few regions from Asia, Africa and South America (Q. Liu et al., 2000; Safitri, 2003). The polymorphism in P47 sequence were classified into three allelic classes mainly the FCR3, K1 and Honduras (recombinant type) (Safitri, 2003). Earlier sequence diversity analysis of the OR region identified repeats with predominantly five or six octamer repeats (Riccio et al., 2005). Also, reports from Brazilian amazon have indicated that the genetic polymorphism in octamer repeat region does not influence specific immune response and immunogenicity(Pratt-Riccio et al., 2008). However, recent evaluation from nine countries, which reports a comprehensive search for novel variants

in repeat regions of P47, have represented octamer repeats variants with repeat numbers more than the usual reported five or six (Tanabe et al., 2012). The impact of such novel variants on immunogenicity and any specific immune response is yet to be identified. There are no reports of PfSERA5 diversity from India which represents an important study centre for complex malaria (Das et al., 2012).

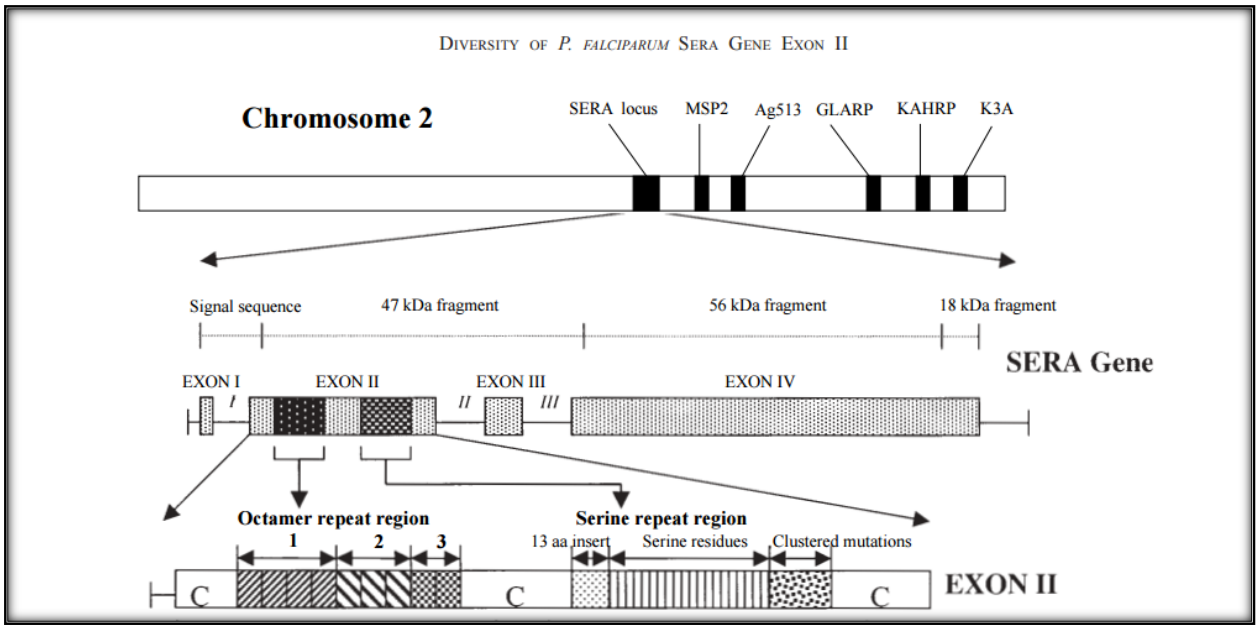



**Figure 5.1 Graphical representation of genetic organisation of PfSERA5 locus and repeat regions in P47 region encoded by exon II (Q. Liu et al., 2000).**

Our study on genetic diversity in PfSERA5 is the first report from Indian isolates that adds to the collection of antigenic repertoire represented worldwide. This contributes to understand variant specific immunity and also evaluate the N-terminal P47 region as a vaccine candidate.

## Results and Discussion

Sequence analysis was done by multiple sequence alignment through Clustal W programme available in BioEdit (Hall, 1999). Phylogenetic tree was generated by neighbour joining method with MEGA Version 5 software (Tamura et al., 2011). Statistical analysis of basic population genetic parameters like haplotype diversity (HD)-measure of uniqueness of a particular haplotype in a given population, and nucleotide diversity ($\pi$)-average number of pairwise differences per site, were also computed for the sequence data obtained with DnaSP (Librado & Rozas, 2009). Antigenicity plots for distinct haplotypes were generated using online web tool Antigenicity Plot Index (Hopp & Woods, 1981).

### *OR region Polymorphism from Indian isolates*

Analysis of this region in Indian isolates showed repetitions, truncations and variations. Almost all Indian isolates analysed showed the standard number of OR i.e. 7 (6 Type I + 1 Type II) (Tanabe et al., 2012). However, three isolates showed exceptions with repeat variation in Type Ie and Type II. They were HF37 (5 repeats of Type Ie), CE145 (1 Type Ie + 3 repeats of Type II) and HF169 (5 repeats of Type Ie + 1 Type II). In addition, samples HF105 and CE145 showed a specific glutamine-arginine (G→R) change within type Ia TGESQT (G/R)N. Truncated OR occurred between 1b and 1c in (23/38) Indian isolates. Out of these, truncated repeat 4/4 -TVGDQAGN, occurred in (21/23) Indian isolates. While two Indian isolates HF170 and HF174 showed 3/5-TVGGQAGN truncated repeat. Truncated repeat 3/5-TVGSPQGS was exceptional to Indian isolate CE145 in the present study (**Figure 5.2A**).

Phylogenetic analysis showed that 36/38 Indian isolates conferred to 5 reported haplotypes (Tanabe et al., 2012) while two samples HF37 and CE145 were novel OR

haplotypes reported in this study (**Figure 5.2B, Table 5.1**). Among the remaining 36 Indian isolates, 18 showed similarity to haplotype Tz93-120 which is the most prevalent OR haplotype reported worldwide. Fourteen Indian isolates showed similarity to reported haplotype Tz03-38 (FCR3), a prevalent haplotype predominant in Tanzania and Ghana. Two Indian isolates, HF170 and HF174 were similar to Thailand haplotype Ti-MS001 (K1) while isolates HF105 and HF169 were similar to low-frequency haplotypes restricted to African continent Tz03-105 and Tz93-085, respectively.

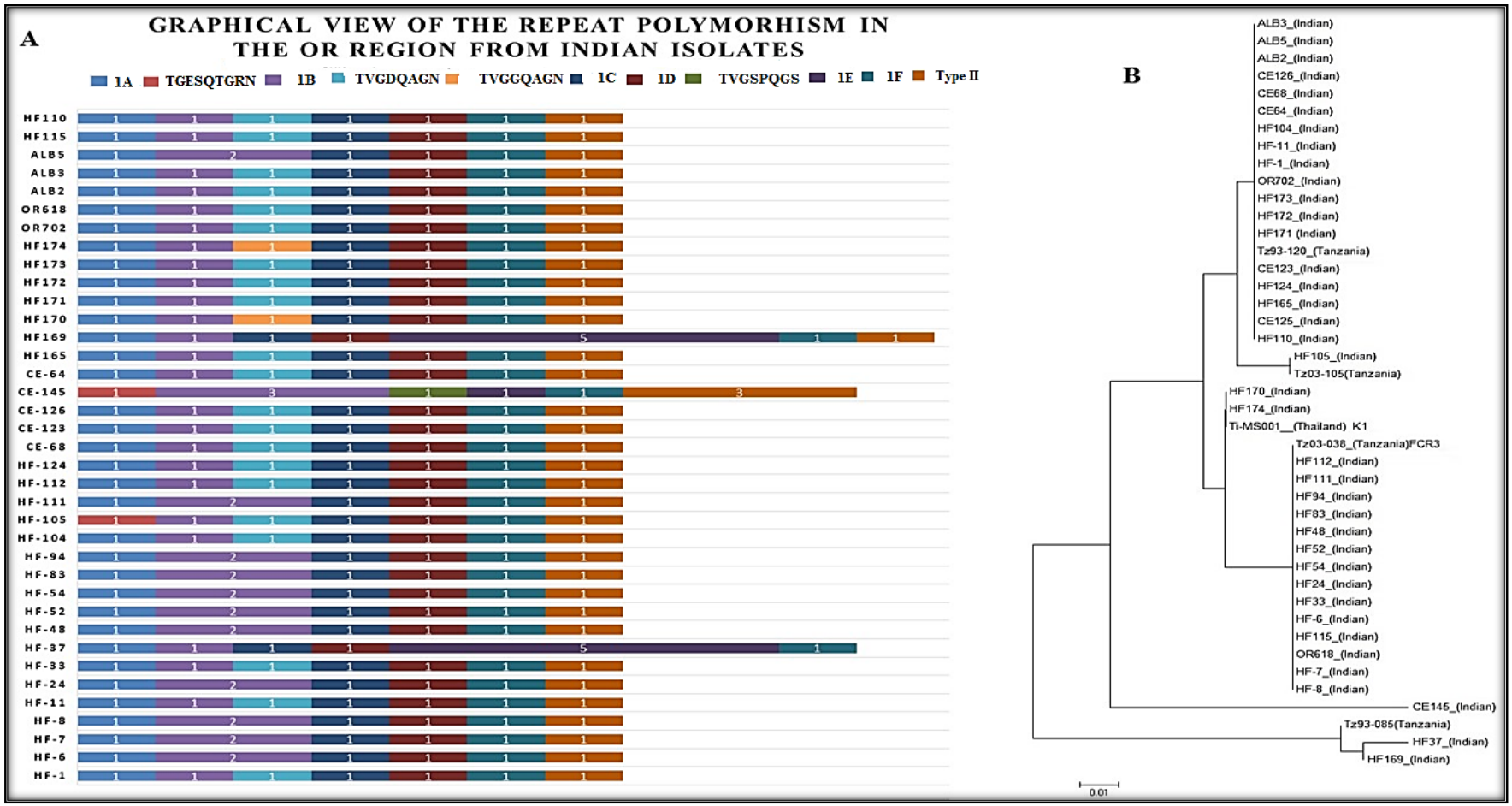



**Figure 5.2 Sequence diversity characterization in the OR region of P47 from Indian isolates: A) Graphical representation of OR polymorphism in Indian isolates. Repeat types: 1A-TGESQTGN variant-TGESQTRN, 1B-TGGGQAGN, Truncated repeat 4/4 - TVGDQAGN, 3/5-TVGGQAGN, 1C-TVGDQAGS, 1D-TGGSPQGS, Truncated 3/5-TVGSPQGS, 1E-TGASPQGS, 1F-TGASQPGS, Type II-SEPSNPVS. B) Phylogenetic analysis of OR region of N-terminal P47 domain of Indian isolates among worldwide isolates.**

**Table 5.1 Haplotype distribution of the P47 domain-OR region from Indian isolates.**

| S. No | Indian Isolate | N | Worldwide OR Haplotype Similarity | Prevalence reported |
|---|---|---|---|---|
| 1. | HF124,HF165,CE125,HF110,HF172,OR702,HF1,HF11,HF104,CE64,CE68,CE126,ALB2,ALB3,ALB5,HF171,HF173, CE123 | 18 | Tz93-120 | Worldwide |
| 2. | OR618,HF112,HF111,HF94,HF83,HF54,HF52,HF48, HF33,HF24,HF8,HF7,HF6,HF115 | 14 | Tz03-038 (FCR3) | Tanzania, Ghana |
| 3. | HF170, HF174 | 2 | Ti-MS001 (K1) | Thailand |
| 4. | HF105 | 1 | Tz03-105 Show singleton mutation (G→R) in Type 1a OR | Low frequency in Tanzania and Ghana |
| 6. | CE145 | 1 | Novel G→R in Type 1a OR, repeat variation in Type 1e &Type II | India |
| 7. | HF37 | 1 | Novel Repeat variation in 1 e and deletion of Type II | India |
| 8. | HF169 | 1 | Tz93-085 Repeat variation in Type 1e | Low frequency in Tanzania |

### *SR region polymorphism in Indian isolates*

The sequence alignment of SR region of the P47 N- terminal region showing amino acid region from 196-274 amino acids is represented in **Figure 5.3A**. The samples were compared with 3D7 standard. Among the Indian isolates, 21 were of K1 type, 14 were of FCR-3 type and 3 were recombinant Honduras type. In our field isolates, Poly-Ser repeat number was in the range of 23-49. Three Indian isolates, HF104, HF33, and HF124 showed minor-frequency single nucleotide polymorphism at amino acid position 217 (S-Serine to N-Aspargine) within the Poly Ser repeat. Phylogenetic analysis of SR haplotypes and their distribution into K1, FCR3 and Honduras is shown in **Figure 5.3B** and **Table 5.2**, respectively. Analysis of Indian isolates for the central protease domain (P50) and C-terminal region (P18) of PfSERA5 indicated absolute genetic conservation. This signified a vital role of these domains in the parasite life cycle.

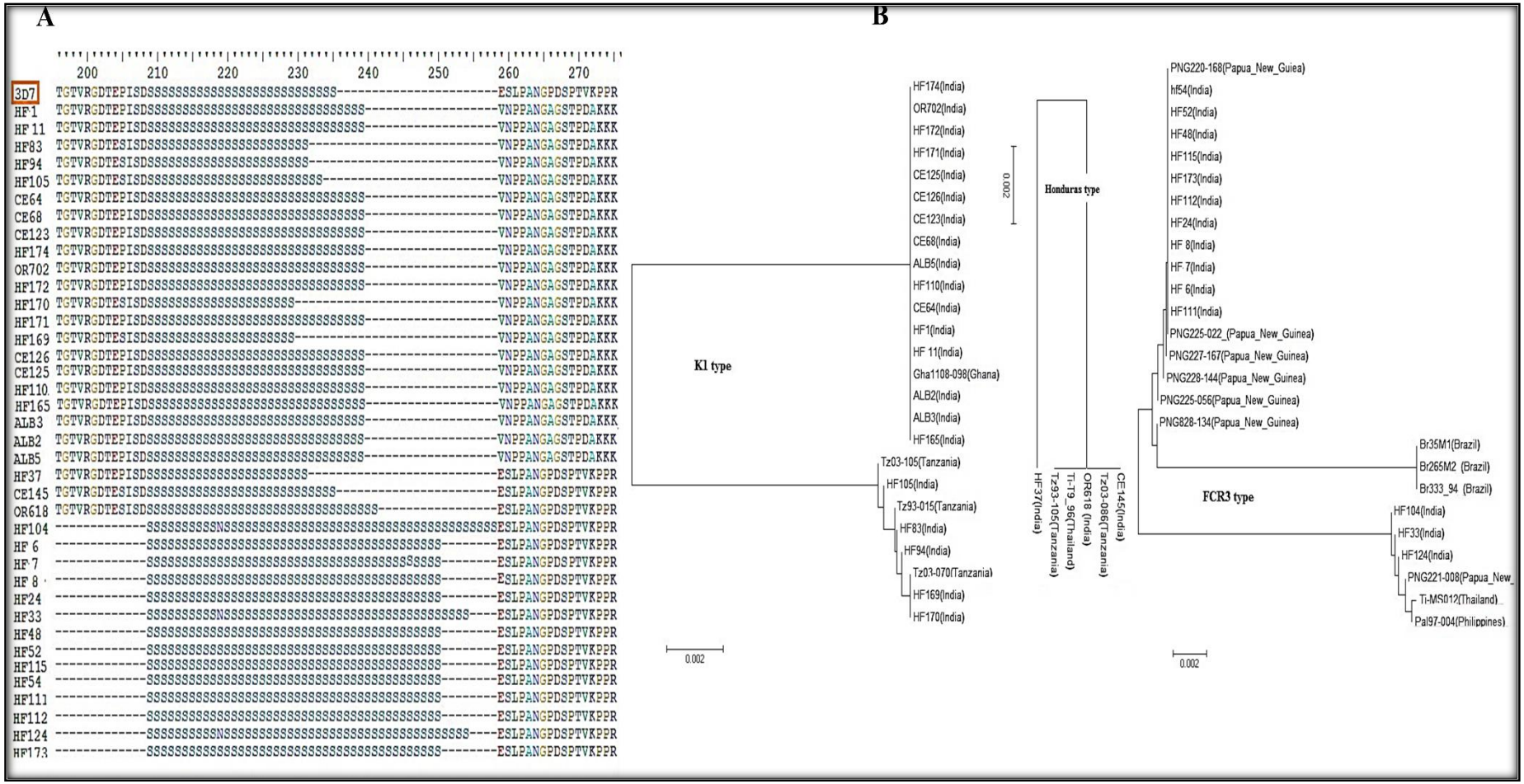



**Figure 5.3 A) Multiple sequence alignment of N-terminal P47 SR region of PfSERA5 gene from Indian isolates using standard reference sequence 3D7 strain highlighted in red box. B) Phylogenetic analysis of SR region of P47 domain of PfSERA5 of Indian isolates.**

**Table 5.2: Haplotype distribution of P47 domain-SR region of PfSERA5 from Indian isolates**

| S.No | Indian Isolate | N | Allelic type | Serine repeat Number | Worldwide SR Haplotype Similarity | Prevalence reported |
|---|---|---|---|---|---|---|
| 1. | ALB2,ALB3, ALB5, HF110, OR702, HF174,HF172,HF171,HF165,CE125, HF1,HF11, CE64,CE68,CE126, CE123 | 16 | K1 | 31 | Gha1108-098 | Ghana |
| 2. | HF105 | 1 | K1 | 25 | Tz03-105 | Tanzania, Ghana |
| 3. | HF169, HF170 | 2 | K1 | 21 | Tz03-070 | Prevalent worldwide |
| 4. | HF83,HF94 | 2 | K1 | 23 | Tz93-015 | Tanzania, Ghana |
| 6. | CE145 | 1 | Hondura s | 27 | Tz03-086 | Tanzania (Single isolate) |
| 7. | HF37 | 1 | Hondura s | 23 | Novel | India |
| 8. | OR618 | 1 | Hondura s | 33 | Ti-T9_96 | Thailand (single isolate) |
| 9. | HF104 | 1 | FCR3 | 49 & S→N at 217$^{th}$ aa | Novel | India |
| 10. | HF33, HF124 | 2 | FCR3 | 42 & S→N at 217$^{th}$ aa | Novel | India |
| 11. | HF48,HF112,HF6,HF7,HF8,HF24,HF 173,HF115,HF111,HF54,HF52 | 11 | FCR3 | 42 | Novel | India |

***Statistical and computational analysis of polymorphic data***

Antigenicity plots of unique Indian OR variants showed differences in pattern of antigenicity in comparison to standard sample isolates (**Figure 5.4**). HD and nucleotide diversity ($\pi$) were obtained for both OR and SR regions for sequence results represented in the present dataset. Results indicated higher HD and value for SR region in comparison to the OR region (**Table 5.3**).

**Table 5.3: Statistical analysis of genetic polymorphism in PfSERA5-P47 repeat region.**

| PfSERA5-P47 repeat region | Haplotype Diversity (HD) | Nucleotide Diversity ($\pi$) |
|---|---|---|
| OR | 0.697± 0.050 | 0.02104 |
| SR | 0.773±0.058 | 0.33299 |

The present study deals with polymorphism of nucleotide sequences of PfSERA5 from Indian field isolates. Sequence results obtained for OR region were represented based on the earlier description of repeat types which consists of two repeat types –Type I (sub divided in to six subgroups 1a-1e), and Type II (Tanabe et al., 2012). Three Indian isolates HF37, CE145, HF169 which showed variations in repeat types exclusively for Type 1e and Type II were earlier reported to be restricted to African isolates alone(Tanabe et al., 2012). HF37 is the only isolate that show deletion of Type II repeat which disrupts the 14 mer stretch QPGSSEPSNPVSSG (amino acid 59-72), conserved till date in all worldwide isolates and believed to be recognized by Mab43E5(Tanabe et al., 2012). This is the first report of such a deletion in the critical sequence of OR region. Antigenicity plot results in our study differ from an earlier report where impact of OR sequence variation on immunogenicity was suggested to be insignificant (Pratt-Riccio et al., 2008). Significance of these observations on immunogenicity needs to be further evaluated with studies on human serum antibodies.

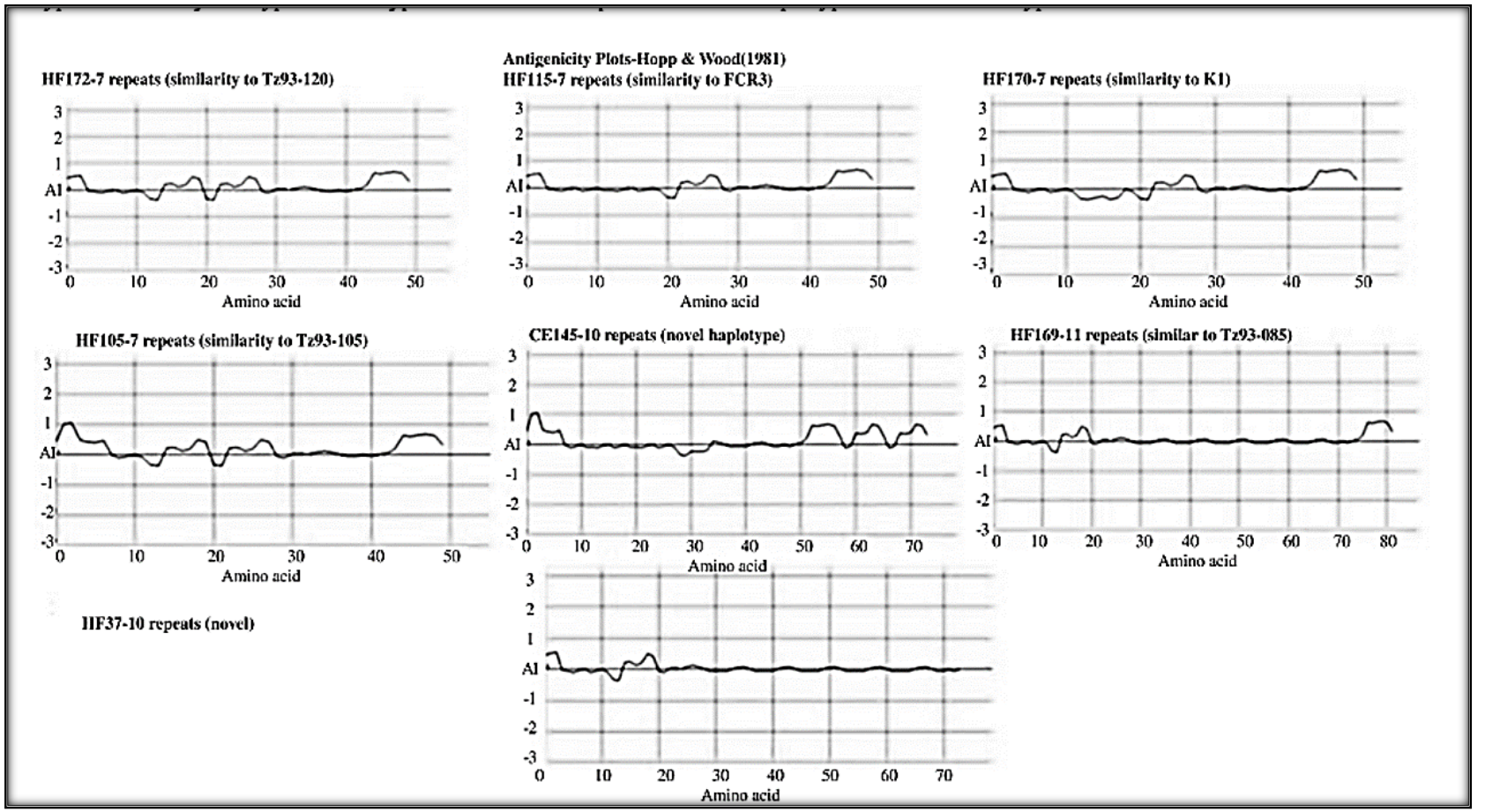



**Figure 5.4 Antigenicity plots for standard (FCR3 & K1) and OR haplotypes from India. The plot displays the variation of the antigenic index as function of amino acid position. Antigenicity plots were generated as predicted by algorithm of Hopp and Woods (1981).**

SR region is a dimorphic locus and is structured into Poly Ser repeats sandwiched between an upstream 13 amino acid (TGTVRGDTEPISD) and downstream 17 amino acid stretch which shows clustered mutations. Based on these variations, the worldwide haplotypes have been grouped into three allelic clusters FCR3, K1 and Honduras (recombinant type). Most of the K1 haplotypes reported in the present dataset are similar to low-frequency African haplotypes. The 14 FCR3 type isolates found in India distribute into 3 novel SR haplotypes unique to these samples. Three Indian isolates CE145, HF37 and OR618 cluster with Honduras type isolate Tz03-086 (Tanzania), Ti-T9_96 (Thailand) and Tz93-105 (Tanzania) are shown in Figure 2B. The HF37 is a novel Honduras haplotype unreported earlier. Statistical results of polymorphism data computed for SR and OR region indicate that genetic diversity in the SR region is higher in comparison to the OR region in Indian field isolates. The overall genetic diversity based on the values for both SR and OR obtained for the present dataset indicates maintenance of medium-range genetic diversity in the N-terminal region of PfSERA5 from Indian isolates. This is evident when compared with earlier reports from other countries (Tanabe et al., 2012) .

**Conclusion**

Our study for the first time characterizes the sequence diversity in PfSERA5 gene from India. The sequence diversity results adds to the antigenic repertoire of PfSERA5 worldwide isolates in the N-terminal region. Several novel OR (2) and SR (4) haplotypes were identified in this study. The predominance of African haplotypes in Indian field isolates suggests probable gene flow from Africa. Further evaluation of these novel mutations by antibody profiling would aid to understand and establish variant specific immunity which have implications in malaria vaccine design. (**Summary of PfSERA5 sequences submitted to the NCBI database is detailed in Appendix II).**

# CHAPTER 6

# *IN SILICO* APPROACH TO ASCERTAIN THE CALCIUM DEPENDENT ROLE OF *Plasmodium falciparum* SERA5

# *In silico* approach to ascertain the calcium dependent role of *Plasmodium falciparum* SERA5

## Introduction

The increasing resistance of malaria towards the existing drugs has greatly hindered the control of this disease (Pandey, Barkan, Sali, & Rosenthal, 2009). The process by which *P.falciparum* merozoites exit from host infected erythrocytes is known as egress (Lew, 2011). Proteases participating in egress of malaria parasite are considered to be potential new targets for antimalarial therapy and (P J Rosenthal, 1998; Philip J Rosenthal, 2004) targeting them requires deeper understanding about their exact functional roles.

Parasite egress from erythrocytes is a tightly regulated and a multi-step process involving rapid degradation of PVM (Parasitophorous Vacuolar Membrane) and erythrocyte membrane (Dvorin et al., 2010; Glushakova et al., 2010). Several *Plasmodial* proteases like falcipains-2 (Sijwali, Koo, Singh, & Rosenthal, 2006), plasmepsins (J. Liu, Gluzman, Drew, & Goldberg, 2005) and a family of nine putative papain like proteases called serine repeat antigen (Aoki et al., 2002; McCoubrie et al., 2007) participate in egress process. PfSERA5 is an abundantly expressed member of the asexual blood stage involved in egress of *P.falciparum* merozoites and is essential for parasite growth (Arastu-Kapur et al., 2008; Blackman, 2008; Delplace, Fortier, Tronchin, Dubremetz, & Vernes, 1987; Yeoh et al., 2007). The molecular events mediated by these proteases in the egress process are yet to be established. PfSERA5 is processed to its respective P47, P56 and P18 subunits by another protease PfSUB-1 (Subtilisin Like Protease-1) (Arastu-Kapur et al., 2008). The central P56 domain is the key player in egress and is indicated to show proteolytic activity and has a 28.8KDa putative enzyme domain and

21.9KDa N-terminal putative prodomain which remain attached together (Alam & Chauhan, 2012) . The crystal structure of the putative enzyme domain is well characterized and shows identity to papain like family of cysteine protease (Hodder et al., 2009). Inspite of all these, the exact function of SERA5 in *P.falciparum* still remains an enigma. Although, putative enzyme domain of SERA5 has a cysteine protease like fold, its active site shows presence of Serine instead of usual cysteine (Hodder et al., 2009). No structural insight into its N-terminal prodomain is available. Recent study support that SERA5 putative enzyme domain has non-catalytic regulatory role in the protease mediated pathway that leads to egress of parasite from host RBC (Stallmach et al., 2015).

Several evidences from literature review report significance of calcium to play a key role in *Plasmodium* for maintaining homeostasis and signalling in the blood-stage malaria (Garcia, 1999). Also, blood stage parasite requires proteolytic activity for processes like invasion, haemoglobin degradation and merozoite escape from RBC. Reports demonstrate in *Plasmodium* species like *P.chabaudi* and *P.falciarum* have cysteine-protease activity stimulated by $Ca^{2+}$ (Farias et al., 2005). Elevation in intracellular calcium, is recently reported to be an important molecular event in parasite egress and it is shown to trigger secretion of proteins from micronemes of *P.falciparum* merozoite and sporozoites of *Toxoplasma gondii* tachyzoites (S. Singh & Chitnis, 2013). Also, it has been observed that elevated intracellular calcium triggers release of PfSUB-1 from exonemes of *P.falciparum* merozoites leading to egress from mature schizonts (Agarwal, Singh, Garg, Chitnis, & Singh, 2013).

In this study, we adopted a two phase approach:

1) To predict the structure of prodomain of PfSERA5 using *in silico* approach as its crystal structure is not available

2) We performed Molecular Dynamic (MD) simulation on PfSERA5 central putative enzyme domain to evaluate effect of calcium on its conformation.

In PfSERA5, though the structure of putative enzyme domain is known the structure of prodomain is not available. This was due to difficulty in co-crystallization at the time of structure determination. Earlier reports show PfSERA5 prodomain to have inhibitory potential on SERA activity (Alam & Chauhan, 2012). However, exact prodomain function is still not clear and hence its structural understanding is very important.

The prodomain model was obtained by *ab inito* modelling using ITASSER web server. Subsequent, stereo-chemical validation of the prodomain 3D model was done by using programmes from SAVES server (The Structure Analysis and Verification Server) like ERRAT and refinement using RAMPAGE analysis available online. MD simulation method was done using GROMACS 4.5.4 package. **The details of these methods have already been discussed in Chapter 2.** Molecular dynamics was performed for PfSERA5 (PDB ID: 2wbf) protein in a water environment and in water containing $Ca^{2+}$ ions at concentrations of 200mM. This concentration was considered based on the crystallization condition of PfSERA5 (PDB id: 2wbf). Appropriate number of calcium ions were added to the system for the required concentration.

## Results and Discussion

### *Ab-initio prediction of Prodomain structure*

The obtained 3D model **(Figure 6.2)** for the submitted prodomain sequence was subjected to stereo-chemical validation using tools from SAVE server. The overall quality factor of prodomain model by ERRAT was found to be 93.85 (expressed as percentage of the protein for which the calculated value falls below 95% rejection limit). Similarly Ramchandran plot of the model as evaluated using RAMPAGE analysis available online showed a) Most of the residues in favoured region: 176 (95.1%). b) Number of residues in allowed region: 5 (2.7%). c) Number of residues in outlier region: 4 (2.2%). This indicates good quality of the obtained model which is shown in Figure 6.1 (A&B).

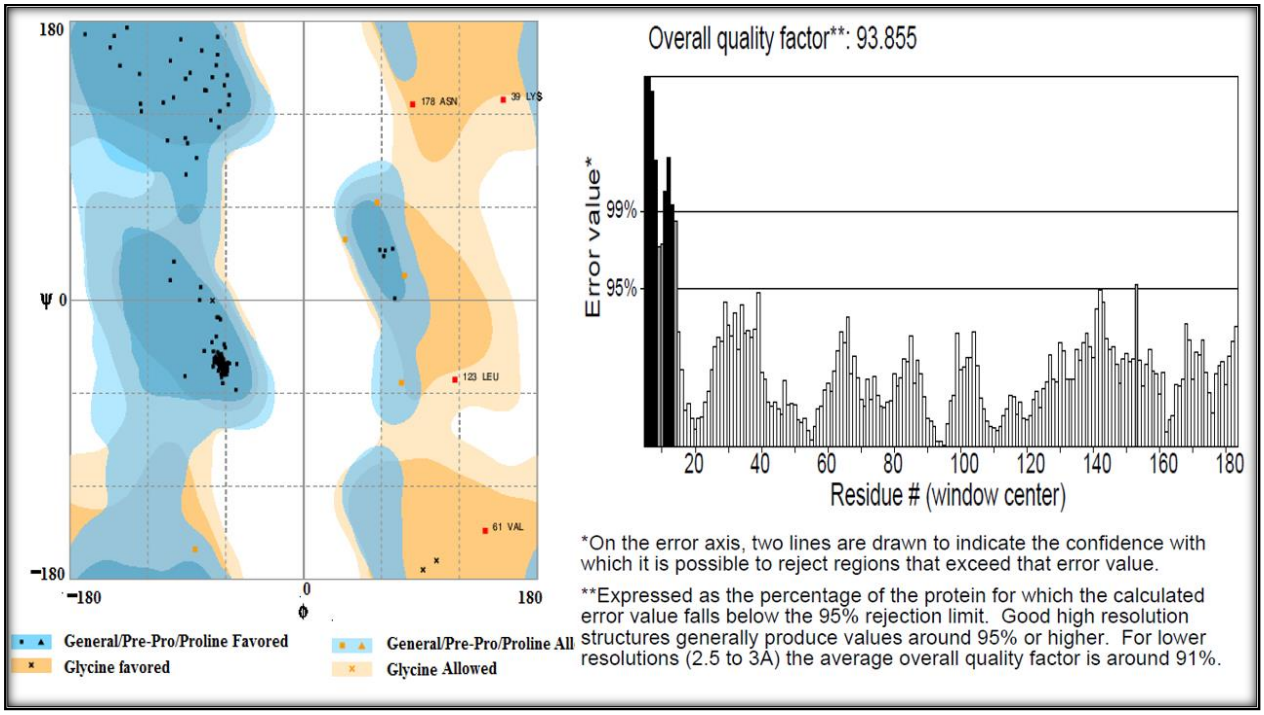



**Figure 6.1: Stereo-chemical validation of PfSERA5 prodomain by A) RAMPAGE: Ramchandran plot which indicates most of the residues in the favourable allowed region. B) ERRAT: Indicates overall quality of the model to be good based on the overall quality factor.**

### *3D homology searches*

The obtained stereo chemically validated prodomain model were searched for 3D structure similarity using DALI server to infer structural and functional insight of the protein model. Results reveal calcium dependent role of PfSERA5 prodomain which gave structural matching hits with proteins like Coelenterazine-binding protein (CBP), obelin, aequolin and sarcoplasmic calcium binding protein etc. in the DALI analysis (**Figure 6.2**).

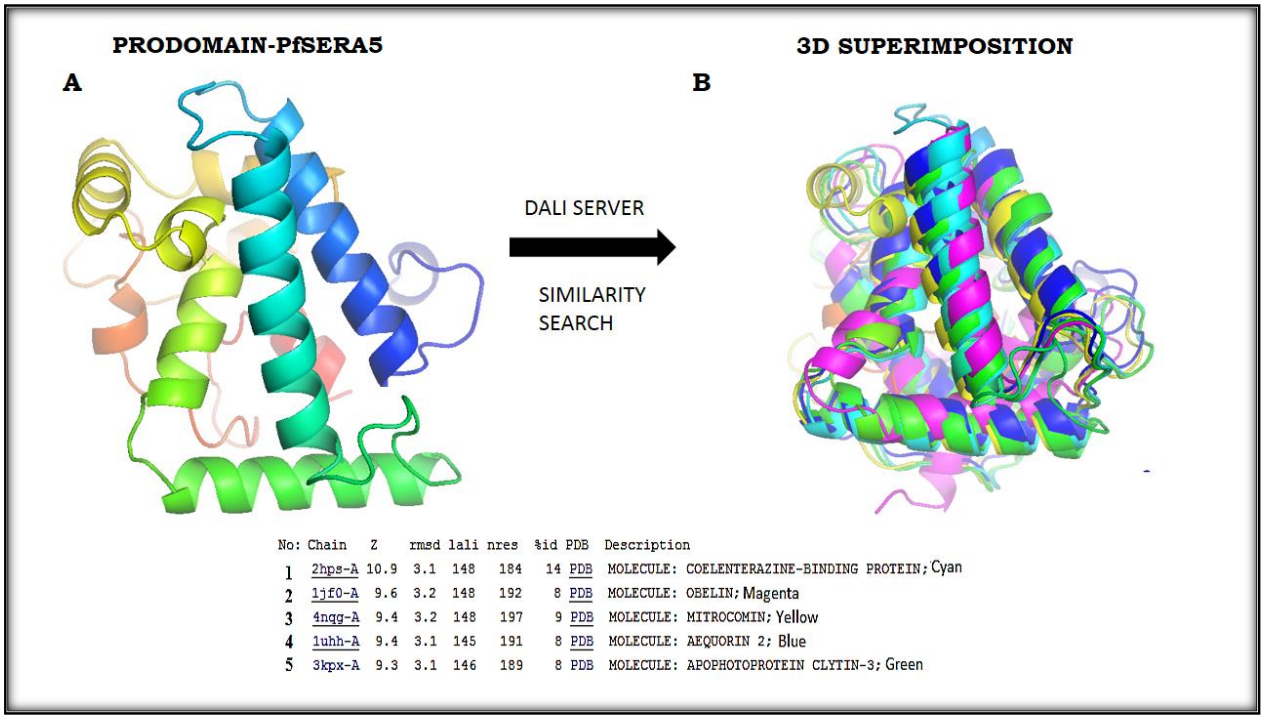



**Figure 6.2 : A) Stereochemically validated prodomain structure B) 3D superimposition of top protein structures with significant Z-score like Coelenterazine binding protein (Cyan), Obelin (Magenta), Mitrocomin (Yellow), Aequarin 2 (Blue) showing structural similarity to Prodomain model of PfSERA5.**

**Analysis of MD simulation**

**a) *RMSD***

The RMSD plot for the crystal structure backbone was obtained as a function of time and the trajectory was analyzed till **25 nanosec**. The RMSD for trajectory in absence of calcium were observed to be unstable till 5ns beyond which the RMSD showed a stable backbone. In comparison, RMSD for trajectory in the presence of calcium was found to be stable all through the simulation run. Both RMSD plots were found to be stabilizing at $3A^0$ (**Figure 6.3**).

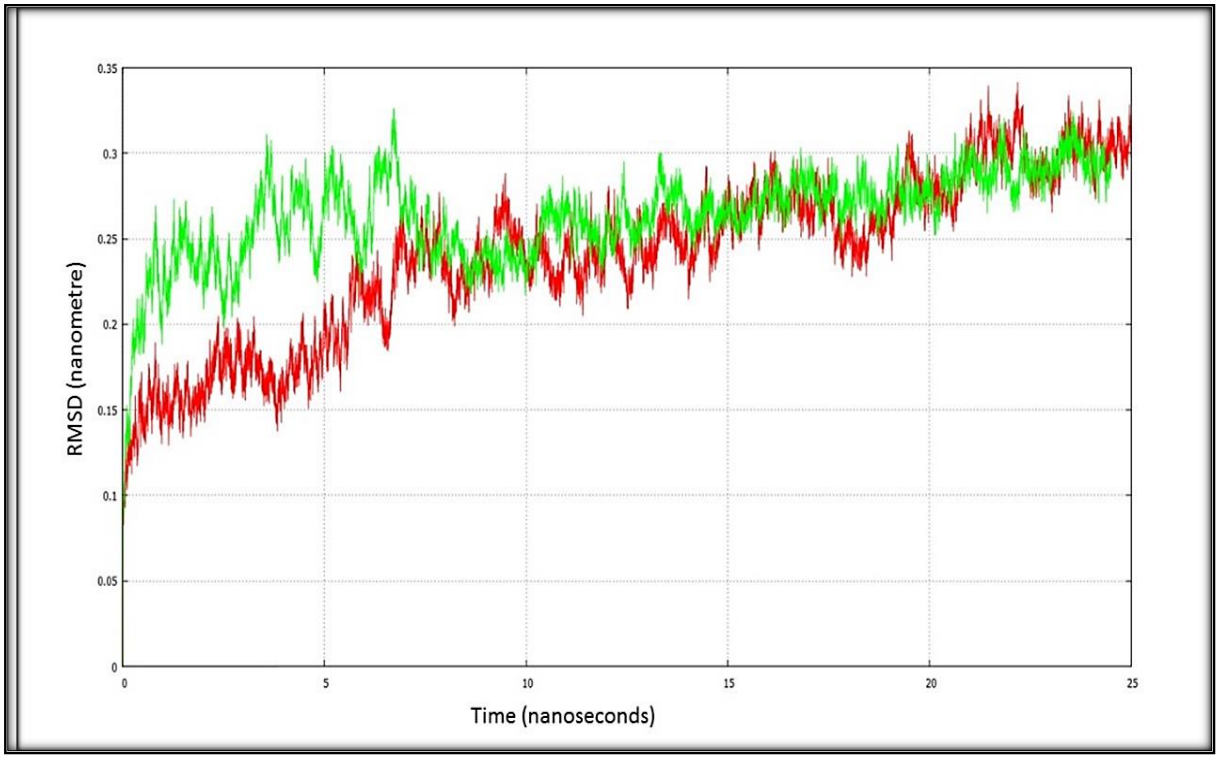



**Figure 6.3 RMSD plots for the two MD simulation trajectories as a function of time. Green-RMSD plot represents in absence of calcium; Red-RMSD plot represents in presence of 200mM calcium.**

**b)** ***Radius of gyration (Rg)***

The plot obtained for radius of gyration (Rg) as a function of time (**25 nanosec**) was observed to be higher for Rg trajectory in presence of calcium (200mM) in comparison to Rg trajectory in absence of calcium (**Figure 6.4**).

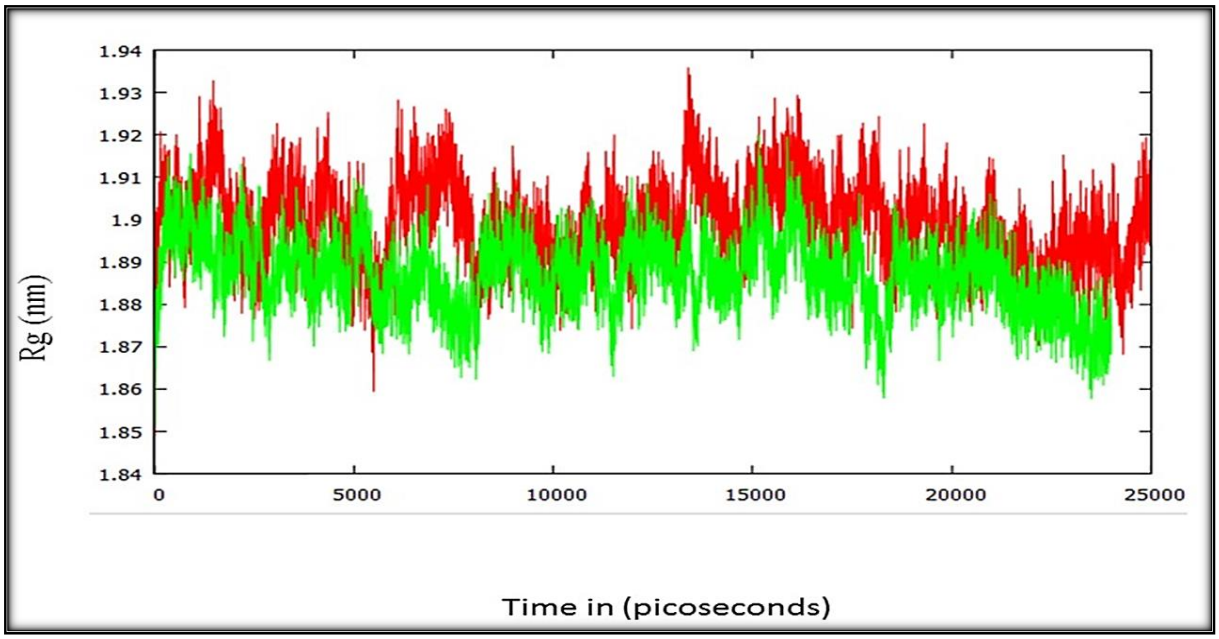



**Figure 6.4 The Rg plots as a function of time were obtained for both conditions presence and absence of calcium. The red colour Rg plot represents for the presence of calcium (200mM) and the green colour Rg plot for the absence of calcium, respectively.**

**c)** ***Root Mean Square Fluctuation (RMSF)***

The RMSF plots was generated for both MD trajectories as a function of position of each residue. The residue based RMSF fluctuations was observed to be higher for 200mM calcium trajectory in comparison to RMSF in absence of calcium (**Figure 6.5**).

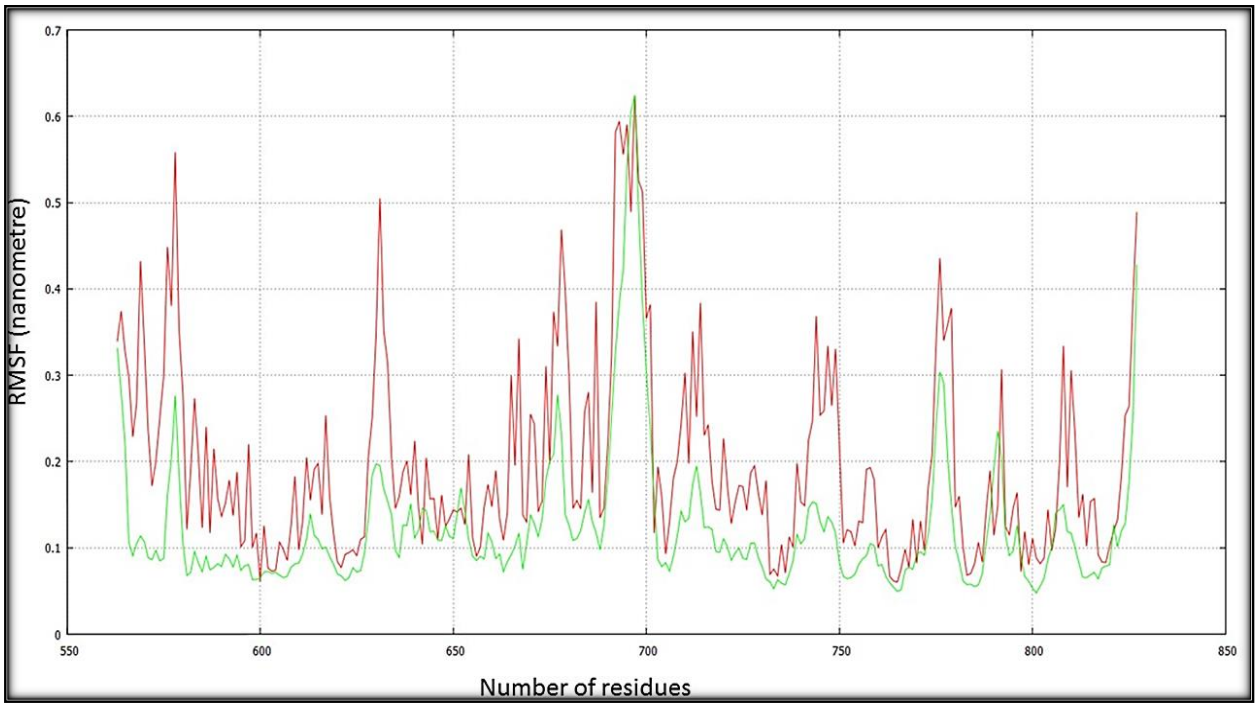



**Figure 6.5 RMSF plots of 200mM calcium (Red) and absence of calcium (Green).**

***Secondary structure annotation of MD simulated structures***

Superimposition of coordinate files in PDB format were done to evaluate conformational changes. Alignment of representative conformations in presence and absence of calcium highlighting the regions that change most are shown. They were obtained as snapshots from MD trajectory at 5ns, 10ns, 15ns, 20ns for both in presence and absence of calcium. Changes were observed with respect to disordered loop region, active site cleft and large loops in the protein structure (**Figure 6.6**).

The coordinate files of MD simulated structures in PDB format for both in the presence and absence of calcium were uploaded in online Poly view 2D tool to generate the secondary structure information in a visual format for thorough interpretation of the local conformational changes of secondary structure annotation and comparisons were made across these two systems. These visible changes have been represented and marked in **Figure 6.7.**

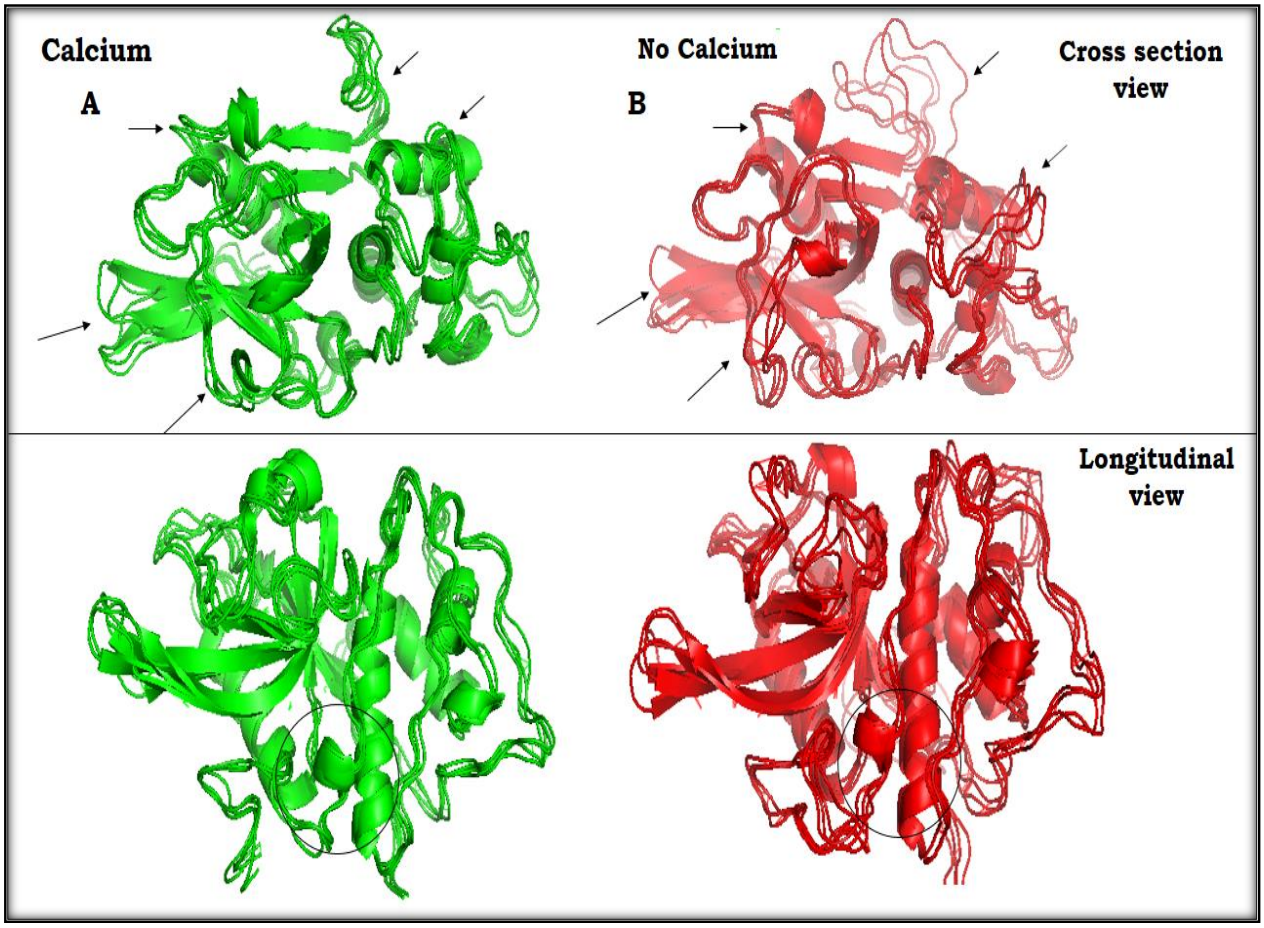



**Figure 6.6 Representing superimposition of snapshots taken from trajectories both: A) In presence (Green) and B) Absence of calcium (Red) at 5ns, 10ns, 15ns, 20ns of MD simulation. Specific differences are marked with arrows and circles.**

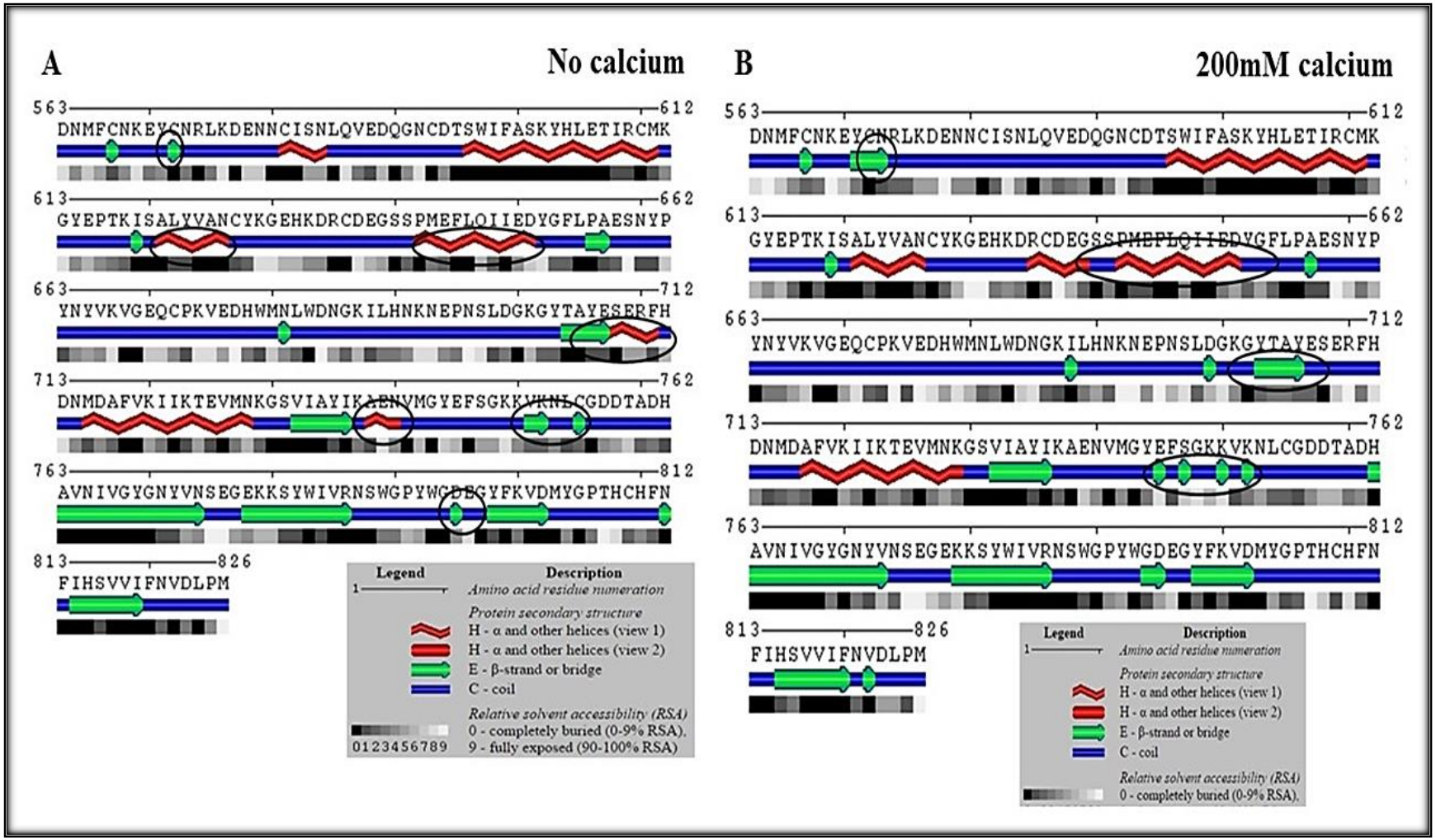



**Figure 6.7 Local conformational changes in secondary structure in absence of calcium A) and presence of calcium B) Differences are marked in dark circles.**

PfSERA5 central domain, is implicated to play an important role in parasite egress. Recent studies on PfSERA5 central domain (P56) demonstrate the processing of a 6KDa (P6) from the C-terminal side of this domain leading to processing of P56 to P50. Mass spectroscopic mapping has revealed this cleavage site to be (839DYYL⬇KAASP846) (Stallmach et al., 2015). However, these studies also show that this P6 is unessential for SERA5 function in contrast to earlier indications of P6 being an important regulator (Kanodia et al., 2014). Besides this, the prodomain of this central region on the N-terminal side is shown to have inhibitory role on the active site (Alam & Chauhan, 2012). Several other roles have been speculated for prodomain like chaperone function, trafficking of the enzyme to its destination and substrate recognition (Alam & Chauhan, 2012). Based on these, the prodomain serves to be an important regulator for PfSERA5 function. Failure to co-crytallize prodomain along with its catalytic domain have marred the study of exact roles of prodomain on activity of SERA putative enzyme domain (Hodder et al., 2009). To understand this, in our study we have attempted to predict the structure of prodomain. Our stereo chemically validated *ab initio* model shows structural homology to proteins having calcium binding property.

Additionally, reports from enzymatic proteins like the GIVA phospholipase A2 have indicated that binding of calcium affects mainly the prodomain region and this binding may extend its effects and have implications on the catalytic domain (Hsu et al., 2008). Also, in *Plasmodial* proteases like PfPCK (Calcium dependent protein kinase from *Plasmodium falciparum*) are sensitive to such calcium induced conformational change and calcium dependent activity(Zhao et al., 1994). Based on these evidences our predicted prodomain structure signify the importance of calcium as a key regulator to understand mechanism of prodomain action on the PfSERA5 putative enzyme activity.

In view of the PfSERA5 processed central fragment being released into the parasitoporous vacuole and indication of high calcium in environment during schizont rupture, it is prudent to hypothesize effect of calcium on PfSERA5 activity and stability (Delplace et al., 1987) (Agarwal et al., 2013). In this context, our Molecular Dynamics (MD) of PfSERA5 putative enzyme domain results clearly indicate local conformational differences as observed in polyview-2D and also higher Rg, RMSF values which substantiate flexible, open conformation in presence of calcium. This probably would be useful in understanding mechanisms of binding or non-enzymatic function of PfSERA5.

**Conclusion**

Our results clearly signify calcium dependency of SERA putative enzyme domain, however it needs to be corroborated with experimental proof for further validation and reconfirmation. Also, structure prediction of prodomain support relevance to calcium dependent role. The present *in silico* analysis of PfSERA5 domain would give insight to design experiments like calcium dependent inhibition studies which would aid to characterize role and biochemical substrates of SERA.

# CHAPTER 7

# CONCLUSIONS &

# FUTURE PERSPECTIVE

## CONCLUSIONS

The work presented in this thesis gives the first report from Indian field isolates on sequence diversity of highly expressed SERA genes in both *P.falciparum* and *P.vivax*.

### Contributions on *Plasmodium vivax*

Our major research focus and priority was to evaluate genetic diversity of highly expressed PvSERA genes which seems prudent contribution in a direction to search for novel vaccine candidates for *P.vivax* malaria. As geographic regions in India represent huge malaria burden from *P.vivax* malaria annually and also report of cases with severe illness and death caused by *P.vivax* malaria are also mostly from India.

The summary of our results on genetic diversity of highly expressed PvSERA4 and PvSERA5 is summarized as follows:

#### *C-terminal region*

1) PvSERA5 C-terminal region showed high genetic diversity with insertions, deletions and SNPs in field sequences in comparison to reference strain (Sal-I). This resulted in reporting of 15 novel haplotypes from India adding to the earlier antigenic repertoire reports available only from Thailand for this C-terminal region.
2) Our results indicate positive selection operating in PvSERA5 C-terminal which is substantiated by statistically significant excess of dN over dS for the present data set strengthening the case for PvSERA5 as potential target for host immunity.

3) We found complete genetic conservation of PvSERA4 C-terminal region in Indian field isolates, which is an interesting finding and in contrasting to earlier reports from Thailand isolates. This needs to be further substantiated by studies from various other geographic regions across the globe.

***Central region***

4) Our results, of genetic diversity characterization of the central regions of the *P.vivax* highly expressed SERA genes both in PvSERA4 and PvSERA5 highlight genetic conservation in multiple field isolates and recommends them to be reliable drug targets for *P.vivax.*
5) This was coupled with *insilico* approach to present structural models of these expressed members like PvSERA4 and PvSERA2, PvSERA5 and its subsequent refinement with MD simulation which will assist in understanding certain unidentified distinctive sequence and structural features of this putative enzyme domain. This will aid in designing future antimalarial chemotherapeutic against this putative enzyme. However, additional functional and biochemical validations are required.

***N-terminal region***

6) Our result of N-terminal region study of both PvSERA4 and PvSERA5 is the first from India and as well its representation worldwide. All haplotypes reported from this study are novel and unique to Indian field isolates.
7) It is interesting to note that diversity analysis of PvSERA4 N-terminal region sheds light on the tetrameric repeats of Salvador strain and the similarity and differences of these repeats as seen in Indian isolates.

8) Our results of PvSERA5 from Indian isolates show novel variations and phylogenetically have haplotypes spread across worldwide strains. Our results point out emerging recombination patterns. This may assist in design of vaccines or to elucidate selection events associated to drug resistance.

**Contributions on *P. falciparum***

I. Worldwide representation of PfSERA5 haplotypes from *P.falciparum* show localization of genetic polymorphism in its immunologically relevant N-terminal region which show repeat polymorphism in the octamer repeat (OR) region and serine repeat (SR) region. This sequence diversity was lacking from malaria endemic region like India. In this regard our results show:

1. Several novel OR (2) and SR (4) are identified in our study. This results adds to the antigenic repertoire of earlier identified 36 OR haplotypes and 67 SR haplotypes from field isolates characterized till date from nine countries.
2. Our finding of predominance of African haplotypes in Indian field isolates suggest probable gene flow. Further evaluation of these novel mutations by antibody profiling would aid to establish variant specific immunity.

II. PfSERA5 is indicated to play a indispensable non catalytic role in parasite egress from blood stage. However, its exact functional role remains enigmatic in egress which is revealed to be a multistep process involving several molecular events. Our *in silico* analysis corroborate this understanding in the following ways:

1. The obtained 3D model of PfSERA5 prodomain by ab-initio prediction show structural similarity to calcium binding proteins strengthening calcium dependent role of PfSERA5 in egress.

2. Our results of MD simulation on putative enzymatic domain of PfSERA5 substantiate conformational differences in presence of calcium. This result ascertain effect of calcium on PFSERA5 activity and stability. Also, establishes the link between two cohesive molecular events in egress 1) high calcium and 2) release of processed central fragment of PfSERA5 in parasitophorous vacuole during egress. However, our *insilico* analysis warrants further experiments to demonstrate role and biochemical substrates SERA in presence of calcium.

## FUTURE PERSPECTIVES

The genetic diversity of highly expressed PvSERA genes represented in our study need to be further studied in detail from various geographic locations which are highly endemic for *P.vivax* malaria. This would aid determining the extent and distribution of immunological relevant diversity of these PvSERA genes. Further, immunological studies would warrant them as novel and potential vaccine candidates to treat *P.vivax* malaria. In *P.falciparum*, although PfSERA5 is well characterized for its vaccine potential novel variants identified from our study and worldwide needs to be further evaluated for their significance in variant specific immunity and immune escape. Structural models of highly expressed PvSERA represented in our study need to be further explored in drug targeting and design of them as novel anti-malaria therapeutics for *P.vivax* malaria. Parasite egress from blood stage is currently a hot topic in malaria research and understanding exact role of SERAs participation in this multistep process remains a conundrum. Our *insilico* analysis which indicates calcium significance for PfSERA5 role in egress would provide lead for further experiments.

# REFERENCES

## References


Adak, T., Kaur, S., & Singh, O. P. (1999). Comparative susceptibility of different members of the Anopheles culicifacies complex to Plasmodium vivax. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *93*, 573–577.

Agarwal, S., Singh, M. K., Garg, S., Chitnis, C. E., & Singh, S. (2013). Ca(2+) -mediated exocytosis of subtilisin-like protease 1: a key step in egress of Plasmodium falciparum merozoites. *Cellular Microbiology*, *15*, 910–21.

Alam, A., & Chauhan, V. S. (2012). Inhibitory potential of prodomain of Plasmodium falciparum protease serine repeat antigen 5 for asexual blood stages of parasite. *PloS One*, *7*, e30452.

Aoki, S., Li, J., Itagaki, S., Okech, B. A., Egwang, T. G., Matsuoka, H., … Horii, T. (2002). Serine repeat antigen (SERA5) is predominantly expressed among the SERA multigene family of Plasmodium falciparum, and the acquired antibody titers correlate with serum inhibition of the parasite growth. *The Journal of Biological Chemistry*, *277*, 47533–40.

Arastu-Kapur, S., Ponder, E. L., Fonovic, U. P., Yeoh, S., Yuan, F., Fonovic, M., … Bogyo, M. (2008). Identification of proteases that regulate erythrocyte rupture by the malaria parasite Plasmodium falciparum. *Nat Chem Biol*, *4*, 203–213.

Arisue, N., Kawai, S., Hirai, M., Palacpac, N. M. Q., Jia, M., Kaneko, A., … Horii, T. (2011). Clues to evolution of the SERA multigene family in 18 Plasmodium species. *PloS One*, *6*, e17775.

Arnott, A., Barry, A. E., & Reeder, J. C. (2012). Understanding the population genetics of Plasmodium vivax is essential for malaria control and elimination. *Malaria Journal*, *11*, 14.

Aucan, C., Traoré, Y., Fumoux, F., Rihet, P., Traore, Y., Fumoux, F., & Rihet, P. (2001). Familial correlation of immunoglobulin G subclass responses to Plasmodium falciparum antigens in Burkina Faso. *Infection and Immunity*, *69*, 996–1001.

Baird, J. K. (2004). Chloroquine Resistance in Plasmodium vivax. *Antimicrobial Agents and Chemotherapy*, *48*, 4075–4083.

Baker, D., & Sali, A. (2001). Protein structure prediction and structural genomics. *Science (New York, N.Y.)*, *294*, 93–96.

Banic, D. M., de Oliveira-Ferreira, J., Pratt-Riccio, L. R., Conseil, V., Gonçalves, D., Fialho, R. R., … Camus, D. (1998). Immune response and lack of immune response to Plasmodium falciparum P126 antigen and its amino-terminal repeat in malaria-infected

humans. *The American Journal of Tropical Medicine and Hygiene*, *58*, 768–74.

Barry, A. E., & Arnott, A. (2014). Strategies for designing and monitoring malaria vaccines targeting diverse antigens. *Frontiers in Immunology*, *5*, 359.

Belnoue, E., Costa, F. T. M., Frankenberg, T., Vigario, A. M., Voza, T., Leroy, N., … Renia, L. (2004). Protective T Cell Immunity against Malaria Liver Stage after Vaccination with Live Sporozoites under Chloroquine Treatment. *The Journal of Immunology*, *172*, 2487–2495.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., & Hermans, J. (1981). Interaction Models for Water in Relation to Protein Hydration. In B. Pullman (Ed.), *Intermolecular Forces SE - 21* (Vol. 14, pp. 331–342). Springer Netherlands.

Blackman, M. J. (2008). Malarial proteases and host cell egress: an "emerging" cascade. *Cellular Microbiology*, *10*, 1925–1934.

Bowie, J. U., Luthy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science (New York, N.Y.)*, *253*, 164–170.

Brito, C. F. A. de, & Ferreira, M. U. (2011). Molecular markers and genetic diversity of Plasmodium vivax. *Memórias Do Instituto Oswaldo Cruz*, *106 Suppl*, 12–26.

Brocchieri, L. (2001). Low-complexity regions in Plasmodium proteins: in search of a function. *Genome Research*, *11*, 195–197.

Carpenter, C. C. J., Pearson, G. W., Mitchell, V. S., & Oaks Jr, S. C. (1991). *Malaria:: Obstacles and Opportunities*. National Academies Press.

Carter, R., & Mendis, K. N. (2002). Evolutionary and historical aspects of the burden of malaria. *Clinical Microbiology Reviews*, *15*, 564–594.

Cino, E., Wong-ekkabut, J., Karttunen, M., & Choy, W.-Y. (2011). Microsecond Molecular Dynamics Simulations of Intrinsically Disordered Proteins Involved in the Oxidative Stress Response. *PLoS ONE*, *6*, e27371.

Colovos, C., & Yeates, T. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science*, *2*, 1511–1519.

Combet, C., Blanchet, C., Geourjon, C., & Deléage, G. (2000). NPS@: Network Protein Sequence Analysis. *Trends in Biochemical Sciences*, *25*, 147–150.

Conway, D. J. (1997). Natural selection on polymorphic malaria antigens and the search for a vaccine. *Parasitology Today (Personal Ed.)*, *13*, 26–9.

Cox, F. E. (2010). History of the discovery of the malaria parasites and their vectors. *Parasites & Vectors*, *3*, 5.

Cox, F. E. G. (2002). History of human parasitology. *Clinical Microbiology Reviews*, *15*(4), 595–612.

Crewther, P. E., Matthew, M. L., Flegg, R. H., & Anders, R. F. (1996). Protective immune responses to apical membrane antigen 1 of Plasmodium chabaudi involve recognition of strain-specific epitopes. *Infection and Immunity*, *64*, 3310–3317.

Das, A., Anvikar, A. R., Cator, L. J., Dhiman, R. C., Eapen, A., Mishra, N., … Valecha, N. (2012). Malaria in India: the center for the study of complex malaria in India. *Acta Tropica*, *121*, 267–73.

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), 161–5. http://doi.org/10.1006/mcpr.1995.0025

de Souza-Neiras, W. C., de Melo, L. M. S., & Machado, R. L. D. (2007). The genetic diversity of Plasmodium vivax--a review. *Memórias Do Instituto Oswaldo Cruz*, *102*, 245–54.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA, USA.: Schrödinger. Retrieved from http://www.pymol.org

Delplace, P., Fortier, B., Tronchin, G., Dubremetz, J. F., & Vernes, A. (1987). Localization, biosynthesis, processing and isolation of a major 126 kDa antigen of the parasitophorous vacuole of Plasmodium falciparum. *Molecular and Biochemical Parasitology*, *23*, 193–201.

Dhingra, N., Jha, P., Sharma, V. P., Cohen, A. A., Jotkar, R. M., Rodriguez, P. S., … Peto, R. (2010). Adult and child malaria mortality in India: a nationally representative mortality survey. *Lancet*, *376*, 1768–74.

Dvorin, J. D., Martyn, D. C., Patel, S. D., Grimley, J. S., Collins, C. R., Hopp, C. S., … Duraisingh, M. T. (2010). A plant-like kinase in Plasmodium falciparum regulates parasite egress from erythrocytes. *Science (New York, N.Y.)*, *328*, 910–912.

Escalante, A. A., Lal, A. A., & Ayala, F. J. (1998). Genetic polymorphism and natural selection in the malaria parasite Plasmodium falciparum. *Genetics*, *149*, 189–202.

Farias, S. L., Gazarini, M. L., Melo, R. L., Hirata, I. Y., Juliano, M. A., Juliano, L., & Garcia, C. R. S. (2005). Cysteine-protease activity elicited by Ca2+ stimulus in Plasmodium. *Molecular and Biochemical Parasitology*, *141*, 71–9.

Farooq, U., Malla, N., & Dubey, M. L. (2009). Genetic polymorphism in msp-2, ama-1 and csp genes in Plasmodium falciparum field isolates from north and north-western India. *Journal of Vector Borne Diseases*, *46*, 109–16.

Ferreira, M. U., Nunes, S., & Wunderlich, G. (2004). Antigenic Diversity and Immune Evasion by Malaria Parasites, *11*, 987–995.

Ferreira, M. U., Zilversmit, M., & Wunderlich, G. (2007). Origins and evolution of antigenic diversity in malaria parasites. *Current Molecular Medicine*, *7*, 588–602.

Fiser, A., Do, R. K., & Sali, A. (2000). Modeling of loops in protein structures. *Protein Science : A Publication of the Protein Society*, *9*, 1753–1773.

Fiser, A., & Sali, A. (2003). ModLoop: automated modeling of loops in protein structures. *Bioinformatics (Oxford, England)*, *19*, 2500–2501.

Fox, B. A., Horii, T., & Bzik, D. J. (2002). Plasmodium falciparum: fine-mapping of an epitope of the serine repeat antigen that is a target of parasite-inhibitory antibodies. *Experimental Parasitology*, *101*, 69–72.

Fox, B. A., Xing-Li, P., Suzue, K., Horii, T., & Bzik, D. J. (1997). Plasmodium falciparum: an epitope within a highly conserved region of the 47-kDa amino-terminal domain of the serine repeat antigen is a target of parasite-inhibitory antibodies. *Experimental Parasitology*, *85*, 121–34.

Gaikwad, S., Ashma, R., Kumar, N., Trivedi, R., & Kashyap, V. K. (2005). Host microsatellite alleles in malaria predisposition? *Malaria Journal*, *4*, 50.

Garcia, C. R. S. (1999). Calcium Homeostasis and Signaling in the Blood-stage Malaria Parasite, *15*(12), 2–5.

Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., … Barrell, B. (2002). Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature*, *419*, 498–511. http://doi.org/10.1038/nature01097

Genton, B., Betuela, I., Felger, I., Al-Yaman, F., Anders, R. F., Saul, A., … Alpers, M. P. (2002). A recombinant blood-stage malaria vaccine reduces Plasmodium falciparum density and exerts selective pressure on parasite populations in a phase 1-2b trial in Papua New Guinea. *The Journal of Infectious Diseases*, *185*, 820–827.

Gilson, P. R., & Crabb, B. S. (2009). Morphology and kinetics of the three distinct phases of red blood cell invasion by Plasmodium falciparum merozoites. *International Journal for Parasitology*, *39*, 91–96.

Glushakova, S., Humphrey, G., Leikina, E., Balaban, A., Miller, J., & Zimmerberg, J. (2010). New stages in the program of malaria parasite egress imaged in normal and sickle erythrocytes. *Current Biology : CB*, *20*, 1117–21.

Greenwood, B. (2002). The molecular epidemiology of malaria. *Tropical Medicine & International Health : TM & IH*, *7*, 1012–21.

Greenwood, B., & Mutabingwa, T. (2002). Malaria in 2002. *Nature*, *415*, 670–2.

Gunasekera, A. M., Wickramarachchi, T., Neafsey, D. E., Ganguli, I., Perera, L., Premaratne, P. H., … Wirth, D. F. (2007). Genetic diversity and selection at the Plasmodium vivax apical membrane antigen-1 (PvAMA-1) locus in a Sri Lankan population. *Molecular Biology and Evolution*, *24*, 939–47.

Haas, L. F. (1994). Pierre Joseph Pelletier (1788-1842) and Jean Bienaime Caventou (1795-1887). *Journal of Neurology, Neurosurgery, and Psychiatry*, *57*, 1333.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Hay, S. I., Gething, P. W., & Snow, R. W. (2010). India's invisible malaria burden. *Lancet (London, England)*, *376*, 1716–1717.

Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, *4*, 435–447.

Hodder, A. N., Drew, D. R., Epa, V. C., Delorenzi, M., Bourgon, R., Miller, S. K., … Crabb, B. S. (2003). Enzymic , Phylogenetic , and Structural Characterization of the Unusual Papain-like Protease Domain of Plasmodium falciparum SERA5 *, *278*, 48169–48177.

Hodder, A. N., Malby, R. L., Clarke, O. B., Fairlie, W. D., Colman, P. M., Crabb, B. S., & Smith, B. J. (2009). Structural Insights into the Protease-like Antigen Plasmodium falciparum SERA5 and Its Noncanonical. *Journal of Molecular Biology*, *392*, 154–165.

Holm, L., & Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Research*, *38*, 545–549.

Hopp, T. P., & Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences*, *78*, 3824–3828.

Horii, T., Shirai, H., Jie, L., Ishii, K. J., Palacpac, N. Q., Tougan, T., … Takahashi, M. (2010). Evidences of protection against blood-stage infection of Plasmodium falciparum by the novel protein vaccine SE36. *Parasitology International*, *59*, 380–6.

Hsu, Y.-H., Burke, J. E., Stephens, D. L., Deems, R. A., Li, S., Asmus, K. M., … Dennis, E. A. (2008). Calcium Binding Rigidifies the C2 Domain and the Intradomain Interaction of GIVA Phospholipase A2 as Revealed by Hydrogen/Deuterium Exchange Mass Spectrometry. *Journal of Biological Chemistry*, *283*, 9820–9827.

Hughes, A. L. (1992). Positive selection and interallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of Plasmodium falciparum. *Molecular Biology and Evolution*, *9*, 381–393.

Inselburg, J., Bathurst, I. C., Kansopon, J., Barr, P. J., & Rossan, R. (1993). Protective immunity induced in Aotus monkeys by a recombinant SERA protein of Plasmodium falciparum: further studies using SERA 1 and MF75.2 adjuvant. *Infection and Immunity*, *61*, 2048–2052.

Inselburg, J., Bzik, D. J., Li, W. B., Green, K. M., Kansopon, J., Hahm, B. K., … Rossan, R. N. (1991). Protective immunity induced in Aotus monkeys by recombinant SERA proteins of Plasmodium falciparum. *Infection and Immunity*, *59*, 1247–1250.

Jongwutiwes, S., Putaporntip, C., Iwasaki, T., Sata, T., & Kanbara, H. (2004). Naturally acquired Plasmodium knowlesi malaria in human, Thailand. *Emerging Infectious Diseases*, *10*, 2211–2213.

Joshi, H., Prajapati, S. K., Verma, A., Kang'a, S., & Carlton, J. M. (2008). Plasmodium vivax in India. *Trends in Parasitology*, *24*, 228–235.

Kanodia, S., Kumar, G., Rizzi, L., Pedretti, A., Hodder, A. N., Romeo, S., & Malhotra, P. (2014). Synthetic peptides derived from the C-terminal 6kDa region of Plasmodium falciparum SERA5 inhibit the enzyme activity and malaria parasite development. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *1840*(9), 2765–2775.

Kiefer, M. C., Crawford, K. A., Boley, L. J., Landsberg, K. E., Gibson, H. L., Kaslow, D. C., & Barr, P. J. (1996). Identification and cloning of a locus of serine repeat antigen (sera)-related genes from Plasmodium vivax. *Molecular and Biochemical Parasitology*, *78*, 55–65.

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*, 194–198.

Kondrashin, A. V, & Rooney, W. (1992). Overview: epidemiology of malaria and its control in countries of the WHO South-East Asia region. *The Southeast Asian Journal of Tropical Medicine and Public Health*, *23*, 13–22.

Kumar, A., Valecha, N., Jain, T., & Dash, A. P. (2007). Burden of malaria in India: retrospective and prospective view. *The American Journal of Tropical Medicine and Hygiene*, *77*, 69–78.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*, 283–291.

Laurie, A., & Jackson, R. (2005). Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics (Oxford, England)*, *21*, 1908–1916.

Lee, M. C. S., & Fidock, D. A. (2008). Arresting malaria parasite egress from infected red blood cells It ' s a mod mod tRNA world, *4*, 161–162.

Lew, V. L. (2011). Malaria : Surprising Mechanism of Merozoite Egress Revealed. *Current Biology*, *21*(9), 314–316.

Li, J., Mitamura, T., Fox, B. A., Bzik, D. J., & Horii, T. (2002). Differential localization of processed fragments of Plasmodium falciparum serine repeat antigen and further processing of its N-terminal 47 kDa fragment. *Parasitology International*, *51*, 343–352.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*, 1451–1452.

Ling, I. T., Ogun, S. A., & Holder, A. A. (1994). Immunization against malaria with a recombinant protein. *Parasite Immunology*, *16*, 63–67.

Liu, J., Gluzman, I. Y., Drew, M. E., & Goldberg, D. E. (2005). The role of Plasmodium falciparum food vacuole plasmepsins. *The Journal of Biological Chemistry*, *280*, 1432–1437.

Liu, Q., Ferreira, M. U., Ndawi, B. T., Ohmae, H., Adagu, I. S., Morikawa, T., … Kawamoto, F. (2000). Sequence diversity of serine repeat antigen gene exon II of Plasmodium falciparum in worldwide collected wild isolates. *The Southeast Asian Journal of Tropical Medicine and Public Health*, *31*, 808–17.

Liu, W., Li, Y., Shaw, K., Learn, G., Plenderleith, L., Malenke, J., … Sharp, P. (2014). African origin of the malaria parasite Plasmodium vivax. *Nature Communications*, *5*.

Lopez, A. C., Ortiz, A., Coello, J., Sosa-Ochoa, W., Torres, R. E. M., Banegas, E. I., … Fontecha, G. A. (2012). Genetic diversity of Plasmodium vivax and Plasmodium falciparum in Honduras. *Malaria Journal*, *11*, 391.

Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., … Richardson, D. C. (2003). Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, *50*, 437–450.

Luthy, R., Bowie, J. U., & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, *356*, 83–85.

Marsh, K., & Howard, R. J. (1986). Antigens induced on erythrocytes by P. falciparum: expression of diverse and conserved determinants. *Science (New York, N.Y.)*, *231*, 150–153.

McCoubrie, J. E., Miller, S. K., Sargeant, T., Good, R. T., Hodder, A. N., Speed, T. P., … Crabb, B. S. (2007). Evidence for a common role for the serine-type Plasmodium falciparum serine repeat antigen proteases: implications for vaccine and drug design.

*Infection and Immunity*, *75*, 5565–74.

McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics (Oxford, England)*, *16*, 404–405.

Mercereau-Puijalon, O., Fandeur, T., Guillotte, M., & Bonnefoy, S. (1991). Parasite features impeding malaria immunity: antigenic diversity, antigenic variation and poor immunogenicity. *Research in Immunology*, *142*, 690–697.

Miller, S. K., Good, R. T., Drew, D. R., Delorenzi, M., Sanders, P. R., Hodder, A. N., … Crabb, B. S. (2002). A Subset of Plasmodium falciparum SERA Genes Are Expressed and Appear to Play an Important Role in the Erythrocytic Cycle. *Journal of Biological Chemistry*, *277*, 47524–47532.

Morimatsu, K., Morikawa, T., Tanabe, K., Bzik, D. J., & Horii, T. (1997). Sequence diversity in the amino-terminal 47 kDa fragment of the Plasmodium falciparum serine repeat antigen. *Molecular and Biochemical Parasitology*, *86*, 249–254.

Mugal, C. F., Wolf, J. B. W., & Kaj, I. (2014). Why time matters: codon evolution and the temporal dynamics of dN/dS. *Molecular Biology and Evolution*, *31*, 212–31.

National Vector Borne Disease Control Programme. (n.d.). NVBDCP-Malaria control Programmes in India. Retrieved July 31, 2015, from http://nvbdcp.gov.in/malaria-new.html

Neafsey, D. E., Galinsky, K., Jiang, R. H. Y., Young, L., Sykes, S. M., Saif, S., … Carlton, J. M. (2012). The malaria parasite Plasmodium vivax exhibits greater genetic diversity than Plasmodium falciparum. *Nature Genetics*, *44*, 1046–50.

Nurisso, A., Daina, A., & Walker, R. C. (2012). A practical introduction to molecular dynamics simulations: applications to homology modeling. *Methods in Molecular Biology (Clifton, N.J.)*, *857*, 137–173.

Ocana-Morgner, C., Mota, M. M., & Rodriguez, A. (2003). Malaria Blood Stage Suppression of Liver Stage Immunity by Dendritic Cells. *Journal of Experimental Medicine*, *197*, 143–151.

Okech, B. A., Nalunkuma, A., Okello, D., Pang, X. L., Suzue, K., Li, J., … Egwang, T. G. (2001). Natural human immunoglobulin G subclass responses to Plasmodium falciparum serine repeat antigen in Uganda. *The American Journal of Tropical Medicine and Hygiene*, *65*, 912–7.

Okech, B., Mujuzi, G., Ogwal, A., Shirai, H., Horii, T., & Egwang, T. G. (2006). High titers of IgG antibodies against plasmodium falciparum serine repeat antigen 5 (sera5) are associated with protection against severe malaria in ugandan children. *The American*

*Journal of Tropical Medicine and Hygiene*, *74*, 191–197.

Palacpac, N. M. Q., Arisue, N., Tougan, T., Ishii, K. J., & Horii, T. (2011). Plasmodium falciparum serine repeat antigen 5 (SE36) as a malaria vaccine candidate. *Vaccine*, *29*, 5837–45. http://doi.org/10.1016/j.vaccine.2011.06.052

Palacpac, N. M. Q., Leung, B. W. Y., Arisue, N., Tanabe, K., Sattabongkot, J., Tsuboi, T., … Horii, T. (2006). Plasmodium vivax serine repeat antigen (SERA) multigene family exhibits similar expression patterns in independent infections. *Molecular and Biochemical Parasitology*, *150*, 353–358.

Pandey, K. C., Barkan, D. T., Sali, A., & Rosenthal, P. J. (2009). Regulatory elements within the prodomain of Falcipain-2, a cysteine protease of the malaria parasite Plasmodium falciparum. *PloS One*, *4*, e5694. http://doi.org/10.1371/journal.pone.0005694

Perera, K. L., Handunnetti, S. M., Holm, I., Longacre, S., & Mendis, K. (1998). Baculovirus merozoite surface protein 1 C-terminal recombinant antigens are highly protective in a natural primate model for human Plasmodium vivax malaria. *Infection and Immunity*, *66*(4), 1500–1506.

Phillips, R. S. (2001). Current status of malaria and potential for control. *Clinical Microbiology Reviews*, *14*, 208–226.

Pratt-Riccio, L. R., Sallenave-Sales, S., de Oliveira-Ferreira, J., da Silva, B. T., Guimarães, M. L., Santos, F., … Banic, D. M. (2008). Evaluation of the genetic polymorphism of Plasmodium falciparum P126 protein (SERA or SERP) and its influence on naturally acquired specific antibody responses in malaria-infected individuals living in the Brazilian Amazon. *Malaria Journal*, *7*, 144.

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*, 430–435.

Price, R. N., Tjitra, E., Guerra, C. A., Yeung, S., White, N. J., & Anstey, N. M. (2007). Vivax Malaria: Neglected and Not Benign. *The American Journal of Tropical Medicine and Hygiene*, *77*, 79–87.

Rajesh, V., Elamaran, M., Vidya, S., Gowrishankar, M., Kochar, D., & Das, A. (2007). Plasmodium vivax: genetic diversity of the apical membrane antigen-1 (AMA-1) in isolates from India. *Experimental Parasitology*, *116*, 252–256.

Ramasamy, R., Ramasamy, M., & Yasawardena, S. (2001). Antibodies and Plasmodium falciparum merozoites. *Trends in Parasitology*, *17*, 194–7.

Ramírez-Soriano, A., Ramos-Onsins, S. E., Rozas, J., Calafell, F., & Navarro, A. (2008).

Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics*, *179*, 555–67.

Ranjit, M. R., & Sharma, Y. D. (1999). Genetic polymorphism of falciparum malaria vaccine candidate antigen genes among field isolates in India. *The American Journal of Tropical Medicine and Hygiene*, *61*, 103–8.

Reeder, J. C., & Brown, G. V. (1996). Antigenic variation and immune evasion in Plasmodium falciparum malaria. *Immunology and Cell Biology*, *74*, 546–554.

Riccio, E. K. P., Zalis, M. G., Guedes, H. C. B., Banic, D. M., Souza, J. M. de, Alecrim, W., … Ferreira-da-Cruz, M. de F. (2005). Genetic polymorphism of the serine rich antigen N-terminal region in Plasmodium falciparum field isolates from Brazil. *Memórias Do Instituto Oswaldo Cruz*, *100*, 47–49.

Richards, J. S., & Beeson, J. G. (2009). The future for blood-stage vaccines against malaria. *Immunology and Cell Biology*, *87*, 377–90.

Rosenthal, P. J. (1998). Proteases of Malaria Parasites : New Targets for Chemotherapy, *4*(1), 49–57.

Rosenthal, P. J. (1998). Proteases of malaria parasites: new targets for chemotherapy. *Emerging Infectious Diseases*, *4*, 49–57.

Rosenthal, P. J. (2004). Cysteine proteases of malaria parasites. *International Journal for Parasitology*, *34*, 1489–1499.

Ross, R. (1897). On some Peculiar Pigmented Cells Found in Two Mosquitos Fed on Malarial Blood. *British Medical Journal*, *2*, 1786–1788.

Roy, S. W., Ferreira, M. U., & Hartl, D. L. (2008). Evolution of allelic dimorphism in malarial surface antigens. *Heredity*, *100*, 103–10.

Safitri, I. (2003). Sequence diversity in the amino-terminal region of the malaria-vaccine candidate serine repeat antigen in natural Plasmodium falciparum populations. *Parasitology International*, *52*, 117–131.

Safitri, I., Jalloh, A., Tantular, I. S., Pusarawati, S., Win, T. T., Liu, Q., … Kawamoto, F. (2003). Sequence diversity in the amino-terminal region of the malaria- vaccine candidate serine repeat antigen in natural Plasmodium. *Parasitology International*, *52*, 117–131. http://doi.org/10.1016/S1383-5769

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*, 406–425.

Sali, A., & Blundell, T. L. (1993). Comparative Protein Modelling by Satisfaction of Spatial

Restraints. *Journal of Molecular Biology*, *234*, 779–815.

Sambrook, J., & Russell, D. W. (2001). Molecular Cloning: A Laboratory Manual, 3rd EditionCold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.

Santamaria, L. (1994). Camillo Golgi as clinical pathologist: epicritical reading of Golgi's works on malaria. *Medicina Nei Secoli*, *6*, 581–608.

Schuler, K., Lu, C., Chang, H. D., Croft, M., Zanetti, M., & Gerloni, M. (2001). Circumvention of MHC class II restriction by genetic immunization. *Vaccine*, *20*, 630–634.

Schultz, L., Wapling, J., Mueller, I., Ntsuke, P. O., Senn, N., Nale, J., … Barry, A. E. (2010). Multilocus haplotypes reveal variable levels of diversity and population structure of Plasmodium falciparum in Papua New Guinea, a region of intense perennial transmission. *Malaria Journal*, *9*, 336.

Sijwali, P. S., Koo, J., Singh, N., & Rosenthal, P. J. (2006). Gene disruptions demonstrate independent roles for the four falcipain cysteine proteases of Plasmodium falciparum. *Molecular and Biochemical Parasitology*, *150*, 96–106.

Singh, S., & Chitnis, C. E. (2013). Flow cytometry-based methods for measurement of cytosolic calcium and surface protein expression in Plasmodium falciparum merozoites. *Methods in Molecular Biology (Clifton, N.J.)*, *923*, 281–290.

Singh, V., Mishra, N., Awasthi, G., Dash, A. P., & Das, A. (2009). Why is it important to study malaria epidemiology in India? *Trends in Parasitology*, *25*, 452–7.

Stallmach, R., Kavishwar, M., Withers-Martinez, C., Hackett, F., Collins, C. R., Howell, S. A., … Blackman, M. J. (2015). Plasmodium falciparum SERA5 plays a non-enzymatic role in the malarial asexual blood-stage lifecycle. *Molecular Microbiology*, *96*, 368–387.

Street, K. (1997). Natural Selection on Polymorphic Malaria Antigens and the Search for a Vaccine, *13*, 0–3.

Sugiyama, T. (1996). Production of recombinant SERA proteins of Plasmodium falciparum in Escherichia coli by using synthetic genes. *Vaccine*, *14*, 1069–1076.

Tajima, F. (1993). Statistical analysis of DNA polymorphism. *Idengaku Zasshi*, *68*, 567–95.

Takala, S. L., & Plowe, C. V. (2009). Genetic diversity and malaria vaccine design, testing and efficacy: preventing and overcoming "vaccine resistant malaria". *Parasite Immunology*, *31*, 560–73.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*, 2731–

2739.

Tanabe, K., Arisue, N., Palacpac, N. M. Q., Yagi, M., Tougan, T., Honma, H., … Horii, T. (2012). Geographic differentiation of polymorphism in the Plasmodium falciparum malaria vaccine candidate gene SERA5. *Vaccine*, *30*, 1583–93.

Targett, G. A. T. (1995). Malaria-advances in vaccines. *Current Opinion in Infectious Diseases*, *8*, 322–327.

Tripathi, R. P., Mishra, R. C., Dwivedi, N., Tewari, N., & Verma, S. S. (2005). Current status of malaria control. *Current Medicinal Chemistry*, *12*, 2643–2659.

Valencia, S. H., Rodriguez, D. C., Acero, D. L., Ocampo, V., & Arevalo-Herrera, M. (2011). Platform for Plasmodium vivax vaccine discovery and development. *Memorias Do Instituto Oswaldo Cruz*, *106 Suppl* , 179–192.

WHO. (2012). *WHO report*. Retrieved from http://www.searo.who.int/entity/malaria/topics/Malaria_factsheetWMD2012.pdf

WHO. (2014). WHO-factsheet. Retrieved July 26, 2015, from http://www.who.int/mediacentre/factsheets/fs094/en/

WHO. (2015). World Health Organization. Malaria Vaccine Rainbow Tables. Retrieved August 4, 2015, from http://www.who.int/immunization/research/development/Rainbow_tables/en/

Wiederstein, M., & Sippl, M. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*, 407–410.

Willcox, M. L., & Bodeker, G. (2004). Traditional herbal medicines for malaria. *BMJ (Clinical Research Ed.)*, *329*, 1156–1159.

Wipasa, J., Elliott, S., Xu, H., & Good, M. F. (2002). Immunity to asexual blood stage malaria and vaccine approaches. *Immunology and Cell Biology*, *80*, 401–414.

Xu, D., & Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical Journal*, *101*, 2525–2534.

Yeoh, S., O'Donnell, R. A., Koussis, K., Dluzewski, A. R., Ansell, K. H., Osborne, S. A., … Blackman, M. J. (2007). Subcellular discharge of a serine protease mediates release of invasive malaria parasites from host erythrocytes. *Cell*, *131*, 1072–83.

Yildiz Zeyrek, F., Palacpac, N., Yuksel, F., Yagi, M., Honjo, K., Fujita, Y., … Coban, C. (2011). Serologic markers in relation to parasite exposure history help to estimate transmission dynamics of Plasmodium vivax. *PloS One*, *6*, e28126.

Zepp, F. (2010). Principles of vaccine design—lessons from nature. *Vaccine*, *28*, C14–C24.

Zhao, Y., Pokutta, S., Maurer, P., Lindt, M., Franklin, R. M., & Kappes, B. (1994). Calcium-binding properties of a calcium-dependent protein kinase from Plasmodium falciparum and the significance of individual calcium-binding sites for kinase activation. *Biochemistry*, *33*, 3714–3721.

# APPENDIX

## APPENDIX I

## PATIENT DATA PROFORMA

S NO. H.Reg.No.: Lab Record No.:

Date of Admission: Date of Discharge:

Name: Age: Sex:

Address:

PF/ PV method:

Complaints with duration:

Any significant Past/ Personal/ Family/ Obstetric history:

Anti – malarial drugs/ any pre medication taken before admission:

## INFORMED CONSENT

I have been asked to take part in a research study "Polymorphism studies of vaccine candidate antigens of malarial parasites *Plasmodium*" carried out by BITS-Pilani, Hyderabad campus. The study team has told me the following things about the study:

1. Why the study is being done

   The study is part of malaria research, to investigate polymorphism in one of the blood stage vaccine candidate antigens ***in Plasmodium*** in field isolates from different parts of India.

2. What will happen to me if I am in the study

   No direct impact on you. The remaining blood sample from your diagnosis will be used for the study.

3. How long I will be in the study

   Once blood sample has been collected no further role for you.

4. The possible risks, discomforts, and the benefits of the study

   There is no risk or discomfort caused to the patient as the blood obtained is the part of the routine step performed in diagnosing patient for presence of malaria. The blood used up for the study will be used for procuring important biological data which will be useful in malarial research

5. Alternatives to being in the study

   The patient is free to unjoin the study and do not give the consent to use the blood sample remaining.

6. How my study records will be kept private

The patients data records like name , sample ID, registration number , age will be properly recorded and stored before starting an experiment and will not be disclosed to anyone other than the researcher involved in the investigation

7. Whether the study will cost me anything
   The study will not cost the patient anything rather the blood provided will be used truthfully in procuring important biological data in the area of research concerned.

I may contact Doctors of the research team at any time if I have questions about the research.

Signing this form means that the research study has been described to me orally in Hindi/English/language known to me. I will have the chance to ask questions about the study. These questions had been answered to my satisfaction before signing this form. In future I have liberty to keep away from the study at any time.

Some blood and serum may be left over which may be used for future research work for gathering some new information on malaria, the results of which are likely to benefit mankind, however I am sure that this will not be used for AIDS testing.

The research team may also do tests to study body factors (genetic) affecting and its response in relation to malaria attacks. Tests may also be carried out to identify parasite genes or immune mechanisms. Any test result that may be useful for taking care of my health will be communicated to me if the test is done within three years after the collection of sample. I can also ask to destroy my stored blood samples if I change my mind within 3 years after the collection of the sample, if not used by that time.

I have been told about what will be done as part of this study. I also have been told about the possible good and bad (benefits and risks) that could happen if I am in this study. After all this,

I choose to be in this study and give my consent for the collection and storage of malaria positive blood collected from me.

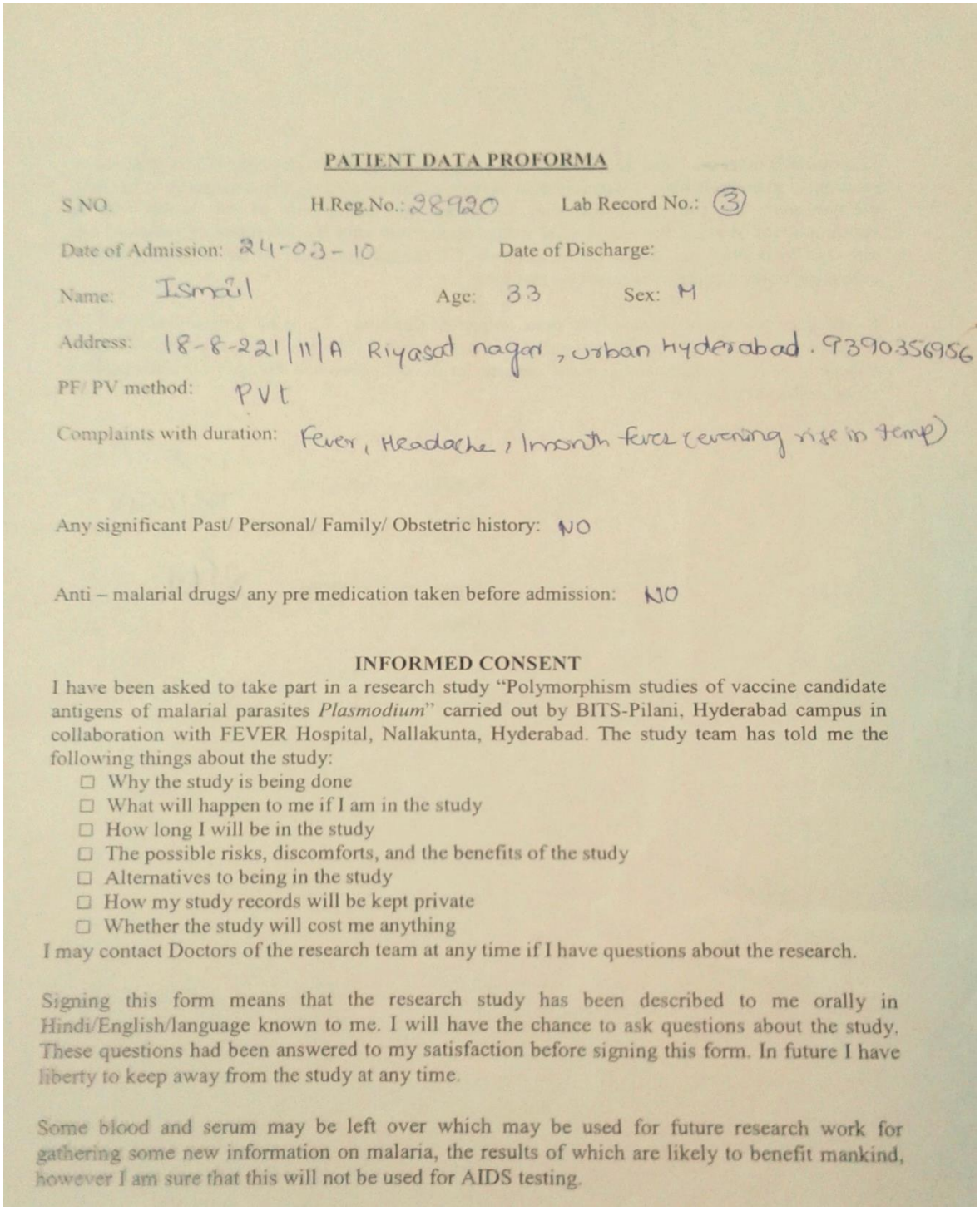
Signature of Participant: _____

Date: __ __ / __ __ / __ __

Signature of Witness: _____

Date: __ __ / __ __ / __ __

**REPRESENTATIVE DULY FILLED PATIENT CONSENT FORM**



**PATIENT DATA PROFORMA**

S NO. H.Reg.No.: 28920 Lab Record No.: (3)

Date of Admission: 24-03-10 Date of Discharge:

Name: Ismail Age: 33 Sex: M

Address: 18-8-221/11/A Riyasat nagar, urban Hyderabad. 9390356956

PF/ PV method: PVt

Complaints with duration: Fever, Headache, 1month fever (evening rise in temp)

Any significant Past/ Personal/ Family/ Obstetric history: NO

Anti – malarial drugs/ any pre medication taken before admission: NO

**INFORMED CONSENT**

I have been asked to take part in a research study "Polymorphism studies of vaccine candidate antigens of malarial parasites *Plasmodium*" carried out by BITS-Pilani, Hyderabad campus in collaboration with FEVER Hospital, Nallakunta, Hyderabad. The study team has told me the following things about the study:

- ☐ Why the study is being done
- ☐ What will happen to me if I am in the study
- ☐ How long I will be in the study
- ☐ The possible risks, discomforts, and the benefits of the study
- ☐ Alternatives to being in the study
- ☐ How my study records will be kept private
- ☐ Whether the study will cost me anything

I may contact Doctors of the research team at any time if I have questions about the research.

Signing this form means that the research study has been described to me orally in Hindi/English/language known to me. I will have the chance to ask questions about the study. These questions had been answered to my satisfaction before signing this form. In future I have liberty to keep away from the study at any time.

Some blood and serum may be left over which may be used for future research work for gathering some new information on malaria, the results of which are likely to benefit mankind, however I am sure that this will not be used for AIDS testing.

## APPENDIX II

**Summary of PvSERA sequences submitter to the NCBI database (First Submission from Indian Isolates)**

**Gene : PvSERA5 (representing C-terminal end)**

**Source: *Plasmodium vivax***

| S.No | Samples | Accession Numbers | **REMARK** |
|---|---|---|---|
| 1. | HS23 | JQ956500 | **First Submission from Indian Isolates** |
| 2. | HS28 | JQ956503 | |
| 3. | CE6 | JQ956488 | |
| 4. | CE7 | JQ956489 | |
| 5. | CE23 | JQ956490 | |
| 6. | CE25A | JQ956491 | |
| 7. | CE33 | JQ956492 | |
| 8. | CE50 | JQ956493 | |
| 9. | CE65 | JQ956494 | |
| 10. | CE67A | JQ956495 | |
| 11. | CE69 | JQ956496 | |
| 12. | CE71 | JQ956497 | |
| 13. | CE93 | JQ956498 | |
| 14. | CE106 | JQ956499 | |
| 15. | HS24 | JQ956501 | |
| 16. | HF25 | JQ956502 | |
| 17. | CE88 | JX188251 | |
| 18. | CE85 | JX188250 | |

**Gene : PvSERA4- (representing Central Protease domain and C-terminal end)**

**Source:** ***Plasmodium vivax***

| S.No | Samples | Accession Numbers | REMARK |
|---|---|---|---|
| 1. | CE147 | KC236394 | **First Submission from Indian Isolates** |
| 2. | CE149 | KC236395 | |
| 3. | CE173 | KC236396 | |
| 4. | HF4 | KC236397 | |
| 5. | CE88 | KC236398 | |
| 6. | CE92 | KC236399 | |
| 7. | CE95 | KC236400 | |
| 8. | CE97 | KC236401 | |
| 9. | CE106 | KC236402 | |
| 10. | HF32 | KC236403 | |
| 11. | CE174 | KC236404 | |
| 12. | HF20 | KC236406 | |
| 13. | HF46 | KC236407 | |
| 14. | CE109 | KC236408 | |
| 15. | CE146 | KC236409 | |
| 16. | HS19 | KC236410 | |
| 17. | HS28 | KC236411 | |
| 18. | HS4 | KC236412 | |

**Gene : PvSERA5- (representing N-terminal end )**

**Source:** ***Plasmodium vivax***

| S.No | Samples | Accession Numbers | **REMARK** |
|---|---|---|---|
| 1. | CE5 | KM982749 | **First Submission from Indian Isolates** |
| 2. | HF14 | KM982750 | |
| 3. | CE6 | KM982751 | |
| 4. | CE39 | KM982752 | |
| 5. | CE181 | JQ956490 | |
| 6. | CE182 | JQ956491 | |
| 7. | CE183 | JQ956492 | |
| 8. | CE185 | KM982753 | |
| 9. | CE195 | KM982754 | |
| 10. | CE198 | KM982755 | |
| 11. | CE199 | KM982756 | |
| 12. | HF177 | KM982757 | |
| 13. | HF178 | KM982758 | |
| 14. | HF182 | KM982759 | |
| 15. | CE184 | KM982760 | |
| 16. | CE196 | KM982761 | |
| 17. | HF175 | KM982762 | |

**Gene : PvSERA4- (representing N-terminal region)**

**Source:** ***Plasmodium vivax***

| S.No | Samples | Accession Numbers | REMARK |
|---|---|---|---|
| 1. | CE12 | KM982736 | **First Submission from Indian Isolates** |
| 2. | CE157 | KM982737 | |
| 3. | CE186 | KM982738 | |
| 4. | CE190 | KM982739 | |
| 5. | CE192 | KM982740 | |
| 6. | CE182 | KM982741 | |
| 7. | CE183 | KM982742 | |
| 8. | HF175 | KM982743 | |
| 9. | HF147 | KM982744 | |
| 10. | CE5 | KM982745 | |
| 11. | CE9 | KM982746 | |
| 12. | CE196 | KM982747 | |

**Gene: PfSERA5- (Representing N-terminal region) Source: *Plasmodium falciparum***

| S.No | Samples | Accession Numbers | REMARK |
|---|---|---|---|
| 1. | CE145 | KC236368 | **First Submission from Indian Isolates** |
| 2. | HF1 | KC236369 | |
| 3. | HF6 | KC236370 | |
| 4. | HF7 | KC236371 | |
| 5. | HF8 | KC236372 | |
| 6. | HF11 | KC236373 | |
| 7. | HF24 | KC236374 | |
| 8. | HF33 | KC236375 | |
| 9. | HF37 | KC236376 | |
| 10. | HF48 | KC236377 | |
| 11. | HF52 | KC236378 | |
| 12. | HF54 | KC236379 | |
| 13. | HF83 | KC236380 | |
| 14. | HF94 | KC236381 | |
| 15. | HF104 | KC236382 | |
| 16. | HF105 | KC236383 | |
| 17. | HF111 | KC236384 | |
| 18. | HF112 | KC236385 | |
| 19. | HF124 | KC236386 | |
| 20. | CE64 | KC236387 | |
| 21. | CE68 | KC236388 | |
| 22. | CE123 | KC236389 | |
| 23. | CE126 | KC236390 | |
| 24. | CE125 | KJ397705 | |
| 25. | HF110 | KJ397706 | |
| 26. | HF115 | KJ397707 | |
| 27. | HF165 | KJ397708 | |
| 28. | HF169 | KJ397709 | |
| 29. | HF170 | KJ397710 | |
| 30. | HF171 | KJ397711 | |
| 31. | HF172 | KJ397712 | |
| 32. | HF173 | KJ397713 | |
| 33 | HF174 | KJ397714 | |
| 34. | OR702 | KJ397715 | |
| 35. | OR618 | KJ397716 | |
| 36. | ALB2 | KJ397717 | |
| 37. | ALB3 | KJ397718 | |
| 38. | ALB5 | KJ397719 | |

**Gene: PfSERA5- (Representing Central and C-terminal region)**

**Source:** ***Plasmodium falciparum***

| S.No | Samples | Accession Numbers | REMARK |
|---|---|---|---|
| 1. | CE64 | KC236348 | **First Submission from Indian Isolates** |
| 2. | HF112 | KC236349 | |
| 3. | CE145 | KC236350 | |
| 4. | CE123 | KC236351 | |
| 5. | CE73 | KC236352 | |
| 6. | HF54 | KC236353 | |
| 7. | HF37 | KC236354 | |
| 8. | HF94 | KC236355 | |
| 9. | HF7 | KC236356 | |
| 10. | CE101 | KC236357 | |
| 11. | HF111 | KC236358 | |
| 12. | CE66 | KC236359 | |
| 13. | CE68 | KC236360 | |
| 14. | HF105 | KC236361 | |
| 15. | HF124 | KC236362 | |
| 16. | HF1 | KC236363 | |
| 17. | HF6 | KC236364 | |
| 18. | HF24 | KC236365 | |
| 19. | HF52 | KC236366 | |
| 20. | HF48 | KC236367 | |

## LIST OF PUBLICATIONS

1. **C.N. Rahul**, Shiva Krishna Katakam, Meerabai, Vijjay Kumar, Sandhya Phadke, Vidya Rajesh, "*Plasmodium vivax:* C-terminal diversity in the blood stage sera genes from indian isolates", ***Experimental Parasitology***, 134, (2013), 82 - 91.
2. **C.N. Rahul**, Shiva Krishna Katakam, Atul P. Pawar, Meerabai, Vijjay Kumar, Sandhya Phadke, Vidya Rajesh, "Genetic and structural characterization of PvSERA-4: Potential implications as therapeutic target for *P.vivax* malaria". ***Journal of Biomolecular Structure and Dynamics***, 32 (4), (2014), 580 - 590.
3. **C.N. Rahul**, K. Shiva Krishna, M. Meera, Sandhya Phadke, Vidya Rajesh, "*Plasmodium vivax*: N-terminal diversity in the blood stage SERA genes from Indian Isolates" ***Blood Cells, Molecules and Diseases***, 55, (2015), 30 – 35.
4. **Rahul C. N.**, Shiva Krishna K, Meera M, Sandhya Phadke, Vidya Rajesh, "Characterization of sequence diversity in *Plasmodium falciparum* SERA-5 from Indian isolates, ***Asian Pacific Journal of Tropical Disease***, 5 (Suppl.1), (2015), 80 - 84**.**
5. **Rahul C. N.**, Shiva Krishna Katakam, Atul P. Pawar, Vidya Rajesh, "Insilico approach to ascertain the calcium dependent role of *Plasmodium falciparum* SERA5", ***Journal of Biomolecular Structure and Dynamics****"* - communicated (Under revision)

## CONFERENCES

1. **C.N.Rahul**, Shiva Krishna K, Vidya Rajesh, "Structural modelling of *Plasmodium vivax* blood stage Serine Repeat Antigens (PvSERA)", CIDSCON 2015 - 5th Annual Conference of Clinical Infectious Diseases Society, New Delhi, 21st - 23rd August, 2015.

2. **C.N. Rahul**, Shiva Krishna K, Ekta Khetan, Sheetal Modi, B.V.S. Manjula, Vidya Rajesh, "Cloning of vaccine candidate antigen PvSERA-4 of malarial parasite *P.vivax* from Indian field Isolates", National Symposium on Immunology – Voyage from serology to molecular biology in Health care, February $16^{th}$ – $18^{th}$, 2012. (Second best poster award).

3. **C.N. Rahul**, Shiva Krishna K, P.B.S. Manjula, P. Vaideesh, A.K.Das, Vidya Rajesh, "Characterization of sequence diversity of the octamer repeat domain of Plasmodium falciparum Serine Repeat Antigen (pfSERA) in Indian Isolates". $XI^{th}$ Symposium on Vector and Vector Borne Diseases, Jabalpur, October $14^{th}$ – $17^{th}$, 2011.

## Biography of Prof. Vidya Rajesh

Prof. Vidya Rajesh is Associate Professor in the Department of Biological Sciences. She is also Associate Dean of Academic Research Division of BITS, Pilani – Hyderabad Campus. Prof. Vidya Rajesh completed her M. Sc. in Microbiology from Nagpur University in the year 1995. She completed her M. E. in Biotechnology (2000) and Ph. D from BITS, Pilani – Pilani campus in the year 2007. The topic of her doctoral thesis was *"Studies on Sequence Diversity and Characterization of the Apical Membrane Antigen of Plasmodium in Indian Isolates".* Prof. Vidya Rajesh is an extremely dedicated and hard core team player with commitment to teaching and research. She has been actively involved and has gained 13 yrs of teaching, research and administrative experience during her tenure in various capacities at BITS. Her current research interest is in the area of molecular genetics of human diseases. Her initial interest was only in the area of genetics of infectious disease like Malaria from Indian isolates and later diversified to Autism and Environmental bioremediation. She has nearly 25 research publications and equal number of conference presentations. She has completed three research projects as principal investigator and has an ongoing DST project as co-investigator. Her DST-SERC project provided the funding for this work on Genetic and Structural characterization of SERA from Indian Isolates. She has extended her work on this area and is also focusing on projects like Thyroid receptor mutations in Autism and analysis of urinary biomarkers in autistic patients. She currently has three Ph.D students working under her supervision. Prof. Vidya Rajesh is also part of other institutional projects like DST – FIST and DST – TBI at BITS, Pilani – Hyderabad campus.

# Biography of C.N.Rahul

Mr. C.N.Rahul is a Ph. D student in the Department of Biological Sciences. He pursued hisM. Sc. in Biochemistry from Osmania University in the year 2008 and passed with distinction from his university. Rahul joined BITS, Pilani Hyderabad campus in a DST SERC project at Department of Biological Sciences and continued his work as UGC-CSIR fellow (JRF and SRF) for his doctoral degree. He is well versed with molecular biology techniques like blood DNA isolation, PCR, cloning and bioinformatics application pertaining to analysis and interpretation of DNA sequence data. He also has experience in handling various other advanced techniques in molecular biology gained during his Ph.D tenure in BITS campus. He has added credible number of publications to the present study and attended seminars (national and International). Currently his research interests focus on malaria epidemiology and DNA sequencing analysis.