HAL
open science

Deciphering the role of G-quadruplexes and their interacting proteins in malaria parasite Plasmodium falciparum

Pratima Gurung

▶ **To cite this version:**

Pratima Gurung. Deciphering the role of G-quadruplexes and their interacting proteins in malaria parasite Plasmodium falciparum. Microbiology and Parasitology. Université Montpellier, 2020. English. NNT : 2020MONTT010 . tel-03328581

HAL Id: tel-03328581

https://theses.hal.science/tel-03328581

Submitted on 30 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biologie Santé

École doctorale Sciences Chimiques et Biologiques pour la Santé (CBS2)

Unité de recherche UMR 5235 CNRS
Laboratory of Pathogen Host Interactions(LPHI)

Deciphering the role of G-quadruplexes and their interacting proteins in malaria parasite *Plasmodium falciparum*

Présentée par Pratima GURUNG
Le 16 Avril 2020

Sous la direction de Jose-Juan LOPEZ RUBIO, Directeur de thèse
et Artur Scherf, co-Directeur de thèse

Devant le jury composé de

Mme Sara RICHTER, Professeur, University of Padua, Padua	Rapporteur
M. Alfred CORTES, Professeur, ISGlobal, CRESIB, Barcelona	Rapporteur
Mme Rosemary KIERNAN, DR, Institut de Genetique Humaine, Montpellier	Examinatrice(President)
Mme Katrin PAESCHKE, Professeur, Univerity Hospital Bonn, Bonn	Examinatrice
M. Jose-Juan LOPEZ RUBIO, CR, Université de Montpellier,Montpellier	Directeur de thèse
M. Artur SCHERF, DR, Institut Pasteur, Paris	co-Directeur de thèse

UNIVERSITÉ DE MONTPELLIER

Acknowledgements

First and foremost, I would like to thank my supervisor Jose-Juan Lopez Rubio and co-supervisor Artur Scherf for giving me the opportunity to be part of the ParaFrap community, for guidance, support, and supervision throughout my Ph.D.

I am grateful to Katrin Paeschke for hosting me in her laboratory to learn Yeast one-hybrid assay. I am also grateful to her team members especially Satya for teaching yeast-based experiments. I am thankful to the Functional Proteomic Platform of Montpellier, especially Serge Urbach for Mass spectrometry analysis.

Next, I would like to thank all the past and present members of our team. I am fortunate to have such awesome team members. Thanks a lot for all your love, support, and guidance. Elodie, thank you for helping me to start this G4 project from scratch, for your kind advice and scientific discussions. Thanks, Ana for being a great mentor, for all scientific and non-scientific discussions, encouragements and motivations throughout these years. Diane, thank you so much for helping me with all the inscriptions and French documents. And thanks Rafa for all the scientific discussions and making sure that we have stock of sweets in our office.

I would like to express my sincere gratitude towards Maryse Lebrun, for her scientific feedback, kind support, and encouragement, especially when I needed the most. Special thanks to all plasmo group members (especially CBB, Mauld, Rachel and Kai) and Arthur for all the discussions and tips on 'how to train the Plasmodium'. Thanks, Kai for all scientific advice and encouragement. Thanks, Coralie for helping me with proteins and HPLC. And thanks Marta, it was nice to meet someone who shares the same enthusiasm and love for Plasmodium. I would like to thank all the members of the LPHI (Diana, Margarida, Justine, Sarah, Seb, Mauld and everyone) and MIVEGEC (Mauld, Lucien. Nada) team along with Luc and admin staff for all their kindness and support.

I acknowledge ParaFrap for their financial support and coursework during my thesis. I would like to thank my thesis committee members: Jerome Dejardin, Ali Hakimi, Olivier Silvie, Mathieu Gissot and ParaFrap scientific advisory board for all their valuable comments, suggestions and criticisms that helped me in improving my project. Special thanks to Patrick Bastein, Slavica and Mila for putting their all efforts to organize all the ParaFrap workshops

and meetings, which not only helped us in understanding parasites but all provided an opportunity to interact with researchers from parasitology field.

I also would like to express my gratitude to my thesis jury members, Sara Richter, Alfred Cortes, Rosemary Kiernan, and Katrin Paeschke.

I would like to thank all my friends in Montpellier and India for their constant love, support, and motivation. Thanks, Akila for helping me settle down in Montpellier and sharing great moments. Thanks, Ksenia for all the small and big talks about lab and life. Thanks to Van, Ammy, Thuy, and Seb for the great company and making sure that I have some fun outside the lab as well. Thanks, Sheena and Vrushali, even though you are in different cities but thanks for being there in every moment of this PhD, my PhD sisters. Thanks, Rakhee and Preeti for taking care of me as a family and keeping the Indian vibe around me. Thanks, Kavita, Abhishek, Rohan, Jyoti, Kushagra for a wonderful time. Even though you guys stayed for a short duration in Montpellier but that was enough to make strong bonds among ourselves. Special thanks to Mukul, with whom I started my journey in this Plasmodium field in Krish's lab and even in JJ's lab as well. Thanks to all my ParaFrap PhD batchmates for constant support.

And last and not the least, special thanks to my parents and sisters for all their love and support. And thanks a lot Kumar Ankush for your kind love, patience, and understanding.

Summary

Malaria is a severe infectious disease that is responsible for half a million deaths worldwide each year. The emergence of resistance against the available antimalarial drugs and non-availability of an effective vaccine calls for the urgency to identify novel antimalarial targets. The severe form of malaria is predominantly caused by *Plasmodium falciparum. P. falciparum* is a protozoan parasite that manifests their life cycle in two distinct hosts: Anopheles mosquitoes, for their sexual development and humans, for their asexual life cycle. The symptomatic phase of the disease takes place when parasites invade and multiply within the human host erythrocytes. In order to thrive in different hosts, *Plasmodium* parasites exhibit uniquely regulated, stage-specific, gene expression, involving genetics and epigenetic factors. However, despite extensive research in epigenetics and transcriptional factors, the challenges in the understanding of the mechanism of gene regulation still remain.

Recently, a new regulation mediated by the non-canonical structure of the nucleic acid sequence, called G-quadruplexes (G4), has gained the attention of the researchers. G4 are guanine-rich nucleic acid sequences that fold into stacks of G-tetrads, which are stabilized by Hoogsteen hydrogen bonding in the presence of cations. G4s are dynamic structures that have a role in fundamental processes including transcription (as a steric hindrance for transcriptional machinery), DNA replication (by stalling replication fork) and telomeric maintenance (as telomeric capping structures). Currently, they have been increasingly exploited as potential drug targets in various anticancer and antiviral therapies. However, our knowledge of the role of G4 in *Plasmodium* is still in its budding state.

Bioinformatics studies have shown that the biased AT-rich genome of *Plasmodium* contains GC-rich motifs that have the propensity to form G4. These motifs are enriched in telomeric and subtelomeric regions; where virulence genes (*var* genes) are present. The *var* gene family is the most characterized multigene virulence family in *P. falciparum.* This family encodes for erythrocyte membrane protein 1 (PfEMP1) that are expressed on the surface of infected erythrocytes. PfEMP1 plays a critical role in malaria pathogenesis through cytoadherence and antigenic variation. G4 are associated with mitotic recombination events amongst *var* genes and telomeric maintenance. Given their role in vital processes for parasite development, G4 stabilizing compounds are now sought as potential antimalarial drugs. So far, most of these ligands are targeted against the telomeric G4. For instance, G4 ligands such as TMPyP4 and

telomestatin have shown to affect the growth of the parasites and inhibit the telomerase activity.

Besides, proteins recognizing G4s are known as G-quadruplex binding proteins (G4-BP). These proteins play a significant role in modulating these structures and their potential functions. They can stabilize the G4, promote the formation of G4 in the GC-rich motif, and can also unwind the existing G4. In order to understand the function of these G4 motifs in *Plasmodium* biology, it will be imperative to study the proteins that interact with these G4 and mediates the biological processes. In pursuit of this, my thesis aims to identify and characterize the G4 binding proteins in these malarial parasites. To this end, we have employed two different and complementary unbiased approaches, Yeast one-hybrid (Y1H) assay, and DNA pull-down based MS approach. Yeast one-hybrid assay is a DNA-protein interaction assay where DNA fragment of interest is used as a bait to identify proteins (prey) that binds to the bait in a yeast based-screen. While in DNA pull-down assay followed by mass spectroscopy, a biotin-labeled DNA probe is used as bait to fish out high-affinity DNA binding protein from the protein extract. Here, we have used a G4 motif as bait in our study. Using different biophysical experiments, we have confirmed that the selected G4 motif forms a stable and parallel type of G4. As a result, we have identified 152 putative G4-BP in *P. falciparum*. Almost half of the identified proteins have a nucleic acid binding function and are implicated in numerous nucleic acid-based processes including nucleic acid metabolism, translation, and transcription, similar to G4-mediated function. In order to validate the obtained results, we have selected three potential candidates (FEN1, GBP2, and putative DNAJ). PfFEN1 and PfGBP2 are identified from Y1H and DNA pull-down assay, while putative PfDNAJ is found common in both the approaches. Since few or no information was available on these proteins, so we decided to first investigate the role of these proteins in *P. falciparum*. To do so, we have generated the inducible knockout parasites line using combined approach of CRISPR-Cas9 and DiCre/Loxp system. The advantage of using this system is that these lines express endogenously HA-tagged protein for respective targeted gene and the same line can be induced by rapamycin to generate the gene knockout in parasites. Thus, we used this generated transgenic line to functionally characterize the selected candidates. Followed by our last objective to validate the G4-protein interaction using two different techniques, namely Electrophoretic mobility shift assay (EMSA) and co-immunoprecipitation followed by high throughput sequencing (ChIP-Seq). EMSA is an *in vitro* assay that is used to study the direct interaction between DNA and recombinant protein

while ChIP-Seq is used to determine the *in vivo* interaction of protein and DNA using endogenously tagged lines.

Thus, using the iKO approach, we have successfully generated the iKO PfGBP2 and PfDNAJb parasite line while we could not generate iKO-PfFEN1 line. Hence, we focused on the other two proteins, PfGPB2 and PFDNAJ in this thesis.

The first candidate, PfGBP2 was earlier demonstrated as telomere binding protein *in vitro.* Using iKO-PfGBP2, which expresses the endogenously HA-tagged protein, we showed that PfGBP2 is expressed throughout the intra-erythrocytic developmental cycle of the parasites. It is present in both cytoplasm and nucleus. ChIP-Seq analysis showed that PfGBP2 binds to the telomere *in vivo* and their binding sites are enriched in G4 forming sequences, which are predicted using G4Hunter. The *in vitro* binding of G4 and PfGBP2 is confirmed by EMSA assay. Corroborating these results, we confirmed that PfGBP2 is a bona fide G4 binding protein in *P. falciparum*. However, the mechanism of binding to G4 is not yet clear. Given that PfGBP2 is a telomere binding protein, we also assessed their role in telomere homeostasis and regulation of expression of *var* gene after culturing for 30 generations. It was observed that the deletion of the PfGBP2 gene does not affect the telomere length and expression of the *var* gene, though slight derepression of the subset of the *var* gene was seen.

The second candidate, PfDNAJb is potential G4 binding protein that was identified by both the unbiased approaches. PfDNAJb is a member of the type-III Hsp40 family. Hsp40 is co-chaperones that acts in concert with its partner Hsp70 to maintain the conformational integrity of the parasite proteome, and are unregulated within parasites under heat stress conditions. Sequence analysis showed that PfDNAJb has conserved J-domain and thioredoxin domain. We showed that PfDNAJb is expressed throughout the IDC of these malarial parasites and is localized in both nuclear and cytoplasmic regions. The pattern in the cytoplasm is reminiscent of ER-resident protein. Furthermore, PfDNAJb is shown to be indispensable for the growth of intra-erythocytic stages of *P. falciparum.* However, the exact mechanism is yet to be elucidated.

To conclude, this study provides the first comprehensive report on the G4-protein interactome in *P. falciparum.* The detailed functional characterization of G4- binding proteins will help in understanding their role in modulation of G4 and their function. Additionally, it

will provide support for the *in vivo* existence of G4 in these parasites. Taken together, the G4–protein interaction network will help in unraveling this G4-mediated regulation in *P. falciparum* and can open avenues to develop new antimalarial drugs.

Résumé

Le paludisme est une maladie infectieuse grave qui est responsable d'un demi-million de décès dans le monde chaque année. L'émergence d'une résistance aux médicaments antipaludiques disponibles et l'absence d'un vaccin efficace imposent d'identifier d'urgence de nouvelles cibles antipaludiques. La forme grave de la malaria est principalement causée par *Plasmodium falciparum*. *P. falciparum* est un parasite protozoaire qui manifeste son cycle de vie chez deux hôtes distincts : les moustiques anophèles, pour leur développement sexuel et les humains, pour leur cycle de vie asexué. La phase symptomatique de la maladie a lieu lorsque les parasites envahissent les érythrocytes de l'hôte humain et se multiplient dedans. . Afin de se développer chez différents hôtes, les parasites *Plasmodium* présentent une expression génétique unique, régulée et spécifique à chaque stade, impliquant des facteurs génétiques et épigénétiques. Toutefois, malgré des recherches approfondies sur l'épigénétique et les facteurs de transcription, les défis à relever pour comprendre le mécanisme de régulation des gènes demeurent.

Récemment, une nouvelle régulation médiée par la structure non canonique de la séquence d'acide nucléique, appelée G-quadruplexes, a attiré l'attention des chercheurs. Les G-quadruplexes (G4) sont des séquences d'acides nucléiques riches en guanine qui se replient en piles de G-quarts, stabilisées par la liaison hydrogène de Hoogsteen en présence de cations. Les G4 sont des structures dynamiques qui jouent un rôle dans les processus fondamentaux, notamment la transcription (en tant qu'obstacle stérique pour la machinerie transcriptionnelle), la réplication de l'ADN (en bloquant la fourchette de réplication) et la maintenance télomérique (en tant que structures de coiffage télomériques). Ils ont été de plus en plus exploités comme cibles potentielles de médicaments dans diverses thérapies anticancéreuses et antivirales. Cependant, nos connaissances sur le rôle du G4 dans le *Plasmodium* sont encore à l'état embryonnaire.

Des études bioinformatiques ont montré que le génome biaisé du Plasmodium, riche en AT, contient des motifs riches en GC qui ont la propension à former des G4. Ces motifs sont enrichis dans les régions télomériques et subtélomériques, où les gènes de virulence (gènes *var*) sont présents. La famille de gènes *var* est la famille de virulence multigénique la plus caractérisée chez *P. falciparum*. Cette famille code pour la protéine 1 de la membrane des érythrocytes (PfEMP1) qui est exprimée à la surface des érythrocytes infectés. PfEMP1 joue

un rôle essentiel dans la pathogénèse du paludisme par sa cytoadhérence et sa variation antigénique. Les G4 sont associées à des événements de recombinaison mitotique parmi les gènes *var* et à la maintenance télomérique. Étant donné leur rôle dans les processus vitaux de développement des parasites, les composés stabilisateurs G4 sont maintenant recherchés comme médicaments antipaludiques potentiels. Jusqu'à présent, la plupart de ces ligands sont ciblés contre le télomère G4. Par exemple, les ligands G4 tels que la TMPyP4 et la télomestatine ont montré qu'ils affectent la croissance des parasites et inhibent l'activité télomérase.

En outre, les G4 sont reconnues par différentes protéines connues sous le nom de protéines de liaison G-quadruplex (G4-BP). Ces protéines jouent un rôle important dans la modulation de ces structures et de leurs fonctions potentielles. Elles peuvent stabiliser les G4, promouvoir la formation de la G4 dans le motif riche en GC, et peuvent également dérouler les G4 existantes. Afin de comprendre la fonction de ces motifs G4 dans la biologie du Plasmodium, il sera impératif d'étudier les protéines qui interagissent avec ces G4 et qui sont les médiateurs des processus biologiques. Dans cette optique, ma thèse vise à identifier et à caractériser les protéines de liaison G4 chez ces parasites du paludisme. À cette fin, nous avons utilisé deux approches non biaisées différentes et complémentaires, le test simple hybrid (Y1H) et l'approche d'une spectroscopie de masse (MS) basée sur la méthode pull-down. L'essai Y1H est un essai d'interaction ADN-protéine dans lequel le fragment d'ADN d'intérêt est utilisé comme appât pour identifier les protéines (proies) qui se lient à l'appât dans un crible basé sur la levure. Lors de l'analyse de l'ADN par extraction suivie d'une MS, une sonde d'ADN marquée à la biotine est utilisée comme appât pour repérer la protéine de liaison à l'ADN de haute affinité de l'extrait protéique. Dans notre étude, nous avons utilisé un motif G4 comme appât. Grâce à différentes expériences biophysiques, nous avons confirmé que le motif G4 sélectionné forme une G4 d'un type stable et parallèle. En conséquence, nous avons identifié 152 G4-BP putatifs chez *P. falciparum*. Près de la moitié des protéines identifiées ont comme fonction la liaison aux acides nucléiques et sont impliquées dans de nombreux processus à base d'acides nucléiques, notamment le métabolisme, la traduction et la transcription des acides nucléiques, similaires à la fonction médiée par la G4. Afin de valider les résultats obtenus, nous avons sélectionné trois candidats potentiels (FEN1, GBP2 et DNAJ putatif). Les PfFEN1 et PfGBP2 sont identifiés à partir de Y1H et de l'analyse de l'ADN pull-down, tandis que le PfDNAJ putatif est commun aux deux approches. Comme nous ne disposions que de peu ou pas d'informations sur ces protéines,

nous avons décidé de commencer par étudier le rôle de ces protéines chez *P. falciparum*. Pour ce faire, nous avons généré la lignée de parasites de knockout inductibles à l'aide de CRISPR-Cas9 et le Dicre/Loxp system. L'avantage d'utiliserce système est que ces lignées expriment de manière endogène la protéine marquée HA du gène d'intérêt respectif et la même lignée peut être induite par la rapamycine pour générer le gène knockout. Ainsi, nous avons utilisé cette lignée transgénique pour caractériser fonctionnellement les candidats sélectionnés. Ensuite, notre dernier objectif était de valider les interactions de la protéine G4 en utilisant deux techniques différentes, à savoir le test de mobilité électrophorétique par déplacement (EMSA) et la co-immunoprécipitation suivie d'un séquençage à haut débit (ChIP-Seq). L'EMSA est un test *in vitro* qui est utilisé pour étudier l'interaction directe entre l'ADN et la protéine recombinante tandis que le ChIP-Seq est utilisé pour déterminer l'interaction *in vivo* entre la protéine et l'ADN en utilisant des lignes marquées de manière endogène.

Ainsi, en utilisant l'approche à médiation iKO dicre, nous avons réussi à générer la lignée parasitaire iKO PfGBP2 et PfDNAJb alors que nous n'avons pas pu générer la lignée iKO-PfFEN1. Nous nous sommes donc concentrés sur les deux autres protéines, PfGPB2 et PFDNAJb dans cette thèse.

Le premier candidat, PfGBP2, a déjà été démontré comme protéine de liaison aux télomères *in vitro*. En utilisant iKO-PfGBP2, qui exprime la protéine marquée HA de manière endogène, nous avons montré que PfGBP2 est exprimée tout au long du cycle de développement intra-érythrocytaire des parasites. Elle est présente à la fois dans le cytoplasme et dans le noyau. L'analyse ChIP-Seq a montré que le PfGBP2 se lie au télomère *in vivo* et que leurs sites de liaison sont enrichis dans les régions riches en GC, qui devraient former G4 en utilisant G4Hunter. La liaison *in vitro* de G4 et de PfGBP2 est confirmée par un test de l'EMSA. En corroborant ces résultats, nous avons confirmé que le PfGBP2 est une véritable protéine d'interaction G4 chez *P. falciparum*. Toutefois, le mécanisme de liaison à la G4 n'est pas encore clair. Étant donné que le PfGBP2 est une protéine de liaison aux télomères, nous avons également évalué son rôle dans l'homéostasie des télomères et la régulation de l'expression des gènes *var* après 30 générations de culture. Il a été observé que la délétion du gène PfGBP2 n'affecte pas la longueur des télomères et l'expression des gènes *var*, bien qu'une légère dérépression du sous-ensemble des gènes *var* a été observée.

Le second candidat, PfDNAJb, est une protéine potentielle d'interaction G4 qui a été identifiée par les deux approches non biaisées. PfDNAJb est un membre de la famille Hsp40 de type III. Hsp40 est une co-chapone qui agit de concert avec son partenaire Hsp70 pour maintenir l'intégrité conformationnelle du protéome du parasite, et qui n'est pas régulée chez les parasites dans des conditions de stress thermique. L'analyse des séquences a montré que PfDNAJb a conservé le domaine J et le domaine de la thiorédoxine. Nous avons montré que PfDNAJb est exprimé à travers le cycle de développement intra-érythrocytaire de ces parasites du paludisme et est localisé dans les régions nucléaires et cytoplasmiques du parasite. La localisation de la protéine dans le cytoplasme rappelle celle d'une protéine du réticulum endoplasmique. En outre, l'ADNJpf s'avère indispensable à la croissance des stades intra-érythocytaires de *P.falciparum*. Toutefois, le mécanisme exact reste à élucider.

Pour conclure, cette étude fournit le premier rapport complet sur l'interactome des protéines G4 chez *P. falciparum*. La caractérisation fonctionnelle détaillée des protéines interagissant avec G4 aidera à comprendre leur fonction precise et leur rôle dans la modulation de la G4. En outre, elle permettra de confirmer l'existence *in vivo* des G4 chez ces parasites. Pris ensemble, le réseau d'interaction des protéines G4 aidera à démêler cette régulation médiée par la G4 chez *P. falciparum* et peut ouvrir des possibilités de développer de nouveaux médicaments antipaludiques.

Table of Contents

Acknowledgements ii
Summary iv
Résumé viii
Table of Contents xii
List of Figures xv
List of Tables xvi
Abbreviations xvii
Chapter 1 Introduction 1
1.1 Malaria 1
1.1.1 The epidemiology of Malaria and human *Plasmodium* species 1
1.1.2 Clinical features of Disease 2
1.1.3 Diagnosis and prevention of Malaria 3
1.1.4 *Plasmodium falciparum* 4
1.2 G-quadruplexes 10
1.2.1 History of G-quadruplexes 10
1.2.2 G4 structural polymorphism 11
1.2.3 G4 binding ligands 12
1.2.4 Computational and experimental identification of G4 12
1.2.5 Validation and characterization of G4 formation 15
1.2.6 Biological role of G4 20
1.3 Role of G4 in human pathogens 27
1.3.1 Viruses 27
1.3.2 Prokaryotic pathogens 29
1.3.3 Eukaryotic pathogens 31
1.4 G-quadruplexes in *P. falciparum* 32
1.4.1 Identification of G4s in *P. falciparum* 32
1.4.2 Potential role of G4 34
1.4.3 G4s ligands as antimalarial agents 36
Objectives of the Thesis 37
Chapter 2 Methods and Materials 39
2.2 *Plasmodium falciparum* culture and transfection 39

2.2.1 Parasite lines 39
2.2.2 Parasite maintenance 39
2.2.3 Generation of inducible Knockout parasite lines 40
2.2.4 Growth phenotype assay 43
2.3 Preparation of nuclear and cytoplasmic extracts 43
2.4 Western blotting 44
2.5 Immunofluorescence assay 45
2.6 Biophysical confirmation of G4 folding of oligonucleotides 45
2.6.1 Absorbance Spectroscopy 45
2.6.2 Circular Dichroism 46
2.7 Approaches for identification of G4 binding proteins 46
2.7.1 Yeast One Hybrid System 46
2.7.2 DNA Pull-down assay and Mass Spectrometry 47
2.8 Cloning, expression and purification of recombinant protein 49
2.9 EMSA 49
2.10 ChIP-Seq 50
2.11 Telomere Restriction Fragment (TRF) Southern Blotting 51
2.12 Quantitative reverse transcription-PCR (qRT-PCR) on *var* genes 52
Chapter 3 Identification of G-quadruplex binding proteins in *P. falciparum* 54
3.1 Introduction 54
3.2 Results 55
3.2.1 Identification and experimental confirmation of the G4 formation of the sequence 55
3.2.2 Identification of G4 binding proteins using unbiased approaches 57
3.3 Discussion 63
Chapter 4 Characterization of telomere binding protein, PfGBP2 in *P. falciparum* 68
4.1 Introduction 68
4.2 Results 69
4.2.1 Bioinformatic analysis and domain organization of a PfGBP2 69
4.2.2 Recombinant PfGBP2 protein binds to G4 *in vitro* 71
4.2.3 Generation of iKO-PfGBP2 parasite lines using combined CRISPR-Cas9 and Dicre/Loxp system 73
4.2.4 PfGBP2 protein is expressed throughout the intraerythrocytic developmental cycle (IDC) of *P. falciparum* 74
4.2.5 PfGBP2 is dispensable for the IDC of *P. falciparum* 76
4.2.6 ChIP-seq displays PfGBP2 binds to telomere and G4FS 78

4.2.7 PfGBP2 does not affect the telomere length homeostasis 83
4.2.8 Loss of PfGBP2 shows slight derepression in some subset of *var* genes 84
4.3 Discussion 85
Chapter 5 Characterization of putative PfDNAJ protein in *P. falciparum* 89
5.1 Introduction 89
5.2 Results 91
5.2.1 Bioinformatic analysis and domain organization of a P. falciparum Hsp40 protein DNAJb 91
5.2.2 Expression and purification of recombinant protein 94
5.2.3 Generation of inducible knockout PfDNAJb parasite line using combined CRISPR-Cas9 and Dicre/Loxp system 94
5.2.4 PfDNAJb is expressed during all the stages of P. falciparum intraerythrocytic life cycle 96
5.2.5 PfDNAJb is essential for the viability of the asexual cycle of the P. falciparum 98
5.3 Discussion 100
Chapter 6 General Discussion and Future Perspectives 104
Potential role of PfGBP2 in *P. falciparum* 106
Potential role of PfDNAJb in *P. falciparum* 108
References 112
Appendices 144

List of Figures

Figure 1.1.Global distribution of malaria cases. 1
Figure 1.2.The life cycle of Plasmodium falciparum. 6
Figure 1.3.The transcriptomic map of intraerythrocytic developmental cycle (IDC) of P. falciparum. 7
Figure 1.4.Schematic of histone post-translational modifications associated with the var gene expression. 9
Figure 1.5.G-quadruplex structures. 11
Figure 1.6.Visualization of DNA G4in nuclei of human U2OS osteosarcoma cancer cells. .. 13
Figure 1.7.Schematic representation of Next generation based approaches. 14
Figure 1.8.The characteristic feature of stable G4 obtained using different biophysical experiments. 17
Figure 1.9.Schematic diagram of different biochemical tecnhiques. 19
Figure 1.10.Representation of ChIP-Seq technique for the G4-protein interaction. 20
Figure 1.11.Biological roles of G4. 20
Figure 1.12.Different mechanism of G4-mediated regulation of transcription. 22
Figure 1.13.Role of G4 and G4 unwinding helicase in replication 23
Figure 1.14.Putative role of G4s at telomeres. 25
Figure 1.15.Putative role of G4 in the translation. 26
Figure 1.16.Schematic model of the role of the pilE G4 in N. gonorrhoeae pilin antigenic variation 30
Figure 1.17.Model describing the role of G4 in mitochondrial DNA (kDNA) in Trypanosoma brucei. 32
Figure 1.18.Distribution of cannical G4FS in Plasmodium genome 33
Figure 2.1.Generation of inducible knockout parasite line 41
Figure 2.2.Construction of donor plasmid 42
Figure 3.1.Identification and experimental confirmation of the G4 formation of the selected G4 sequence. .....
Figure 3.2. Yeast one hybrid assay to identify G4-BP. 59
Figure 3.3.DNA-pull down assay, followed by LC-MS/MS to identify G4-BP. 61
Figure 3.4.Distribution and biological function of potential G4-BP identified using Y1H assay and DNA pull-down assay. 63
Figure 4.1.Bioinformatic analysis of domain organization of PfGBP2 and its homologs. 71
Figure 4.2.Expression and purification of recombinant PfGBP2 to perform EMSA 73
Figure 4.3.Generation of P. falciparum transgenic line expressing floxed PfGBP2-HA 75
Figure 4.4.Induction of KO in iKO-PfGBP2 parasite line. 77
Figure 4.5.Optimisation of ChIP-Seq protocol. 79
Figure 4.6.Genome-wide mapping of PfGBP2 on the P. falciparum genome. 81
Figure 4.7.Highly enriched motifs under the PfGBP2 peak dataset with [log (Q)>50]. 83
Figure 4.8.Telomere Restriction Digestion Southern blotting. 84
Figure 4.9.Quantitative RT-PCR (qRT-PCR) analysis for the var genes expression in rapamycin induced PfGBP2 KO line and control iKO-PfGBP2 parasite line. 85
Figure 5.1. Bioinformatic analysis of Type III Hsp40s and PfDNAJb in P. falciparum. 93
Figure 5.2.Expression of recombinant PfDNAJb protein in E.coli BL21 DE3 pLysS. 94

Figure 5.3.Generation of P. falciparum transgenic line expressing floxed PfDNAJb-HA gene 95
Figure 5.4.Localization of PfDNAJb during the intraerythrocytic developmental cycle of P. falciparum. 97
Figure 5.5.. Induction of iKO-PfDNAJb line to generate KO gene of PfDNAJb using rapamycin.....99
Figure 5.6.Supplementary figures of Chapter 5 102

List of Tables

Table 3-1.List of primers or oligonucleotides used in Y1H assay and DNA pull down assay 144
Table 3-2.List of candidates obtained from Yeast one-Hybrid system 145
Table 3-3.List of candidates obtained from DNA pull down assay followed by MS.....150
Table 4-1.List of primers and gblock used in chapter 4, characterization of PfGBP2 159
Table 4-2.MS analysis of recombinant protein PfGBP2 161
Table 4-3.List of primers used for qRT-PCR to verify the expression of var gene.....162
Table 4-4.qRT-PCR results for the expression of var genes expression 165
Table 5-1.List of primers and gblock used in the chapter 5, characterization of PfDNAJb ..167
Table 6-1.List of candidates obtained in silico search of RGG motif in P. falciparum.....169

Abbreviations

ACT	Artemisinin-based combination therapy
ApiAP2	Apicomplexa specific AP2
AT	Adenine and thymine
bp	Base pair
CD	Circular dichroism
ChIP	Chromatin immunoprecipitation
DMS	Dimethyl sulfate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
EMP1	Erythrocyte membrane protein 1
EMSA	Electrophoretic mobility shift assay
ER	Endoplasmic reticulum
G	Guanine
G4	G-quadruplex
G4-BP	G4 binding protein
G4FS	G4 forming sequence
GC	Guanine and cytosine
GO	Gene ontology
gRNA	Guide RNA
G-tetrad	Guanine tetrad

hpi	Hours post treatment
HRP	Horseradish peroxidase
HRP	Homology region
Hsp	Heat shock protein
IDC	Intraerythrocytic developmental cycle
iKO	Inducible knockout
IRS	Indoor residual spraying
ITNs	Insecticide treated nets
Kb	Kilo base
KO	Knockout
LAMP	Loop-mediated isothermal amplification
Mb	Mega base
MS	Mass spectrometry
NCL	Nucleolin
NHEIII	Nuclease hypersensitivity element III
NMR	Nuclear magnetic resonance
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphaste buffered saline
PCR	Polymerase chain reaction
ppm	Parts per million
qRT-PCR	Quantitative real time PCR

RBC	Red blood cells
RDT	Rapid diagnostic test
RNA	Ribonucleic acid
SD	Synthetically defined medium
SDS-PAGE	Sodium Dodecyl sulfate–Polyacrylamide gel electrophoresis
SP	Sulfadoxine-pyrimethamine
TAF	TATA binding protein associated factor
TDS	Thermal difference spectrum
TF	Transcription factor
TRF	Telomere restriction fragment
TSS	Transcription start site
UTR	Untranslated region
VEGF	Vascular endothelial growth factor
WHO	World health organization
Y1H assay	Yeast One- Hybrid assay

Chapter 1

Introduction

Chapter 1 Introduction

1.1 Malaria

1.1.1 The epidemiology of Malaria and human *Plasmodium* species

Malaria is a vector borne infectious disease, responsible for an estimated 228 million cases and 405,000 deaths worldwide in 2018 alone (WHO, 2019). The most vulnerable groups affected by malarial infection are pregnant women and children under the age of 5 years. According to the report, 85% of the malarial cases emerged from occurred within 19 countries in sub-Saharan Africa and India. More than half of these cases are accounted in six countries: Nigeria (25%), the Democratic Republic of the Congo (12%), Uganda (5%), Ivory Coast (4%), Mozambique (4%), and Niger (4%) (Figure 1.1). Although there is a decline in deaths due to malaria in some of the WHO African and the South-East Asia Region compared with 2010, this decline has slowed down since 2016.

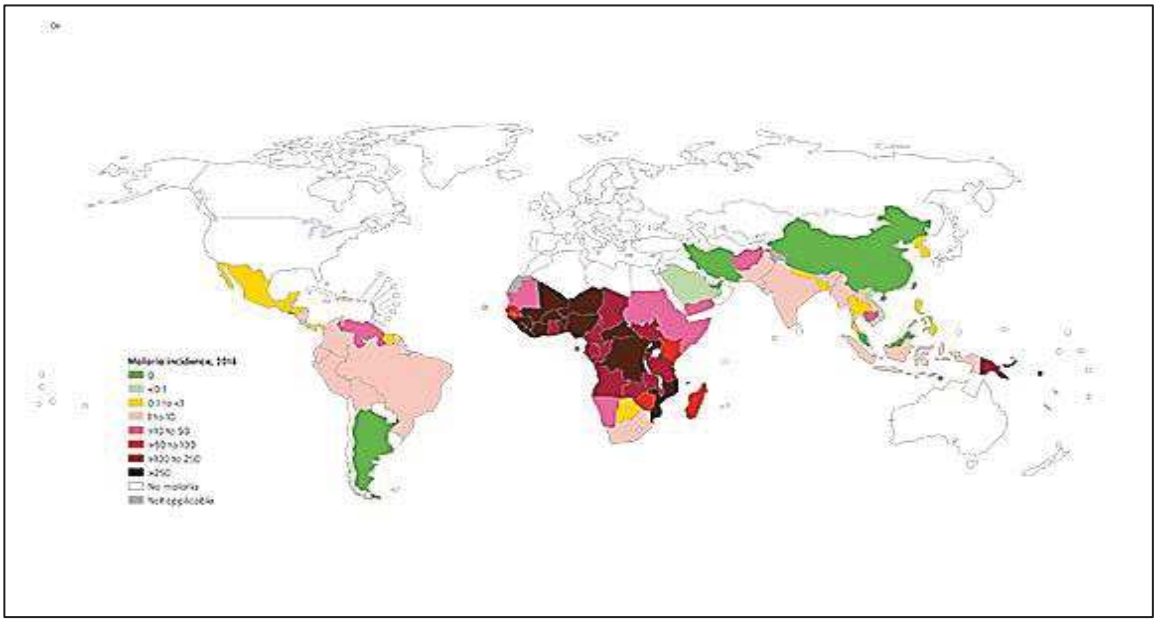


Figure 1.1.Global distribution of malaria cases.
Most of the malarial cases are confined to tropical countries. Different colors of the region represents the geographic distribution of malaria case incidence rate by country, according to the WHO report 2019.The incidence rate is calculated by cases per 1000 population at risk.(Source (WHO, 2019))

Malaria is caused by a protozoan parasite belonging to the genus *Plasmodium* which is transmitted by the bite of the infected female mosquitoes of the Anopheles genus.

Plasmodium spp are early divergent eukaryotic organisms of the phylum Apicomplexa. As of now, there are around 120 *Plasmodium spp* that infect mammals, birds and reptiles, while only 6 of them are capable of infecting human beings: *Plasmodium falciparum*, *P. vivax*, *P. malariae*, *P. ovale curtisi*, *P. ovale wallikeri* and *P. knowlesi* (Ashley, Pyae Phyo and Woodrow, 2018). *P. falciparum and P. vivax* are the predominant malaria causing parasites worldwide. The *P. falciparum* infected malarial cases are more prevalent in sub-Saharan African Region accounting for 99.7% of estimated malaria cases, the Eastern Mediterranean Region (71%), the Western Pacific Region (65%) and the South-East Asia region (50%), whereas *P. vivax* malaria cases are more common in South-East Asian region (53%) with the majority of the cases in India (47%) (WHO, 2019).

1.1.2 Clinical features of Disease

Malaria is mainly characterized by paroxysms (acute fever that is typically preceded by episodes of chills and fever), vomiting, headache and anaemia. The periodic spikes of fever correspond to the erythrocytic cycle length of the parasites: *P. malariae* causes fever every 72 hours, *P. knowlesi* every 24-28 hours and the remaining three species have 48-hour cycles (Singh and Daneshvar, 2013).

Malaria is classified into two categories based on the severity of its manifestations: uncomplicated and severe malaria (Bartoloni and Zammarchi, 2012). Uncomplicated malaria displays non-specific symptoms at different stages including a cold stage (sensation of cold, shivering), hot stage (fever, headaches), vomiting and sweating stage (sweating and tiredness). These cycles last for 6-10 hours. , tertian parasites (*P. falciparum, P. vivax and P. ovale*) causes the oncoming of symptomatic stage every second day while quartan parasite (*P. malariae*) every third day. In most cases, after a certain period of clinical symptoms and treatment, the parasite density decreases with eventual resolution of the infection (Cowman *et al.*, 2016). However some cases may progress to severe malaria.

P. falciparum infection causes severe malaria in non-immune patients which further causes cerebral malaria, severe anaemia, respiratory distress and acute renal failure. Due to the rapid progression of these complications, a combination of few of these conditions could be fatal. In *P. falciparum* malaria, a major factor of the pathogenesis is the adhesion of mature parasites in deep vasculature, eg. brain, lungs and placenta in pregnant women. The accumulation of infected erythrocytes causes endothelial activation, microvascular obstruction and induction of host inflammatory responses (Cowman *et al.*, 2016).

1.1.3 Diagnosis and prevention of Malaria

1.1.3.1 Diagnosis

The WHO criteria for diagnosis of malaria is fever and the presence of parasites in the blood. The parasites are detected by light microscopy analysis of a blood smear or by rapid diagnostic test (RDT). The RDT is based on the detection of specific proteins (antigens) produced by the malaria parasites in whole or peripheral blood of infected individuals. Uncomplicated malaria patients have a parasitaemia in the range of 1,000- 50,000 parasites per microlitre of blood except non-immune individuals and young children, who can display symptoms with less than 1000 parasites. The limit of detection by the microscopy is 50 parasites per microlitre while the RDT can detect 50-1000 parasites per microlitre. However, the sensitivity of these techniques is not adequate to predict clinical relapse cases or in cases of low parasitemia. In these cases, nucleic acid-based techniques such as PCR and loop-mediated isothermal amplification (LAMP) have strong advantages as they can detect as few as 22 parasites per milliliter (Phillips *et al.*, 2017).

1.1.3.2 Prevention

The prevention of the disease can be accomplished by three different ways: vector control, chemoprophylaxis and vaccines.

Vector control measures

More than 40 species of *Anopheles* are known to transmit human malaria. Vector control measures are mainly based on insecticide treated nets (ITNs) and indoor residual spraying (IRS). These measures helped in averting two-thirds of malarial cases in Africa between 2000 -2015. However, this decline has stalled since 2016 with insecticide resistance (esp. pyrethroids) in these vectors (WHO, 2019). This situation demands for the development of other approaches such as use of CRISPR Cas9 to generate transgenic mosquitoes.

Antimalarial drugs and chemoprophylaxis

Various combinations of antimalarial drugs are being recommended by the WHO for the treatment of malaria. Artemisinin-based combination therapy (ACT) is considered as the first line of defense against *Plasmodium*. In ACT, artemisinin derivatives are combined with partner drugs such as lumefantrine, mefloquine, sulfadoxine/pyrimethamine, amodiaquine, piperaquine and chlorproguanil, in order to reduce the chance of development and spread of resistance to either of the drugs.

As a malaria prevention strategy, chemoprophylaxis is recommended for pregnant women, young children and travelers. To protect pregnant women in moderate and high endemic regions like Africa, WHO has recommended intermittent preventive treatment in pregnancy (IPTp) with the antimalarial drug sulfadoxine-pyrimethamine (SP) (WHO, 2019). Seasonal malaria chemoprevention campaigns are run where combinations of sulfadoxine–pyrimethamine plus amodiaquine are given to children under 5 years old (Cairns *et al.*, 2012). And for travelers, atovaquone–proguanil and doxycycline are prescribed (Ashley, Pyae Phyo and Woodrow, 2018).

1.1.3.3 Vaccines

To date, there is no licensed vaccine that has been used against malaria, but different strategies have been followed to develop vaccines targeted against pre-erythrocytic, erythrocytic and sexual stages. RTS, S/ASO1 (trademarked as Mosquirix™) is one of the most promising vaccines against the pre-erythrocytic stages of the *Plasmodium* in human hosts. This was developed by GSK and the PATH's Malaria Vaccine Initiative, MVI with the support of the Bill & Melinda Gates Foundation. This vaccine is composed of a recombinant protein with parts of the circumsporozoite protein (CSP) of *P. falciparum* combined with the hepatitis B virus surface antigen in AS01 adjuvant system (Mahmoudi and Keshavarz, 2017; Phillips *et al.*, 2017). In phase III clinical trial, RTS,S/ASO1 exhibited relatively less protection in infants while there was rapid decline in the observed higher rate of efficacy in older children (Partnership, 2015; Mahmoudi and Keshavarz, 2017). Currently, it has been recommended by the WHO for pilot introduction in selected areas of three African countries.

1.1.4 *Plasmodium falciparum*

Plasmodium falciparum is the etiological agent of the severe form of malaria, which belongs to the phylum of Apicomplexa. The phylum Apicomplexa includes a large group of unicellular parasites which infect humans and animals such as *Plasmodium*, *Cryptosporidium*, *Toxoplasma, Babesia* and *Theileria*. This phylum is characterized by the presence of an apical complex which is necessary for host invasion and motility in the zoite forms of these parasites (van Dooren and Striepen, 2013). Most of the members of this group (except *Cryptosporidium*) harbor an unusual organelle, called the apicoplast, which is acquired through secondary endosymbiosis. These organelles are essential for the survival of parasites and are involved in several metabolic pathways such as synthesis of fatty acids (via

FASII pathway), isoprenoids , iron-sulfur clusters and heme synthesis (Lim and McFadden, 2010).

1.1.4.1 Life cycle of P. falciparum

P. falciparum has a complex life cycle that alternates between two hosts: female *Anopheles* mosquitoes and human (Figure 1.2). Infection begins with the injection of *Plasmodium* sporozoites into the host dermis by an infected female mosquito during a blood meal. These motile forms of parasites enter the host's bloodstream and cross the liver sinusoidal barrier to invade hepatocytes by the process known as traversal (Tavares *et al.*, 2013). Over subsequent 2-10 days, sporozoites develop into a liver stage (or exo-erythrocytic forms) to produce several thousands of merozoites. These merozoites are eventually released into the bloodstream by budding of parasite-filled vesicles called merosomes (Sturm *et al.*, 2006). Free merozoites invades erythrocytes in less than 2 minutes via multi-step process involving initial contact, reorientation of the merozoites, irreversible attachment to erythrocytes via moving junction formation, and the formation of the parasitophorous vacuole (Cowman *et al.*, 2016). Within infected erythrocytes, a single merozoite replicates by schizogony and undergoes successive transitions from ring (0-24h), trophozoites (24-36h) and schizont (36-48h) stage to eventually release as 16 to 32 new merozoites, which in turn infect new erythrocytes. This repeated infection of erythrocytes causes the symptoms and pathologies of malaria. During this exponential growth parasites replicate asexually while some of them escape the asexual multiplication and differentiate into sexually committed cells: the female and male gametocytes. These gametocytes are taken up by the mosquitoes in another meal. Inside the mosquito midgut, each male gametocyte undergoes three rounds of mitosis to produce 8 motile microgametes and female gametocyte matures into a macrogamete. The fertilization of a female and male gametocytes fuse to form zygote which elongates into an ookinete. These ookinetes exit from the lumen of the gut as an oocyst, which further replicates to form sporozoites. Sporozoites then travel through haemolymph to the salivary glands of the mosquito, from where they are ready to be injected into the next human host (Aly, Vaughan and Kappe, 2009).

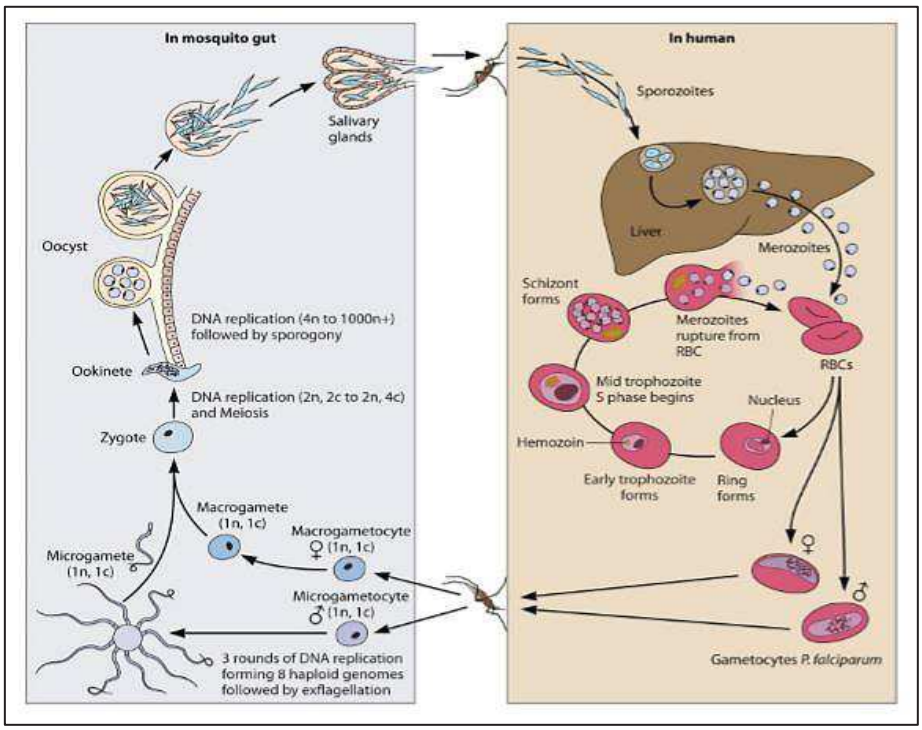


Figure 1.2.The life cycle of Plasmodium falciparum.
P. falciparum completes life their life cycle in two different hosts: the female Anopheles mosquito (left) and human (right). The asexual cycle is initiated by the injection of the parasites into the human by an infected mosquito during their blood meal. These parasites migrate to the liver where they generate several thousands of merozoites that are released into bloodstream to invade erythrocytes, thus commencing the symptomatic phase of the disease. Some of these asexual parasites differentiate into sexual forms, which are taken up by another mosquito. Inside the mosquito, the sexual phase of the cycle takes place, where the gametocytes undergoes meiotic and mitotic replication to form sporozoites (Adapted from (Lee, Symington and Fidock, 2014))

1.1.4.2 Plasmodium falciparum genome and their regulation

Genome

The *P. falciparum* clone 3D7 genome sequencing was first published in 2002 (Gardner *et al.*, 2002). The 22.8Mb nuclear genome is distributed amongst 14 chromosomes with size ranging from ~0.643 to 3.29Mb. The *Plasmodium* genome is highly AT rich with overall (A+T) composition of ~80.6%, which even rises to ~90% in introns and intergenic regions. *P. falciparum* contains approximately 5300 genes, of which more than 60 % encodes for proteins with weak or no homology to other eukaryotes (Gardner *et al.*, 2002).

Regulation of gene expression

Given its complex life cycle in two completely distinct environmental niche in human and *Anopheles* mosquitoes, these parasites require a tightly regulated gene expression. Several transcriptomics studies were performed to understand the gene expression variation across the *Plasmodium* life cycle. Bozdech *et al*. reported that during intraerythrocytic developmental cycle (IDC), a cascade of continuous gene expression with no clear boundaries or sharp stage transitions were observed (Figure 1.3). Moreover, the genes are expressed "just-in-time" when they are required and only once per IDC (Bozdech *et al.*, 2003). Le Roch *et al* reported that the functionally related genes shared similar mRNA and protein expression profiles. Moreover, there is a significant difference between the mRNA and protein abundance in different cycles of the malarial parasites (Roch *et al.*, 2004). Hence, these data hints towards the existence of transcriptional and post-transcriptional regulations in *Plasmodium*.

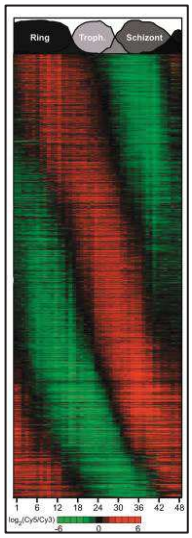


Figure 1.3.The transcriptomic map of intraerythrocytic developmental cycle (IDC) of P. falciparum.
The IDC Phaseogram depicting the transcriptional profile of 2,712 genes. This shows that the cascade of continuous expression lacks clear boundaries or sharp transitions (adapted from (Bozdech et al., 2003)

Transcription regulation

Whole genome sequencing of *P. falciparum* has boosted the research on exploring the transcriptional machinery of these parasites. The important component of basal transcriptional machinery such as RNA pol II and most of its subunits were identified in *Plasmodium spp* (Coulson, Hall and Ouzounis, 2004; Callebaut *et al.*, 2005). Although 4 TATA binding protein (TBP) associated factor (TAF) have been identified as the only known components of TFIID complex, these parasites lacks a classical TFIID with a histone fold domain, suggesting divergent nature of the complex (Callebaut *et al.*, 2005).

Specific transcription factors (TFs) are essential for the recruitment and activation/repression of the gene transcription. Remarkably, *P. falciparum* has only about 30 specific TFs, 27 of which belong to the Apicomplexa specific AP2 (ApiAP2) TF family (Coulson, Hall and Ouzounis, 2004; Balaji *et al.*, 2005). These ApiAP2 TFs are differentially expressed in a stage specific manner, suggesting that they might be involved in the progression and differentiation process of parasite's life cycle (Balaji *et al.*, 2005). For instance, AP2-G is

involved in sexual differentiation (Kafsack *et al.*, 2014), AP2-L is critical for liver stage development of parasites (Iwanaga *et al.*, 2012), AP2-Sp is involved in sporozoite development (Yuda *et al.*, 2010), and AP2-O activates ookinete stage specific genes (Yuda *et al.*, 2009).

However, the low number of transcription factors and lack of TAFs in *Plasmodium* highlights the existence of alternative mechanisms of gene regulations.

Epigenetic regulation

Epigenetic regulation primarily refers to the heritable changes mediated by DNA and histone modifications, without altering the DNA sequence. In *P. falciparum*, these regulations are involved in various processes such as gametogenesis, invasion and antigenic variation of virulence genes of the parasites.

Regulation of gametogenesis:

As discussed earlier, ApiAP2-G TF is a master regulator of sexual differentiation (Kafsack *et al.*, 2014). The PfAP2-G locus is also mediated by histone mark (H3K9me3) and PfHP1. PfHP1 binds specifically to H3K9me3 resulting in inhibition of PfAP2-G expression (Flueck *et al.*, 2009). Downregulation of PfHDA2 and PfHP1 activates PfAP2-G and induces gametocytogenesis in malarial parasites (Brancucci *et al.*, 2014; Coleman *et al.*, 2014).

Invasion:

The invasion process is regulated by a wide repertoire of proteins that interact with host erythrocytes such as MSPs, RONs , AMA1, PfRh and PfEBA. The *P. falciparum* erythrocyte binding antigens (PfEBAs) and reticulocyte binding like homologues (PfRHs) are the main families involved in tight junction formation that determine the use of the so-called alternative invasion pathways (Cowman *et al.*, 2017). Due to the redundant roles of these proteins, parasites can switch from one phenotype to another by changing the expression of these key invasion ligands to evade the host immune recognition. Besides genetic changes sequence polymorphisms and copy number variations (CNV), they are shown to undergo epigenetic changes (Cortés, 2008; Duraisingh and Skillman, 2018). The PfRh4, a member of RBL family that is involved in sialic acid-independent invasion, was shown to undergo epigenetic switching(Stubbs *et al.*, 2005; Gaur *et al.*, 2006; Awandare *et al.*, 2018).

Regulation of *P. falciparum* virulence genes:

The best known example of epigenetic regulation is the epigenetic regulation of clonally variant gene expression of *var* genes. The *var* gene family is the most characterized multigene virulence family in *P. falciparum.* This family encodes for erythrocyte membrane protein 1 (PfEMP1) that are expressed on the surface of infected RBCs. PfEMP1 plays a critical role in malaria pathogenesis through cytoadherence and antigenic variation (Scherf, Lopez-Rubio and Riviere, 2008; Cowman *et al.*, 2016). Out of 60 *var* genes, only one is expressed at any given time in these parasites (Scherf, Lopez-Rubio and Riviere, 2008). Moreover, the switching of *var* genes helps these parasites to evade the host immune response. Several studies have shown that the histone post-translational modifications affect the *var* gene expression. For instance, active *var* genes are highly enriched for H3K4me2 and H3K4me3 during ring stage while silent *var* genes are associated with H3K9me3 (Lopez-Rubio *et al.*, 2007) (Figure 1.4).

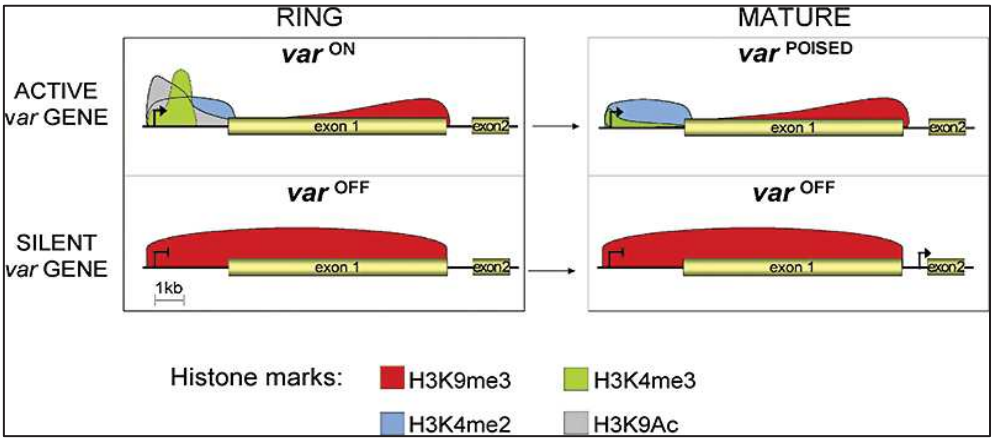


Figure 1.4.Schematic of histone post-translational modifications associated with the var gene expression.
The histone modifications (5' UTR and exon 1 of var gene) and transcriptional states of var (var2csa) genes in ring and mature stages are shown. 5' UTR of active var gene (var^{ON}) in ring stage are highly enriched with H3K4me2, H3K4me3 and H3K9Ac while in late stages, previously active var gene is transiently silenced (poised-varpoised) but still maintains the enrichment of H3K4me2.The silent var gene (var^{OFF}) are enriched in H3K9me3 throughout the asexual life cycle (source-(Lopez-Rubio et al., 2007)).

These epigenetic marks of histones are introduced by epigenetic modifiers such as histone methyltransferases (HMT) and histone acetylases/deacetylases (HAT/HDAC). Class III (*Pf*Sir2A and *Pf*Sir2B) and Class II (PfHDA2) HDAC proteins are involved in regulating the repressive marks on silent *var* genes and disruption of these enzymes causes loss of monoallelic *var* gene expression (Duraisingh *et al.*, 2005; Tonkin *et al.*, 2009; Coleman *et al.*, 2014). Histone methyltransferase *Pf*SET2 specifically marks *var* genes and its deletion results in the transcription of all *var* gene cluster (Jiang *et al.*, 2013; Ukaegbu *et al.*, 2014).

Recently, several studies have reported the existence of another mechanism of regulation for *var* gene expression (Smargiasso *et al.*, 2009; Stanton *et al.*, 2016). This mechanism is based on non-canonical structure of DNA, particularly G4. This thesis elucidates the role of these G4 in the biology of *P. falciparum*.

1.2 G-quadruplexes

1.2.1 History of G-quadruplexes

The G-quadruplex (G4) is one of the widely studied non-canonical secondary structures. It is formed in guanine rich nucleic acid sequences. It was first reported in 1910 that concentrated solutions of guanylic acid forms a gel (Bang, 1910). Fifty years later, in 1962, Ralph *et al* reported that tri- and tetra nucleotides of deoxyriboguanylic acid, unlike other three bases (thymidine, deoxyadenosine , deoxycytidine), can form a stable base pair interactions with other guanines to form stable aggregates (Ralph, Connors and Khorana, 1962). In 1962, Gellert *et al*, performed the first X- ray diffraction study to reveal that the guanylic acids assembles into tetrameric structures. These G4s are composed of guanine tetrad, which are stabilised by Hoogsteen base pairing (Gellert, Lipsett and Davies, 1962). These tetrads stack over each other to form a four-stranded G4 under physiological salt conditions (Gellert, Lipsett and Davies, 1962; Sen and Gilbert, 1988, 1990) (Figure 1.5A and 1.5B). Being started as a structural curiosity, now the researchers have reported their role in various fields, including biology, medical biology, chemistry and nanotechnology.

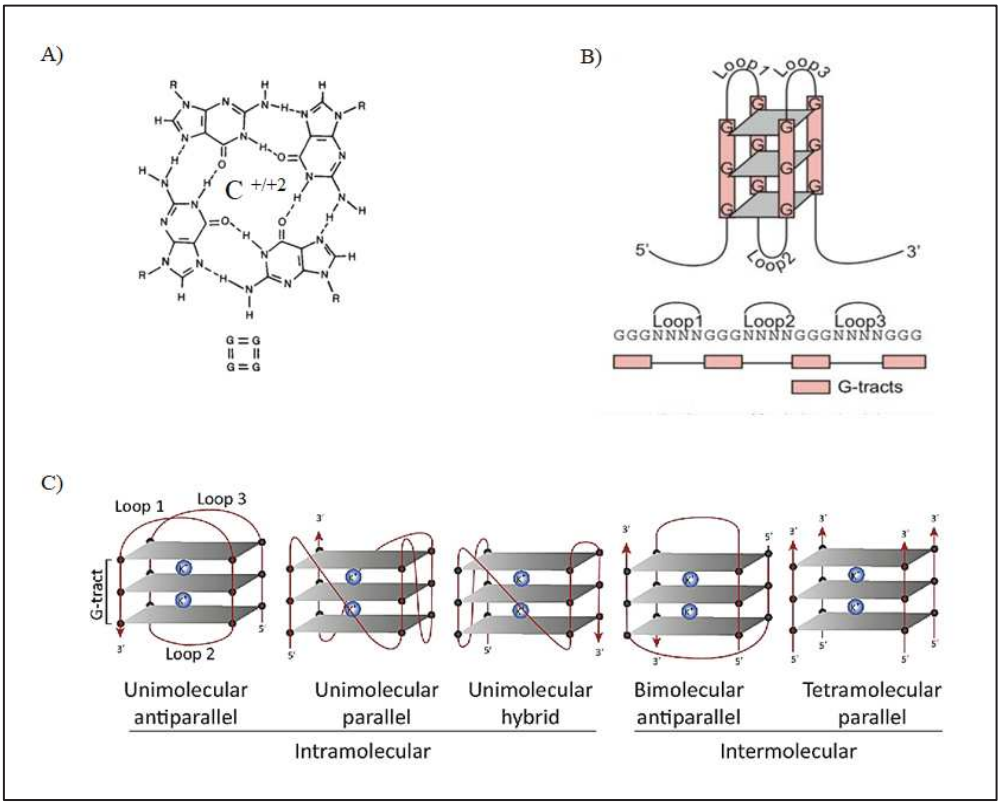


Figure 1.5.G-quadruplex structures.
The chemical structure of G-tetrad is formed by the association of four guanines that are stabilized by Hoogsteen hydrogen bonding in the presence of cation (monovalent or divalent) at the centre of the G-tetrad. b) Schematic representation of an intramolecular G4 structure and its motif. c) Schematic representation of different topologies of G4 (source ref (Capra et al., 2010; Kwok et al., 2017)).

1.2.2 G4 structural polymorphism

G4 is composed of stacked G-tetrads (Guanine-tetrad) and intervening sequence forming loops in the presence of cations. G-tetrad is yielded by interaction of four guanine bases via Hoogsteen bonding. G4s can exist within the same strand (intramolecular) or multiple strands (intermolecular) of nucleic acid (Burge *et al.*, 2006). The topology and the stability of these structures are dependent on several factors including the length and the composition of the G4 motif, the size of the loops, strand directionality and the nature of monovalent cations (Rachwal, Brown and Fox, 2007; Guédin *et al.*, 2010). Based on the strand orientation, G4s can adopt variable topologies: parallel, antiparallel, and hybrid-type structures (Sen and Gilbert, 1990) (Figure 1.5C). These non-canonical structures can destabilize the double helix structure as many of these G4s are thermodynamically more stable than double stranded DNA and their unfolding kinetics are significantly slower (König *et al.*, 2013).

The representative sequence for a canonical intramolecular G4 is ‘ $G_xN_yG_xN_yG_xN_yG_x$’ , where G_x represent guanine repeats (G ≥3) and ‘N_y’ stands for any nucleotides other than G (Figure 5B). The length of the G-tract (G_x) stabilises and favors the intramolecular parallel complexes (Rachwal, Brown and Fox, 2007). For ‘N_y’, y stands for the number of residues involved in the loop formation. The short single loop residue favors parallel while more than one residue can adopt either of parallel or antiparallel structures (Hazel *et al.*, 2004; Guédin *et al.*, 2010). The stabilization of loop length and G4 also depend on the nature of the cation (Guédin *et al.*, 2010). Recently, both monovalent and divalent cations have been reported to affect G4. The order of ions that can stabilize G4 is : $K^+ > Rb^+ > Na^+$ (but not Li^+ or Cs^+) and $Sr^{2+} > Ba^{2+} > Ca^{2+} > Mg^{2+}$ (Bhattacharyya, Arachchilage and Basu, 2016).

Different sequences can adopt distinctive topologies, however given sequence can also form different conformations. The best example of such complexities is human telomere sequence, 5’- $(GGGTTA)_3GGG$ -3’ which exhibits all three types of topologies including parallel, anti-parallel and hybrid structures under slightly different conditions (Sen and Gilbert, 1990; Phan, 2010).

1.2.3 G4 binding ligands

In the past few years, several small molecules have been discovered that are capable of binding the G4s. These molecules generally consist of an aromatic surface that π stacks onto the G-tetrads, a positive or basic groups to interact with intervening loops or grooves, and steric structure to avoid intercalation with dsDNA (Monchaud and Teulade-Fichou, 2008). The information on existing molecules targeting G4 are available on the G4LDB database (http://www.g4ldb.org/) (Li *et al.*, 2013). This web resource complies systematic data on existing G4 ligands and which could provide valuable information in designing and discovery of G4 targeting ligands or drugs. The most widely used G4 ligands are BRACO-19, TMPyP4,telomestatin, Phen-DC_3, and Pyridostatin (Monchaud and Teulade-Fichou, 2008).

1.2.4 Computational and experimental identification of G4

The identification of G4 motifs and the factors that determine their structure laid a framework to develop the computational algorithms to identify the putative G-quadruplex forming sequence (G4FS) in different organisms (Huppert, 2008) . Several algorithms have been published including Quadparser, G4FSFinder, G4Hunter and QGRS Mapper (Huppert and

Balasubramanian, 2005; Kikin, D'Antonio and Bagga, 2006; Bedrat, Lacroix and Mergny, 2016; Hon *et al.*, 2017). They are based on different types of search such as pattern matching techniques, sliding window approaches and score calculation (Ravichandran, Ahn and Kim, 2019) .

In silico studies in numerous organisms have displayed that these G4FS are not present randomly but they have defined locations in the genome. They are enriched in 5' UTRs, promoters, CpG islands, exon/intron junctions and nuclease-hypersensitive sites (Marsico *et al.*, 2019; Ravichandran, Ahn and Kim, 2019).

However, the formation of G4 *in vivo* has always been controversially discussed (Rhodes and Lipps, 2015). To provide evidence for the existence of G4 *in vivo*, several G4 recognizing antibodies and G4 binding proteins have been developed and identified, respectively. The role of G4 binding protein will be discussed later in this chapter.

The first G4 specific antibody, Sty49 was able to recognize telomeric G4s present in the macronuclei of ciliate *Stylonychin lemnae* with Kd of 3-5 nM (Schaffitzel *et al.*, 2001). Later, two antibodies BG4 and 1H6 were developed to visualize G4s in mammalian cells. BG4 antibody binds with high affinity to intra- and intermolecular G4s. It showed a punctate nuclear staining and their localization were enriched at the chromosomal ends, thus confirming the presence of G4s at human telomeres (Biffi *et al.*, 2013) (Figure 1.6)**.** The antibody 1H6 is able to bind different types of G4-DNA, exhibiting strong granular nuclear staining in both human and murine cell lines (Henderson *et al.*, 2017). Besides these, several other protein chemical probes have also been developed that can recognize G4s and be used in living cells (Kwok *et al.*, 2017).

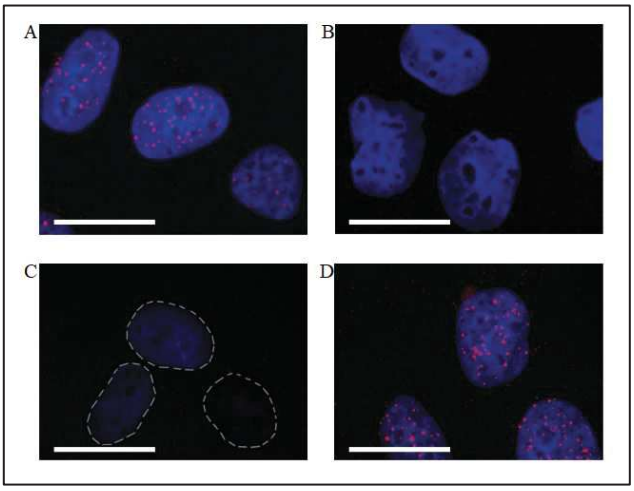


Figure 1.6.Visualization of DNA G4in nuclei of human U2OS osteosarcoma cancer cells.

a) Immunofluorescence showing BG4 loci (red) in U2OS osteosarcoma cell nuclei. b) Preincubation of the BG4 antibody with pre-folded G4 oligonucleotides results in loss of BG4 foci in these cells. c) DNase I treatment results in the loss of BG4 foci in U2OS cells. d) Transfection with pre-folded G4 oligonucleotides increases the BG4 foci number. Nuclei are stained with DAPI. The scale bars correspond to 20 µm (Adapted from (Biffi et al., 2013))

Furthermore, the advent of the Next Generation Sequencing (NGS) has also benefited the G4 community by experimentally demonstrating the formation of G4s and their location within cells. This involves two different approaches - antibody based pull-down approach and the polymerase stalling approach (Hänsel-Hertsch *et al.*, 2016; Marsico *et al.*, 2019) (Figure 1.7).

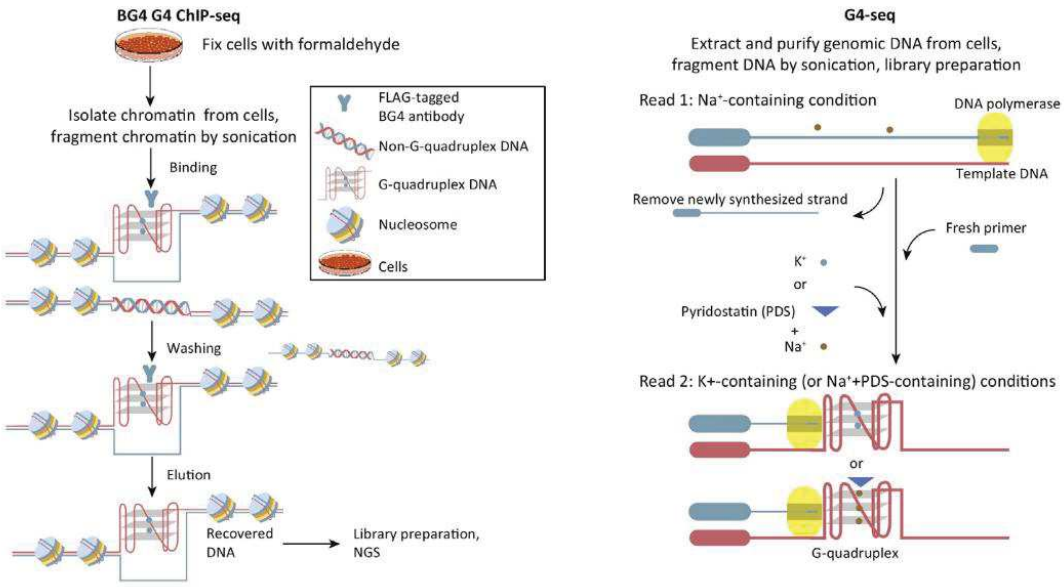


Figure 1.7.Schematic representation of Next generation based approaches.
G4 specific antibody (BG4) based pull down assay approach and b) G4-seq approach. (Source (Kwok et al., 2017))

The BG4 antibody-based ChIP-Seq revealed ~10,000 G4s in regulatory and nucleosome-depleted regions within human chromatin. The positive and dynamic relationship between G4s and transcriptional activity of G4-bearing genes was observed, thereby suggesting G4s as epigenetic markers in the human genome (Hänsel-Hertsch *et al.*, 2016).

Recently, the whole-genome experimental mapping of G4s was performed by polymerase stalling based high-throughput sequencing method (named as G4-Seq). This study experimentally mapped the formation and localization of G4s in 12 different species including eukaryotes and prokaryotes. The strong G4 enrichment was observed at promoter and transcriptional start site (TSS) regions in studied mammalian species such as human and mouse (Marsico *et al.*, 2019).

1.2.5 Validation and characterization of G4 formation

Next, to validate and characterize the topologies of predicted or identified G4 motifs, various biophysical and biochemical methods have been developed. Some of these most widely used methods are described below:

1.2.5.1 Biophysical methods

Biophysical methods are employed to identify G4 formation by the nucleic acid sequence and investigate the structural properties of the G4s. However, these methods are limited to studying short oligonucleotides and does not account for the effect of flanking sequence, beyond the central G4 motif, which can studied using biochemical methods. Various biochemical methods that are widely employed are explained below and Figure 1.8 displays the characteristic feature shown by G4s using these methods:

Thermal Difference Spectrum (TDS) is a differential absorbance curve, which corresponds to the difference in UV absorbance of nucleic acid sequence at high (unfolded) temperature and the low (folded) temperature. Each nucleic acid structure has its own distinct TDS shape. The signature spectrum of a G4 consists of two positive peaks near 240 nm and 270 nm and a negative peak at 295 nm (Mergny *et al.*, 2005).

UV spectroscopy or UV melting experiment can measure the thermostability of G4 at 295 nm. The existence of an inverted transition at 295 nm supports the presence of G4 but it must be complemented with other methods such as TDS and circular dichroism (CD) to confirm G4 formation (Mergny, Phan and Lacroix, 1998).

Circular dichroism (CD) is employed to investigate the topology of G4 as well as the effect of ligands, cations, and chemical modification on G4. The CD spectrum of parallel G4 has a positive peak near 260 nm and negative peak near 240 nm whereas an anti-parallel G4 exhibits positive peak near 290 nm and a negative peak near 260 nm. The third type of G4, i.e hybrid G4 exhibits two positive peaks (~295 nm and 260 nm) and negative peak near 245 nm (del Villar-Guerra, Trent and Chaires, 2018). This technique can also be used to confirm the obtained results of TDS and Tm.

^{1}H Nuclear magnetic resonance (1D NMR) allows the determination of high-resolution structures and their dynamic, kinetic and intermolecular interaction studies. In contrast to nucleic acid with Watson-Crick base pairing exhibiting peaks between 12-14 parts per million (ppm), G4 exhibits the presence of peaks between 10 and 12 ppm which is consistent with imino protons bound by Hoogsteen hydrogen bonds at 298 K. For example, three tetrad

G4 exhibits 12 sharp guanine peaks at 10-12 ppm. NMR can also demonstrate the existence of multiple G4 conformations by a single G rich sequence (Adrian, Heddi and Phan, 2012).

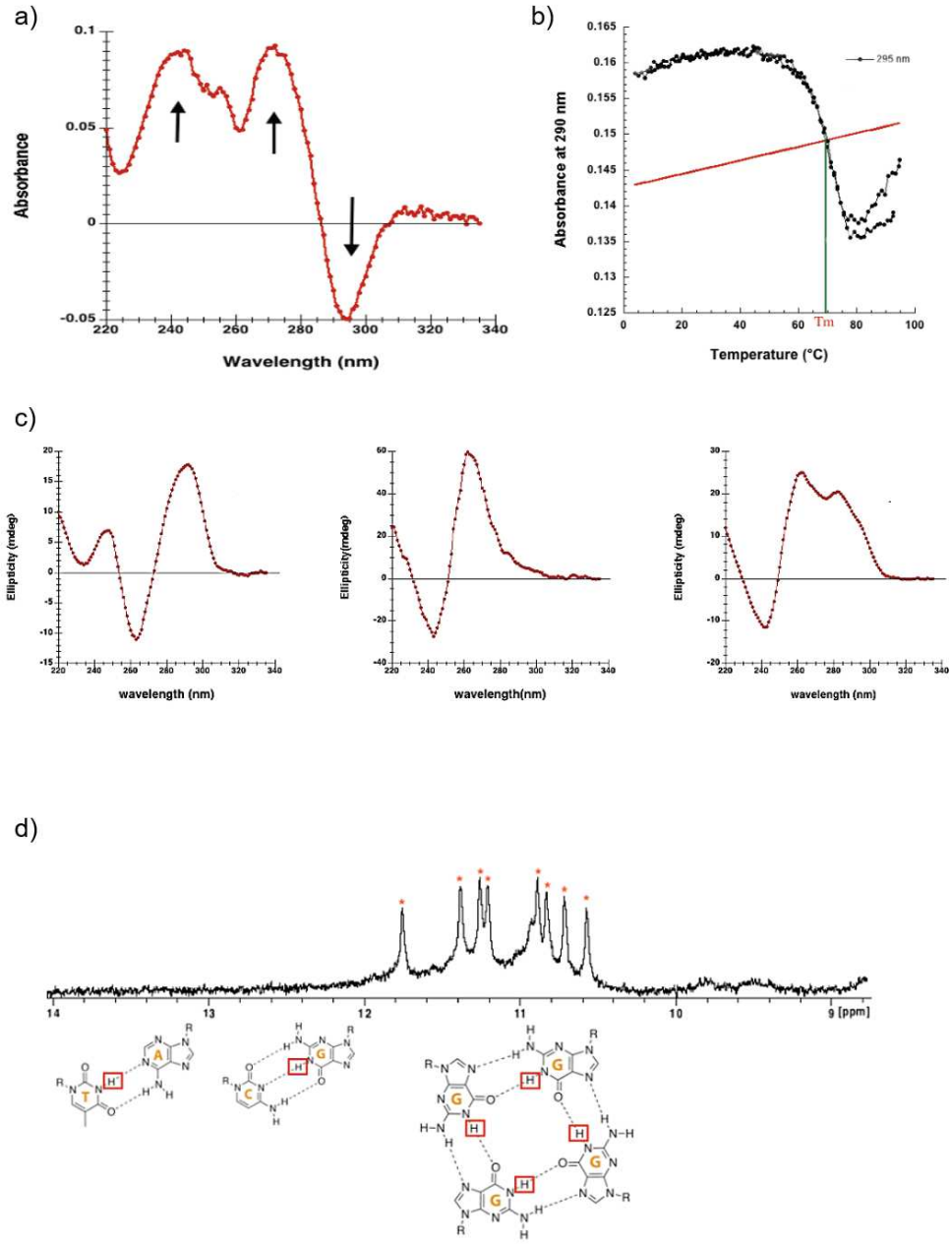


Figure 1.8.The characteristic feature of stable G4 obtained using different biophysical experiments.
TDS displays the signature peak of G4: two positive peaks near 240 nm and 270 nm and a negative peak at 295 nm b) UV melting depicts the measurement of the thermostability of G4 at 295 nm. c) The CD spectrum of anti-parallel G4 (positive peak near 290 nm and a negative peak near 260 nm), the parallel G4 (positive peak near 260 nm and negative peak near 240 nm), and hybrid type of G4 (two positive peaks (~295 nm and 260 nm) and negative peak near 245 nm) (left to right).d) 1D NMR spectra representing the imino peaks between 12 -14 ppm for Watson-Crick base-pairs [AT bp shows peak between 13-14 ppm while GC bp is in 12-13 ppm range], whereas the peaks of G4 exist between 10 -12 ppm. (Adapted from (Bedrat, 2017))

1.2.5.2 Biochemical methods

These methods are usually used to investigate the G4 formation in long nucleic acid sequences including flanking regions of G4 motif. These methods can also be employed in

characterizing the interaction between G4 and their binding proteins or ligands. Figure 1.9 displays the principle of the below mentioned techniques.

Dimethyl sulfate (DMS) footprinting can help in deducing the guanine bases involved in the G4 formation. Since N7 position of the guanine that are involved in Hoogsteen bonds become inaccessible to methylation, this leads to the formation of methylation protection pattern (decrease in the intensity or absence of the corresponding G band) when treated with DMS followed by piperidine cleavage (Buket *et al.*, 2019).

DNA polymerase stop assay provides a simple and rapid method to confirm that the G-rich sequence is capable of forming G4. This assay is based on the principle that the presence of secondary structure acts as a roadblock and pauses or stops the DNA polymerase at the 3'-end of the G4FS and thereby hinders its procession along the template (Wu and Han, 2019).

Electrophoretic Mobility Shift Assay (EMSA) is based on the principle that molecules in different size and charge display different electrophoretic mobilities when resolved in native polyacrylamide gel. Folded G4 or G4-complex with proteins/ligands tends to move slower than unfolded or free G4. This assay is highly sensitive and can be carried out using radioisotope or biotin labeled nucleic acid sequence (Buket *et al.*, 2019).

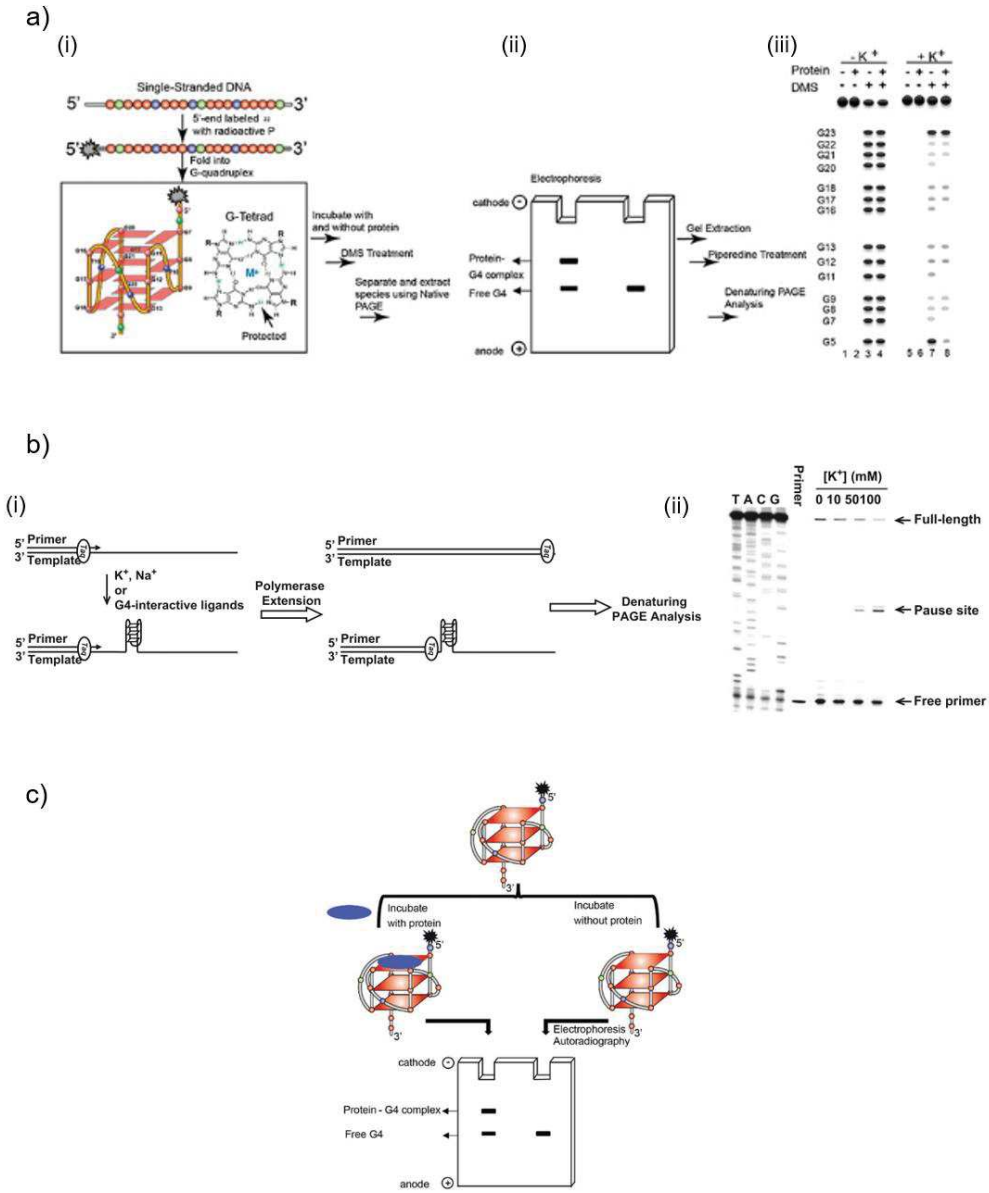


Figure 1.9.Schematic diagram of different biochemical tecnhiques.
a) DMS footprinting experiment, b)DNA polymerase stop assay and c) EMSA (Source (Buket et al., 2019; Wu and Han, 2019)

1.2.5.3 Genome wide in vivo G4-protein interaction using ChIP-Seq

The *in vivo* binding of G4-protein can be inferred from chromatin immunoprecipitation followed by high-throughput DNA sequencing (ChIP-Seq). In this experiment, antibodies against the proteins that binds to the G4s *in vivo* (Spiegel, Adhikari and Balasubramanian, 2019) (Figure 1.10).

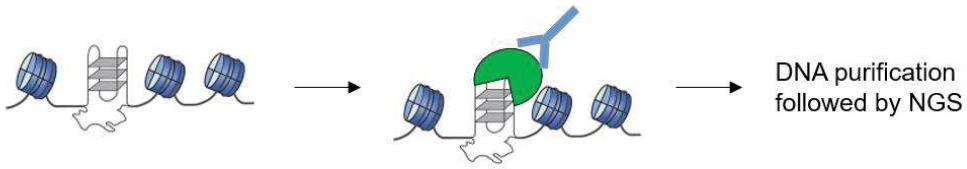


Figure 1.10.Representation of ChIP-Seq technique for the G4-protein interaction.
ChIP-Seq is performed by using antibody against tagged protein, followed by protocol similar to normal ChIP-Seq

1.2.6 Biological role of G4

Computational and experimental studies have provided several evidences of the prevalence of G4s in key regulatory regions of multiple genomes. Their existence in regulatory regions points towards their important role in various fundamental genomic processes, including transcription, DNA replication, and telomere maintenance (Rhodes and Lipps, 2015) (Figure 1.11).

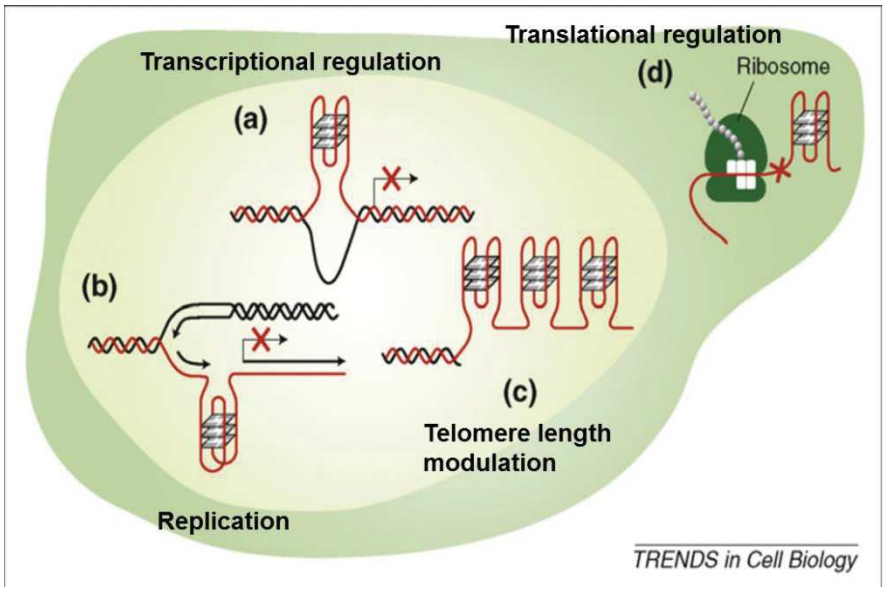


Figure 1.11.Biological roles of G4.
G4 are involved in transcription, replication, translation and telomere maintenance (Adapted from (Lipps and Rhodes, 2009))

1.2.6.1 Transcription

The prevalence of G4s within promoter and TSS sites of a gene suggests its role in regulation of gene expression. The direct evidence of effect of G4 on promoter activity during transcription was shown by the studies on the oncogene c-MYC (Siddiqui-Jain *et al.*, 2002). The nuclease hypersensitivity element III (NHE III), upstream of P1 promoter of *c-myc* gene forms a G4 that acts as a negative regulator of c-MYC expression which is further stabilized by the addition of G4-stabilizing compounds (TMPyP4). Whereas the disruption of the structure by single base mutation, lead to the increase in c-MYC expression. Some of the models described to understand the effect of G4 on transcription are enlisted below (Kim, 2019) (Figure 1.12).

a) G4 as roadblocks to RNA polymerase activity

The G4s have been shown to impede the transcription process *in vitro* by blocking RNA pol activity based on their location on the strand (sense or anti -sense) (Belotserkovskii *et al.*, 2010, 2013). For example, $(GGA)_4$ repeats downstream of TSS of human c-myb proto-oncogene can form G4. In the presence of K^+, the formation of this G4-complex inhibits the c-myb transcription elongation by arresting T7 RNA polymerase (Broxson, Beckett and Tornaletti, 2011).

b) G4 as docking sites for transcriptional factors

Some of the G4s can interact with transcription factors differently, either upregulating or blocking a transcription. For example, a zinc-finger transcription factor SP1, binds with higher affinity to two-tetrad G4 formed at the promoter of oncogene *c-KIT*. Genome-wide SP1 ChIP-on-chip data displayed a significant overlap between G4FS and SP1 binding sites (Raiber *et al.*, 2012). Nucleolin (NCL) is a common biomarker of a variety of cancers that can efficiently interact with G4. NCL serves as both transcriptional activator and repressor of human vascular endothelial growth factor (VEGF) gene and c-MYC gene, respectively. The NCL specifically binds to G4 present within poly-purine/pyrimidine tract of the promoter of human VEGF. Overexpression of NCL also increases the VEGF mRNA levels and thus functions as transcription activator (Uribe *et al.*, 2011). In contrast, when NCL interacts with G4 formed within c-Myc NHEIII region, it represses *c-myc* promoter activity (González *et al.*, 2009a).

c) G4-based regulation of transcription by nucleosome exclusion mechanism

A vast majority of actively transcribed genes in yeast and humans reveal a distinct nucleosome-depleted region upstream of TSS (Bai and Morozov, 2010). Interestingly, stable G4s are found to be highly enriched in nucleosome-depleted regions in human, yeast and *C.elegans* (Halder, Halder and Chowdhury, 2009; Wong and Huppert, 2009; Hänsel-Hertsch *et al.*, 2016). This indicates that stable G4s could also alter the deposition of chromatin modifiers and/or modify chromatin structure and stability. The presence of these structures could locally exclude nucleosomes to maintain open conformation that can further regulate gene expression.

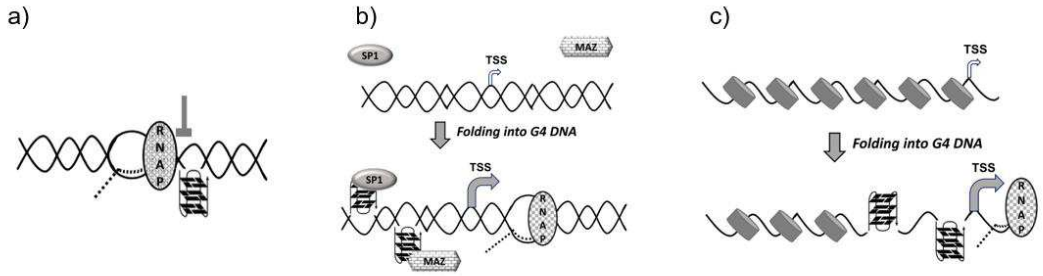


Figure 1.12.Different mechanism of G4-mediated regulation of transcription.
a) G4 as a roadblock in RNA polymerase progression, b)G4 as a docking sites for transcription factors (Sp1 and MAZ), and c) G4 as nucleosome exclusion region at upstream of TSS. (Adapted from (Kim, 2019))

Similarly, several G4-binding ligands such as TMPyP4, NMM, 360A and PhenDC3 have also reported the role of G4s in transcriptional regulation. The genome wide expression studies in the presence of G4 ligands have shown widespread changes in the transcription of numerous genes specifically G4-harboring genes in yeast, human, and mouse (Hershman *et al.*, 2008; Verma *et al.*, 2008; Halder *et al.*, 2012).

1.2.6.2 Replication

DNA replication is an essential process that ensures the correct duplication of the genome. G4s have recently emerged as key regulators in this replication process. They seem to have a dual role in the regulation of replication: as roadblocks to replication that leads to genomic and epigenomic instability, and as a component of replication origins.

a) G4 as a potential source of genomic and epigenomic instability

G4 formation impedes the progression of replication forks *in vivo*. Several studies have demonstrated the role of helicases and polymerases to ensure the smooth replication of these potential G4-based blocks (Brosh Jr, 2013; Wickramasinghe *et al.*, 2015) (Figure 1.13). The

first evidence was shown by *dog-1* (deletion of guanine rich DNA) gene, which encodes FANCJ helicase that can resolve G4s in *C. elegans*. Knockout of *dog-1* gene causes persistent replication stalling at G4s (Cheung *et al.*, 2002). The direct demonstration of the role of FANCJ in promoting G4-DNA replication was obtained in an *in vitro* system reproducing nascent strand progression (Castillo Bosch *et al.*, 2014). Similarly, *Saccharomyces cerevisiae* Pif1 helicase, responsible for unwinding G4 *in vitro*, is involved in maintaining the genomic integrity (Paeschke, John A. Capra and Zakian, 2011). Deletion of *Pif1* gene induces rearrangements in a G4 containing human CEB1 mini-satellite repeats that were inserted into the *S. cerevisiae* genome, primarily when G4s are present on the leading strand (Lopes *et al.*, 2011). Similar effect was observed when wild type strains were treated with potent G4 ligand, Phen-DC_3 that efficiently inhibits G4s unwinding by Pif1 *in vitro* (Piazza *et al.*, 2010).

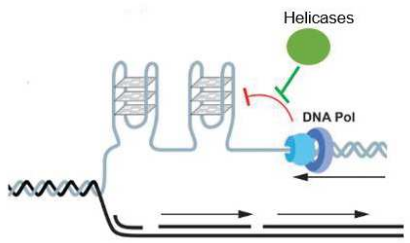


Figure 1.13.Role of G4 and G4 unwinding helicase in replication

The absence of REV1 helicase in chicken DT10 cell line allows the presence of a G4 block on the leading DNA strand thereby creating a post replication gap finally resulting in epigenetic instability. This blockage caused uncoupling of DNA replication and recycling of parental histones, which ultimately resulted into the loss of epigenetic marks over 4.5kb upstream region from G4. This was observed by the effect of G4 located 3.5kb downstream from the TSS that induced the repression of actively expressed locus, BU-1 along with the loss of H3K4me3 mark in the promoter region (Schiavone *et al.*, 2014).

b) G4 as a component of replication origins

Genome-wide analyses of replication initiation profiling revealed that the metazoan origins were enriched near CpG islands (Sequeira-Mendes *et al.*, 2009; Prorok *et al.*, 2019). Interestingly, 80% of all these origins mapped by small nascent DNA strand (SNS) sequencing (SNS-Seq) technique overlap with putative G4s in mouse and human cells. However, not all the mapped G4s overlap with replication origin, suggesting that G4s are not the sole determinants of origin specificity (Picard *et al.*, 2014).

1.2.6.3 Genome stability: Repair and recombination

The G-rich sequences are particularly prone to oxidative damage due to the presence of sensitive guanine bases. For instance**,** under hypoxic condition, oxidative stress induces the guanine base damage in G4 containing promoter sequence, which in turn recruits the DNA repair proteins (Ogg1 and Ref-1/APE1) that are involved in Base excision repair (BER) pathway (Clark *et al.*, 2012).

Besides, G4s are involved in recombination. Mani *et al* showed that G4FSs are present within the recombination hotspots in mammalian cells (Mani *et al.*, 2009).Their locus are significantly correlated with the binding sites of transcription factors such as c-Rel, NF-kappa-B subunits and Evi-1, which further suggests the role of G4s in transcription mediated recombination events (Mani *et al.*, 2009). Another study showed that during a meiotic event, the GC-rich region associated with DNA double strand breaks folds into G4. This intermolecular pairing is promoted by Hop1 in *Saccharomyces cerevisiae*, thereby revealing an important link between G4 and Hop1 in meiotic chromosome synapsis and recombination (Kshirsagar *et al.*, 2017).

1.2.6.4 Telomere maintenance

The telomeres are nucleoprotein complexes that are present at the ends of linear chromosomes. They protect the chromosome ends from degradation, end-to-end fusions, and being recognized as sites of DNA damage (Zakian, 2012) (Figure 1.14).. Most of the telomeric DNA consists of tandem repeats of guanine rich sequence which terminates as a 3′ single-stranded (ss) DNA overhang and is oriented 5’ to 3’ toward the chromosome terminus. These guanine-rich-overhang is a conserved feature from prokaryotes to eukaryotes with the propensity to form G4. Direct *in vivo* experimental evidence for the presence of G4s in telomere was shown by the telomeres of *Stylonychia* macronuclei using specific antibodies Sty49 against G4 (Schaffitzel *et al.*, 2001). These antiparallel G4 are resolved during DNA replication, suggesting that G4 might act as a telomeric capping structure.

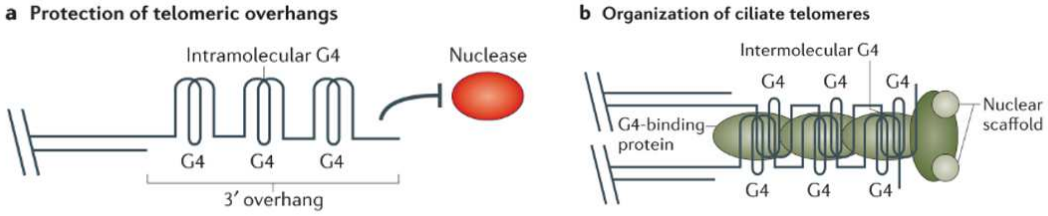


Figure 1.14.Putative role of G4s at telomeres.
a) G4 protects the telomere ends from nucleases, b)G4 take part in the organization of ciliate telomeres.(Source (Bochman, Paeschke and Zakian, 2012))

Furthermore, it has been reported that the formation of telomeric G4 is dependent on two telomere-binding proteins, TEBPα and TEBPβ *in vivo*. TEBPα binds to the telomeric G-overhang and recruits TEBPβ, which further promotes G4 formation (Paeschke *et al.*, 2005). During S phase, TEBPβ are phosphorylated and likely results into the dissociation of the complex and unfolding of G4. This would make telomeres accessible for the end-replication machinery.

However, there are also some factors that disrupt these telomeric G4 and help in telomere replication and elongation. For example, human POT1 (protection of telomere protein 1) binds to telomeric G-strand and disrupt telomeric G4, thereby allowing elongation by telomerase (Zaug, Podell and Cech, 2005). Helicases such as yeast Sgs1p and human RecQ (WRN and BLM) unwinds telomeric G4 in ATP dependent manner and promotes telomeric DNA replication. Deletion of these yeast and human helicases exhibit premature aging in yeast and cause human disease (Werner's and Bloom's Syndrome), respectively (Sun, Bennett and Maizels, 1999; Mohaghegh, 2001).

The G4-binding ligands act as a potent inhibitor of telomerase activity, which are used as potential cancer therapeutic agents (Rezler, Bearss and Hurley, 2002). For instance, telomestatin exhibits both high G4 affinity and telomerase inhibitory potency. Treatment with Telomestatin causes gradual telomere shortening and growth arrest in human cancer cells (Rezler, Bearss and Hurley, 2002).

1.2.6.5 Translation

Besides DNA, the G-rich sequences in RNA are also capable of forming G4s. These G4-RNA are usually more stable than G4-DNA (Millevoi, Moine and Vagner, 2012). In human

mRNAs, G4-RNA are highly enriched within 5' and 3' UTRs of the primary transcripts, two regions that are essential in post-transcriptional regulation of gene expression (Huppert *et al.*, 2008) (Figure 1.15).

The 5' UTR is a key element for the initiation of translation. Existence of G4-RNA in the 5' UTR of several mRNAs reduces their translation efficiency (Bugaut and Balasubramanian, 2012). This might be due to steric hindrance caused by G4-RNA for the binding of the translation pre-initiation complex (eIF4F complex) or progressing of 40S ribosomal subunit along the 5'-UTR (Gray and Hentze, 1994). Besides this repressive activity, 5'UTR G4-RNA can also activate the translation. G4 present in IRES region within 5'UTR of mRNA of human vascular endothelial growth factor (hVEGF) and fibroblast growth factor 2 (hFGF-2) mRNA acts as a positive translational element (Bonnal *et al.*, 2003; Morris *et al.*, 2010).

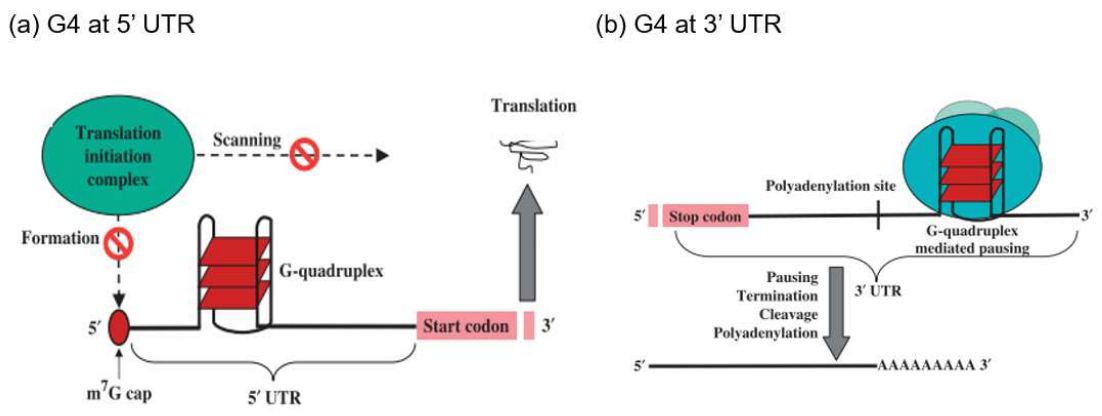


Figure 1.15.Putative role of G4 in the translation.
Location of the G4 regulates the translationary machineries. G4 at 5' UTR blocks the translation by acting as a roadblock for the translation initiation complex while G4 at 3' UTR results into truncated mRNA molecules. (Source (Huppert et al., 2008)

Additionally, several reports have shown that the G4-RNA located within 3'UTR of mRNA can also affect the translation efficiency. For instance, G4-RNA in the 3'UTR of the proto oncogene PIM-1 acts as a transcription repressor (Arora and Suess, 2011). Additionally, Fragile X mental retardation protein (FMRP), known G4-RNAbinding protein, can bind to the G4 of APP ORF to inhibit APP mRNA translation (Westmark and Malter, 2007).

1.3 Role of G4 in human pathogens

Over the past few years, G4s has gained momentum as an important pathogenicity factor for various microbial pathogens infecting humans. These pathogens include a wide range of organisms such as bacteria, protozoa, and viruses. Within these pathogens, the G4 motifs could influence virulence phenotypes such as antigenic variation and viral latency. Investigations of the G4-mediated regulations in pathogen virulence may unravel their pathogenic mechanisms and aid to develop novel therapeutic interventions. Similar to anti-cancer strategies where telomeric and oncogenic G4s are targeted by G4-specific ligands (Quarfloxcin) (Drygin *et al.*, 2009), these ligands could also be used to target DNA or RNA of pathogens.

1.3.1 Viruses

Viruses replicate within their hosts by exploiting host's replication and protein synthesis machineries. The G4FSs are also present in several viral genomes (Metifiot *et al.*, 2014; Lavezzo *et al.*, 2018), for instance human immunodeficiency virus (HIV- 1) (Sundquist and Heaphy, 1993), Epstein–Barr virus (EBV) (Norseen, Johnson and Lieberman, 2009), and Herpes Simplex virus (HSV) (Biswas *et al.*, 2016). Recent emerging evidence proposes the involvement of these G4s in viral replication and recombination, regulation of gene expression, and in key steps of viral life cycles. Thus, targeting of these G4s by G4 ligands have demonstrated consequent anti-viral effects (Ruggiero and Richter, 2018).

Human immunodeficiency virus (HIV- 1)

HIV-1 is responsible for the acquired immune deficiency syndrome (AIDS), which currently affects ~ 35 million people worldwide. It has a single stranded RNA genome and belongs to the family of *Retroviridae* family. It contains two copies of genomic RNA that are converted into dsDNA by viral retrotranscriptase. This newly formed viral dsDNA integrates into the host's genome to form proviral genome, which encodes for viral proteins and genomic RNA. In this provirus state, the expression of viral genes and the full length genome are regulated by HIV-1 long terminal repeat (LTR). Perrone *et al* reported that three mutually exclusive G4 (LTR-II, LTR-III and LTR-IV) are present in U3 region of the LTR promoter. Amongst them, LTR-III and IV acts as a regulatory mechanism in HIV-1 promoter activity (Perrone, Nadai, Poe, *et al.*, 2013). LTR-III decreases the LTR promoter activity while this activity is enhanced by LTR-IV G4 formation (De Nicola *et al.*, 2016). These G4 activities are further

mediated by the binding of two nuclear proteins: nucleolin and human ribonucleoprotein (hnRNP) A2/B1 (Tosoni *et al.*, 2015; Scalabrin *et al.*, 2017).

Besides LTR, G4s are also found in the most conserved region of *nef* gene (Perrone, Nadai, Poe, *et al.*, 2013). Nef is an essential factor that is involved in proviral DNA synthesis and pathogenesis (Aiken and Trono, 1995).

Furthermore, the enhanced stabilization of HIV G4 using G4 ligands is shown to have anti-viral effects at different levels. For instance, BRACO-19 stabilizes both DNA and RNA G4 to inhibit the LTR promoter activity and the reverse transcriptase activity, respectively (Perrone, Nadai, Frasson, *et al.*, 2013; Perrone *et al.*, 2014).TMPyP4, another G4-ligand, stabilizes G4 that impairs the Nef-mediated enhancement of HIV-1 infectivity (Perrone, Nadai, Poe, *et al.*, 2013).

Herpes virus

Herpes virus species belongs to the family of *Herpesviridae* with long linear double stranded DNA genomes. The genome-wide bioinformatics analysis revealed high density of G4FSs density among herpes virus species (Biswas *et al.*, 2016). These G4FS are enriched in repeat regions, regulatory regions, and immediate early genes, thus suggesting potential regulatory role of these G4s (Biswas *et al.*, 2016; Frasson, Nadai and Richter, 2019). The experimental evidence of G4s has been reported for Herpes Simplex virus 1 (HSV-1), Epstein–Barr virus (EBV), Kaposi's sarcoma associated herpes virus (KSHV) and human herpes virus 6 (HHV-6).

HSV-1 establishes life-long persistent infections that affect more than half of the world population. The HSV-1 genome has a very high GC rich content (68%) and contains multiple clusters of highly stable G4. These G4s are mainly located in the repetitive regions of the genome (Artusi *et al.*, 2015). In HSV-1 infected cells, G4s were visualized with G4-specific monoclonal antibody (1H6). Interestingly, the formation and localization of G4 are dependent on virus cycle. Viral G4 formation peaks during viral replication and can localize in different host cell's compartments according to the viral genome movement (Artusi *et al.*, 2016). Treatment with the G4 ligand, BRACO-19 stabilizes these G4 and inhibits virus production. In addition, it also affects the viral DNA replication by inhibiting the Taq polymerases processing (Artusi *et al.*, 2015). This work provided the proof of concept for using G4-ligands as new therapeutic agents.

Epstein–Barr virus (EBV) can induce infectious mononucleosis and cancers. EBV encodes a genome maintenance protein (nuclear antigen 1, EBNA1) that is essential for replication and maintenance of the genome during latent infection by EBV in proliferating cells. EBNA1 recruits the cellular origin recognition complex (ORC) through an RNA G4-dependent interaction with EBNA1 linking region 1 and 2 (LR1 and LR2) (Norseen, Johnson and Lieberman, 2009). The interaction between EBNA1 and ORC has shown to be disrupted with the treatment of G4-binding compound BRACO-19, thus inhibiting the EBNA1-dependent DNA replication. BRACO-19 also acts as a potent inhibitor of EBNA1 metaphase chromosome attachment (Norseen, Johnson and Lieberman, 2009).

Interestingly, the G4 formed in EBNA1 mRNA itself acts as cis-acting regulators of viral mRNA translation, producing ribosome dissociation. Such existence of G4s in mRNA tightly governs the synthesis of EBNA1 mRNA in such a way that it is sufficiently high to maintain viral infection but at the same time below the threshold levels for host immune recognition (Murat *et al.*, 2014). The stabilization of these G4s with Pyridostatin, decreases the EBNA1 synthesis, which further highlights the importance of G4s within virally encoded transcripts as unique regulatory signals for translational control and immune evasion (Murat *et al.*, 2014). These findings suggest alternative therapeutic strategies based on targeting RNA structures within viral ORFs.

1.3.2 Prokaryotic pathogens

Prokaryotic pathogens lack a defined nucleus, cytoskeletal elements (microtubules) and membrane bound organelles. The literature describing putative roles for G4s in these pathogens is noticeably smaller and less persuasive. So far, in prokaryotes, G4 motifs have been identified in *Escherichia coli*, *Deinococcus radiodurans*, *Mycobacterium tuberculosis*, and *Xanthomonas and Nostoc sps* (Rawal *et al.*, 2006; Beaume *et al.*, 2013; Rehm *et al.*, 2015; Perrone *et al.*, 2017). G4 destabilizing helicases such as Pif1 and RecQ have also been identified in *E.coli* (Mendoza *et al.*, 2016).

One of the most compelling studies for the functional role of G4 is pilin antigenic variation system, which is employed by prokaryotic pathogen *Neisseria gonorrhoeae* (Figure 1.16). *N. gonorrhoeae* is responsible for the sexually transmitted infection gonorrhea. The pilin antigenic variation (AV) system of N. *gonorrhoeae* mediates non-reciprocal DNA recombination between one of the 19 silent gene copies (pilS) and the pilin expression locus (pilE). The recombination initiation site is a 16bp GC-rich motif, which is located upstream

of the pilE locus (Cahoon and Seifert, 2009). This motif forms a parallel intramolecular G4 *in vitro* and mutation in any of the GC base pairs, but not AT base pair, results into the inhibition of pilin AV. The recombinant RecA was shown to bind to this pilE G4 *in vitro*, suggesting that these G4 could recruit RecA to promote strand exchange during recombination (Kuryavyi *et al.*, 2012). The Neisseria RecQ helicase, similar to other RecQ families, can efficiently unwind the pilE G4 while the deletion of RecQ HRDC domain decreases the frequency of pilin AV (Cahoon *et al.*, 2013). Moreover, the replacement of the pilE G4 with other predicted G4 motifs did not allow pilin AV. Also, the integration of pile G4 motif in the upstream of the pilS copy did not direct any gene conversion to the silent locus (Cahoon and Seifert, 2009). This data establishes the role for G4 in pilin AV, which could be most likely to initiate the programmed recombination process.

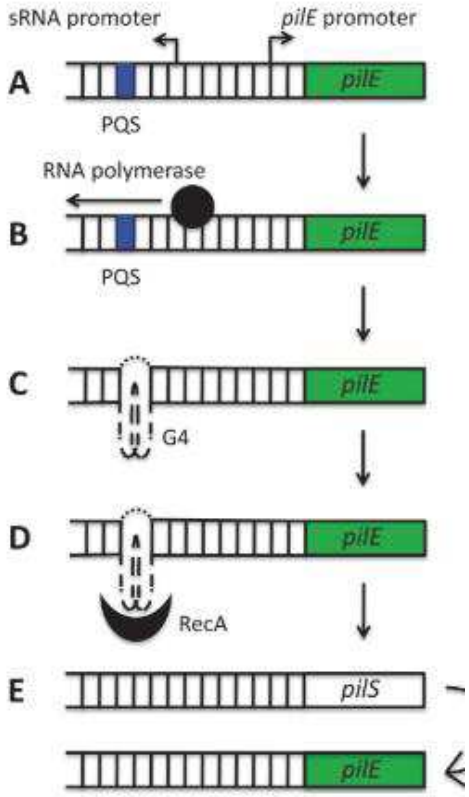


Figure 1.16.Schematic model of the role of the pilE G4 in N. gonorrhoeae pilin antigenic variation
Source (Harris and Merrick, 2015)

Another prokaryote, the *Mycobacterium tuberculosis*, with 65% GC-rich genome, possesses ~10000 G4FS. A majority of the stable G4s are present in the promoter regions of genes with definite functional categories (Rawal *et al.*, 2006; Perrone *et al.*, 2017). Additionally, the G4-specific helicase (MtDinG), and G4 aptamer (PPK2 G9) that inhibits a polyphosphate kinase

2 activities were identified in these pathogens (Shum *et al.*, 2011; Thakur *et al.*, 2014). Furthermore, treatment with G4 ligands BRACO-19 and c-exNDI are able to inhibit Mtb growth *in vivo*. Thus, this data supports the presence of G4 in *M. tuberculosis* as well as suggests the possibility of using G4s as novel targets to develop antitubercular agents (Perrone *et al.*, 2017).

1.3.3 Eukaryotic pathogens

In general, protozoan pathogens are a unicellular eukaryotic parasite that has virulence factors and pathogenic mechanisms analogous to prokaryotic and viral pathogens. This section, particularly, focuses on the role of G4s in early diverging eukaryotes such as *Trypanosoma brucei* and *Plasmodium falciparum*.

Trypanosoma brucei is responsible for African sleeping sickness in humans. *Trypanosoma* genome displays strong G4 enrichment in promoter and TSS regions, which is similar to higher species such as human and mouse (Marsico *et al.*, 2019). The authors suggested the presence of G4s in regulatory regions might be implicated in transcriptional regulation within these parasites.

Along with G4-DNA, G4-RNA has also been reported in *Trypanosoma spp*. Being kinetoplastids, *Trypanosoma* carries kinetoplast that contains the mitochondrial DNA (kDNA). kRNA transcribed from these kDNA undergoes an unusual mRNA editing process. In such pan-editing process, hundreds of U-nucleotides are either inserted or deleted from primary transcripts to generate functional kRNAs (Aphasizhev and Aphasizheva, 2014). These sites are specified by the guide RNAs and their positions are positively correlated to the position of intramolecular G4 in the pre-kRNA (Leeder *et al.*, 2016). During the pan-editing process, editosome unwinds upto 50% of G4-RNA to favour the formation of pre-kRNA:gRNA hybrid RNAs for maturation into functional kRNA (Leeder *et al.*, 2016). When these G4-RNA in the pre-kRNA is resolved, it also allows the formation of a kDNA/kRNA hybrid (HQ) with the non-coding strand of the maxicircle DNA, thus preventing transcription and activating kDNA replication. Whereas when these G4-RNA are formed on the neo-synthesized pre-kRNA, it leads to active transcription and inactive maxicircle replication (Leeder, Hummel and Göringer, 2016) (Figure 1.17).. Hence, this data highlights the new role of G4s in the *Trypanosomatidae* biology.

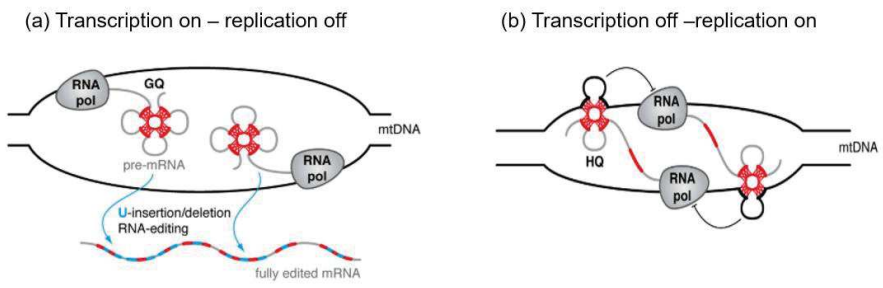


Figure 1.17.Model describing the role of G4 in mitochondrial DNA (kDNA) in Trypanosoma brucei.
a) Presence of G4 (GQ) in the neo-synthesized prekRNA, it defines transcription on/replication off state. RNA editing is required to resolve this structure and generate translatable mRNA. b) While the formation of G4 (HQ) between the new formed RNA and the nontemplate strand of the maxicircle DNA, results in transcription-OFF/replication ON state. (adapted from (Leeder, Hummel and Göringer, 2016))

Furthermore, a screen of carbohydrate naphthalene diimide derivatives (carb- NDI), previously reported G4 ligand, demonstrated the antiparasitic activity against *T. brucei* and limited toxicity in human, making them potential drug candidates against *T. brucei* (Belmonte-Reche *et al.*, 2018). G4-binding fluoroquinolone compound quarfloxin, also known as a Pol I inhibitor, is shown to have trypanocidal activity with IC_{50} of only 155nM (Kerry *et al.*, 2017).

Plasmodium falciparum is another example of divergent eukaryotic parasite where the evidence of G4s has been reported. The detailed information on the role of G4 in these malarial parasites will be explained in the next section.

1.4 G-quadruplexes in *P. falciparum*

P. falciparum is a malaria causing parasite that exhibits unique gene regulations and antigenic variations. They have a highly AT-rich genome (>80%) with uniquely localized GC-motifs that have the propensity to form G4.

1.4.1 Identification of G4s in *P. falciparum*

The first bioinformatic G4FS screen of the *Plasmodium falciparum* genome was published in 2009 (Smargiasso *et al.*, 2009). Using QGRS-Mapper, they identified 891 G4FS in the *Plasmodium* genome. The majority of these G4FS are located within the telomeric regions (828/891) due to the GC-rich (GGGTT(T/C)A) repeats. While the rest of the 63 G4FS are located in non-telomeric regions. 16 of these 63 G4FS are present in the upstream region (predominantly within B region) of *var* genes, which were confirmed to adopt G4 *in vitro*.

The *var* genes encode for the major surface antigen PfEMP1 whose expression is regulated by epigenetic silencing and switching (Scherf, Lopez-Rubio and Riviere, 2008). The study suggested that the presence of G4 within *var* genes could be involved in the regulation of these virulence genes (Smargiasso *et al.*, 2009). Later on, using an updated version of the *P. falciparum* genome, 80 G4FS were identified within the *var* genes (Stanton *et al.*, 2016). The same study also anticipated the strong association between G4FS and recombination events among virulence genes (Stanton *et al.*, 2016).The *in vivo* existence of G4s was shown using G4-pecific antibodies, 1H6 and BG4 (Harris and Merrick, 2015).

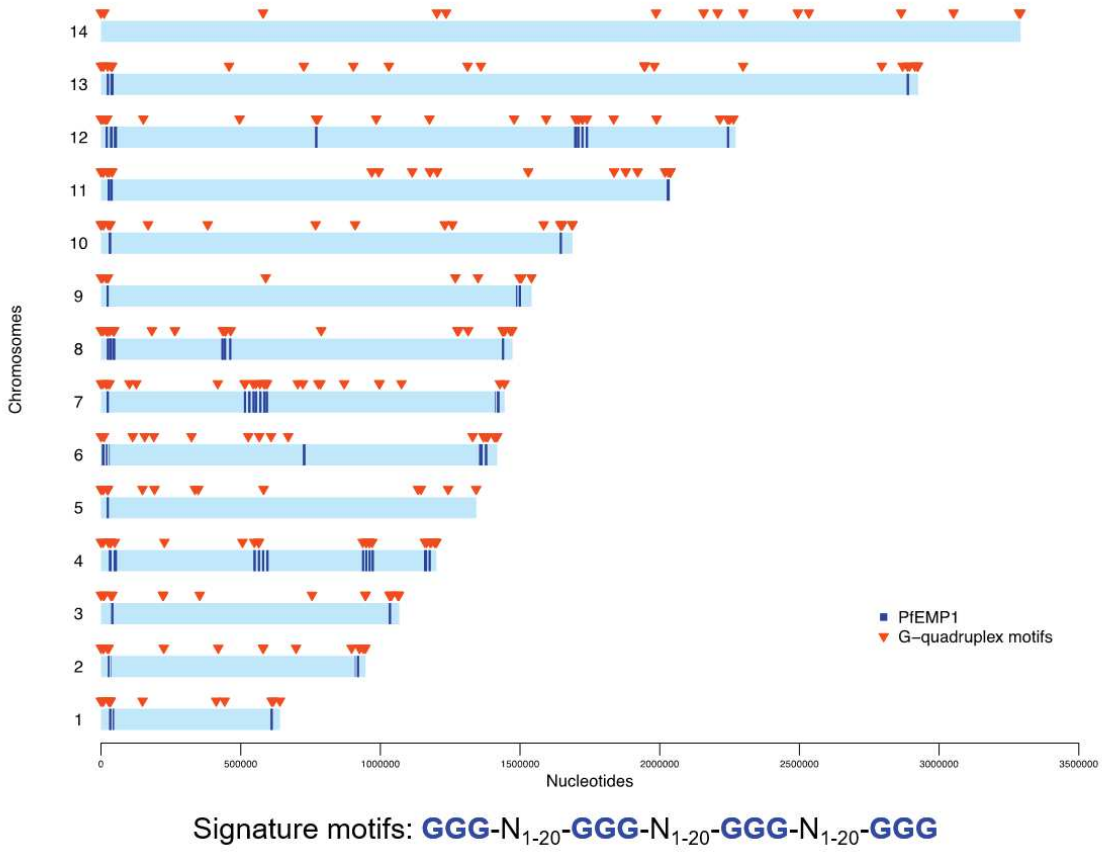


Figure 1.18.Distribution of cannical G4FS in Plasmodium genome
(adapted from (Stanton et al.*, 2016))*

In another study, Bhartiya *et al* performed a G4FS screening within the genome of 6 related species of *Plasmodium* (*P. falciparum*, *P. vivax*, *P. knowlesi*, *P. chaubadi*, *P. berghei*, and *P. yoelii*) using Quadfinder (Bhartiya *et al.*, 2016). They observed that amongst all investigated genomes, the *P. yoelii* and *P. chaubadi* displays a lowest number of G4FS and apparently *P. falciparum* contains a significantly low number of G4FS than *P. vivax* and *P. knowlesi*. They also showed that high densities of G4FSs are present in mitochondrial genome as compared to the nuclear and apicoplast genome of *P. falciparum*. It was the first study that showed the

G4FS present within the UTR regions of the genes can also adopt G4 formation *in vitro,* thus providing the evidence of G4-RNA in *Plasmodium* (Bhartiya *et al.*, 2016).

A genome-wide experimental mapping of G4s in multiple species including *P. falciparum* was published recently (Marsico *et al.*, 2019). In this study, they identified 173 stable G4 in the *Plasmodium* genome using previously described G4-Seq approach. These G4s are highly enriched throughout the genome except in intragenic regions (Marsico *et al.*, 2019).

In an unpublished data (from our lab), Gazanion E. *et al* identified a total number of 1,763 to 145 G4FS with a threshold of the score from 1.2 to 1.75 using an advanced G4 algorithm, named G4Hunter (Bedrat, Lacroix and Mergny, 2016). These identified G4FS, particularly at *var* gene, showed a conserved pattern within the closely related *Laverania Plasmodium* parasites. Furthermore, this data also highlights the significant correlation of G4FS with the nucleosome depleted regions and recombination hotspots within *Plasmodium* genome (Gazanion *et al.*, 2020).

1.4.2 Potential role of G4

As discussed earlier, G4s play an important role in different biological processes. The presence of G4s in *Plasmodium* genome has intrigued the researchers to explore the potential role of G4s in these parasites as well. However, this expedition on the role of G4s is still in early stages in these malaria parasites. Here, I have discussed some of the potential roles that have been postulated so far.

Recombination and antigenic variation:

PfEMP1 is an essential antigen that takes part in cytoadherence and is responsible for disease pathogenesis. This protein is encoded by *var* genes. At any given point, one *var* gene is expressed out of 60 copies. Recombination and/or switching of these *var* genes helps the parasites to avoid the host immune response. Bioinformatics studies have shown the predominance of G4FS within the regulatory regions of *var* genes, suggesting their role in *var* gene regulation (Smargiasso *et al.*, 2009; Stanton *et al.*, 2016; Gazanion *et al.*, 2020). Besides, a strong correlation between the G4FS and the recombination breakpoint, which further supports the role of G4s in the generation of diversity within virulence genes (Stanton *et al.*, 2016).

Recently, the two homologs of RecQ helicase, PfWRN and PfBLM, have been identified in Plasmodium. PfRecQ helicases were shown to influence the G4-related phenotypes. Interestingly, deletion of *PfWRN* gene showed increase in recombination events in *var* genes

with high indel mutation rate and generation of extra chimeric *var* genes in these parasites (Claessens *et al.*, 2018).

Transcription regulation:

Role of G4s in the transcription regulation are very well characterized in other eukaryotes (Rhodes and Lipps, 2015). Likewise, few studies have also proposed the potential role of G4 in the regulation of gene expression in Plasmodium. Classens et al showed that the loss of PfRecQ helicase, known G4-unwinding enzyme in other organisms, causes a major transcription change in several genes including var genes throughout the genome (Claessens *et al.*, 2018).

Similarly, *Gazanion et al* (unpublished data) demonstrated the genome-wide expression of several genes was affected by the treatment with Pyridostatin, G4 stabilising compound). The most affected genes include half of the members of ApiAP2 TF, genes involved in ribosome biogenesis and genes that are part of metabolic pathways related to DNA biology(Gazanion *et al.*, 2020).

Telomere maintenance:

Similar to other eukaryotes, the Plasmodium telomeric sequence also comprise of GC repetitive motifs (GGGTTYA, where Y is T/C). De Cian et al showed that the G-rich telomeric sequence can form G4 *in vitro*. These telG4 can be stabilized by the known G4 ligands such as Phen-DC3, telomestatin, TMPyP4, similar to human telG4 (De Cian *et al.*, 2008). Moreover, the telomerase activity of Plasmodium telomerase was inhibited by TMPyP4 and telomestatin treatment at an IC50 of 4 µM and 37.5 µM, respectively (Calvo and Wasserman, 2016).

Translation:

Recently, the presence of post-transcriptional and translational regulation was reported using RNA-Seq and Ribosome profiling in these parasites (Bunnik *et al.*, 2013; Caro *et al.*, 2014). Bhartiya et al analyzed these data (Bunnik *et al.*, 2013; Caro *et al.*, 2014) and observed no correlation between mRNA abundance data of G4FS harboring genes and their respective ribosome occupancy (Bhartiya *et al.*, 2016). Thus, suggesting towards the regulatory dynamics of G4FS containing genes in these parasites. Furthermore, G4FS harboring genes showed lower translational efficiencies in ring and merozoite stage than non-G4FS genes. However the exact mechanism remains largely unknown.

1.4.3 G4s ligands as antimalarial agents

In Plasmodium, several studies reported the use of G4 ligands as antimalarial agents. Most of these ligands are targeted against the telomeric G4. For instance, TMPyP4 and telomestatin have shown to affect the growth of the parasites and inhibit the telomerase activity (Calvo and Wasserman, 2016). Bisquinolinium derivates of 1,8-naphthyridine and pyridine (3AQN, 6AQN and 360A) were shown to induce structural changes and enhance the stability of the Pf telG4 *in vitro*. These ligands exhibited promising antimalarial activity in both Chloroquine sensitive and resistant strains with minimal effect on human cells in drug sensitivity assay(Anas *et al.*, 2017). The exposure of these ligands resulted in disruption of telomere homeostasis and transcriptional activation of multigene families (*var*, *rifin*, and *stevor*) and upregulation of genes involved in DNA repair and recombination such as Mre11, H2A, and RPA1 (Anas *et al.*, 2017).

Recently, G4-binding compound, a fluoroquinolone called Quarfloxin was examined as an antimalarial drug. Quarfloxin is previously used as an anticancer drug in Phase II clinical trials and has already shown to have trypanocidal activity with an IC_{50} of only 155nM (Drygin *et al.*, 2009; Kerry *et al.*, 2017). In *P. falciparum,* Quarfloxin inhibited the parasite growth with EC_{50} of 114nM, especially at ring and trophozoite stages but the mechanism is not yet clear (Harris *et al.*, 2018).

Objectives of the Thesis

Even though the research on G4s in *Plasmodium* is still in the early stages, but it is clear that these non-canonical structures have a potential role in gene regulation and antigenic variation of virulence genes in *P. falciparum*. To gain more insight into the role of G4s, the goal of the thesis is to identify and unravel the role of proteins that can interact with these G4s and hence modulate their function in these malarial parasites.

To achieve this goal, my thesis aimed for the following objectives:

1. Identification of G4 binding proteins in *P. falciparum* using two different and complementary unbiased approaches: Yeast one-hybrid assay and DNA-pull down assay with subsequent mass-spectrometry (MS) analysis.
2. Functional characterization of selected potential G4-binding proteins in *P. falciparum* by generating inducible Knockout parasite lines.
3. Validation of the interaction between the selected candidate and G4 by performing EMSA and ChIP-Seq.

Chapter 2

Methods and Materials

Chapter 2 Methods and Materials

2.2 *Plasmodium falciparum* culture and transfection

2.2.1 Parasite lines

The following *P. falciparum* lines were used in this dissertation.

a) *P. falciparum* NF54 was used to grow parasites for the DNA pull-down assay.
b) *P. falciparum* 3D7 p230pDiCre line was used to generate inducible knockout strains (provided by E. Knuepfer) (Knuepfer *et al.*, 2017) .

2.2.2 Parasite maintenance

2.2.2.1 In vitro culture conditions

Parasites were cultured in A^+ human erythrocytes in RPMI 1640 medium (Gibco Life Technologies, 52400 RPMI 1640, HEPES) supplemented with 5% human serum and 0.5% Albumax II, 0.2 mM hypoxanthine (C.C.Pro GmbH) and 25 µg/mL gentamicin (Sigma).The cultures were kept at 37°C under a controlled trigaz atmosphere (3% CO_2 , 5% O_2 and 92% N_2). Synchronization of parasites were done by a sequential combination of Percoll(Saul *et al.*, 1982) and sorbitol treatment (Lambros and Vanderberg, 1979). Parasite development was monitored by giemsa staining (Moll *et al.*, 2013).

2.2.2.2 Thawing of parasites

Glycerol frozen parasites were warmed for 1 min in a 37°C water bath, followed by slow addition of 0.2V of solution A (12% NaCl, w/v) and incubated for 5 min at RT. After the incubation, 10V of solution B (1.6% NaCl, w/v) was added and then centrifuged. Parasites were washed with washing media (RPMI medium with gentamycin) and resuspended in culture media with desired hematocrit (Moll *et al.*, 2013).

2.2.2.3 Freezing of culture

Young ring stage (< 8 hr) parasites were used to freeze the parasite culture. Parasites were centrifuged and pre-warmed freezing solution (6.2 M Glycerol, 0.14 M sodium lactate, 0. 5 mM KCl in PBS) was added dropwise in varying amount: 0.4, 1.2 and 2.4 volume of the

parasite pellet with an interval of 5 min incubation at RT. Aliqout of 1ml of the above mixture was stored in cryovial at -80 ^{0}C.

2.2.2.4 Synchronization of parasites

Sorbitol treatment

In this method, young ring stage (<10 h) parasites were centrifuged and the pellets were mixed vigorously with 9V of warm 5% (w/v) sorbitol. *(Vigorous vortexing results into the rupture of old RBC and mature forms of parasites).*This mixture was incubated for 8 min at 37°C and then centrifuged. The obtained pellets containing only ring stages were washed and resuspended into the new culture media (Lambros and Vanderberg, 1979).

Percoll treatment

Late stage parasites (late trophozoites and schizont stages) were used in this synchronization method. 3ml of media containing suspended infected erythrocytes pellet were carefully poured on the top of 70% Percoll solution in a 15 ml falcon. After centrifugation at 2200 rpm for 11 min (without brake), the top layer were removed carefully and washed with washing media. This washed pellets containing majority of the late schizonts were resuspended into the new culture media with desired hematocrit. (Saul *et al.*, 1982; Radfar *et al.*, 2009) In order to get tightly synchronized parasites, these percoll obtained mature schizonts were treated with sorbitol after 1 hour of incubation.

2.2.2.5 Harvesting of parasites by Saponin lysis

Infected erythrocytes were centrifuged and pellets were suspended in 0.15% cold Saponin solution(w/v) in PBS.(Moll *et al.*, 2013) This mixture was incubated on ice for 5 min and then spun at 3000 rpm for 5 min at 4 °C to separate the saponin pellet from the lysate. The pellet was washed twice with PBS and can be stored at -80 °C.

2.2.3 Generation of inducible Knockout parasite lines

In order to generate the inducible Knockout parasite strains, we have used the combined technology of CRISPR-Cas9 and DiCre/Loxp system (Knuepfer *et al.*, 2017). The strategy of this system is described in Figure 2.1.

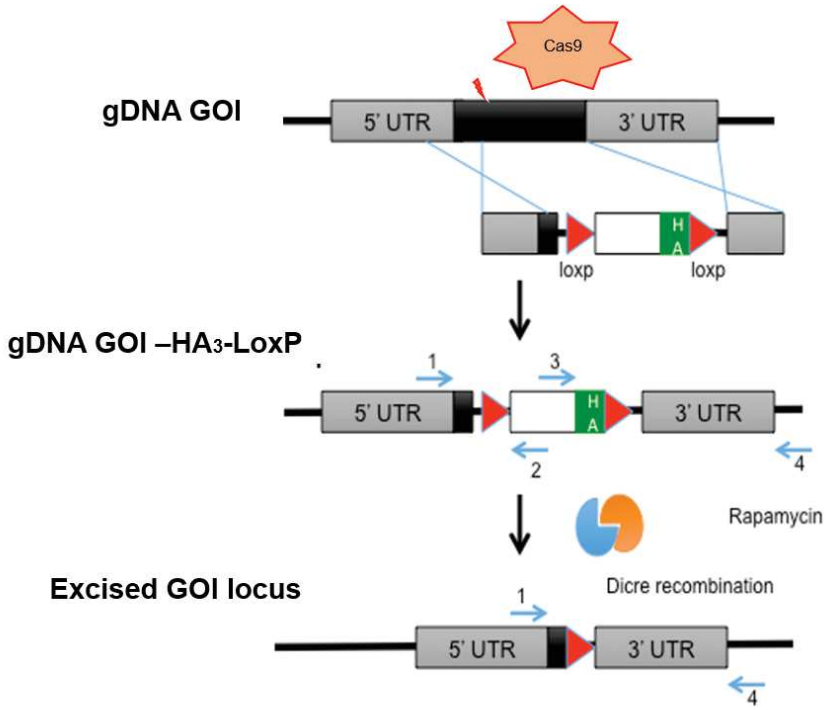


Figure 2.1.Generation of inducible knockout parasite line
Using CRISPR-Cas9 system, the endogenous copy of the gene of interest (GOI) is replaced by the modified gene cassette consist of recodonized gene with fused C-terminal HA tag and two loxp sites, along with flanking 5' and 3' homology region. This results in the formation of endogenously HA tagged transgenic line. When this obtained transgenic line is treated with rapamycin, the presence of rapamycin activates the dimerization of DiCre recombinase , which then further recognizes the loxp sites to provoke excision of the gene and results into the generation of KO line. Primers (1-4) are used for verification of the transgenic line: GBP2 (1- PG120, 2- PG121, 3- PG122 and 4- PG123) and DNAJb (1- PG174, 2- PG178, 3- PG176 and 4- PG182)

2.2.3.1 Construction of plasmids used in the parasite transfections

The modified pLN vector (consisting of multiple cloning site and HA_3loxp region) was used as a template for donor plasmid (pLN-don-HA_3loxp) where 5' homology region, modified gene coding region and 3' homology regions were cloned using infusion cloning (Figure 2.2). PCR amplified 5' HR spanning ~ 300 bp upstream of start codon to coding region of respective gene until the insertion site of loxpintron was cloned into *Apa*I and *Eco*RV; Gblock consisting of loxpintron and recodonised gene until stop codon were synthesized by geneart gene synthesis (ThermoFisher) and cloned into *Bse*R1 and *Mlu*I; PCR amplified 3' HR containing 3'UTR of the gene was cloned into *Spe*I and *Xho*I. All the PCR reactions were performed using *Pfu*Ultra II fusion DNA polymerase and resulting plasmids were verified using Sanger sequencing. The gRNA (GBP2-PG101/102 and DNAJb- PG145/146 & PG147/148) were selected for the respective gene using CHOPCHOP tool and were cloned at the *Bbs*I site of pDC- Cas9 U6-hdhfr/yfcu vector (gifted by Ellen knuepfer).

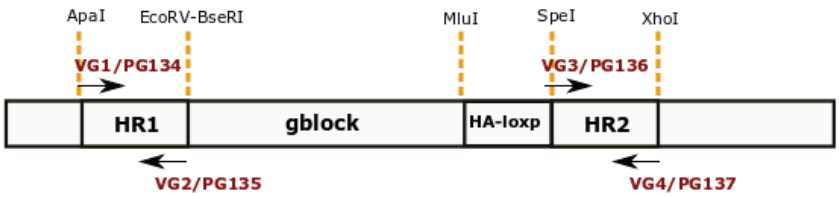


Figure 2.2.Construction of donor plasmid
The schematic diagram represents the cloning strategy for GBP2 (VG labelled primers) and DNAJb (PG labelled primers).

The details of the primers that were used in the constructions of above mentioned plasmids were provided in Table 4.1 and 5.1.

2.2.3.2 Transfection protocol

All transfections were performed by electroporation of synchronized ring stage of Pf 3D7 p230pDiCre line. For electroporation, 100 µl of infected RBCs (> 10% parasitemia) were resuspended in 300 µl of Cytomix (120 mM KCl, 0.15 mM $CaCl_2$, 2 mM EGTA, 5mM $MgCl_2$, 10 mM K_2HPO_4/KH_2PO_4, 25 mM HEPES) containing mixture of plasmid DNA.(Moll *et al.*, 2013) The plasmid DNA mixture consists of 60 µg of pDC2 Cas9/gRNA/hDHFR/yFCU plasmid and 60 µg of linearized pLN-loxp donor plasmid (linearized with *Apa*I and *Xho*I). Electroporation were performed using a Bio-Rad Gene Pulser at settings of 310 V, 950 µF and 200 Q in 0.2 cm cuvette. Time constants were 9-12 ms. These electroporated samples were immediately mixed with culture media and grown in agitation. After 4 hours of post-transfection, the culture medium was replaced with fresh media supplemented with 2.5 nM of WR99210, which was withdrawn after 5 days. Once the parasites were observed on giemsa staining, limiting dilution cloning was performed to obtain the individual clones with desired endogenous locus modification.

In order to induce the DiCre mediated knockout of the gene, early ring stage parasites were treated with rapamycin (20 nM) for normally 4 hours, unless stated otherwise. After treatment, parasites were washed and returned to culture.

2.2.3.3 Limiting dilution cloning

Limiting dilution cloning was performed in 96 well plates at 2% hematocrit in culture media. Each well contained 200ul of medium with an average of 0.6 parasites. Media was changed on day 7. 10 and 12 and parasites were observed with giemsa staining. On day 12, cultures with positive slides were expanded to isolate DNA for genotyping.

2.2.3.4 Confirmation of transgenic parasites with PCR genotyping

Parasites were genotyped to check for correct modification at the gene locus. Three PCRs were performed on genomic DNA: a) Box 1 PCR - region between 5' end of the WT locus (outside the homology region) and within the modified locus was amplified, b) Box 2 PCR – region between the modified locus (close to HA tag) and the 3' end of WT locus (outside the homology region), and c) whole locus- region between 5' end and 3' end of the WT locus, outside the homology regions. The location of the verification primers [GBP2 (1- PG120, 2- PG121, 3- PG122 and 4- PG123) and DNAJb (1- PG174, 2- PG178, 3- PG176 and 4- PG182)] was shown in Figure 2.1 and the details of all the primers are listed in Table 4.1 and 5.1. PCR reaction mixture were prepared using Gotaq Green master mix (Promega) according to the manufacturer's guidelines with elongation temperature as 62 ^{0}C and time varied according to the size of the amplicon, i.e 1 min per kb.

The genomic DNA were extracted from saponin lysed parasites using NucleoSpin Blood QuickPure kit (Macherey-Nagel, 740569.250) according to the manufacturer's guidelines.

2.2.4 Growth phenotype assay

This assay was performed in trophozoites stages of synchronised parasite culture with initial parasitaemia of 0.01% in 5% hematocrit. During the 6-day assay, media was changed and slides were prepared every 48hrs. The assay was performed in triplicates and the parasitemia was counted under microscopy. Data was represented as mean of triplicates along with calculated error as standard deviation.

2.3 Preparation of nuclear and cytoplasmic extracts

The nuclear lysate were prepared as previously described with some modifications (Chêne *et al.*, 2012). A total of 1×10^9 parasites were isolated from different stages of infected

erythrocytes (rings, trophozoites and schizonts) by saponin lysis. The harvested or -80 °C stored pellet were resuspended in 1 ml of lysis buffer (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.65% NP-40) supplemented with 1× protease inhibitor complex (PIC) (Complete, Roche) and incubated for 30 min at 4°C. Total parasite lysis was obtained by 50 strokes for rings and 25 strokes for late stages (trophs and schizonts) in a prechilled Dounce homogenizer. The lysate was spun at 14 000 rpm for 10 min at 4°C. The collected supernatant containing the cytoplasmic fraction were recovered, aliquoted and stored at −80°C. The nuclei pellet was washed thrice with 1× phosphate-buffered saline (PBS) and then resuspended in 100 µl of extraction buffer (20 mM HEPES pH 7.8, 600 mM KCl, 1 mM EDTA, 2 mM $MgCl_2$, 25 µM $ZnCl_2$, 1 mM DTT) supplemented with 1X PIC and sonicated for 10 cycles (30sec ON/OFF) at 4°C using Bioruptor (Diagenode). The lysate was then spun at 14 000 rpm for 10 min at 4°C. The supernatant containing the nuclear fraction was recovered, aliquoted and stored at −80°C. The purity of the extracts was checked by western blotting, probing the membrane with α-aldolase (cytoplasmic fraction) and α-H3 core (nuclear fraction) antibodies.

2.4 Western blotting

Saponin lysed parasites (equivalent to 10^8 parasites per lane) were lysed using 1× Laemmli Buffer supplemented with 1× protease inhibitors and 50 mM DTT. These lysates were resolved on a 10 - 15 % SDS-PAGE gel at a constant voltage of 80 V for 20 min (until it crosses stacking layer of the gel) and then increased to 150 V for 1h. After electrophoresis was complete, the proteins were transferred onto a PVDF membrane using Towbin buffer pH 8.3 [25 mM Tris, 192 mM glycine, 20% (v/v) methanol] at a constant voltage of 30V for overnight. The membrane was blocked with 3 % BSA/TBST for 1h at RT and probed using primary antibody (rat α-HA 3F10 and mouse α-His) in 3% BSA/TBST for either 1h at RT or overnight at 4 °C. After this incubation, membrane was washed thrice with 1× TBST buffer (1× Tris-buffered saline with 0.1% Tween 20). Blots were incubated for 1h at RT with HRP-conjugated secondary antibody in 3% BSA/TBST, followed by three washes with TBST. The blots were developed with Clarity Western ECL substrate (Biorad).

2.5 Immunofluorescence assay

Immunofluorescence assay was performed on smears of infected erythrocytes. The smear was fixed with 4% PFA for 15 min at RT, followed by neutralization with 0.1M glycine/PBS for 10 min. Cells were permeabilised with 0.1% Triton-X-100/PBS for 15 min. Washing with 1× PBS was performed after every step. Cells were blocked with 1.5 % BSA/PBS for 60 min, followed by incubation with primary antibody (rat α-HA 3F10) in 0.15% BSA/PBS for either 1h at RT or overnight at 4 °C. After three washes, cells were incubated with Alexa conjugated secondary antibodies, diluted in 0.15% BSA/PBS for 1h at RT. Slides were washed to remove the unbound antibodies and were stained with Hoechst (diluted in 1: 4000 in 1× PBS) for 10 min. The cells were mounted with ProLong™ Gold antifade reagent (Invitrogen, P10144) and coverslip was placed. Images were captured and processed by Zen Blue software (Zeiss).

2.6 Biophysical confirmation of G4 folding of oligonucleotides

Patrizia Alberti as a part of collaboration performed this assay. Oligonucleotides were purchased from Eurogentec (Belgium) and were resuspended at a strand concentration of 4 µM , unless stated, in cacodylic acid buffer (10mM cacodylic acid, pH 7.2 adjusted with LiOH), containing 100mM KCl, followed by carrying out the measurements for biophysical experiments in 1cm quartz cells.

2.6.1 Absorbance Spectroscopy

Absorption spectra and thermal denaturation profiles (absorbance as a function of temperature) were acquired on an UVIKON XS spectrophotometer. For thermal denaturation profiles, the absorbance was recorded at different wavelengths (such as 245, 260, 273, 295 and 335 nm), while cooling and then heating the samples between 5°C and 95°C at a rate of 0.2°C/min. Melting temperatures (T_m) were graphically calculated as the intercept between the melting curves and the median lines between low and high-temperature absorbance linear baselines. Thermal difference spectra (TDS) were obtained by subtracting the absorption spectra at 2°C (low temperature) from the one at 90°C (high temperature), the spectrum were recorded after annealing from 90°C to 2°C at 0.2°C/min (Mergny *et al.*, 2005). UV/Vis spectra were recorded on a SAFAS spectrometer.

2.6.2 Circular Dichroism

Circular dichroism (CD) spectra were recorded by using a Jasco J-810 spectropolarimeter at 20°C or 5°C, after annealing from 90°C to 2°C at 0.2°C/min. The oligonucleotide was dissolved at a strand concentration of 3 or 6 µM in cacodylic acid buffer (as mentioned above).

2.7 Approaches for identification of G4 binding proteins

2.7.1 Yeast One Hybrid System

The Y1H assay was performed according to the Matchmaker^{TM} Gold Y1H System (Clontech) guidelines (Figure 3.2). WT (PG1/2) and mutated G4 (PG15/16) sequence, listed in Table 3.1. The *BbsI* linearized pAbai reporter vector containing bait (WT_G1 or mut_G4) and reporter gene (Aureobasidin resistance (AbA^{r}) gene, *AUR-1C*) were integrated into the Y1HGold yeast genome at the inactive *ura*3-52 locus to generate yeast 'bait reporter strain', called as Y1HGold^{bait} using LiAc/ss-DNA/PEG transformation protocol. This generated Y1HGold^{bait} is a result of homologous recombination between wild-type *ura3* gene of pAbAi vector and inactive *ura3-52* gene of the Y1HGold strain. The transformed colonies is able to grow in the absence of Uracil on SD/-Ura plates whereas the untransformed will not grow. The minimal inhibitory concentration of Aureobasidin A (AbA) is determined for this generated Y1HGold [bait+pAbAi] strain. 7 µg cDNA library plasmid of *Plasmodium falciparum* 3D7 [cloned in pGAD-HA vector as a prey plasmid library was gifted by Jamal Khalife, Institut Pasteur de Lille] was transformed in Y1HGold [bait+pAbAi] strain, according to the instructions in Yeastmaker^{TM} Yeast Transformation system 2 (Clontech). The obtained colonies were streaked twice on selective plates and the prey plasmids were isolated from overnight cultures. Lysis was performed using DNA lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris/HCl pH 8.0, 1 mM EDTA) and glass beads in a FastPrep instrument (MP Biomedicals, Santa Ana, CA, USA) for 1 min at 4 °C, followed by phenol/chloroform extraction and ethanol precipitation. Plasmids were re-transformed in *E.coli* (XL10 Gold), and overnight cultures were used to isolate plasmids by NucleoSpin Plasmid kit (Machery Nagel, 740588.250). These rescued plasmids were sent for sequencing using primer Matchmaker 5' and 3' insert_seq, provided by kit. The obtained sequence was used as an input in BLAST tool of Plasmodb database against *Plasmodium falciparum 3D7* to

identify the G4 binding proteins, listed in Table 3.2) (Aurrecoechea *et al.*, 2008). This work was done under the supervision of Katrin Paeschke at ERIBA, Groningen.

2.7.2 DNA Pull-down assay and Mass Spectrometry

Biotinylated oligonucleotides (WT_G4 -PG33 and Mut_G4 -PG34) were synthesized by Eurofins genomics (Table 3.1). These oligos (1200 pmol /500 µl) were folded to achieve G4 by heating at 95°C and followed by cooling at RT for overnight. The nuclear lysate of *P. falciparum* NF54 were prepared as previously described earlier.

DNA pull-down assay using biotinylated oligos and nuclear lysate were performed as described in published protocol with some modifications (Jutras, Verma and Stevenson, 2012)(Figure 3.3). In this assay, the 5' biotinylated WT_G4 was folded into G4 by heating at 95 °C for 5 min, and then gradually cooling at room temperature for overnight in the presence of K^+ ions. Even though 5' biotinylated mut_G4 does not form any G4, we followed the same folding protocol as WT_G4 in order to use it as control bait. Here, we performed the subtractive based pull-down assay. First, the lysate was incubated with pre-folded 5' biotinylated mut_G4 followed by WT_G4. The extracted nuclear lysate were treated with 100 µg of avidin for 30mins at 4°C and was cleared by pre-incubation with 0.1 ml of the washed beads with BS/THES buffer [22mM Tris-Cl (pH 7.5), 10mM HEPES-KOH (pH 7.5), 8.9% Saccharose, 62mM NaCl, 5mM CaCl2, 50mM KCl,1mM EDTA, 12% Glycerol, 1mM DTT and 1× Protease inhibitor complex (PIC)] for 60 min at 4°C in rotating shaker. 2 µl of the supernatant was kept as input control. During this incubation, 200µl Dynabeads M-280 streptavidin (Invitrogen) were washed thrice with 1000 µl 2xBW buffer with 1X PIC. Beads were resuspended in 0.5 ml 2xBW buffer [10mM Tris-Cl (pH 7.5), 1mM EDTA (pH 8.0), 2M NaCl], mixed with 0.5 ml biotinylated DNA (1200 pmol) and incubated for 1h at RT on a rotator. After 1h, the DNA concentration was measured in the supernatant. If the concentration of unbound DNA in all samples was not equal, the incubation was extended for another 1h. The immobilized DNA (~1200 pmol) was washed thrice with 1ml TE buffer and all the washes were stored. The beads were then blocked with 1ml of 0.1% (w/v) bovine serum albumin (BSA) in 2xBW buffer for 15 min at 4°C on a rotator, followed by the washing with BS/THES buffer and once with 1 ml BS/THES buffer supplemented with 5 µg oligos. These beads were incubated with pre-cleared nuclear lysate in 0.2ml BS/THES buffer supplemented with 50 mM potassium acetate and 100× non-specific oligos as compared to

the bound DNA for overnight at 4°C on a rotator. To elute the proteins, the beads were incubated with increasing concentrations of NaCl in elution buffer (200 mM, 400 mM, 600 mM, and 800 mM NaCl in 25 mM Tris-HCl, pH 7.5). In each elution, beads were incubated with 0.6 ml elution buffer for 4 min at RT on a rotator. Eluted fractions were dialyzed and concentrated with Dialysis buffer (25mM Tris-Cl, 6mM DTT and 1mM iodacetamide) using Vivaspin 10MWCO tubes. These dialyzed samples were used for LC-MS/MS.

MS step and analysis were done by the MS facility platform, as a part of collaboration.Eluted proteins were digested in solution. Briefly, each sample was diluted (final volume 100 µl) in TEAB 100 mM. One microliter of DTT 1 M was added and incubation was performed for 30 min at 60 °C. A volume of 10 µL of IAA 0.5 M was added (incubation for 30 min in the dark). Enzymatic digestion was performed by addition of 1 µg trypsin (Gold, Promega, Madison USA) in TEAB 100 mM and incubation overnight at 30 °C. After digestion, peptides were purified and concentrated using OMIX (Agilent Technologies Inc.) according to the manufacturer's specifications. Peptides were dehydrated in a vacuum centrifuge. After resupsension in formic acid (0.1%, buffer A) samples were loaded onto a 50 cm reversed phase column (75 mm inner diameter, Acclaim Pepmap 100® C18, Thermo Fisher Scientific) and separated with an Ultimate 3000 RSLC system (Thermo Fisher Scientific) coupled to a Q Exactive HF (Thermo Fisher Scientific) via a nano-electrospray source, using a 125 min gradient of 2 to 40% of buffer B (80% ACN, 0.1% formic acid) and a flow rate of 300 nl/min.

MS/MS analyses were performed in a data-dependent mode. Full scans (375 – 1,500 m/z) were acquired in the Orbitrap mass analyzer with a 60,000 resolution at 200 m/z. For the full scans, 3e6 ions were accumulated within a maximum injection time of 60 ms and detected in the Orbitrap analyzer. The twelve most intense ions with charge states ≥ 2 were sequentially isolated (1e5) with a maximum injection time of 45 ms and fragmented by HCD (Higher-energy collisional dissociation) in the collision cell (normalized collision energy of 28%) and detected in the Orbitrap analyzer at 30,000 resolution. Raw spectra were processed using the MaxQuant environment (Cox and Mann, 2008) and Andromeda for database search with label-free quantification (LFQ) and match between runs (Cox *et al.*, 2011). The MS/MS spectra were matched against the UniProt Reference proteomes (Proteome ID UP000001450 and UP000005640) of *Plasmodium falciparum* and Human and 250 frequently observed contaminants as well as reversed sequences of all entries (MaxQuant contaminant database). Enzyme specificity was set to trypsin/P, and the search included cysteine

carbamidomethylation as a fixed modification and oxidation of methionine, and acetylation (protein N-term) as variable modifications. Up to two missed cleavages were allowed for protease digestion. FDR was set at 0.01 for peptides and proteins and the minimal peptide length at 7.

2.8 Cloning, expression and purification of recombinant protein

Since *Plasmodium falciparum* has AT rich genome, so in order to efficiently express the plasmodial protein in *E. coli*, we had first recodonized the coding region of respective gene of interest (PfGBP2 and PfDNAJ), followed by synthesizes of the recodonized genes as gblocks from Eurofins IDT service. These gblocks were amplified using gene specific primers (GBP2- PG111/112 and DNAJb- PG132/133), listed in Table 4.1 and 5.1. The PCR amplified region was cloned into the *Nde*I and *BamH*I sites of a pET15b expression vector (Novagen, Germany) by infusion cloning. The resulting construct was confirmed by restriction digestion and sequencing. The N-terminal 6× His tagged protein was expressed in BL21 (DE3) pLysS cells grown in medium supplemented with 100 µg/mL Ampicillin and 25 µg/mL of chloramphenicol. Expression was induced at an optical density of 0.7-0.9 at 600 nm, with 0.5- 1 mM isopropyl β-D-thiogalactoside (IPTG) at 25 °C for 5h.

All the purifications steps were carried out at 4 °C. Cell lysis was performed in lysis buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 5% *(v/v)* glycerol and 20 mM of imidazole) by sonication (0.5s ON, 0.5s OFF at 60% pulse intensity, 3 min) using a Branson Digital Sonifier 450-D. After centrifugation of the lysate (14000 rpm, 30 min), the supernatant was loaded onto HisPur Cobalt Spin column and purification was carried out according to the manufacturer's instruction. The eluates were run through Superdex 200 16/600 (10 – 60kDa) size exclusion chromatography. The positive fraction was combined and concentrated using a Amicon Ultra-15 centrifugal filters. Expression and purification of His-tagged protein was confirmed by Coomassie staining and Western blot analysis with α-his antibody.

2.9 EMSA

Electromobility shift assays (EMSAs) were performed using biotinylated oligos and recombinant protein as described earlier with some modifications (Takahama *et al.*, 2011). Biotin labelled oligos were purchased from Eurofins genomics and listed in Table 3.1, same

primers that were used for DNA pull-down assay. The oligonucleotides were diluted to 100 nM in 10 mM Tris-HCl, pH 7.5 with or without 150 mM KCl. Quadruplex formation or duplex nnnealing was performed by heating the samples at 95 °C and then cooling it at room temperature for overnight. DNA-protein binding reaction was carried out in a final volume of 15 uL using 100 nM of the labeled oligonucleotide and a varying concentration of recombinant protein in a binding buffer (50 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 0.5 mM DTT, 0.1 mg/mL bovine serum albumin, and 100 mM KCl). After the samples were incubated for 1h at 25 °C, they were resolved on a 10% native PAGE gel with or without 20 mM KCl. The native PAGE gel was pre-ran at 100 V for 1h at 4 °C. Electrophoresis was performed at 100V for 1h at 4 °C in 0.5× TBE buffer supplemented with or without 20 mM KCl. The DNA bound proteins were transferred to the Amersham Hybond- N^+ nylon membrane by electroblotting at 130 mA for 30 min using Trans-Blot Semi-dry blot transfer cell (Biorad). The interactions were detected using a Thermo Scientific™ Pierce™ LightShift chemiluminescent EMSA kit following the manufacturer's guidelines.

2.10 ChIP-Seq

ChIP-Seq was performed according to the previously published protocol with some modfications (Lopez-rubio, Siegel and Artur, 2013). Infected RBCs were crosslinked using 1% formaldehyde for 10 min at RT. 0.125 M glycine was added for quenching the cross-linking reaction. The samples were washed thrice using 1X PBS (chilled) before proceeding with saponin lysis. Sample was resuspended in cold lysis buffer (10 mM Hepes (pH 7.9), 10 mM KCl, 0.1 mM EDTA (pH 8.0), 0.1 mM EGTA (pH 8.0),1 mM DTT, PIC and 10 % NP40) followed by lysis the parasites with 200 strockes for rings and 100 strokes for late stages. The lysed sample was spun and the pellet was resuspended in SDS lysis buffer (1% SDS, 10 mM EDTA (pH 8.0), and 50 mM Tris–HCl (pH 8.1)) and sonicated in Bioruptor with settings (24 cycle of sonication 10 sec on/off) to obtain the chromatin size of 200-400 bp. Sample was split into two 1) IgG as pre-clear step, 2) IgG done in independent sample. Pre-clearing was performed for 2 h at 4^0C using recombinant protein A conjugated sepharose beads with continuous gentle inverting. The sample was divided into eight aliquots (250 ul/aliquot) and incubated with α-HA or IgG antibody for 12 h at 4^0C. Sample were then collected with protein A agarose beads for 2 h at 4^0C. This protein A-antibody-chromatin complex was washed with different buffers with varying concentration of salts and eluted using ChIP buffer (1% SDS and 0.1 M $NaHCO_3$). Both IP sample and input were reverse crosslinked

using 0.2 M NaCl for 6 h at 65^0C , followed by RNAse treatment (2 h at 37^0C), Proteinase K treatment (2 h at 45^0C). Finally DNA was purified using phenol chloroform precipitation.

ChIP-sequencing libraries for all the samples were prepared from purified DNA using TruSeq ChIP Library Preparation Kit (illumine), as described by the manufacturer. Libraries were then quality control checked by a bioanalyser and qPCR prior to pooling and sequencing. The platform used was the recent Illumina NovaSeq 6000 system, flowcell SP (ID AHJHNGDRXX). Sequencing of 150 bp paired-ends yielded on average $70x10^6$ reads per library. Fastq files were obtained by demultiplexing the data using bcl2fastq software (Illumina), prior to downstream analysis. Mapping was performed using BWA mem and peaks were calculated using the command "callpeak" from MACS2 and the input as "control" sample. Intersection of peaks from the different replicate samples using Bedtools produced the final dataset. Motif surveys were performed using MEME (version 5.1.1 (Release date: Wed Jan 29 15:00:42 2020 -0800) The R package ChipQC and ggplots2 were used to plot the various datasets. All the files used for the analysis are provided in Annexure. This ChIp-Seq experiment and analysis was done in collaboration with Ana Rita Gomes.

2.11 Telomere Restriction Fragment (TRF) Southern Blotting

To perform TRF southern blotting, iKO-PfGBP2 parasite line was cultured in the presence of rapamycin/ DMSO and parasites were collected after 4 weeks and 8 weeks. The genomic DNA were extracted from saponin lysed parasites using NucleoSpin Blood QuickPure kit (Macherey-Nagel, 740569.250) according to the manufacturer's guidelines.

Two micrograms of genomic DNA were digested with 10 units of each enzyme: *AluI, DdeI, MboII, RsaI* overnight at 37°C. Digestions were precipitated and run in 1% agarose gels at 70 V for 5 h and transferred to a Hybond N+ membrane overnight through capillarity (Sambrook et al. Molecular Cloning). The DNA was then cross-linked to the membrane in a UV oven for 1 min and prehybridized in 6X SSC 0.1% SDS 2.5% skimmed milk for 1 h at 42°C. 10 pmol of biotinylated telomeric probes (PG184-5' GGGTTTAGGGTTTAGGGTTTAGGGTTTA 3' and PG185-5'GGGTTCAGGGTTCAGGGTTCAGGGTTCA 3') were added and incubated overnight, followed by four times 30 min washes of 6X SSC with 0.1% SDS. The biotin labeled probes was detected using the LightShift Chemiluminescent EMSA Kit and imaging in a BioRad ChemiDoc. Rafael Martins, as part of collaboration, performed this experiment.

2.12 Quantitative reverse transcription-PCR (qRT-PCR) on *var* genes

To perform qRT-PCR, samples were prepared by culturing transgenic parasites in the presence of rapamycin/ DMSO for two month. Parasites were harvested at 14-17h, where peak expression of *var* genes is observed, using saponin lysis. After harvesting with saponin, parasite pellet was resuspeneded in Trizol. and RNA was extracted using trizol solution, as described in previously published protocol(Moll *et al.*, 2013).

Quantitative PCR was performed on cDNA using specific primers for each *var* gene, as previously described (Salanti *et al.*, 2003) with few modifications from Dzikowski *et al* (Dzikowski, Frank and Deitsch, 2006) . The list of the primers is provided in Table 4.6. DNase-treated RNA samples were reverse-transcribed into cDNA, using SuperScript III first-strand synthesis SuperMix (Invitrogen), according to manufacturer's instructions. Alternatively, RNA samples were run without RT enzyme, to check for genomic DNA contamination. Target genes in cDNA samples were quantified using PowerUp SYBR Green Master Mix (Applied Biosystems) and normalized using housekeeping genes Serine-tRNA-ligase gene. The results were expressed as relative copy number. This experiment was performed along with Diane-Ethana Benet.

Chapter 3

Identification of G-quadruplex binding proteins in *P. falciparum*

Chapter 3 Identification of G-quadruplex binding proteins in *P. falciparum*

3.1 Introduction

G4s are secondary structures that are formed within the guanine rich nucleic acid sequences. These structures have been implicated in multiple biological processes, where they act as enhancers, repressors and/or blockades (Rhodes and Lipps, 2015). Different proteins recognize these structures that are known as G-quadruplex binding proteins (G4-BP). These proteins play a significant role in modulating these structures and their potential functions. G4-BP can be categorized into three groups (Qiu *et al.*, 2015). Group I proteins (G4-stabilizers) binds to the existing G4s and increases their stability. Group II proteins (G4-inducers) bind to the guanine rich sequence and induce the formation of G4. Whereas, Group III proteins (G4-destabilizers) destabilizes the existing G4s. Several examples of G4-BPs have already been discussed in previous chapter 1 and Table A displays a list of some proteins based on their categories.

Type of G4 protein	Examples	References
G4-stabilizers	MyoD, NPM1, MutSa	(Etzioni *et al.*, 2005; Larson *et al.*, 2005)
G4-inducers	Est1p, Nucleolin, CNBP	(Zhang *et al.*, 2010); Guo, Hurley and Sun, 2009; Borgognone, Armas and Calcaterra, 2010)
G4-destabilizers	BLM, hnRNP A2/B1, Sgs1 helicase, Pif1 helicase	(Sun *et al.*, 1998), (Scalabrin *et al.*, 2017), (Sun, Bennett and Maizels, 1999), (Paeschke, John A Capra and Zakian, 2011)

Table A: Different types of G4 proteins

Over the past few decades, the increasing number of studies have been reported on identification of G4-binding proteins in numerous organisms (Sissi, Gatto and Palumbo,

2011; Wang *et al.*, 2012a; Brázda *et al.*, 2014; Von Hacht *et al.*, 2014; Mcrae, Booy, Padilla-Meier, *et al.*, 2017; Brázda *et al.*, 2018). Recently, the first database on G4 binding proteins, called G4 Interacting Protein Database (G4IPDB), was published. This platform provides comprehensive information on more than 200 G4-BPs (Mishra *et al.*, 2016). Based on the amino acid composition of known human G4-BPs, it was revealed that arginine/glycine-rich region, named as RGG domain is often present in these G4-BPs (Thandapani *et al.*, 2013; Brázda *et al.*, 2018; Huang *et al.*, 2018).

Several techniques including homology-based search and DNA pull-down assays have been frequently employed to identify the G4-BPs. Despite the availability of vast information on G4-BPs in numerous organism, yet only RecQ helicase have been reported that influences G4 related phenotypes in *P. falciparum* (Stanton *et al.*, 2016; Claessens *et al.*, 2018).

RecQ helicases (BLM and WRN) are known G4-unwinding proteins that have been extensively studied in different prokaryotes and eukaryotes (Mendoza *et al.*, 2016). Recently, the biological roles of the RecQ homologs (PfBLM and PfWRN) have been investigated in *P. falciparum* (Claessens *et al.*, 2018). The knockdown of PfWRN increased the indel mutation rate that led to chromosomal abnormalities while no significant change was observed PfBLM knockout line. The PfWRN-k/d line displayed the G4-related phenotypes such as unusual *var* gene recombination patterns and high sensitivity to G4 ligand (TMPyP4). Overall, the RecQ helicases are shown to affect the transcription, replication and antigenic variation in these malarial parasites (Claessens *et al.*, 2018).

Hence, in order to gain a deeper understanding of the possible roles of G4-BPs in *Plasmodium*, we conducted this study to identify putative G4-BPs in these parasites. We have identified putative G4-BPs in *P. falciparum* using two different unbiased approaches (yeast one-hybrid assay and DNA pull-down assay, followed by LC/MS). A putative G4-forming sequence was selected and confirmed to form a stable G4, which was used as bait in these approaches to fish out G4-BP.

3.2 Results

3.2.1 Identification and experimental confirmation of the G4 formation of the sequence

In this study, we have selected a putative G4 motif (GGGATTTGGGAGGGGGGG), named as WT_G4 which was identified as G4FS by use of bioinformatics analysis by two different algorithms Quadparser and G4Hunter algorithm. (Unpublished data from Elodie Gazanion's work in our lab) (Figure 3.1a). This G4 motif is located in the upstream of transcription start sites (TSSs) of putative AP2 transcription factor (Pf3D7_0934400), a member AP2 transcription factor family. Pf3D7_0934400 displays dynamic mRNA expression pattern with maximum expression during late stages (trophozoites and schizonts) of the asexual stages of the parasites (Figure 3S.1).

Before using this WT_G4 motif as bait in the unbiased approaches, it was necessary to characterize and confirm the formation of the G4 by this WT_G4 motif. In addition, we also generated the mutated version (hereon called as mut_G4) motif by replacing the central G to A (GAGATTTGAGAGGGAGGG), in order to disrupt the potential G4 formation. This mut_G4 was used as control bait in the unbiased approaches. We performed different biophysical experiments (TDS, UV melting profile, and CD) to analyze the ability of the WT_G4 and its mutated form-to-form G4.

Thermal differential spectra (TDS) of WT_G4 displayed two positive peaks at 240 nm and 275 nm, and a negative peak around 295 nm, as compared to mut_G4. These exhibited peaks correspond to the TDS signature that is specific for the G4 (Figure 3.1b). To examine the stability of the G4, we recorded the melting profiles of the selected sequences at 295 nm. As shown in Figure 3.1c, the WT_G4 displayed the inverted transition with Tm of around 68 °C, hence confirming the formation of stable G4. To determine the topology of the G4 formed by the WT_G4, the CD spectra of the WT_G4 was recorded at 20 °C. The obtained data showed that the WT_G4 displays a positive peak at 260 nm and a negative peak at 240 nm, signature peaks of parallel type of G4 (Figure 3.1d). Hence, CD analysis reveals the formation of parallel G4 by WT_G4 in the presence of potassium ions.

Taken together, these biophysical experiments demonstrate that the WT_G4 forms a stable parallel G4 whereas mut_G4 does not form any G4.

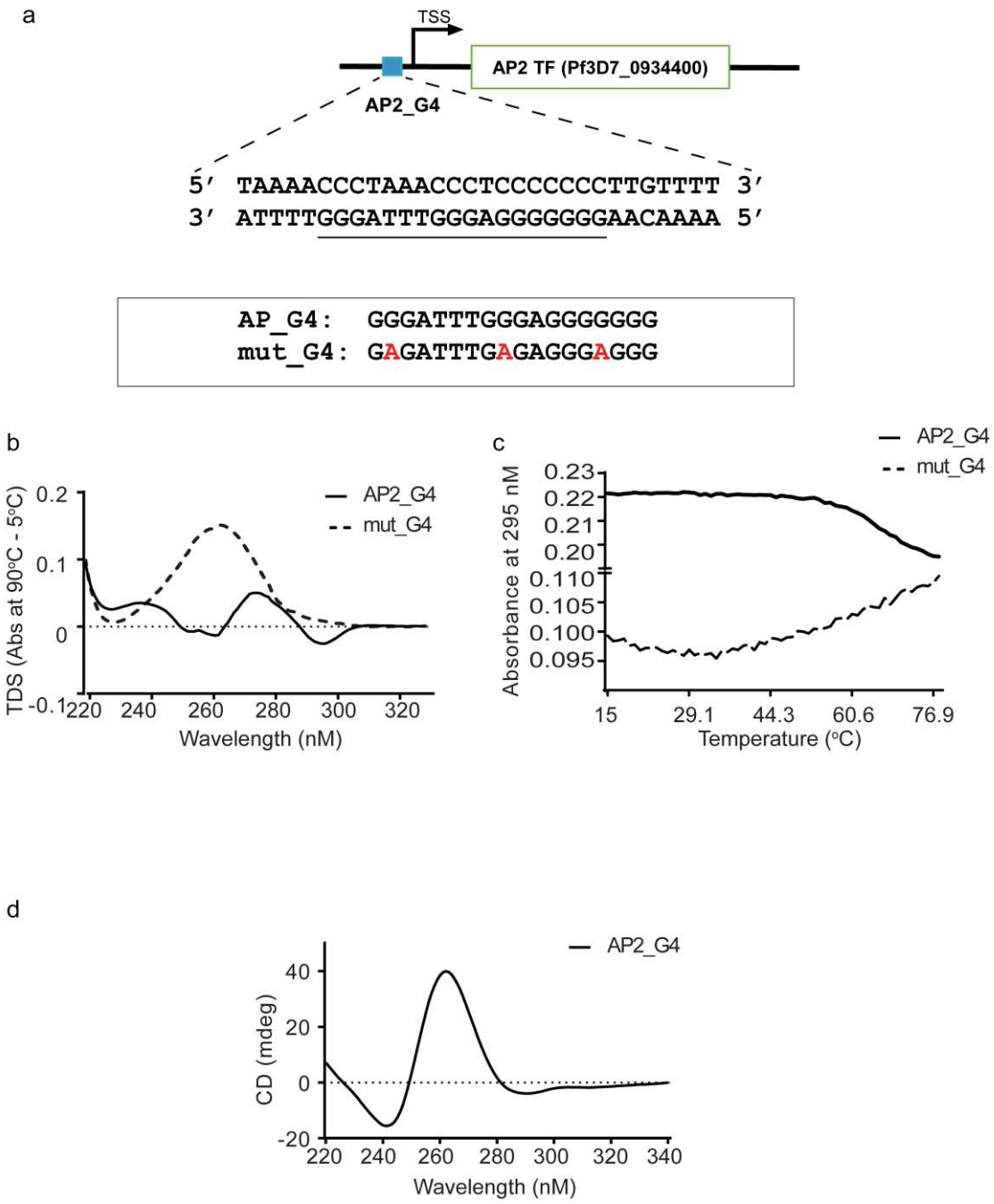


Figure 3.1.Identification and experimental confirmation of the G4 formation of the selected G4 sequence.
a) schematic representation of the location of the predicted G4 motif (blue box) upstream of the AP2 gene, Pf3D7_0934400. b) Thermal differential spectra was measured and plotted for WT_G4 and mut_G4. WT_G4 displayed the TDS signature characteristics of G4. c) Thermal melting followed by absorbance at 295 nm of WT_G4 and mut_G4 were recorded. WT_G4 displays an inverted transition curve, characteristic of G4. d) CD spectroscopy of WT_G4 displayed the peak characteristic of parallel G4. The sequence of WT_G4 oligonucleotide is 5' GGGATTTGGGAGGGGGGG 3' and mut_G4 oligonucleotide is 5' GAGATTTGAGAGGGAGGG 3'.

3.2.2 Identification of G4 binding proteins using unbiased approaches

To identify the G4-BPs in asexual stages of *P. falciparum*, we employed two complementary unbiased approaches: Yeast one-hybrid (Y1H) assay and DNA pull-down assay, followed by mass spectrometry. The selected G4 forming WT_G4 was used as bait whereas non-G4 forming mut_G4 was used as control bait for these experiments.

Yeast one-hybrid assay is a DNA-protein interaction assay where DNA fragment of interest is used as a bait to identify proteins (prey) that binds to the bait in a yeast based-screen. Here, we conducted the **Y1H screen** according to the Matchmaker Gold Y1H library screening system (Clontech, 2010). Using yeast transformation protocol, pAbAi vector containing bait sequence (WT_G4/mut G4) and reporter gene (AUR1-C) was integrated into the genome of Y1HGold strain at inactive ura3-52 locus by homologous recombination. As a result, we obtained the transformed colonies containing genome integrated bait (Y1HGold^{bait}) that grew even in the absence of Uracil on SD/-Ura plates due to integrated active wild-type *URA3* whereas the untransformed strain did not grow (Figure 3.2a). Before the Y1H screen, this Y1HGold^{bait} strain was tested for the background expression of *AUR1-C*. *AUR1-C* provides resistance against antibiotic Aureobasidin A. We determined that the concentration of 850 ng/ml of Aureobasidin A (AbA), a cyclic depsipeptide antibiotic, can suppress the basal expression of *AUR1-C* that might be induced by the non-specific binding of bait by the yeast endogenous proteins in generated Y1HGold^{bait} strain. Therefore, we used 850 ng/ml of AbA in our Y1H screen. This generated Y1HGold^{bait} were co-transformed with pGAD-cDNA library of *P. falciparum* that expresses plasmodial protein fused with GAL activation domain. When the positive interacting prey protein (plasmodial protein fused with GAL-AD) bound to the G4 motif in the Y1HGold^{bait} strain, the fused GAL4 AD activated the expression of AUR1-C enzyme on SD/-Leu agar medium containing 850 ng/ml Aureobasidin A (AbA) (Figure 3.2b). Consequently, this screen yielded 95 positive colonies from about 8.6 million transformants. The prey plasmids were isolated from the obtained positive colonies and then sequenced. After analyzing and removing the redundant entries from the obtained sequencing data, we identified 18 putative non-redundant candidates in *P. falciparum* that binds solely to the G4 forming WT_G4 as compared to 61 candidates that bind to non-G4 forming mut_G4 (Figure 3.2c). The majority of the mut_G4 binding proteins comprise DNA binding proteins, ribosomal proteins, and uncharacterized proteins. The complete list of obtained candidates is provided in Table 3.2.

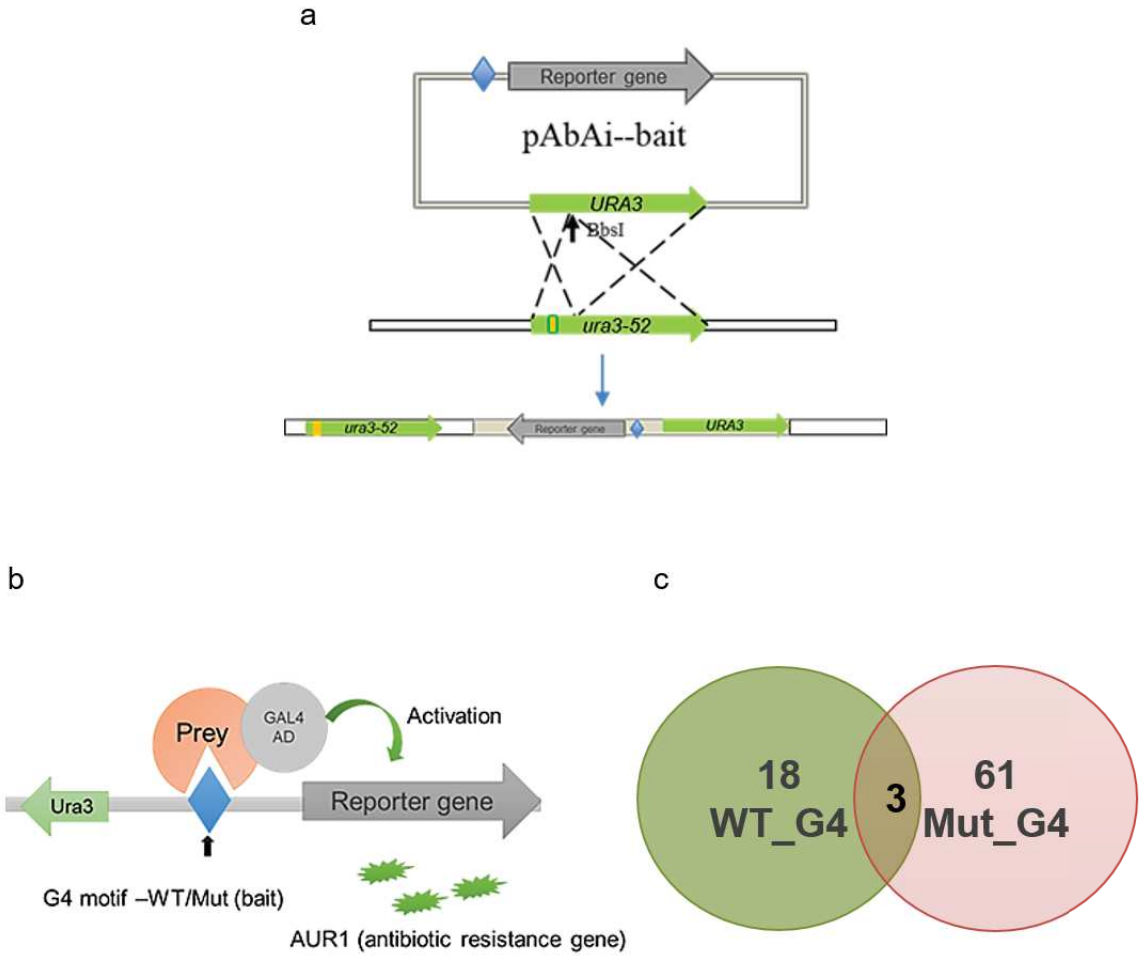
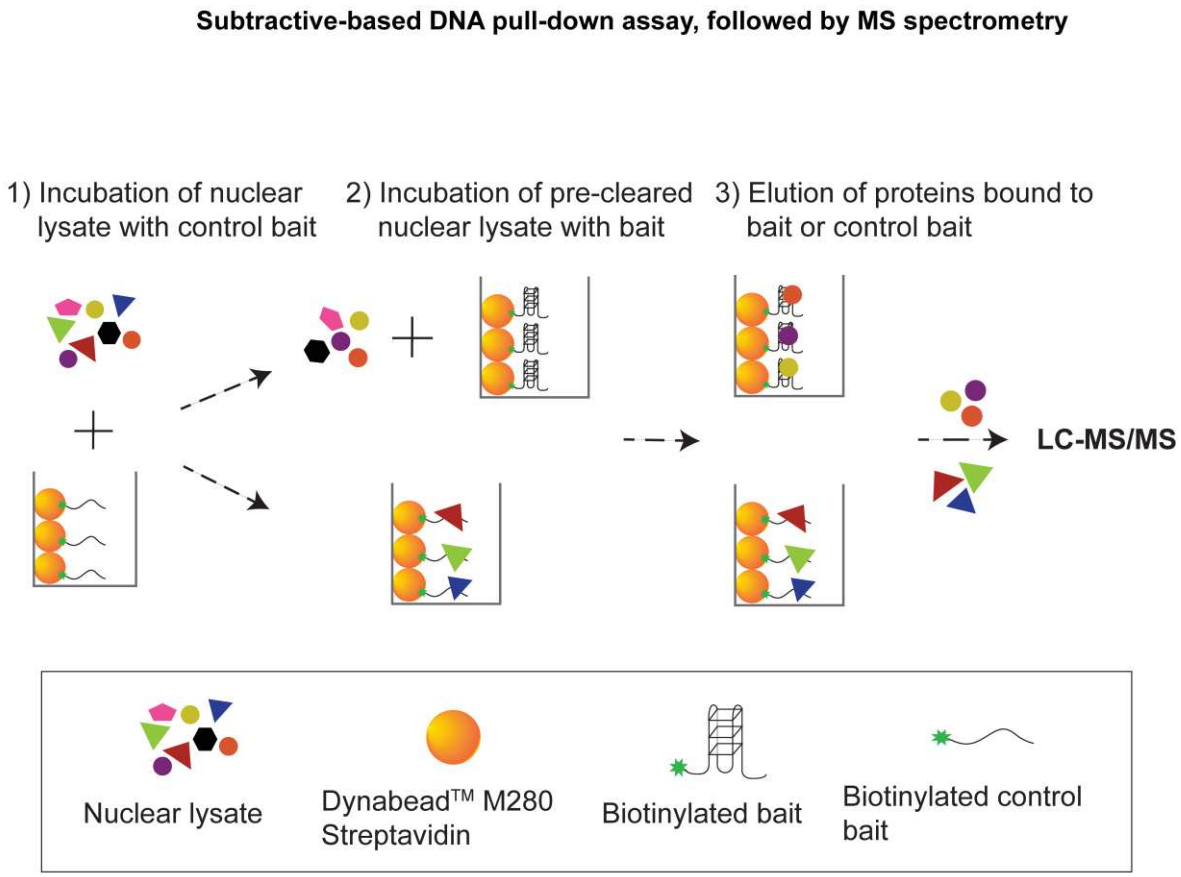


Figure 3.2. Yeast one hybrid assay to identify G4-BP.
a) Schematic representation of overview of Y1H assay protocol. pAbAi vector containing bait sequence (WT_G4/mut G4) is integrated into the genome of Y1HGold strain at inactive ura3-52 locus by homologous recombination with the wild type URA3 gene present in the pAbAi vector. As a result, transformed Y1HGold strain, named as Y1HGold^{bait}, are able to grow in the absence of uracil on SD/-Ura agar plates. b) This generated Y1HGold^{bait} are co-transformed with pGAD-cDNA library of P. falciparum that expresses plasmodial protein fused with GAL activation domain on SD media lacking Leucine. The successful DNA-protein interaction results in the expression of AUR1-C. d) Venn diagram depicting the number of obtained potential candidates from Y1H assay using WT_G4 and mut_G4 as a bait.

To complement the Y1H assay, we implemented the **DNA pull-down assay** approach followed by LC-MS/MS (Figure 3.3a). In this assay, the WT_G4 and mut_G4 oligos were subjected to fold into G4 by heating them at 95°C for 5 min followed by overnight cooling at room temperature in the presence of potassium ions. These folded oligos were incubated with

the nuclear lysate of the blood stages of *P. falciparum* NF54 in a subtractive-based manner. In this subtractive based pull-down assay, nuclear lysate of blood stages were first incubated with the mut_G4, followed by the incubation of pre-cleared lysate with WT_G4. Thereafter, the proteins bound to the WT_G4 and mut_G4 were eluted and used for LC-MS/MS. The MS analysis resulted into the list of several proteins, which were then further curated on the basis of the number of unique peptides and MS/MS spectra. After curating the list by removing contaminants including proteins from other organisms, we have selected 134 potential G4-BP proteins (Table 3.3).To identify the potential role of the proteins, the gene ontology (GO) analysis were performed. GO term (Molecular function) analysis of all potential G4-BP revealed that 52 % of the candidates are predicted to bind to nucleic acid sequence and are implicated in various biological processes. In these G4-BP with nucleic acid binding function, 66 % in translation (due to large number of ribosomal proteins in this list), 11 % with unknown function, 9 % are involved in nucleic acid metabolism, 6 % in transcription, and the remaining of the proteins are involved in replication, chromatin organization, telomere maintenance, cellular homeostasis, and component assembly (Figure 3.3b). However, detailed domain analysis is needed to investigate the role of these candidates.

a



b

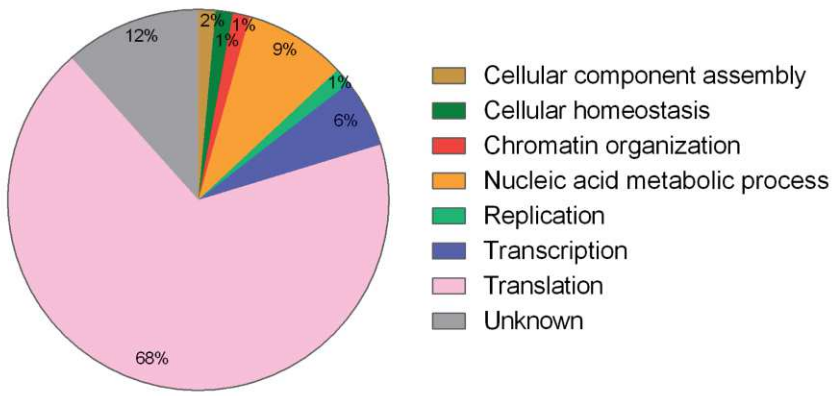


Figure 3.3.DNA-pull down assay, followed by LC-MS/MS to identify G4-BP.
a) Schematic representation of principle of DNA pull down assay. First, biotin labeled mut_G4 is incubated with the nuclear lysate of the blood stages of P. falciparum, followed by incubation of this precleared lysate with the folded WT_G4. After

overnight incubation, eluates are collected and analyzed by LC-MS/MS. b) Pie chart represents the GO term analysis of potential G4-BP that have nucleic acid binding function

Overall, a total of 152 potential G4-BP are identified in *P. falciparum* using two complementary unbiased approaches (Figure 3.4a). Three common proteins (PF3D7_1223300, PF3D7_0919900, and PF3D7_0823800) were identified between these two approaches. GO analysis was performed for all the identified proteins (Figure 3.4b). ~29 % of these identified G4-BP have unknown functions. For the remaining of the proteins, the proteins are associated with translation, transcription, nucleic acid metabolism, and cellular homeostasis.

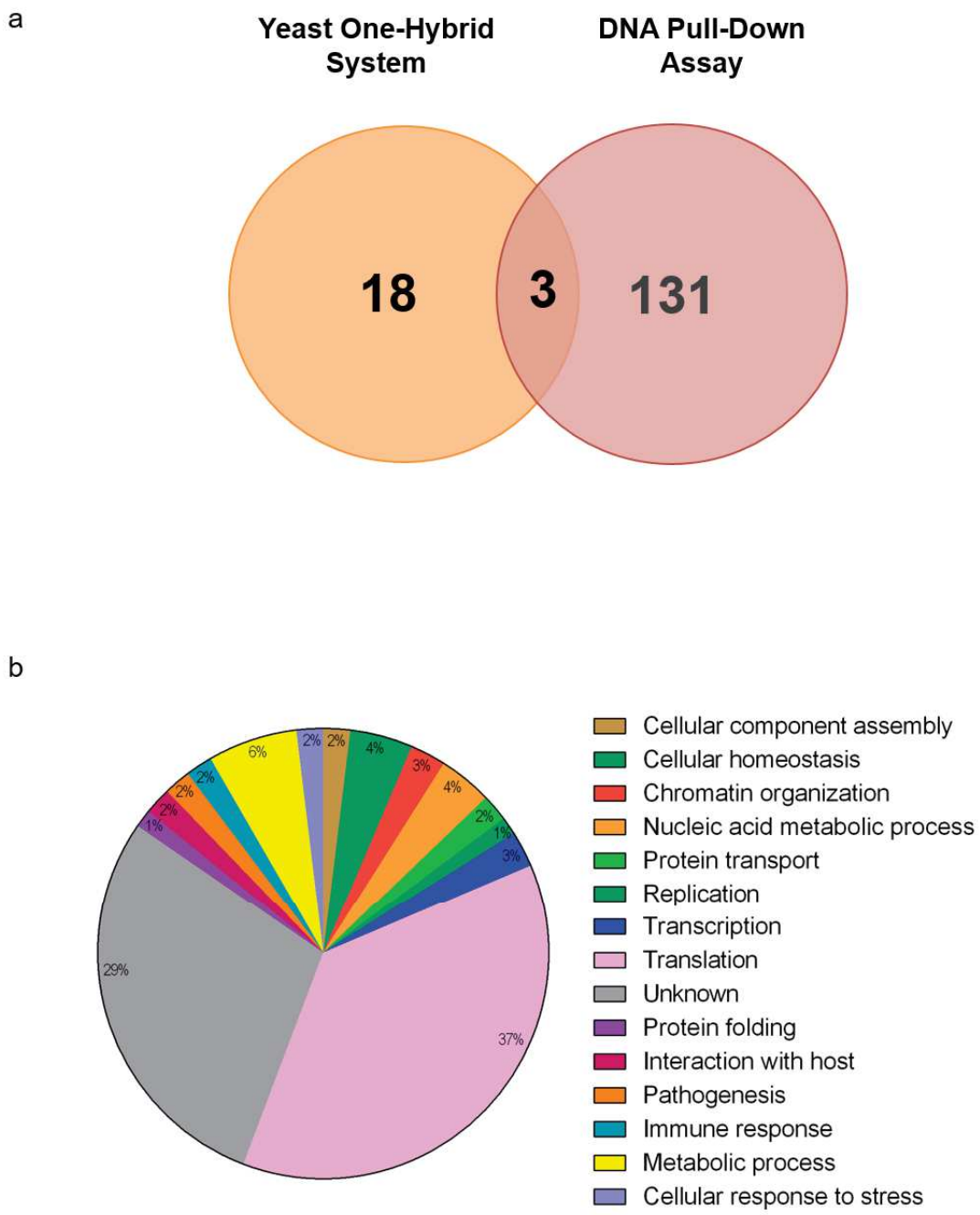


Figure 3.4.Distribution and biological function of potential G4-BP identified using Y1H assay and DNA pull-down assay. a) Venn diagram displays the distribution of identified G4-BP by the two unbiased approaches. b) Pie chart displays classification of the potential G4-BP in 15 different biological functions based on putative functions indicated in UniProt.

3.3 Discussion

Several computational analyses and *in vitro* experiments have demonstrated the existence of G4 in the genomes of *P. falciparum* (De Cian *et al.*, 2008; Smargiasso *et al.*, 2009; Stanton *et al.*, 2016; Anas *et al.*, 2017; Harris *et al.*, 2018; Marsico *et al.*, 2019), however, there is lack

of *in vivo* evidence of these G4. Identification and characterization of G4-BP will help in understanding the biological roles of these structures and will further support the existence of G4 in these parasites.

In this study, we have identified 152 G4-binding proteins using two different and complementary approaches. The reason behind the use of two different approaches was to eliminate the possible limitations that are usually associated with the respective techniques. For instance, the limitations of the Y1H assay include coverage of cDNA library, high rates of false positives/false negatives and/or improper protein folding due to the presence of activation domain with prey protein. Whereas the DNA pull-down assay can be affected by the low endogenous expression of the protein or presence of contaminants including bovine serum albumins or keratins.

However, we have found only three common proteins in these two approaches. This could be due to several reasons. 1) Improper protein folding of prey proteins in Y1H assay. 3) As G4s are primarily associated with the nuclear-based biological process, we used extracted nuclear proteins to perform DNA pull-down assay. While the total cDNA library of *P. falciparum* 3D7 was used in the Y1H assay. Therefore, it would be important to consider these limitations for future studies.

Despite these limitations, the GO analysis of the obtained candidates revealed that the majority of the identified proteins are involved in nucleic acid metabolism, translation, transcription, and cellular homeostasis. Altogether these data highlights the involvement of the obtained protein in different nuclear-based biological processes, supporting the possible multifaceted role of G4 in these parasites, similar to other organism (Rhodes and Lipps, 2015).

Moreover, the homologs of some the identified G4-BPs such as p1/s1 nuclease, FEN1, Replication A1 protein, DNA gyrase, and HMGB1 are reported to be associated with G4, hence proving the data's correlation (Vallur and Maizels, 2010; Qureshi *et al.*, 2012; Lv *et al.*, 2013; Zhou *et al.*, 2013; Amato *et al.*, 2018).

Furthermore, in order to validate some of these obtained interacting proteins, we decided to investigate three candidates, based on their identification method, protein properties and

predicted biological roles in *P. falciparum*. These three selected candidates are Flap endonuclease 1 (FEN1), PfGBP2, and putative DNAJ protein.

First, PfFEN1 is a nuclease possessing both 5'-3' exonuclease and 5' flap-endonuclease activity (Casta *et al.*, 2008). In our Y1H assay, we identified that the C-terminal of the PfFEN1 interacts with the selected G4. Interestingly, human FEN1, a homolog of PfFEN1,is shown to cleave the substrate bearing G4 DNA 5' flaps (Vallur and Maizels, 2010), proposing that PfFEN1 a promising candidate for our study. To validate PfFEN1 and G4 interaction, we tried to express full-length and truncated (containing C-terminal region which was identified in Y1H assay) recombinant PfFEN1 protein. Unfortunately, we faced difficulties in expressing soluble full length and truncated PfFEN1 protein, as mentioned in the previously published report (Casta *et al.*, 2008). Therefore, we tried to generate an inducible transgenic PfFEN1 line to examine *in vivo* G4 binding of PfFEN1. Even though FEN1 is predicted to be dispensable for the growth of the parasites in *P. falciparum* (Zhang *et al.*, 2018), our several attempts of transfections to generate inducible KO line was failed. Hence, we focused on the other two selected candidates.

Second, PfGBP2, G-strand binding protein 2, is shown to be a telomere binding protein (Calvo and Wasserman, 2015). In our study, PfGBP2 was detected to bind to G4 in all three replicates of pull-down assay, along with some binding to the mutated sequence as well. However, it was not surprising to us as some part of our selected G4 sequence resembles to the telomeric sequence, which explains the presence of the protein in control sample as well. From the previous studies, it was shown that the telomeric sequence of *Plasmodium* can form G4, suggesting that PfGBP2 could also be involved with the G4 (De Cian *et al.*, 2008; Calvo and Wasserman, 2016).

Third, PfDNAJ is a putative protein that belongs to the family of heat shock proteins, Hsp40 (Rug and Maier, 2011). Remarkably, we identified this candidate in both of our unbiased approaches. Even though, it was identified only in one replicate of the DNA pull-down assay, it was detected to bind exclusively to the native G4 as compared to mutated G4. Thus, it is interesting to investigate the role of heat shock protein in G4 biology.

The biochemical and functional characterization of PfGBP2 and DNAJ protein is discussed in the following chapters 4 and 5, respectively.

Chapter 4

Characterization of telomere-binding protein, PfGBP2 in *P. falciparum*

Chapter 4 Characterization of telomere binding protein, PfGBP2 in *P. falciparum*

4.1 Introduction

Telomeres are nucleoprotein complexes that are present at the termini of eukaryotic chromosome ends (Wellinger and Sen, 1997). They protect the chromosome extremities from degradation and end-to end fusion, ensure complete replication of the DNA at chromosome termini and contribute in nuclear architecture (Zakian, 1995; Wellinger and Sen, 1997; Lu *et al.*, 2013). The telomeric DNA comprise of double stranded tandem repeats of conserved eukaryotic GGGTTA motif followed by 3' single stranded G-rich overhang (Meyne, Ratliff and Moyzis, 1989). These 3' single stranded G-rich overhang can adopt to form G4 (Schaffitzel *et al.*, 2001; Demkovičová *et al.*, 2017). Several proteins that interact with these telomeric G4 have been identified, thereby supporting the existence of G4 in telomeres. For instance, yeast RAP1 and ciliate's TEBPα and TEBPβ binds to the telomeric DNA and promotes the formation of G4 (Giraldo and Rhodes, 1994; Paeschke *et al.*, 2005). In contrast, human POT1 is a single stranded G-rich DNA binding protein that disrupts the formation of telomeric G4 and hence, restores the proper elongation by telomerase *in vitro* (Zaug, Podell and Cech, 2005).

Similar to other eukaryotes, telomeres also play an essential role in *P. falciparum.* In Plasmodium, telomeres are composed of repetitions of degenerate telomeric motif, GGGTTYA, (where Y is T or C), which is different from the human telomeric motif (GGGTTA) (Vernick and McCutchan, 1988; Scherf, 1996). Additionally, the telomeres are proposed to regulate the subtelomeric gene families including *var* genes (Duraisingh *et al.*, 2005).

Two recent studies demonstrated that the 3' G-overhang of *P. falciparum* telomeres can form stable G4 under physiological conditions *in vitro* (De Cian *et al.*, 2008; Calvo and Wasserman, 2016). Moreover, the treatment with G4 stabilizing ligands, Telomestatin and TMPyP4, inhibited the parasite telomerase activity *in vitro* (Calvo and Wasserman, 2016). Although, there is no information available on the plasmodial proteins that can interact with

these telomeric G4s, few telomere binding proteins have been identified in *P. falciparum* (Calvo and Wasserman, 2015; Bertschi *et al.*, 2017; Sierra-Miranda *et al.*, 2017).

Here, we are focused on PfGBP2, one of the identified telomere binding protein in *P. falciparum*. Calvo et al showed that the PfGBP2 is a cyto-nuclear protein that can bind to telomeric DNA *in vitro* (Calvo and Wasserman, 2015). However, there is a lack of information on the functional role of this protein. Intriguingly, we have identified this PfGBP2 protein as a potential G4-BP in our previously mentioned DNA pull-down screen in chapter 3. Therefore, to examine whether it can bind to the G4 and have G4-related phenotypes, we have performed two extensively used G4-protein interaction experiments: EMSA and ChIP-Seq (Giraldo and Rhodes, 1994; Muniyappa, Anuradha and Byers, 2000; Zaug, Podell and Cech, 2005; Sanders, 2010; Williams *et al.*, 2017; Götz *et al.*, 2019). To this end, we expressed recombinant PfGBP2 protein to examine *in vitro* binding or protein to G4 using EMSA. We also generated inducible knockout (iKO) PfGBP2 parasite line using combined approach of CRISPR-Cas9 and DiCre/Loxp system, recently developed for the *Plasmodium* (Knuepfer *et al.*, 2017). The advantage of this iKO line is we can tag the protein with HA to study the expression and localization of the protein and the same strain can be induced with rapamycin to generate knockout line. So, to investigate *in vivo* binding of protein to G4 motifs, we performed ChIP-Seq using this generated HA tagged line.

4.2 Results

4.2.1 Bioinformatic analysis and domain organization of a PfGBP2

PfGBP2 (or G-strand binding protein) is a putative single strand telomere binding proteins, which was named after its homolog *C. parvum* CpGBP1 (Calvo and Wasserman, 2015). PfGBP2 contains two conserved RNA recognition motifs (RRMs) of approximately 80 amino acid residue long, which is known to bind single-stranded RNAs (Keene and Query, 1991). Multiple sequence alignment of PfGBP2, its orthologues (CpGBP1 and CrGBP1) and RRM containing telomere-binding proteins in yeast and human are shown in Figure 4.1a. This analysis revealed that the PfGBP2 contains conserved RRMs domains similar to its orthologues. However, unlike its orthologues, PfGBP2 contains long inter RRM region, which is highly rich in arginine residue (30 %). The hinge region of PfGBP2 also contains two RGG motifs. These RGG motifs are shown to mediate nucleic acid binding and protein-protein interactions (Thandapani *et al.*, 2013).

We next examine the homologs of this protein in other closely related Plasmodium species, called *Laverania* species. *Laverania* species are subgenus of *Plasmodium* that can infect apes and cause malaria. *Plasmodium falciparum* is derived from an ancestral species that infect Great Apes. Thus, investigating the role of protein in these species will help provide insight into the importance of the protein with aspect to the evolution. As shown in Figure 4.1b, these two RRM domains and hinge region is highly conserved among closely related species except *P. billcollinsi G01,* suggesting the crucial role of GBP2 in across the Laverania species.

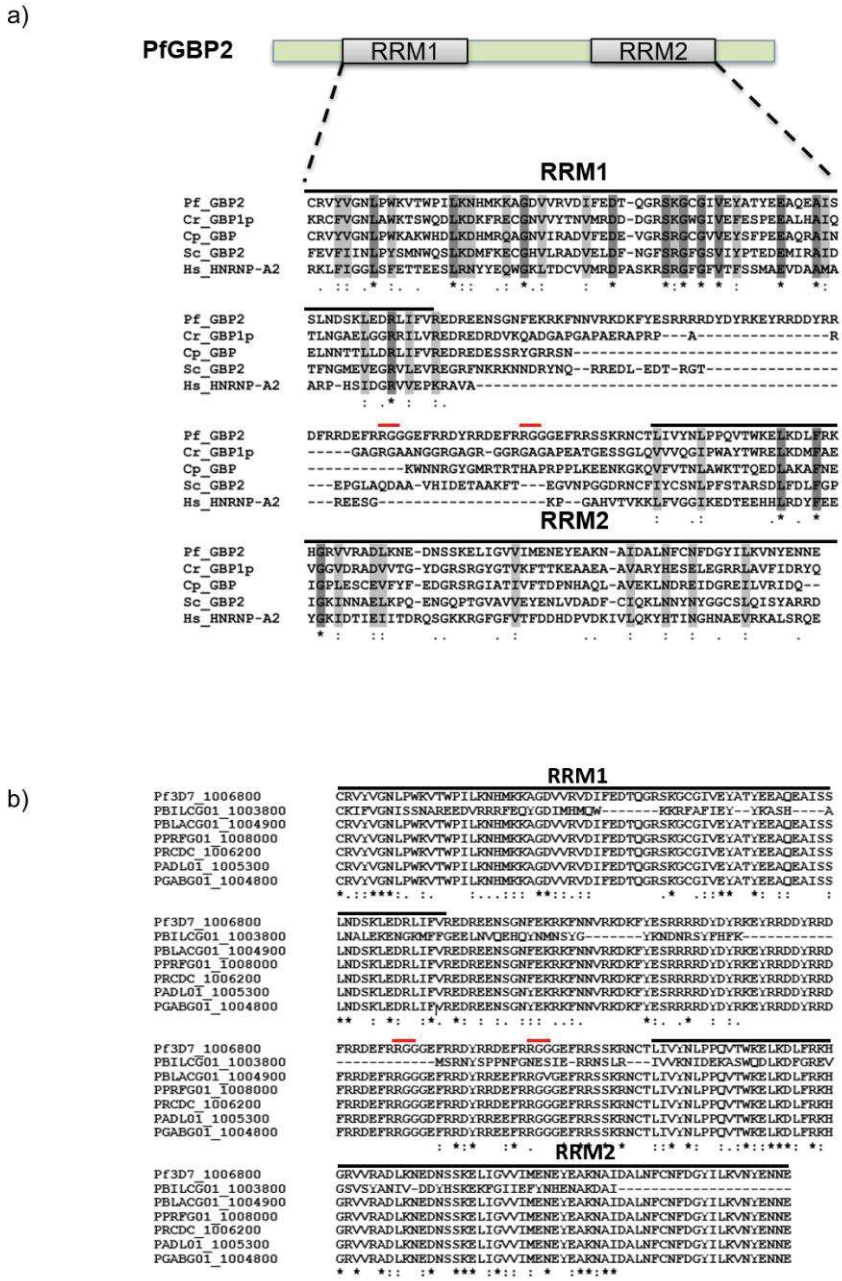


Figure 4.1.Bioinformatic analysis of domain organization of PfGBP2 and its homologs.
a) Schematic representation of the domain architecture of PfGBP2 with sequence alignment of PfGBP2 with its homologs displaying conserved RNA-recognition motif domains (light grey blocks labelled as RRM1 and RRM2) and RGG motif (red bar). Aligned protein sequence includes P. falciparum PfGBP2 (Q8IJX3), Chlamydomonas reinhardtii CrGBP1p (Q39568) , Cryptosporidium parvum CpGBP (Q86PS0), S. cerevisiae ScGBP2 (P25555), and H. sapiens HnRNP-A2 (P22626). b) Multiple sequence alignment of PfGBP2 and its homologues in Laverania species P. falciparum 3D7(Pf3D7_1006800), P. billcollinsi G01 (PBILCG01_1003800), P. blacklocki G01 (PBLACG01_1004900), P. praefalciparum strain G01 (PPRFG01_1008000), P. reichenowi CDC (PRCDC_1006200), P. adleri G01 (PADL01_1005300) and P. gaboni strain G01 (PGABG01_1004800). Multiple sequence alignment of PfGBP2 and its homologs in Laverania species using MUSCLE algorithm at www.phylogeny.fr. (*) indicates identical amino acid while (;) indicates conserved amino acid residues

4.2.2 Recombinant PfGBP2 protein binds to G4 *in vitro*

As described in chapter 3, PfGBP2 was identified as G4-BP in the DNA pull-down assay and MS analysis. So, to examine whether PfGBP2 binds directly to the G4, we performed an electrophoretic mobility shift assay (EMSA) using full-length recombinant PfGBP2 protein.

To this end, recodonized *PfGBP2* gene is cloned into Pet15b expression vector to produce N-terminal His-tagged recombinant PfGBP2 protein. The expression of recombinant PfGBP2 in *E. coli* BL21 DE3 pLysS strain is induced with 1.0 mM IPTG at 25 °C for 5h. The expressed protein is purified using HisPur Cobalt Spin column, followed by Superdex 200 16/600 (10 – 60kDa) size exclusion chromatography. The successful purification of N-terminal His-tagged PfGBP2 protein is confirmed by the detection of His- tagged PfGBP2 with an expected size of ~ 31 kDa using Coomassie staining and western blot analysis with α-his antibody (Figure 4.2a). However, immunoblot displayed the presence of an additional protein, which was later confirmed to be non-specific proteins of *E. coli* origin through mass spectroscopy analysis (Table 4.2).

To perform EMSA, 100 nM of biotinylated wt/mut_G4 motif is incubated with different concentration of purified recombinant PfGBP2 (0- 818 nM). The wt/mut_G4 is the same motif that was used as the bait in previously described unbiased approach. The results revealed that the PfGBP2 binds the WT_G4, whereas no or low significant binding was observed with mut_G4, even in the presence of 818 nM of recombinant PfGBP2. Moreover, the binding of PfGBP2 and WT_G4 displayed a displayed a shift in the G4-protein complex with the disappearance of folded G4 on native PAGE (Figure 4.2b). Thus, indicating that the PfGBP2 binds to G4 *in vitro*.

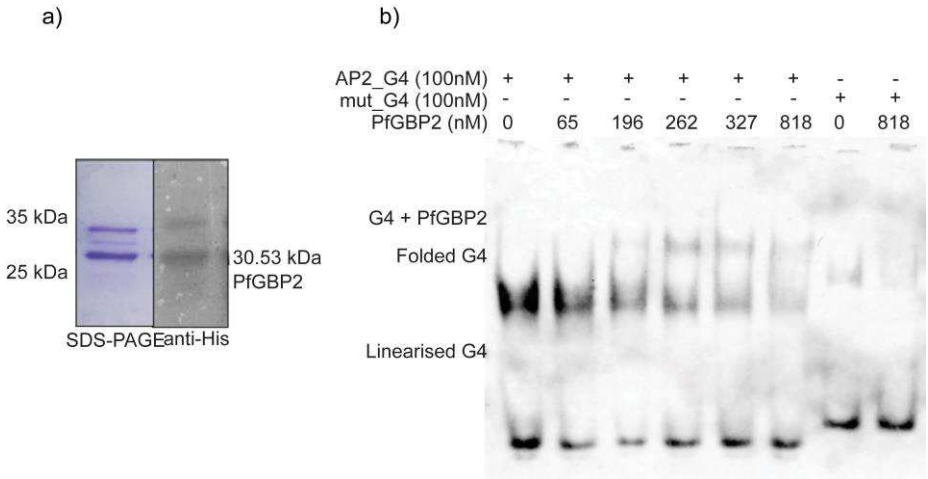


Figure 4.2.Expression and purification of recombinant PfGBP2 to perform EMSA
a) *Coomassie staining of SDS-PAGE and western blot analysis displays the expression of N-terminal His-tagged recombinant PfGBP2 with an expected size of ~31 kDa. NEB pre-stained protein ladder was shown on the left side of the SDS-PAGE. b) EMSA of the recombinant PfGBP2 protein with biotinylated wild type/mutated G4 oligonucleotide displays the binding of PfGBP2 to WT_G4 oligonucleotide.WT_G4 oligonucleotide was used in lane 0- 5. Lane 0 shows two bands for two different forms of WT_G4: folded G4 (upper band) and linearized unfolded form (lower band).Lane 2 – 5 contains increasing amount of proteins (0-818 nM) with 100 nM of wt G4 oligonucleotide. The shift in the band representing binding of PfGBP2 and biotinylated wt G4 are shown by box. Lane 6 – 7 contains mut_G4 oligonucleotide with 0 and 818 nM of PfGBP2 protein, respectively.*

4.2.3 Generation of iKO-PfGBP2 parasite lines using combined CRISPR-Cas9 and Dicre/Loxp system

To dissect the functional role of the PfGBP2 in the asexual stages of the *P. falciparum*, we generated an iKO-PfGBP2 line using the combined approach of CRISPR-Cas9 and DiCre/Loxp system (Knuepfer *et al.*, 2017). In this iKO-PfGBP2 parasite line, we have replaced the wild type gene with a modified PfGBP2 gene containing two loxp sites and C-terminal HA tag in *P. falciparum* p230p 3D7 parasite line (Figure 4.3a). After successful transfection, the transgenic parasites are cloned using limiting dilution and the integration of the recodonized PfGBP2-Ha_3loxp gene is verified using PCR, Sanger sequencing and western blot analysis. The amplified product of 814 bp for the HR1 and 949 bp for the HR2 region is detected exclusively in the transgenic line using integrative PCRs with primer pairs as shown in Table 4.1 (Figure 4.3b). The western blot analysis demonstrated the endogenous expression of HA-tagged floxed –PfGBP2 with an expected protein size of 32.5 kDa in the transgenic line (Figure 4.3c). Furthermore, Sanger sequencing confirms the integration of the recodonized PfGBP2-HA3loxp. The primers used in the PCR analysis and Sanger sequencing are listed in Table 4.1.

4.2.4 PfGBP2 protein is expressed throughout the intra-erythrocytic developmental cycle (IDC) of *P. falciparum*

To examine the expression pattern and localization of PfGBP2 during the intra-erythrocytic stages of *P. falciparum*, we performed an immunofluorescence assay (IFA) using α-HA antibody. IFA demonstrated that the PfGBP2 is expressed throughout the asexual stages of IDC of the parasites. PfGBP2 is localized in both: cytoplasmic and nuclear region. PfGBP2 displays a diffuse cytoplasmic pattern along with a punctuate pattern in nuclei, predominantly in schizont stages (Figure 4.3d).

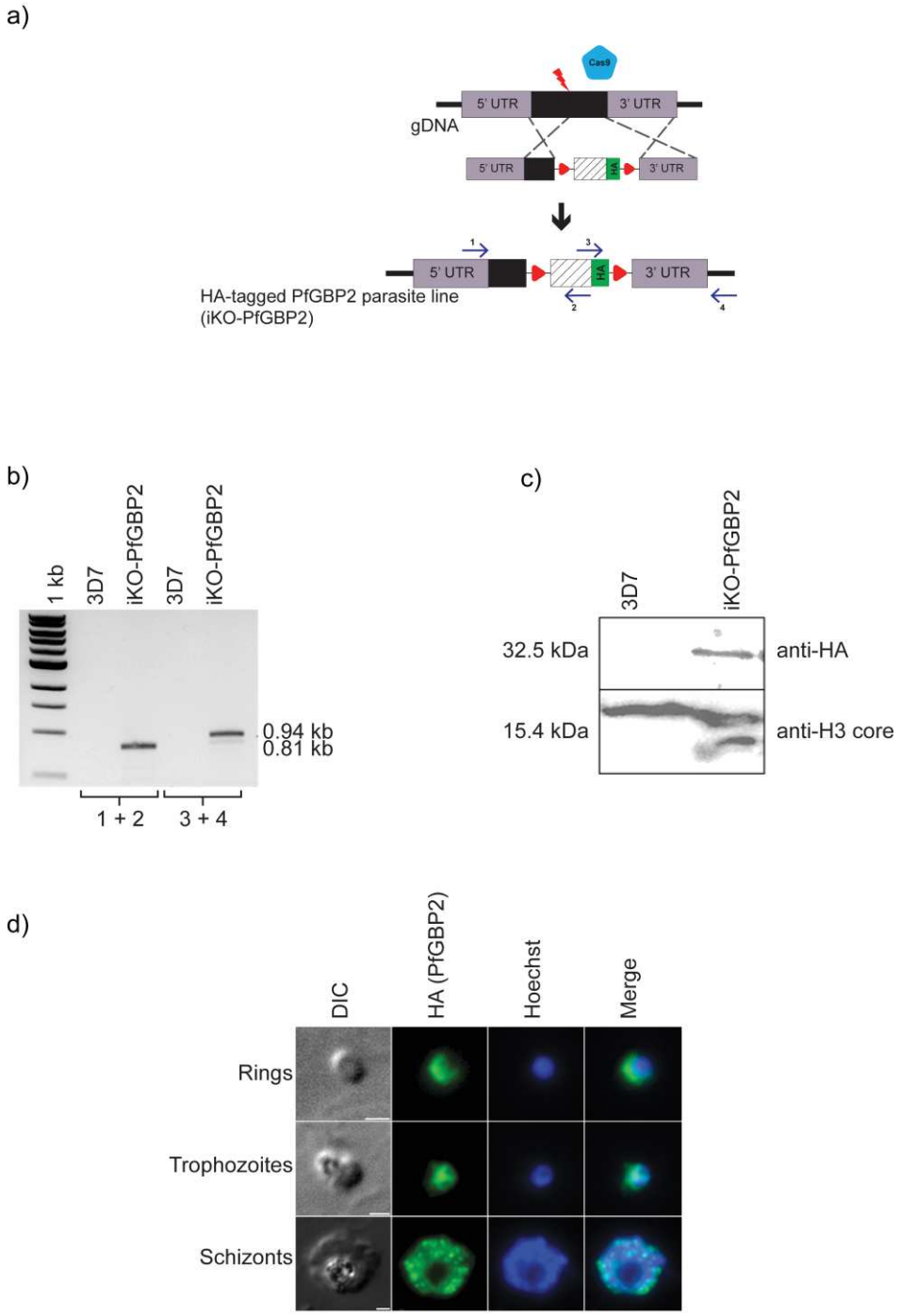


Figure 4.3.Generation of P. falciparum transgenic line expressing floxed PfGBP2-HA

a) Schematic illustration of the generation of P. falciparum 3D7 230p PfGBP2-HA3 loxP line. A donor plasmid with modified PfGBP2 gene containing artificial loxP intron within the recodonised gene (white box) fused to a HA3 tag (green box) followed by loxP sequence. This cassette is flanked by homologous regions 1 and 2 (grey colored box). The 490 bp downstream to start codon and 3'UTR regions are amplified using PCR to obtain HR1 and HR2 region, respectively. Two loxp regions are represented by red color arrowhead. The position of primers used for the confirmation of modified locus are shown as blue colored arrows. The gRNA cleavage site by the Cas9 endonuclease is represented by red flash symbol. b) Diagnostic PCR analysis of genomic DNA of the control parental line (as 3D7-WT) and transgenic line (iKO-PfGBP2), confirming the successful modification at endogenous locus. Expected size of the amplicons are indicated on the right and the lane 1 represents the NEB 1kB ladder. 2-3 lanes are for box1 verification (primer pair 1+2) and 4-5 lanes are for box 2 verification (primer pair 3 + 4). c) Western blot of wild type and transgenic line are probed with α-HA, confirming the expression of HA tagged PfGBP2 line. α-H3 core was used as loading control. The molecular weight of the HA tagged PfGBP2 (~32.5 kDa) and loading control Pf H3 (~15.4 kDa) are represented on the left hand side of the blot. d) IFA

displays the expression of iKO-PfGBP2 throughout the intraerythrocytic cycle. DIC, differential interference contrast. HA tagged PfGBP2 were visualized by α-HA antibody 3F10. Nuclei is stained with DAPI. The scale bar = 1 µm.

4.2.5 PfGBP2 is dispensable for the IDC of *P. falciparum*

To scrutinize the role of PfGBP2 during the IDC of *P. falciparum*, excision of the PfGBP2 gene was induced in the iKO-PfGBP2 parasite line by the addition of 20 nM rapamycin at the 4h ring stage followed by removal 4h later (Figure 4.4a). PCR analysis showed that the excised gene of ~ 1.4 kb length is detected exclusively in the rapamycin-treated parasites while the gene in control parasites (DMSO treated) remained intact. (Figure 4.4b). In addition, the loss of PfGBP2 protein is further confirmed by the western blot analysis using α-HA antibody (Figure 4.4c).

To determine the effect of loss of protein on the growth of the parasites, we performed a growth phenotype assay. Monitoring the parasitaemia of rapamycin and dimethyl sulfoxide (DMSO) treated parasites over two consecutive cycles, displayed subtle difference in the growth of parasites lacking PfGBP2 when compared to the similarly treated parental lines (Figure 4.4d). Thus, highlighting the dispensability of the PfGBP2 for the asexual erythrocytic cycle of these parasites.

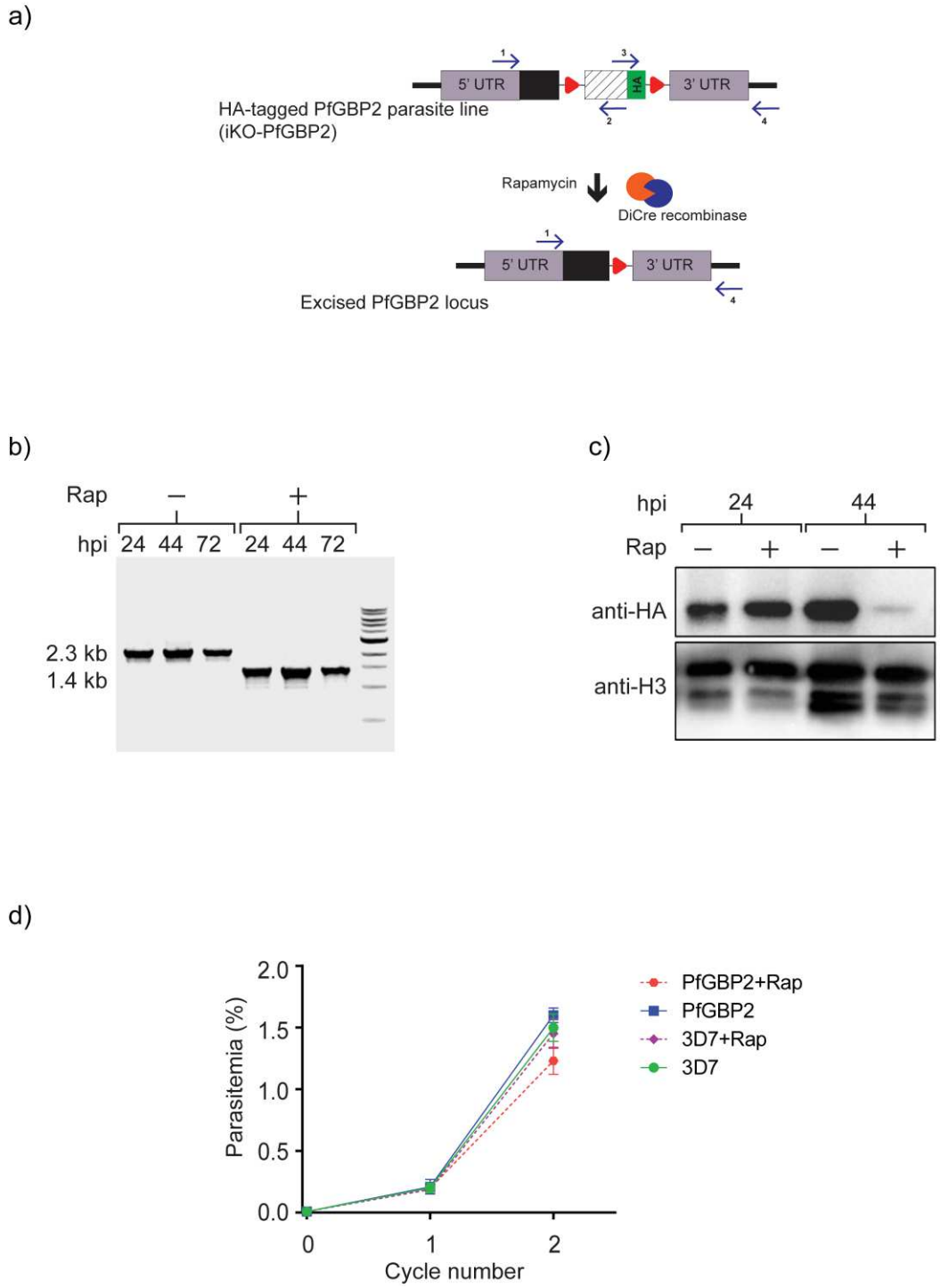


Figure 4.4.Induction of KO in iKO-PfGBP2 parasite line.
a) Schematic representation of conditional deletion of the PfGBP2 gene when induced with rapamycin. Rapamycin induces the dimerisation of dicre-recombinase subunits, which further interacts with the loxp site and induces recombination to generate truncated gene. b) PCR analysis of rapamycin and DMSO treated parasites showed the successful recombination of the floxed gene in rapamycin treated parasites. Primer pair (1 and 4) for the PfGBP2 gene amplified a 2.3 kb band in the DMSO treated transgenic line and a truncated band of ~1.4 kb after excision. c) Western blot analysis of rapamycin and DMSO treated parasites with α-HA antibody showed that depletion of the PfGBP2 protein in rapamycin treated parasites as compared to the control parasites. α-H3 was used as loading control. Parasites were highly synchronized and harvested at different stages of the erythrocytic cycle: trophozoites(24h), schizonts (44h) and re-invaded rings (72h) for genomic DNA and protein lysate preparation. d) Growth phenotype assay of parental line and iKO-PfGBP2 was performed over two consecutive cycles, with DMSO or rapamycin treatment. The disruption of the PfGBP2 gene showed no significant change in

growth as compared to the DMSO treated parasites. Means and standard error were displayed for three independent replicates.

4.2.6 ChIP-seq displays PfGBP2 binds to telomere and G4FS

Next, to investigate whether PfGBP2 can also binds to the G4 *in vivo*, we performed the ChIP-Seq assay using iKO-PfGBP2 parasites that endogenously express HA-tagged PfGBP2 protein. We performed two experiments with addition of pre-clearing step with IgG antibody to reduce the background noise generated by the non-specific binding to the antibody. In experiment 1 (rep 1), the PfGBP2 samples underwent a pre-clearing (i.e. addition of IgG) prior to HA pull-down step while in experiment 2 (rep 2), an IgG control was prepared independently from the IP sample

4.2.6.1 ***Pre-clear step is important for the reduction of non-specific binding to the α-HA beads in P. falciparum Chip-seq***

Due to the multitude of Chip-Seq protocols available, we first established whether a pre-clear step prior to the HA pull-down would decrease the likelihood of non-specific binding to the α-HA beads. As shown in Figure 4.5a, a sharp increase in the number of mappable reads that fall within the called peaks was observed for rep 1, thus suggesting that the pre-clear step is efficient in reducing noise generated by non-specific binding.

4.2.6.2 ***Quality assessment of the ChIP experiment***

Replicates showed consistent peak sets

To confirm the consistency between the two replicates, we compared the datasets obtained from above mentioned, rep1 and rep2. As a result, we found that 89 % of the peaks detected in the second replicate (rep 2) overlapped with the first replicate (rep 1), thus confirming the reproducibility of the ChIP replicates (Figure 4.5b). Hence, we worked exclusively with the intersection of these datasets in our study.

PfGBP2 binds preferentially to narrow regions in the genome

Regarding the distribution of sizes of the obtained peaks, we observed that they were on average ~233bp with most peaks concentrated with the majority of peaks being less than 300 bp (Figure 4.5c).

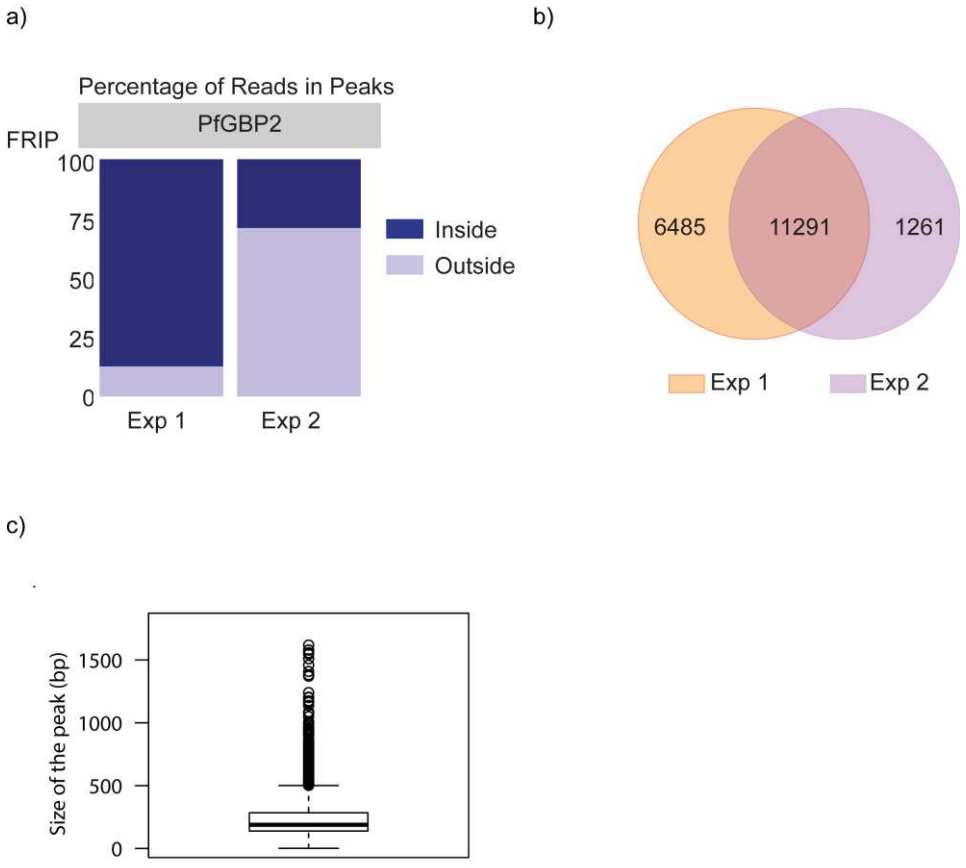


Figure 4.5.Optimisation of ChIP-Seq protocol.
a) Calculation of percentage of reads that map within called peaks. In experiment 1 (rep 1) the PfGBP2 samples underwent a pre-clear step while in experiment 2 (rep 2) an IgG control was prepared independently from the IP sample. b) Venn diagram represents the intersection of the peaks detected in each of the experiments (rep1 and rep 2). c) Box plot represents the size distribution of obtained peaks and shows that PfGBP2 binds to narrow regions in the genome. Min. 1 ; 1st Qu. 139; Median 189; Mean 233,4 ; 3rd Qu. 284 ; Max. 3436.0 (in bp).

4.2.6.3

Enrichment of G4 forming sequences in the peak dataset.

To determine the binding regions of PfGBP2 over the *Plasmodium* genome, we analyzed the genome-wide distribution of the enriched regions, focusing on the most significant peaks (logQ>50). The ChIP-Seq analysis revealed that GBP2 is likely to have affinity for multiple sites throughout the genome. However, we noticed that the peaks with very high scores are mostly prevalent on the edges of the chromosomes (Figure 4.6a). This result was encouraging given that previous studies showed the enrichment of G4 motifs in the edges of the chromosomes covering telomeric and subtelomeric regions (Smargiasso *et al.*, 2009; Stanton *et al.*, 2016; Gazanion *et al.*, 2020).

As a result, we next aimed to verify whether indeed the ChIP dataset was enriched in G4 forming sequences (G4FS). We collated the genomic sequences under the peaks (logQ>50) into a common FASTA file (akin to a "fake" genome) and subsequently used the web version of G4Hunter application to check if this dataset was enriched in G4FS comparatively to the *P. falciparum* genome. G4Hunter is a widely used G4 prediction tool that gives the quadruplex propensity score of a given nucleic acid sequence on the basis of G-richness and G-skewness of a sequence (Brázda *et al.*, 2019).

As per calculated by the G4Hunter application, the 23.3 Mb of Pf genome with a % GC content of 19.3% harbors 6033 genomic sites with propensity to form G4. Comparatively, the peaks FASTA file had 1949 potential G4FS over 0.51 Mb. This resulted in an increased %GC content to 26.8% and corresponded to an enrichment of 15 fold (hypergeometric p-value= 2,994 e-8) (https://github.com/AnimaTardeb/G4Hunter). The screenshots of the results of G4Hunter analysis are shown Figure 4.6b,c. Therefore, the ChIP-Seq analysis reveals the *in vivo* binding of the PfGBP2 to the G4FS within the genome of *P. falciparum.*

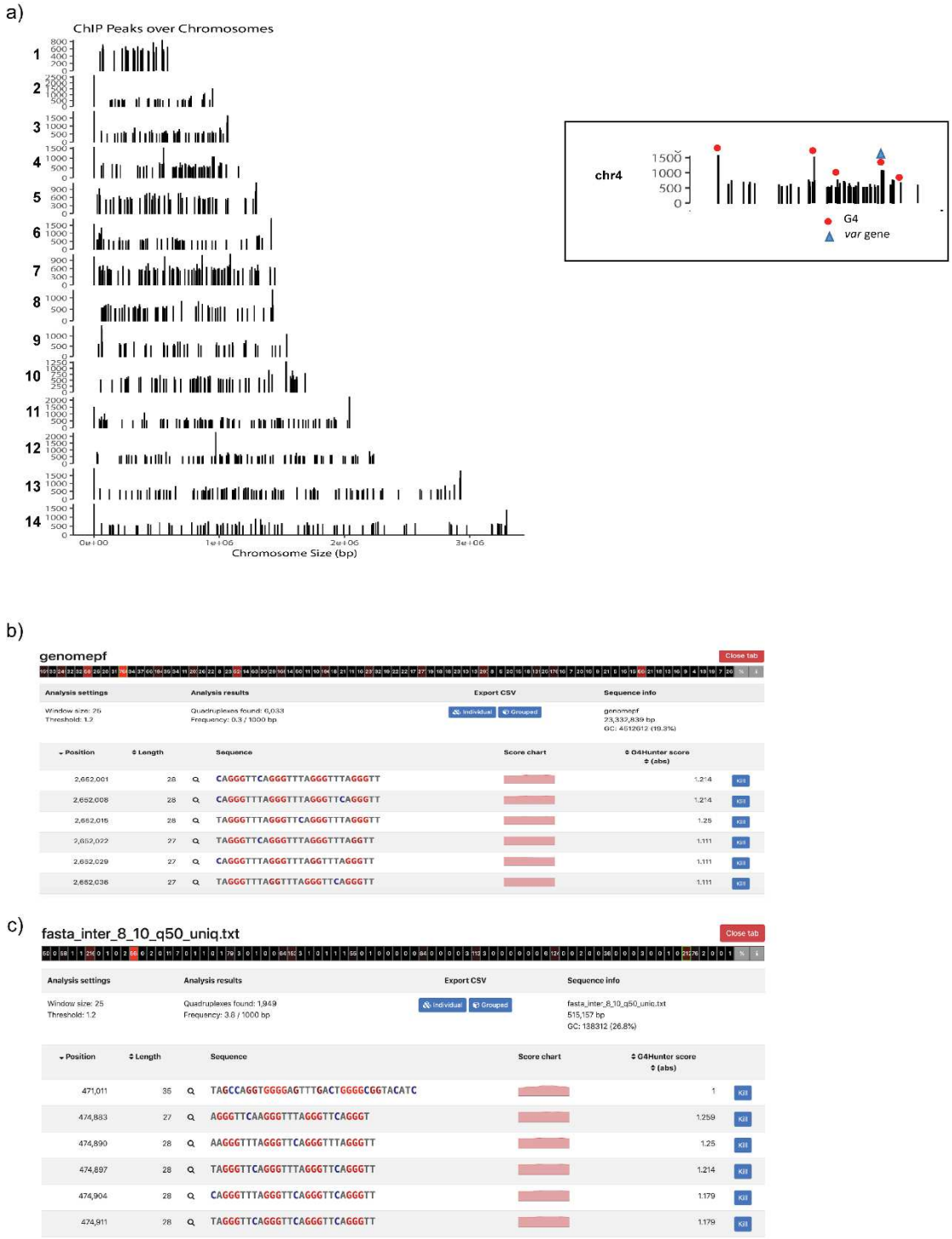


Figure 4.6.Genome-wide mapping of PfGBP2 on the P. falciparum genome.
a) Intregative genomic view of ChIP-Seq analysis of iKO-PfGBP2 depicting the distribution of the peaks (black) with -log(Q)>50 in all the chromosomes (1-14) with an enlarged image of chr4 displaying the GBP2 peaks and predicted G4FS using G4Hunter. b) and c) Screenshot of G4Hunter search displays some results obtained for Pf genome (b) and fake genome(containing the concatenated ChIP-Seq peaks with –log(Q)>50)

Telomeric motifs are enriched in the ChIP-Seq dataset (MEME search).

Given the high number of obtained peaks, we next examined whether PfGBP2 binds to motifs other than G4FS. Thus, we probed the sequences under the peaks with the MEME suite tool to identify overrepresented motifs in the dataset. We focused on the filtered log (Q)>50 dataset and identified nine high significantly enriched motifs (Figure 4.7). Interestingly, the first hit corresponded to a motif (GGGTTYA, where Y is T or C) that has been described as consensus sequence of telomeric repeat sequence in *P. falciparum* (Vernick and McCutchan, 1988; Scherf, 1996). In addition, the ninth hit also corresponds to the telomeric repeat sequence. This data corroborates with the previous study that showed the *in vitro* interaction of PfGBP2 with telomeric DNA (Calvo and Wasserman, 2015). Taken together, the data confirms that the PfGBP2 is a *bona fide* telomere-binding protein in *P. falciparum*.

4.2.6.4

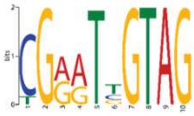
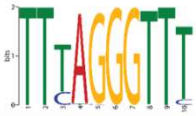
Rank	Logo	E-value	Sites	width
1		4.7e-350	915	10
2		7.7e-227	592	10
3		1.4e-065	242	10
4		2.9e-042	328	10
5		1.8e-032	325	10
6		2.8e-030	741	8
7		1.3e-018	1000	10
8		3.4e-016	96	10
9		1.3e-021	271	10

Figure 4.7.Highly enriched motifs under the PfGBP2 peak dataset with [log (Q)>50].
The list represents the top 9 highly reached motif under PfGBP2 significant peak in dataset with [log(Q)>50] . These highly enriched peaks were obtained from MEME search. The e-value describes the statistical significance of the motif. The sites represents the number of sites that contributed to the construction of the motif. And the width describes the length of the motif.

4.2.7 PfGBP2 does not affect the telomere length homeostasis

Given that PfGBP2 binds to the telomeric region of *P. falciparum*, we next investigated whether PfGBP2 depletion affects telomere length homeostasis. We performed telomere restriction fragment (TRF) length Southern blotting analysis on the genomic DNA from iKO-PfGBP2 parasites cultured in the presence of rapamycin or DMSO for four and eight weeks. Southern blot analysis showed no significant difference in the telomere length of rapamycin-induced PfGBP2-KO line in comparison to the control parasites (iKO-PfGBP2^{DMSO}) within

30 generation of culturing (Figure 4.8). Moreover, the mean telomere length of both parasites was found to be ~ 1.5 kb long, which is consistent with the average length of telomeres in *P. falciparum* 3D7 (Figueiredo *et al.*, 2002). Hence, indicating that the PfGBP2 does not play any role in the telomere length homeostasis in *P. falciparum* within 30 generation of culturing.

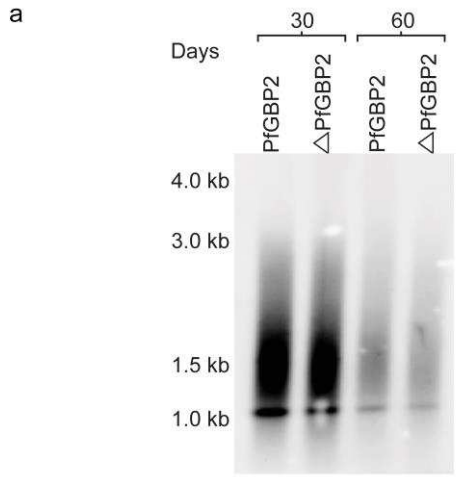


Figure 4.8.Telomere Restriction Digestion Southern blotting.
TRF Southern blot displaying the telomere length from iKO-PfGBP2 lines that were treated with rapamycin (△PfGBP2) or DMSO (PfGBP2) for 30 days and 60 days.

4.2.8 Loss of PfGBP2 shows slight derepression in some subset of *var* genes

Previous studies implicated the role of G4 in the regulation of *var* gene expression, so in order to examine whether the deletion of PfGBP2 has any effect on the expression of the *var* genes, we assessed the *var* genes expression on rapamycin-induced PGBP2 knockout parasites by qRT- PCR. The list of the primers used in the study is given in Table 4.3. As shown in Figure 4.9, identical *var* gene (PF3D7_0711700) is expressed in both the rapamycin treated and control iKO-PfGBP2 parasites at similar levels. On the other hand, the most of the remaining *var* genes showed some derepression in the rapamycin-induced PfGBP2-KO line, however the overall level of expression is low (Figure 4.9) (Table 4.4).

a)

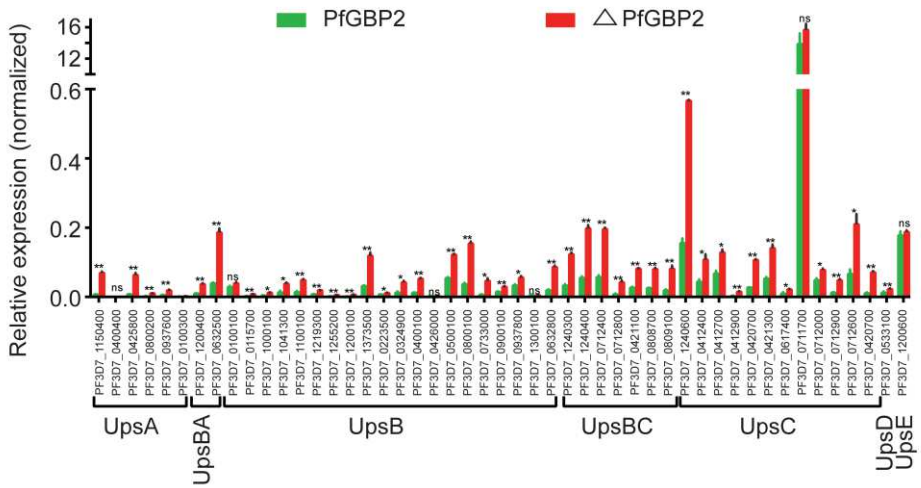


Figure 4.9.Quantitative RT-PCR (qRT-PCR) analysis for the var genes expression in rapamycin induced PfGBP2 KO line and control iKO-PfGBP2 parasite line.
*Bar graph displays the quantitation of var gene transcripts in rapamycin (△PfGBP2) or DMSO (PfGBP2) treated iKo-PfGBP2 parasites with slight derepression of several var genes in rapamycin-induced PfGBP2 KO parasites. Data are represented as mean ± SD (n =2). ns non-significant, ,** ≤ 0.05,* ≤ 0.01.*

4.3 Discussion

Several telomere binding proteins are known to bind G4 that are formed in the G-rich telomeric sequences (Giraldo and Rhodes, 1994; Paeschke *et al.*, 2005; Zaug, Podell and Cech, 2005; Wang *et al.*, 2012b). Likewise in *Plasmodium falciparum*, telomere binding proteins were implicated to be involved with G4 (Bertschi *et al.*, 2017; Sierra-Miranda *et al.*, 2017), though none of the studies established the direct interaction with the G4FS in these malarial parasites.

This study is the first report, to our knowledge, exhibits the interaction between telomere binding protein and G4 in *P. falciparum*. Previous study has shown that PfGBP2 binds to the telomeric sequence of *P. falciparum in vitro* (Calvo and Wasserman, 2015), however no functional implications have been elucidated. Based on ChIP-Seq analysis, we here confirmed the *in vivo* binding of PfGBP2 to the telomeric sequence of these parasites. Besides telomeres, PfGBP2 binding sites are enriched with the G4FS. Strikingly, the most significant identified binding motif of PfGBP2 (GGGTTYAGGG, where Y is C or T) is similar to the G4 motif (GGGATTTGGGAGGGGGGG) that was used in EMSA analysis to

validate the PfGBP2 and G4 interaction. Hence, taken together, this study confirms that PfGBP2 is a bona fide G4 binding protein in *P. falciparum.*

Telomere-associated proteins such as PfSir2A and PfSir2B regulates the transcription of *var* genes in *P. falciparum* (Figueiredo and Scherf, 2005; Tonkin *et al.*, 2009). The same study has shown that both PfSir2A and PfSir2B showed an effect on the expression of *var* gene while lacking of PfSir2A, but not PfSir2B affected the telomere length. Thus, it could be possible that if we culture these PfGBP2 lacking parasites for longer period, more than two months (30 generations) as it was in our case, we might observe the increased derepression in *var* gene expression. Besides, as it has been previously reported that G4 are associated with the mitotic *var* gene recombination (Claessens *et al.*, 2014; Stanton *et al.*, 2016), PfGBP2 may be involve in the recombination and diversification of *var* genes. It will be interesting to examine the exact function of PfGBP2 in future studies.

Given that PfGBP2 is shown to localize in both cytoplasm and nucleus of the asexual blood stages of the parasites, this dual localization might attributes to the non-nuclear based functions of PfGBP2. This is further supported by the presence of conserved RNA recognition motif (RRM) in this protein. RRMs are found within several RRM containing proteins that play a crucial role in telomere biogenesis and RNA metabolism (Maris, Dominguez and Allain, 2005). Human nuclear ribonucleoprotein hnRNP A2/B1 containing RRM domain was reported to bind the telomeric G4 and unfolds the G4 to promote telomerase activity (Wang *et al.*, 2012a). On the other hand, in HIV-1, hnRNP A2/B1 was shown to act in concert with Rev protein to promote nucleo-cytoplasmic trafficking of viral RNA and nuclear viral RNA retention (Gordon *et al.*, 2013). Hence, we speculate that PfGBP2 might have an additional role in RNA metabolism in *P. falciparum*.

In conclusion, we have characterized the role of PfGBP2 during the intra-erythrocytic development of *P. falciparum*. PfGBP2 is expressed throughout the intra-erythrocytic cycle of the protein. It is present in both nucleus and cytoplasm. The loss of PfGBP2 is not detrimental for the growth of asexual stages of the parasites. PfGBP2 is shown as *bona fide* G4-binding protein using EMSA and ChIP-Seq analysis. Although it is not clear yet whether PfGBP2 promotes the G4 formation or stabilizes the G4, which needs to be addressed in future studies. Moreover, this study provides the evidence for the existence of G4 in *P.*

falciparum and open new avenues for dissecting the molecular mechanisms underlying the G4 functions in these malarial parasites.

Chapter 5

Characterization of putative PfDNAJ protein in *P. falciparum*

Chapter 5 Characterization of putative PfDNAJ protein in *P. falciparum*

5.1 Introduction

Plasmodium falciparum manifests a complex life cycle alternating between two different hosts, female *Anopheles* mosquitoes and humans. During its life cycle, these parasites are subjected to various stresses including temperature fluctuations while alternating between two different hosts and an increase in temperature associated with febrile malaria episodes in infected patients. Under such heat stress conditions, several members of heat shock proteins, functioning as molecular chaperones, are shown to be up-regulated in these parasites (Kumar *et al.*, 1991; Biswas and Sharma, 1994; Bonnefoy *et al.*, 1994; Das *et al.*, 1997; Watanabe, 1997; Banumathy *et al.*, 2003). Besides, a significant fraction of 2 % of the *Plasmodium* genome encodes for these molecular chaperones (Acharya, Kumar and Tatu, 2007). These chaperones are involved in maintaining the conformational integrity of the parasite proteome, crucial for its survival within its host. Strikingly, amongst all chaperones, Hsp40s constitute approximately 5 % of around 400 parasite derived proteins that are predicted to be exported to the infected erythrocytes (Hiller *et al.*, 2004; Sargeant *et al.*, 2006).

P. falciparum genome encodes for at least 49 Hsp40 proteins (PfHsp40) as compared to only 6 Hsp70 proteins (PfHsp70) (Botha, Pesce and Blatch, 2007; M. Njunge *et al.*, 2013). Hsp40 acts as a co-chaperone that interacts and modulates Hsp70 activity through a conserved J-domain (Hennessy *et al.*, 2005). These proteins, PfHsp40s are classified into four types-I, II, III and IV, based on the presence of four domains: a highly conserved J-domain, a GF (Gly/Phe rich region) domain, a zinc-finger domain, and a less conserved C-terminal domain. Each of these domains has a distinctive role. The highly conserved J-domain comprise of the HPD (His-Pro-Asp) motif that is involved in the stimulation of the ATPase activity of Hsp70 (Hennessy *et al.*, 2000). Based on the presence of a conserved J-domain, Hsp40 acts as a co-chaperone that interacts and modulates Hsp70 activity through a conserved J-domain (Hennessy *et al.*, 2005). The GF domain is proposed to be significant in the regulation of the substrate-binding activity of Hsp70 (Botha, Pesce and Blatch, 2007). The Zn-finger domain comprises four cysteine repeat sequences (CXXCXGXG) that binds to two zinc ions and are involved in the stabilization of the tertiary structure of Hsp40 (Banecki *et al.*, 1996). Among

all the domains, the C-terminal domain is the least conserved region. The C-terminal domains are implicated in substrate binding and in dimerization in Hsp40 proteins (Han and Christen, 2003; Wu *et al.*, 2005).

The type-I Hsp40 proteins contain all the four domains. Type-II proteins lack the Zn-finger domain, whereas type-III proteins harbor only the J-domain. On the other hand, the type-IV proteins are distinct from others as they comprise of the J-like domain, where the highly conserved HPD motifs are disrupted (Botha, Pesce and Blatch, 2007).

Out of the 49 Hsp40 proteins, 2 are type-I Hsp40s containing all four domains, 9 are type-II Hsp40s lacking the Zn-finger domain, 26 are type-III Hsp40s harboring only the J-domain, and lastly, 12 type-IV Hsp40s, which comprise J-like domain with disrupted HPD motifs are identified in *P. falciparum*. However, only a few of them have been experimentally characterized (M. Njunge *et al.*, 2013). (i) The two type-I proteins (PFD0462w- PfJ1 and PF14_0359) have been analyzed and found to be upregulated during heat stress [6,17]. They are proposed to function as housekeeping co-chaperones within malarial parasites [12]. (ii) Five type-II proteins (PFA0660w, PFB0090c, PFE0055c, PF11_0099, PFL0565w, and PFB0595w) have been characterized. The first three proteins contains PEXEL/HT motif and are predicted to be exported beyond the parasite vacoular membrane into the infected erythrocytes (Sargeant *et al.*, 2006). The PEXEL (Plasmodium export element) or HT (host targeting signal) motifs are signal motifs that essential for the export of *Plasmodium* proteins into the host cell (Sargeant *et al.*, 2006). Despite exhibiting a high degree of sequence identity among themselves, only the loss of PFA0660w is lethal for the parasites while others are dispensable (Maier *et al.*, 2008). PF11_0099 (Pfj2) is predicted to have PEXEL/HT motif that overlaps with the helix II of the J domain and thus not considered as the true PEXEL/HT motif containing protein (Botha, Pesce and Blatch, 2007). This protein is constitutively expressed in erythrocytic stages though its expression decreases after heat shock (Watanabe, 1997). On the other hand, the latter two type-II proteins (PFL0565w, and PFB0595w) are non-PEXEL proteins. PFL0565w (Pfj4) is located in the cytoplasm and nucleus of the late asexual stages and is upregulated under heat shock (Watanabe, 1997). It is proposed to be involved in translational regulation through its interaction with PfGCN20 (Vazquez de Aldana, Marton and Hinnebusch, 1995; LaCount *et al.*, 2005). Next, cytosolic PFB0595w are demonstrated to stimulate the *in vitro* ATPase activity of PfHsp70-1 and are observed to be upregulated under heat shock similar to its predicted partner PfHsp70-1. (Njunge *et al.*,

2015). (iii) Type-III Hsp40s comprises of the largest and most diverged class of Hsp40s, yet little information is known about them. They are suggested to perform specialized functions in other species (Kelley, 1998; Young *et al.*, 2004). PF10_0378 (Pfj3) exhibited an increased expression level during heat shock stress (Watanabe, 1997). Other two proteins, PF13_0102 and PF08_0032 are predicted to be residing in the endoplasmic reticulum (ER) and are homologous to Sec63 and ERdj5, respectively (M. Njunge *et al.*, 2013). (iv)Type-IV Hsp40s include members of the RESA family (PFA0110w, PFA0675w, PF11_0509, and PF11_0512) containing PEXEL/HT signal sequence. Two type-IV proteins (PF11_0034 and PF11_0509) are indispensable for the growth of the parasites while the deletion of PF10_0381 resulted in the lack of surface knobs on infected erythrocytes and resulted in loss of cytoadherence to endothelial cells (Maier *et al.*, 2008).

Here, we examined putative PfDNAJ, one of the type-III Hsp40 proteins (named as PfDNAJb in our study). PfDNAJb (PF3D7_0823800) is a putative DNAJ protein that has been identified as G4- interacting protein by the unbiased approach, Y1H system and DNA pull-down assay. To this end, we performed characterization of PfDNAJb using inducible knockout parasite line.

5.2 Results

5.2.1 Bioinformatic analysis and domain organization of a P. falciparum Hsp40 protein DNAJb

Until now, there are around 26 proteins that have been identified within the type-III Hsp40s group. Phylogenetic analysis and domain identification of all the members showed that this group includes several distinct proteins with varied length (115 - 2738 amino acids in length) that shares the common conserved J-domain (Figure 5.1a). Some members of this group contain PEXEL/HT motifs and are predicted to be transported to the host erythrocytes(Maier *et al.*, 2008; M. Njunge *et al.*, 2013). Amongst all the members, PF3D7_0823800 (PfDNAJb) is a unique protein that exhibits DNA polymerase family B signature motif and thioredoxin domain along with conserved J-domain (Figure 5.1b). These domains are found to be highly conserved among the closely related species called *Laverania* species, suggesting that this protein is crucial for the parasites for their survival (Figure 5.6a).

In addition, PfDNAJb is predicted to comprise three transmembrane domains and a nuclear localization signal (NLS). The NLS motif is predicted to be present at the N-terminal of the protein using NLStradamus and cNLS Mapper program (Kosugi *et al.*, 2009; Nguyen Ba *et al.*, 2009).

Sequence analysis of PfDNAJb revealed that the conserved J-domain comprise of four helices (I-IV) with highly conserved HPD (His-Pro-Asp) tripeptide in the loop region between helices II and III. This HPD motif is crucial for the J-domain based stimulation of ATPase activity of Hsp70s (Tsai and Douglas, 1996; Genevaux, Georgopoulos and Kelley, 2002). Other key residues Y7, L10, Y25, F47, L57, and D59 that are structurally important for the J-domain functions are also well conserved in PfDNAJb (Genevaux, Georgopoulos and Kelley, 2002; Nicoll *et al.*, 2007) (Figure 5.1b). Next, we searched for the orthologues of the PfDNAJb using this conserved J-domain as a query in *H. sapiens, S. cerevisiae and E. coli*. The BLASTp and phylogenetic tree analysis revealed that despite sharing more than 44% identity with its orthologues, the J-domain of PfDNAJb is relatively diverged from its orthologues (Figure 5.1c).

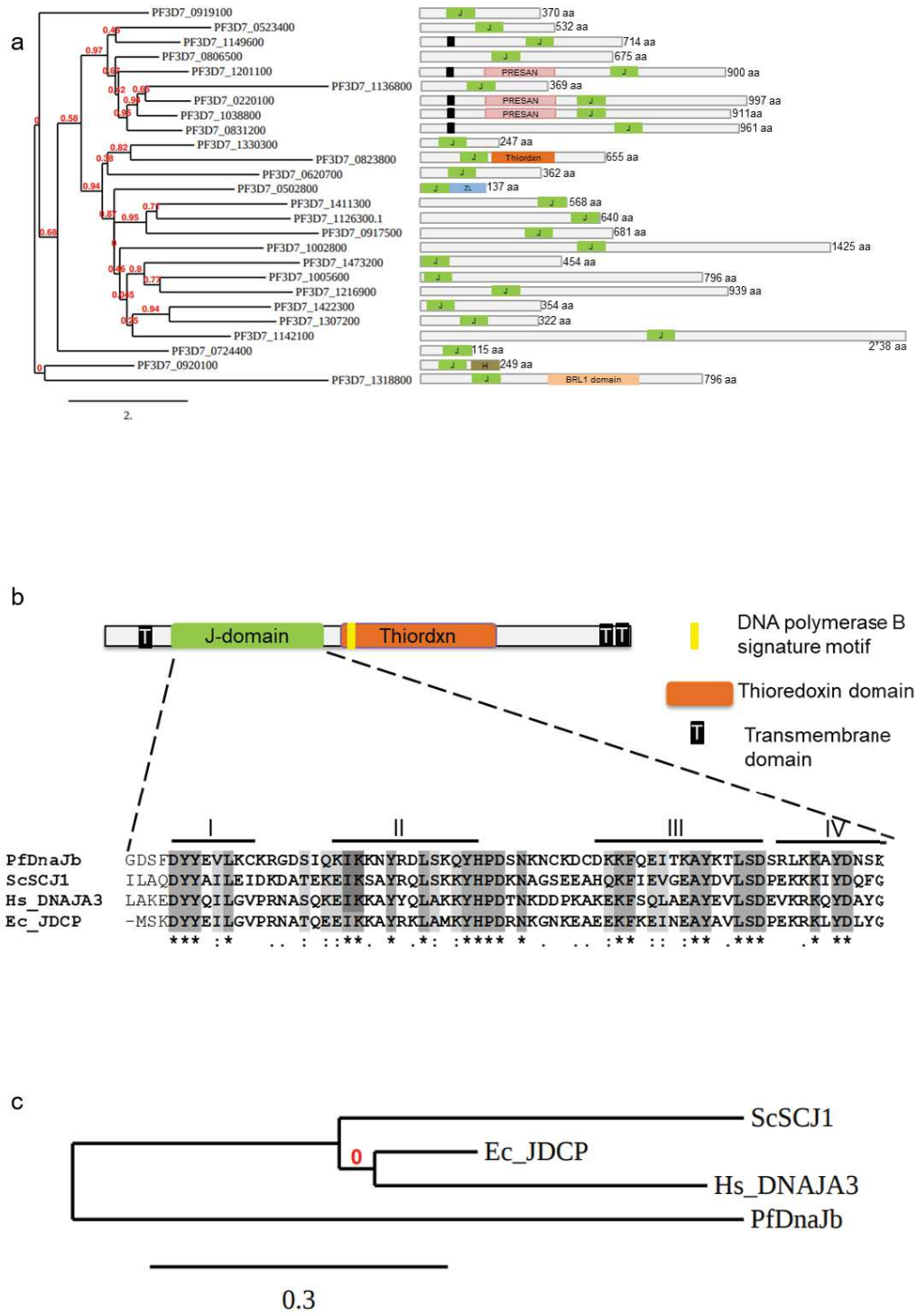


Figure 5.1. Bioinformatic analysis of Type III Hsp40s and PfDNAJb in P. falciparum.
a) Schematic representation of the domain architecture and phylogenetic tree analysis of all the type III Hsp40s of P. falciparum. Each protein sequence is represented by grey bar with the number of amino-acids displayed on the right end of the bar. Different domains are depicted by blocks within the grey bar using different colors. Green block (labeled as J) represents J- domain, black block for PEXEL/HT signal motif, light pink block (labeled as PRESAN) for PRESAN domain, orange block (thiordxn) for thioredoxin domain, yellow block represents DNA polymerase B signature motif, grey colored H labeled block represents HSCB domain, and light orange block represents BRL1 domain of Sec63 orthologues, Black bar (labeled as T) denotes transmembrane domain. These domains are identified using InterPro and ScanProsite tool. b) Schematic representation of different domains of PfDNAJb (PF08_0032 /PF3D7_0823800) with sequence alignment of J

domain of PfDNAJb and its homologs. Aligned protein sequence includes P. falciparum DNAJ (XP_001349286.1) , S. cerevisiae SCJ1 (CAA41529.1), H. sapiens DNAJA3 (AAH32100.1), E. coli J domaincontaining protein (WP_000031254.1). Four helixes of J domains are marked by bold line (-) and labelled as I, II, III and IV. (*) indicates identical amino acid while (;) indicates conserved amino acid residues. c) Phylogram representing J domain of PfDNAJb and its homologs. Sequences were aligned and phylograms (a and c) were generated in web server phylogeny.fr (www.phylogeny.fr) in "one-click" mode

5.2.2 Expression and purification of recombinant protein

To determine the interaction between PfDNAJb and G4 using EMSA.We first tried to express recombinant protein of PfDNAJb, the recodonized *PfDNAJb* gene was cloned into a bacterial Pet15b expression vector to produce N-terminal His-tagged recombinant protein in *E. coli*. Different temperature conditions and IPTG concentrations is used to induce the expression of the recombinant protein in *E.coli* BL21 DE3 pLysS strain. However, despite several attempts, the recombinant expression of this protein could not be achieved (Figure 5.2). This could be due to the high toxicity effect of the protein on the *E. coli* strain due to the presence of transmembrane domain in PfDNAJb.

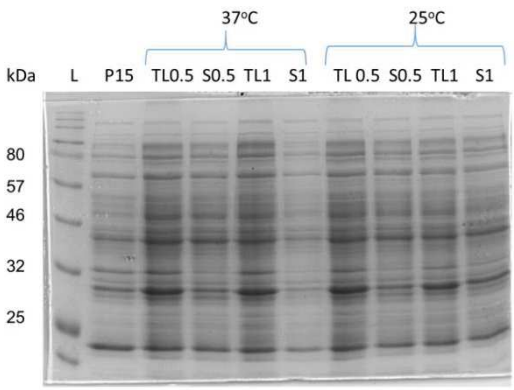


Figure 5.2.Expression of recombinant PfDNAJb protein in E.coli BL21 DE3 pLysS.
Expression vector pET15b containing PfDNAJb was transformed in E.coli BL21 DE3 pLysS. Different expression conditions such as temperature (37°C for 3 h and 25 °C for 5 h) and IPTG concentration (0.5 and 1 mM) are used to induce the expression of recombinant protein (~75 kDa). Total bacterial lysates and supernant lysate are loaded. SDS-PAGE shows no expression of PfDNAJb when compared to the total lysate of bacterial culture containing empty Pet15b vector. Protein marker was loaded on the left end of the SDS PAGE gel.

5.2.3 Generation of inducible knockout PfDNAJb parasite line using combined CRISPR-Cas9 and Dicre/Loxp system

Next, to investigate the role of PfDNAJb in the asexual stages of the *P. falciparum*, we generated an inducible knockout strain of PfDNAJb using the combined approach of CRISPR-Cas9 and Dicre/Loxp system, (Knuepfer *et al.*, 2017). In this iKO-PfDNAJb

parasite line, we have replaced the wild type gene with a modified PfDNAJb gene containing two loxp sites and C-terminal HA tag in *P. falciparum* p230p 3D7 parasite line (Figure 5.3a). After successful transfection, the transgenic parasites are cloned using limiting dilution and the integration of the recodonized PfGBP2-Ha$_3$loxp gene is verified using PCR, Sanger sequencing and western blot analysis. The successful replacement of wild type *PfDNAJb* gene with modified *PfDNAJb-HA$_3$loxp* in *P. falciparum* 3D7 p230p strain to obtain iKO-PfDNAJb line was analyzed by integrative PCRs using primer pairs as shown in Table 5.1. The genomic DNA of the transgenic line amplified the amplicons of 860 bp and 600 bp with primer pair (PG174 +PG175) and (PG176+PG182), respectively whereas no such band was observed in parental line, as expected (Figure 5.3b). Immunoblot with α- HA antibody demonstrated an expected protein band of ~ 72kDa specific to the generated transgenic line (Figure 5.3c).

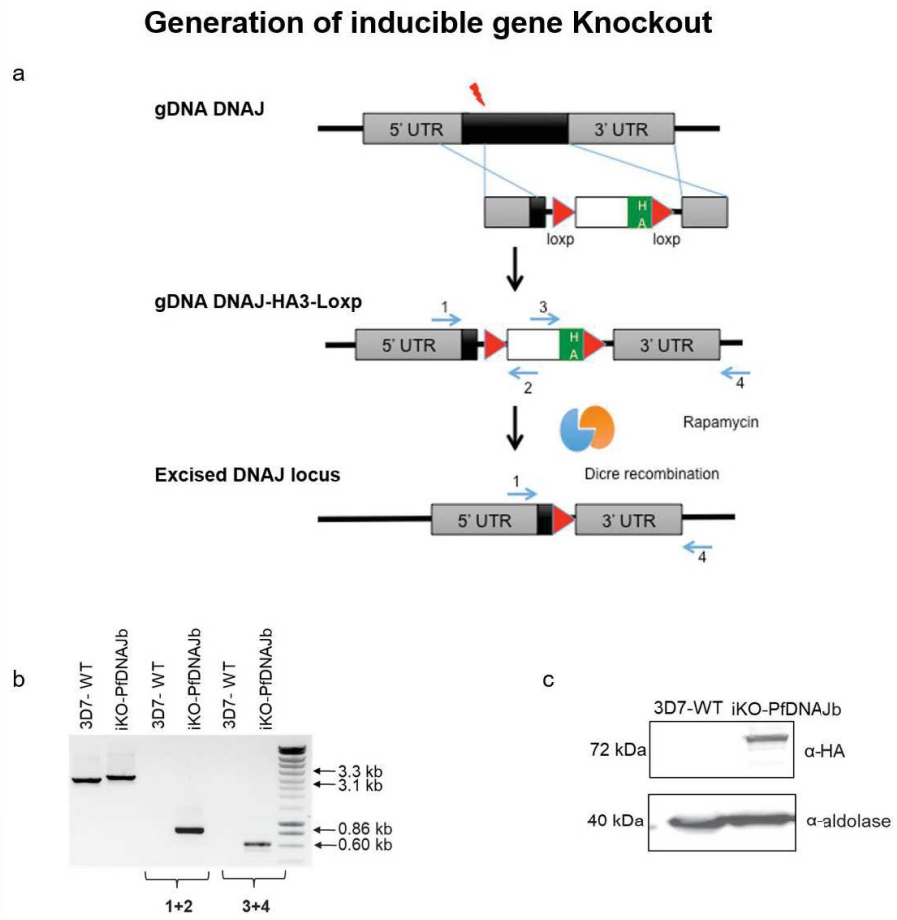


Figure 5.3.Generation of P. falciparum transgenic line expressing floxed PfDNAJb-HA gene
a) Schematic illustration of the generation of P. falciparum 3D7 230p PfDNAJjb-HA3 loxP line. A donor plasmid with modified PfDNAJb gene containing artificial loxP intron within the recodonised gene (white box) fused to a HA3 tag (green box) followed by loxP sequence, flanked by homologous regions 1 and 2 (grey colored box). The 5'UTR plus 100bp of cds region and 3'UTR regions are amplified using PCR to obtain HR1 and HR2 region, respectively. Two loxp regions are represented by red color arrowhead. The positions of primers used for the confirmation of modified locus are shown as blue

colored arrows. The gRNA cleavage site by the Cas9 endonuclease is represented by red flash symbol. b) Diagnostic PCR analysis of genomic DNA of the control parental line (Pf 3D7 230p as 3D7-WT) and transgenic line (iKO-PfDNAJb), confirms the successful modification at endogenous locus. Expected size of the amplicons is indicated on the right and the right hand corner lane contains the NEB 1kB ladder. 1-2 lanes are for whole locus PCR verification (primer pair 1 + 4), 3-4 lanes are for box1 verification (primer pair 1+2/PG174 +PG175) and 5-6 lanes are for box 2 verification (primer pair 3 + 4/PG176+PG182). c) Western blot of wild type and transgenic line are probed with αHA, confirming the expression of HA tagged PfDNAJb line. α-aldolase is used as loading control. The molecular weight of the HA tagged PfDNAJb (~72 kDa) and Pf aldolase (~40 kDa) are represented on the left hand side of the blot.

5.2.4 PfDNAJb is expressed during all the stages of P. falciparum intraerythrocytic life cycle

To study the expression pattern and subcellular localization of PfDNAJb in generated endogenously HA-tagged PfDNAJb transgenic parasites, we performed an immunofluorescence assay (IFA) using α-HA antibody 3F10. IFA revealed the HA specific signal throughout the asexual stages (ring, trophozoites, and schizonts) of these parasites (Figure 5.3a). IFA also displayed that PfDNAJb is localized in cytosolic compartments of the parasites, with a pattern reminiscent of endoplasmic reticulum (ER) resident proteins, PfGRP170 and PfSec12 (Lee *et al.*, 2008; Kudyba *et al.*, 2019) (Figure 5.4a) .

Given that PfDNAJb comprise of a DNA polymerase B signature motif and predicted NLS motif, it was anticipated to be localized in the nuclear region. But it was not clear in IFA, hence to examine the nuclear localization of this protein, we performed the subcellular fractionation of the parasite pellet to yield cytoplasmic and nuclear fraction. The parasite pellet was obtained from the late trophozoite stage parasite, where the maximum expression of mRNA was reported (López-Barragán *et al.*, 2011)(Figure 5.6b). The cytosolic (α-aldolase) and nuclear (α-H3) markers demonstrated the clear fractionation of the compartments. As shown in Figure 5.4b, immunoblot confirms the presence of PfDNAJb in both cytoplasmic and nuclear fraction in the troohozoite stage harvested lysate.

Surprisingly, an additional protein band with higher molecular weight was observed in the cytoplasmic fraction while the nuclear fraction contains only one band that has same size as the expected PfDNAJb protein (Figure 5.4b). The appearance of two differentially sized protein bands for PfDNAJb suggest potential post-translational modifications.

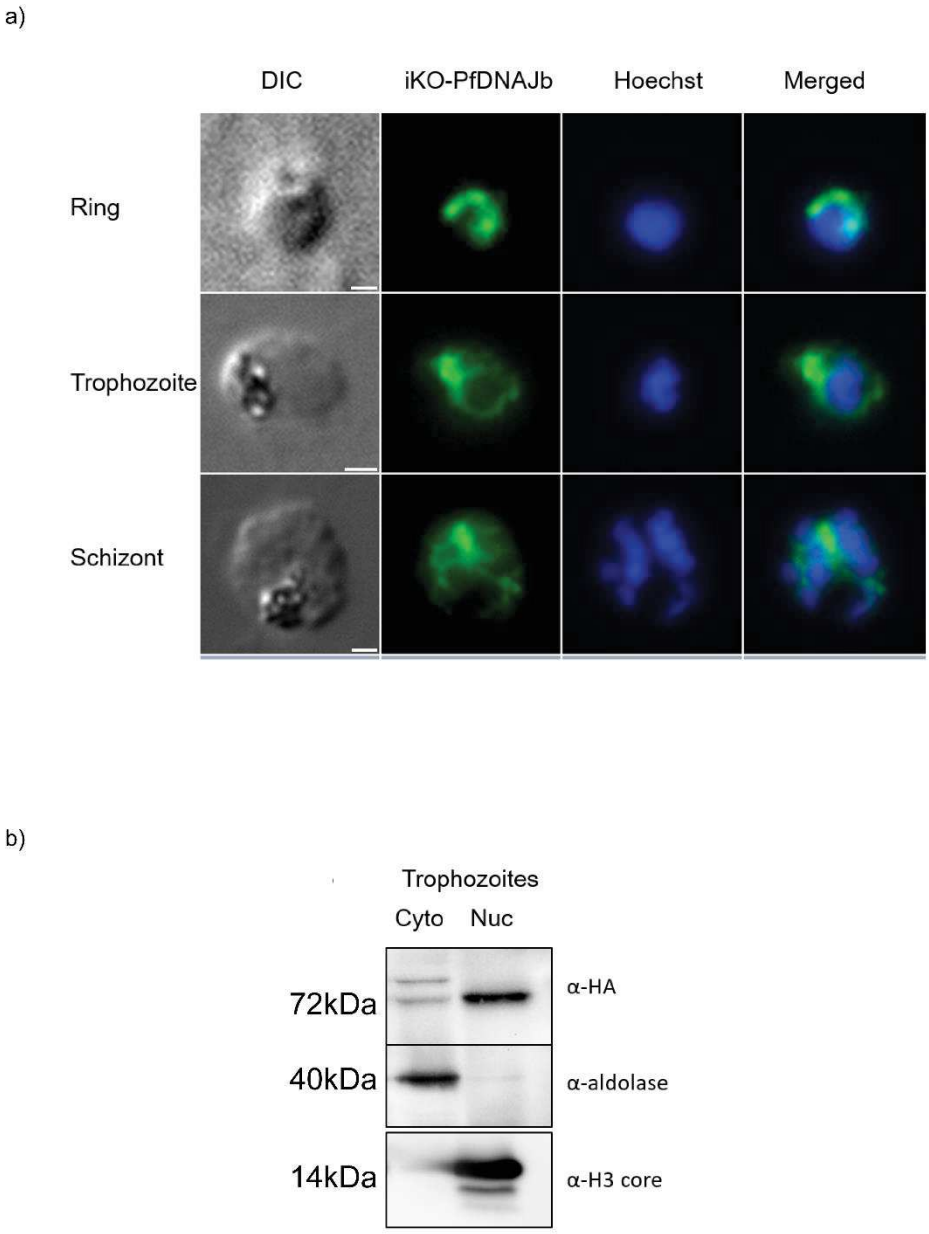


Figure 5.4.Localization of PfDNAJb during the intraerythrocytic developmental cycle of P. falciparum.
a) Expression and localization of PfDNAJb-HA throughout the asexual life cycle are observed by immunofluorescence assay. IFA shows that PfDNAJb is expressed throughout the asexual cycle. PfDNAJb shows a diffuse cytoplasmic pattern ,a reminiscent of ER resident protein. DIC, differential interference contrast. HA tagged PfDNAJb is visualized by α-HA

antibody 3F10. Nuclei is stained with Hoeschst. The scale bar for Ring = 0.5 µm and for troph and schizonts = 1µm. b) Western blot analysis shows the presence of PfDNAJb in both cytoplasmic (cyto) and nuclear (nuc) fractions from mixed trophozoite stage. The quality of fractionalization is checked by α-aldolase (middle panel) and α-H3 (lower panel) antibody as cytoplasmic and nuclear markers, respectively. Equal amount of cytoplasmic and nuclear proteins extractions are loaded on the blot.

5.2.5 PfDNAJb is essential for the viability of the asexual cycle of the P. falciparum

Next, to investigate the role of PfDNAJb during the development of the asexual cycle of the *P. falciparum*, highly synchronized young ring-stage parasites of iKO-PfDNAJb line were treated with DMSO or 20 nM rapamycin for 4h. Treatment with rapamycin induces the Dicre mediated recombination to excise the floxed endogenous iKO-*PfDNAJb* gene (Figure 5.5a). PCR analysis of genomic DNA from rapamycin-treated parasite culture demonstrated that the gene is successfully excised. Moreover, this disruption was observed immediately within 16h rapamycin treatment (Figure 5.5b). More than 90% loss of the protein was observed in the rapamycin-treated sample in western blot analysis with the α-HA antibody (Figure 5.5c). Examination of the parasites by Giemsa stained blood smears indicated that the rapamycin-treated iKO-DNAJb parasites were arrested at the late blood stages (late trophozoites and schizonts) in comparison to the control parasites that grew normally by invading new erythrocytes (Figure 5.5d),

To determine the effect of deletion of Pf*DNAJb* gene on parasite viability, growth phenotype assay was performed. The rapamycin or DMSO treated iKO-PfDNAJb parasites were cultured for two consecutive cycles and their parasitaemia was monitored by standard Giemsa staining. As shown in Figure 5.5e, the rapamycin-treated iKO-PfDNAJb parasites were blocked within the same cycle of the treatment whereas the DMSO treated transgenic and parental lines grew normally. Thus, confirming that the disruption of the PfDNAJb gene exhibited the arresting of the late-stage parasites and ultimately leading to the subsequent death of the parasites. Hence, data shows that the PfDNAJb is essential for the growth of asexual stage parasites.

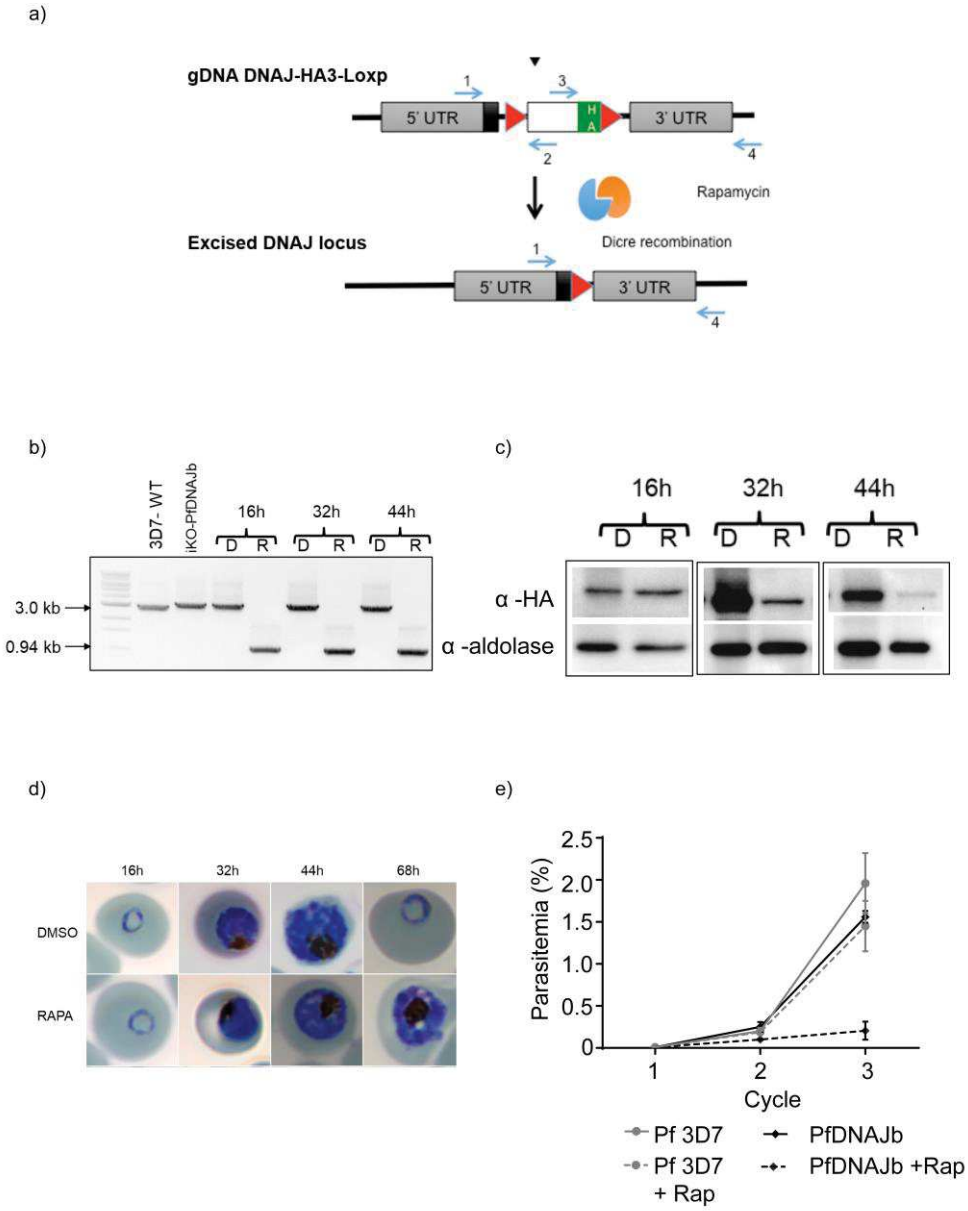


Figure 5.5.. Induction of iKO-PfDNAJb line to generate KO gene of PfDNAJb using rapamycin.
a) Schematic representation of conditional deletion of the PfDNAJb gene when treated with rapamycin. Rapamycin induces the dimerisation of dicre-recombinase subunits, which further interacts with the loxp site and induces recombination to generate truncated gene. b) PCR analysis of rapa and DMSO treated parasites showed the successful recombination of the floxed gene in rapa-treated parasites. Primer pair (1 and 4) for the PfDNAJb gene amplified a 3.0 kb band in the parental line, ~3.2 kb band in DMSO treated transgenic line and a truncated band of ~0.94 kb after excision. c) Western blot analysis of rapa and DMSO treated parasites with α-HA antibody showed the depletion of the PfDNAJb protein in rapamycin treated parasites as compared to the control parasites. α-aldolase used as loading control. Highly synchronized parasites at different stages of the erythrocytic cycle: ring (16h), trophozoites (32h) and schizonts (44h) used for genomic DNA and protein lysate preparation. d) Images of giemsa stained blood smear of iKOPFDNAJb lines, with DMSO or rapa treatment, are captured under 100X oil immersion microscope. The images the abruption of the growth progression in rapa-treated parasites as compared to control parasites during intraerythrocytic developemental cycle. e) Growth phenotype assay of parental line and iKO-DNAJb was performed over two consecutive cycles, with DMSO or rapamycin treatment. The

disruption of the PfDNAJb gene resulted into the inhibition of the growth as compared to the DMSO treated parasites. Means and standard error were displayed for three independent replicates.

5.3 Discussion

Heat shock proteins (Hsp40s) are part of molecular chaperones that interacts with Hsp70s to play a significant role in the establishment and development of *P. falciparum* within their host, especially under heat stress conditions. In addition, they are also involved in proper protein trafficking and remodeling of the host cells (M. Njunge *et al.*, 2013). This family is characterized by conserved J-domain and is classified into four types of Hsp40s. Several studies have been reported on the role of different types of Hsp40s, yet no or less information is available on the largest type of Hsp40 i.e. Type-III proteins.

This is the first study, to our knowledge, where the functional characterization of Type-III Hsp40 member of *P. falciparum* has been investigated. PfDNAJb is an unusual member of Type-III Hsp40s that contains both the J-domain and the thioredoxin domain. Multiple sequence alignment shows that all the residues that are crucial for the interaction between J-domain and Hsp70 are highly conserved in PfDNAJb (Tsai and Douglas, 1996).

In higher eukaryotes, similar type of Hsp40 proteins comprising both the J-domain and the thioredoxin domains has been reported. For instance, human ERdj5 and its mouse homolog JPDI has both J-domain and thioredoxin domain. These proteins interact with BiP (Hsp70) in an ATP-dependent manner through their J-domain (Cunnea *et al.*, 2003; Hosoda *et al.*, 2003). While the thioredoxin domain of ERdj5 is involved in the cleavage of the disulfide bonds present in the misfolded proteins, which further accelerates the ER-associated degradation (ERAD) pathway through its association with BiP protein in the endoplasmic reticulum (Ushioda *et al.*, 2008). In Plasmodium, PfBiP (Hsp70), a homolog of human BiP, is also proposed to be involved in proper protein folding and ER-associated degradation (Pesce and Blatch, 2014). Hence, it could be possible that similar to ERdj5, PfDNAJb might interact with PfBiP and play a role in pathway similar to ERAD.

Interestingly, PfDNAJb is the only member of type-III Hsp40 that contains a conserved (YxDTDS) DNA polymerase family B signature motif. This motif is a conserved region that is found in the DNA polymerases that are involved in accurate replication of DNA (Argos,

1988)·(Copeland and Wang, 1993). The presence of this motif and the fact that PfDNAJb was detected in the nuclear region of *P. falciparum*, suggests that the PfDNAJb could be involved in the replication process of the parasites.

Identification of PfDNAJb as a potential G4-interacting protein raises curiosity on the link between co-chaperones and G4 biology. However, a number of studies have shown that Hsp40s such as Zuo1, CbpA1, and Mdj1, can interact with DNA (Zhang *et al.*, 1992; Bird *et al.*, 2006; Ciesielski *et al.*, 2013). Besides, Zuo1 has been shown to bind with G4 and are proposed to involved in DNA repair or replication in *S. cerevisiae* (Zhang *et al.*, 1992; Silvia Götz, 2017). Thus, it would be interesting to examine another possible role of PfDNAJb as a G4-BP along with their inherent role as co-chaperones in *P. falciparum*.

In conclusion, we have characterized PfDNAJb, a distinct member of type-III Hsp40s, which is essential for the intra-erythrocytic developmental cycle of *P. falciparum*. We have demonstrated that PfDNAJb localizes to nuclear region and in cytosolic region, as reminiscent of ER resident proteins within the asexual life stage of *P. falciparum*. This study postulates that the PfDNAJb may play an important role in a chaperone system and/or DNA replication of these malarial parasites. Besides, they might play an important role as G4 biology, although further studies are required to validate the interaction between PfDNAJb and G4.

Supplementary figures

a

```
                                I          II                  III      IV
PF3D7_0823000        DYYEVLKCKRGDSIQKIKKNYRDLSKQYHPDSNKNCKDCDKKFQEITKAYKTLSDSRLKKA
PBLACG01_0823700     DYYEVLKCKRGDSVQKIKKNYRDLSKQYHPDSNKNCKDCDKKFQEITKAYKTLSDSRLKKA
PBILCG01_00828200    DYYEILKCKRGDSIQKIKKNYRDLSKQYHPDSNKNCNDCDKKFQEITKAYKTLSDSRLKKA
PADL01_0823000       DYYEVLKCKRGDSIQKIKKNYRDLSKQYHPDSNKNCKDCDKKFQEITKAYKTLSDSRLKKA
PGABG01_082300       DYYEVLKCKRGDSIQKIKKNYRDLSKQYHPDSNKNCKDCDKKFQEITKAYKTLSDSRLKKA
PPRFG01_0825300      DYYEVLKCKRGDSIQKIKKNYRDLSKQYHPDSNKNCKDCDKKFQEITKAYKTLSDSRLKKA
PRCDC_0823100        DYYEVLKCKRGDSIQKIKKNYRDLSKQYHPDSNKNCKDCDKKFQEITKAYKTLSDSRLKKA
                     ****:********:**********************:************************
                                                Thioredoxin-like superfamily
PF3D7_0823000        YDNSKGKVLKLIESNSINLNMKNYVDLVENSNDYWIIQIYSDTDSLCLSFSNIWEESFDKY
PBLACG01_0823700     YDNSKGKVLKLIESNSINLNMKNYVDLVENSNDYWIIQIYSDTDSLCLSFSNIWEESFDKY
PBILCG01_00828200    YDNSKGKVLKLIESNSINLNMKNYVDLVENSNDYWIIQIYSDTDSLCLSFSNIWEESFDKY
PADL01_0823000       YDNSKGKVLKLIESNSINLNMKNYVDLVENSNDYWIIQIYSDTDSLCLSFSNIWEESFDKY
PGABG01_082300       YDNSKGKVLKLIESNSINLNMKNYVDLVENSNDYWIIQIYSDTDSLCLSFSNIWEESFDKY
PPRFG01_0825300      YDNSKGKVLKLIESNSINLNMKNYVDLVENSNDYWIIQIYSDTDSLCLSFSNIWEESFDKY
PRCDC_0823100        YDNSKGKVLKLIESNSINLNMKNYVDLVENSNDYWIIQIYSDTDSLCLSFSNIWEESFDKY
                     ************************************************************
PF3D7_0823000        HEYISFGRINISTDKKLIKQKVPFNVKIYPTIFILSPDGTYQLYSNIFNATSKDFQNFI
PBLACG01_0823700     HEYISFGRINISTDKKLIKQKVPFNVKIYPTIFILSPDGTYQLYSNIFNATSKDFQYFI
PBILCG01_00828200    HEYISFGRINISTDKKLIKQKVPFNVKIYPTIFILSPDGTYQLYSNIFNATSKDFQNFI
PADL01_0823000       HDYISFGRINISTDKKLIKQKVPFNVKIYPTIFILSPDGTYQLYSNIFNATSKDFQNFI
PGABG01_082300       HDYISFGRINISTDKKLIKQKVPFNVKIYPTIFILSPDGTYQLYSNIFNATSKDFQNFI
PPRFG01_0825300      HEYISFGRINISTDKKLIKQKVPFNVKIYPTIFILSPDGTYQLYSNIFNATSKDFQNFI
PRCDC_0823100        HEYISFGRINISTDKKLIKQKVPFNVKIYPTIFILSPDGTYQLYSNIFNATSKDFQNFI
                     *:***************************************************  **
```

b

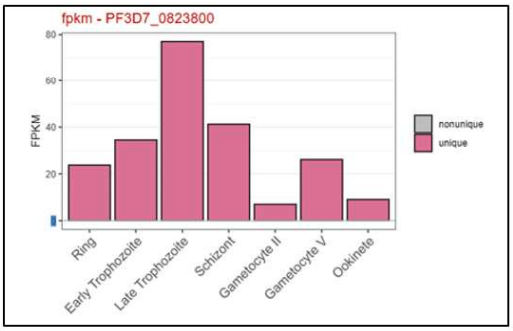


Figure 5.6.Supplementary figures of Chapter 5

a) Sequence alignment displays the alignment of J and thioredoxin domain of PfDNAJb with its Laverania species using MUSCLE algorithm at www.phylogeny.fr. P. falciparum 3D7(Pf3D7_0823000), P. billcollinsi G01 (PBILCG01_00828200), P. blacklocki G01 (PBLACG01_0823700), P. praefalciparum strain G01 (PPRFG01_0825300), P. reichenowi CDC (PRCDC_0823100), P. adleri G01 (PADL01_0823000) and P. gaboni strain G01 (PGABG01_082300).Four helixes of J-domains are marked by bold line (-) and labelled as I, II, III and IV. Thioredoxin domain are highlighted with dashed line (--) and YSDTDSLCL – DNA polymerase B signature motif is underlined with red color. (*) indicates identical amino acid while (;) indicates conserved amino acid residues. b) mRNA expression profile of PfDNAJb during the 48h intraerythrocytic cycle from RNA-Seq data , adapted from PlasmoDB((Aurrecoechea et al., 2008)

Chapter 6

General Discussion and Future perspectives

Chapter 6 General Discussion and Future Perspectives

Given that G4 play a diverse role in different biological processes of a cell, they have been increasingly exploited as potential drug targets in various anticancer and antiviral therapies (Balasubramanian, Hurley and Neidle, 2011; Ruggiero and Richter, 2018). Over the last decades, several parasitologists have started paying attention to understand the role of G4s in the *Plasmodium falciparum*. In these parasites, the G4 are found in the key regulatory regions of the genome (Smargiasso *et al.*, 2009; Bhartiya *et al.*, 2016; Stanton *et al.*, 2016; Harris *et al.*, 2018; Marsico *et al.*, 2019). They are postulated to be involved in antigenic variation of virulence genes and telomere maintenance (De Cian *et al.*, 2008; Smargiasso *et al.*, 2009; Calvo and Wasserman, 2016; Stanton *et al.*, 2016). Besides, several G4 ligands such as TMPyP4, telomestatin, and Quarfloxin are shown to have antimalarial activity, yet their exact mechanism needs to be elucidated (De Cian *et al.*, 2008; Calvo and Wasserman, 2016; Anas *et al.*, 2017; Belmonte-Reche *et al.*, 2018; Harris *et al.*, 2018).

To gain a deeper understanding of the biological role of these G4, my PhD focused on the identification and characterization of G4 binding proteins. In this study, we identified 152 potential G4 binding proteins using two different and complementary unbiased approaches: Y1H assay and DNA pull-down assay, followed by LC-MS/MS. We detected 74 (~ 48%) proteins implicated in G4 related functions including nucleic acid metabolism, transcription, replication, chromatin organization, and translation, this classification is based on gene ontology analysis. 45 (27 %) of unknown proteins are identified. Besides, we detected several proteins related to the cellular metabolism, pathogenesis, protein folding, transport, cellular component assembly, and chaperone system. It could be possible that these proteins are directly or indirectly associated to the selected G4 motif. Although, some chaperones like Zuo1 and CbpA are shown to be associated with G4 (Zhang *et al.*, 1992; Bird *et al.*, 2006). Thus, it would be interesting to examine the role of such proteins in G4 biology of parasites.

Overall, we detected 76 (50.3 %) proteins that are predicted to have nucleic acid binding property based on the GO terms. Four of them (DNA gyrase, DNAJb, Aspartate--tRNA ligase, and FEN1) are found binding exclusively to the WT G4 as compared to the mutated G4. The rest of the candidates were detected to have binding to both WT G4 and mutated G4,

but with different affinities (Table 3.2 and 3.3). For instance, Tudor staphylococcal nuclease and putative RNA binding protein are detected to have more binding to WT G4 whereas Alba family proteins prefers mutated G4 binding over WT G4. Note that, these binding affinities/preferences are assumed based on the obtained data on Y1H assay, unique peptide number and/or MS/MS count; thorough affinities are yet to be measured. Nevertheless, it is plausible that the G4-binding proteins can also interact with non-G4 forming sequences due to the presence of nucleic acid binding regions.

Helicases such as RecQ, and Pif are widely established as G4-binding proteins that unwinds the G4 (Mendoza *et al.*, 2016), however we could not detect such helicases in our study. This was anticipated result as our approaches were primarily based on identification of protein that binds to the G4 rather than destabilizing the formation, so it could be possible that we lost those proteins prior to elution. However, we did detect a DEAD-box containing RNA helicase DDX23 in Y1H assay. Interestingly, the DEAD-box containing RNA helicases such as Dbp2, Ded1, and Mss116 were also identified as G4 binding proteins in yeast using affinity purification and LC-MS/MS (Gao *et al.*, 2019). Gao *et al* demonstrated that these proteins bind to both G4-DNA and G4-RNA *in vitro*. However, the DEAD-box helicases exhibited different mechanism of binding, these DEAD-box helicases bind to the G4-DNA without altering the G4 whereas they destabilizes G4-RNA in an ATP-independent manner. Therefore, it could be possible that the Plasmodium RNA helicase DDX23 also have similar binding mechanism for G4-DNA and G4-RNA.

Strikingly, we also obtained several the RNA binding proteins including 48 ribosomal binding proteins. We believe that the presence of these RNA binding proteins is due to the fact that there is not much difference in the RNA and DNA G4 *in vitro*. This is also the reason that we have such a large number of ribosomal proteins in our dataset. Besides, Several studies on identification of G4 binding proteins have reported the presence of several ribosomal proteins in their screen (Von Hacht *et al.*, 2014; Vlasenok *et al.*, 2018). However, the detailed investigation is required to dissect the binding preference of these proteins.

Apart from using unbiased approach, several *in silico*-based approaches are well established to identify G4 binding proteins. These approaches include homology-based search, use of GO term annotations, literature survey and use of well-characterized motifs that are shared by G4-binding proteins. Recently, Brazda *et al* analyzed the amino acid composition of selected

human G4-binding proteins and revealed that these G4-binding proteins share a 20 amino acid long RG rich motif/ domain (RGRGRGRGGGSGGSGGRGRG) (Brázda *et al.*, 2018). Therefore, I tried to search for this motif in *P. falciparum* using PlasmoDB, a functional genomic database for *Plasmodium* (Aurrecoechea *et al.*, 2008). Consequently, we identified 196 proteins that harbors RGG motif. When compared with our results from unbiased approach, we found 12 proteins common in both the list (Table 6.1). Interestingly, we also found PfGBP2 in this list, hence supporting our choice to select it as one of the candidate. However, this approach can be useful only if supported by other experiments.

Furthermore, the possible interactions between G4 and the proteins could be investigated using biochemical approaches such as EMSA, CD, pull-down or dot blot assays with expressed recombinant proteins. The *in vivo* genome wide association of G4 and protein can be determined using chromatin immune-precipitation experiments like ChIP-seq or ChIP-on-chip. Furthermore, it will be interesting to scrutinize the biological roles of these proteins by generating a conventional or inducible knockout parasite lines. Such studies will shed light on the multifaceted functions of G4 in parasites.

Taken together, the data highlights that several potential G4-BP are identified in our unbiased approaches but other experiments are requisite to validate their accurate binding and affinity to native G4 over mutated G4/other nucleic acid sequence. Furthermore, it could be possible that the obtained candidates bind indirectly to G4 through some protein-protein interactions. Though, it could be legitimate to say that we might also have missed other potential candidates due to limitations of these approaches and the usage of only one G4 motif. Therefore, for future studies, it would be advantageous to improve the approach using several measures. For instance, (i) use of several different G4 motifs as bait, (ii) performing the pull-down experiments with whole parasite lysate, and (iii) use of G4-stabilizing compound such as pyridostatin during Y1H assay.

Potential role of PfGBP2 in *P. falciparum*

In this study, we have showed that indeed PfGBP2 binds to the telomere sequence *in vivo* in *P. falciparum*. PfGBP2 contains two-conserved RNA recognition domain, which are conserved in all *Laverania* species except *P. billcollinsi G01*.

PfGBP2 was identified as a potential G4-binding proteins in our DNA pull-down based approach, however it was detected in both the samples, wt and mutated G4. I believe the reason to obtain this protein in control sample is due to the resemblance of our bait sequence to the telomeric repeat motifs (GGGTTYA, where Y is T or C). Thus, although mutated G4 bait (5' GAGATTTGGGAGGGAGGG 3') does not form any structure, it can still be recognized by the telomere binding protein due to the sequence similarity (Vernick and McCutchan, 1988; Scherf, 1996). As anticipated, we indeed found the part of our bait sequence (GGGAYTTGGG) as one of the most significant motif in our ChIP-Seq dataset.

We here confirmed that PfGBP2 is a *bona fide* G4-binding protein using EMSA and ChIP-Seq. But I think we can still improve the results by adopting few measures in future studies. For EMSA, we can add more controls such as (a) use of unlabeled G4 motif that will compete with the biotinylated G4 to interact with PfGBP2 and (b) use of α-His in the EMSA against His-tagged recombinant protein to observe super-shift. In ChIP-Seq, we confirmed *in vivo* binding of PfGBP2 to the G4FS by scanning the sequence using G4Hunter and comparing the motifs. For future experiments, we can select some of these binding motifs and characterize the formation of G4 using previously described biophysical approaches such as TDS, CD and/ or NMR (Chapter 1). In addition, we can also cross-validate the interaction of PfGBP2 with these sequences using EMSA or CD.

Given that PfGBP2 interacts with G4, next step is to determine the mechanism of this interaction. Based on the obtained data, we can speculate two possibilities; it can either act as a G4-stabilizer or G4-inducer. To address this, we can incubate different concentration of proteins with folded G4-DNA or unfolded G4-DNA to perform EMSA. Then, by comparing the binding affinity of the protein to folded G4-DNA with the unfolded G4-DNA, we can deduce the nature of the interaction (Muniyappa, Anuradha and Byers, 2000).

This leads us to our next query: how does PfGBP2 recognize the nucleic acid sequence? PfGBP2 contains two RRM domains separated with a hinge region, while their functions are yet to be elucidated. RRM domains are shown to interact with nucleic acid sequence and proteins (Maris, Dominguez and Allain, 2005). On the other hand, hinge region has been implicated in nucleocytoplasmic shuttling and protein dimerization (Cartegni *et al.*, 1996; Fan and Steitz, 1998). It would be interesting to determine the roles of these domains in PfGBP2. On analyzing the sequence, we found that the hinge region of the PfGBP2 contains

two RGG motifs, which were shown to interact with G4 (Mcrae, Booy, Padilla-meier, *et al.*, 2017; Brázda *et al.*, 2018). Further experiments will be required to examine the role of this motif in the G4 binding.

As discussed earlier, telomere binding protein, PfSir2B is shown to regulate the transcription of *var* genes in *P. falciparum* but doesn't affect the length of the telomeres (Tonkin et al., 2009). As seen in PfGBP2, we could not detect any change in the telomere length in the absence of protein but we did detect slight derepression in the expression of some subset of *var* genes. These results were obtained by culturing PfGBP2 lacking parasites for 1 months, so in order to examine whether keeping them for longer time in culture has any effect. Thus, we are still culturing the parasites and will examine the change in expression of *var* genes.

Potential role of PfDNAJb in *P. falciparum*

In this study, we revealed that the PfDNAJb is a distinct member of type III Hsp40 proteins with an unusual combination of two different domains, J domain and thioredoxin domain. PfDNAJb is found to be essential for the intra-erythrocytic development of *P. falciparum* as the loss of protein exhibited a growth arrest at the late stages, which ultimately led to the death of the parasites.

PfDNAJb was identified as potential G4 binding protein in our screen, but unfortunately, I could not finish the validation of their interaction with G4. Thus, one of the future prospects of my thesis is to validate the PfDNJAb and G4 interaction using EMSA and ChIP-Seq. As described earlier, we were unable to express the N-terminal His tagged recombinant protein, which could be due to the toxicity associated the transmembrane proteins. Thus, for future attempts to express recombinant protein, it would be beneficial to adopt different strategies for the expression of transmembrane containing protein such as (a) use of improved bacterial strains like C41 and C43 (Miroux and Walker, 1996), (b) use of large size affinity tags such as GST and MBP (Young, Britton and Robinson, 2012) , (c) use of eukaryotic organism such as insects or HEK/CHO cells.

On the other hand, generated endogenously HA tagged iKO-PfDNAJb can be used for ChIP-Seq to map the genome-wide occupancy of PfDNAJb in *Plasmodium* genome. Although I have already initiated this experiment, but due to the constraint of time, I will not be able to

finish the experiment. Nevertheless, with the help of ChIP-seq analysis, we will be able to identify the DNA binding sites of the protein, which will further assist in deducing the biological roles of the protein.

Besides, we observed that PfDNAJb exists in two different forms having different molecular weight, which could be due to some post-translational modifications (PTM). The unmodified form with the estimated size of protein is present in both cytosolic and nuclear region whereas the likely modified form (with higher molecular weight) is present exclusively in the cytosolic fraction of the late trophozoite stage. Based on this observation, we can say that the PTM might be ascribing the exclusive cytosolic-based function of the protein. In *P. falciparum*, phospho-proteomic studies have indicated the presence of several sites of phosphorylation in this protein including serine residues (at position 6, 9,13,24) and tyrosine residue (at position 10) (Treeck *et al.*, 2011; Lasonder *et al.*, 2012). However, the difference of molecular weight observed on the immunoblot is more than 10-15 kDa, suggesting the likelihood of presence of other PTMs that can increase the molecular weight of the protein. Therefore, we searched for PTM in PfDNAJb using ScanProsite tool (de Castro *et al.*, 2006). As foreseen, we did detect several sites of N-glycosylation and lipidation. Although the existence of glycosylation modification is debatable in *P.falciparum* (Doerig *et al.*, 2015), but the presence of consensus sequence for N-glycosylation (NXS/T, where x≠ P) and our previous observation i.e. increased molecular weight of the PfDNAJb, strengthens the high probability of existence of this modification. On the other hand, the presence of predicted transmembrane domains in PfDNAJb favors the lipidation (Nadolski and Linder, 2007). However, further experiments will be required to identify the bona fide PTM and their significant role in the function of PfDNAJb. Approaches such as Mass spectroscopy and western blot analysis with PTM specific antibodies can be employed to determine these PTMs.

Recently, Kudyba *et al* have shown that the PfGRP170, ER localized Hsp110, is essential for the asexual growth of the *P. falciparum* (Kudyba *et al.*, 2019). Using a combination of mass spectroscopy based approaches such as Co-IP and BirA mediated proximity ligation assay, they identified the potential interactors of PfGRP170. Moreover, they have also demonstrated that the PfGRP170 interacts with PfBiP (Hsp70). Intriguingly, while scanning their list of PfGRP170 interactors, we also found PfDNAJb as one of the potential interactors in both the approaches. Furthermore, we showed that the deletion of PfDNAJb results in growth arrest at

the late asexual stages leading to the parasite death, which is similar to the lethal growth defect shown by the disruption of PfGRP170 (Kudyba *et al.*, 2019). Thus, suggesting that perhaps the PfDNAJb acts in concert with PfGRP170 in multiple functions.

In eukaryotic and prokaryotic chaperone systems, it is established that Hsp70/DnaK, Hsp40/DnaJ, and Hsp110/GrpE proteins form a complex, where they collaborate with each other to refold the misfolded proteins in an ATP dependent manner (Schroder *et al.*, 1993; Sharma *et al.*, 2010; Shorter, 2011). Hence, it could be possible that PfDNAJb acts in concert with PfBiP and PfGRP170 to establish a similar disaggregase system. To investigate this collaborative function, first it is important to confirm the interaction between these three proteins. Kudyba *et al* has already demonstrated that PfGRP170 interacts with PfBiP and their data showed that PfDNAJb can also be a potential partner of PfGRP170 (Kudyba *et al.*, 2019). Besides, it would be interesting to examine whether PfDNJAb can also interact with PfGRP170 and/or PfBiP. To this end, we can perform similar mass spectroscopy-based approach (Co-IP and BirA mediated proximity ligation assay) as mentioned in Kudyba *et al* study, to fish out the proteins that interacts with PfDNAJb. Hence, it will further help in unraveling the multifaceted role of PfDNAJb in *P. falciparum*.

References

References

Acharya, P., Kumar, R. and Tatu, U. (2007) 'Chaperoning a cellular upheaval in malaria: Heat shock proteins in Plasmodium falciparum', *Molecular and Biochemical Parasitology*, 153(2), pp. 85–94. doi: 10.1016/j.molbiopara.2007.01.009.

Adrian, M., Heddi, B. and Phan, A. T. (2012) 'NMR spectroscopy of G-quadruplexes', *Methods*, pp. 11–24. doi: 10.1016/j.ymeth.2012.05.003.

Aiken, C. and Trono, D. (1995) 'Nef stimulates human immunodeficiency virus type 1 proviral DNA synthesis.', *Journal of virology*, 69(8), pp. 5048–56.

Aly, A. S. I., Vaughan, A. M. and Kappe, S. H. I. (2009) 'Malaria Parasite Development in the Mosquito and Infection of the Mammalian Host', *Annual Review of Genetics*, 63, pp. 195–221. doi: 10.1038/jid.2014.371.

Amato, J. *et al.* (2018) 'HMGB1 binds to the KRAS promoter G-quadruplex: A new player in oncogene transcriptional regulation?', *Chemical Communications*, 54(68), pp. 9442–9445. doi: 10.1039/c8cc03614d.

Anas, M. *et al.* (2017) 'Investigating Pharmacological Targeting of G-Quadruplexes in the Human Malaria Parasite', *Biochemistry*, 56(51), pp. 6691–6699. doi: 10.1021/acs.biochem.7b00964.

Aphasizhev, R. and Aphasizheva, I. (2014) 'Mitochondrial RNA editing in trypanosomes: small RNAs in control', *Biochimie*, 100, pp. 125–131. doi: 10.1016/j.physbeh.2017.03.040.

Argos, P. (1988) 'A sequence motif in many polymerases', *Nucleic Acids Research*, 16(21), pp. 9909–9916. doi: 10.1093/nar/16.21.9909.

Arora, A. and Suess, B. (2011) 'An RNA G-quadruplex in the 3′ UTR of the proto-oncogene PIM1 represses translation', *RNA Biology*, 8(5), pp. 1–5. doi: 10.4161/rna.8.5.16038.

Artusi, S. *et al.* (2015) 'The Herpes Simplex Virus-1 genome contains multiple clusters of repeated G-quadruplex: Implications for the antiviral activity of a G-quadruplex ligand',

Antiviral Research. Elsevier B.V., 118, pp. 123–131. doi: 10.1016/j.antiviral.2015.03.016.

Artusi, S. *et al.* (2016) 'Visualization of DNA G-quadruplexes in herpes simplex virus 1-infected cells', *Nucleic Acids Research*, 44(21), pp. 10343–10353. doi: 10.1093/nar/gkw968.

Ashley, E. A., Pyae Phyo, A. and Woodrow, C. J. (2018) 'Malaria', *The Lancet*, 391(10130), pp. 1608–1621. doi: 10.1016/S0140-6736(18)30324-6.

Aurrecoechea, C. *et al.* (2008) 'PlasmoDB: a functional genomic database for malaria parasites', *Nucleic Acids Research*, 37, pp. 539–543. doi: 10.1093/nar/gkn814.

Awandare, G. A. *et al.* (2018) 'Plasmodium falciparum strains spontaneously switch invasion phenotype in suspension culture', *Scientific Reports*, 8(1), pp. 1–10. doi: 10.1038/s41598-018-24218-0.

Bai, L. and Morozov, A. V (2010) 'Gene regulation by nucleosome positioning', *Trends in Genetics*, pp. 476–483. doi: 10.1016/j.tig.2010.08.003.

Balaji, S. *et al.* (2005) 'Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains', *Nucleic Acids Research*, 33(13), pp. 3994–4006. doi: 10.1093/nar/gki709.

Balasubramanian, S., Hurley, L. H. and Neidle, S. (2011) 'Targeting G-quadruplexes in gene promoters: A novel anticancer strategy?', *Nature Reviews Drug Discovery*, 10(4), pp. 261–275. doi: 10.1038/nrd3428.

Banecki, B. *et al.* (1996) 'Structure-function analysis of the zinc finger region of the DnaJ molecular chaperone', *Journal of Biological Chemistry*, 271(25), pp. 14840–14848. doi: 10.1074/jbc.271.25.14840.

Bang, I. (1910) 'Untersuchungen über die Guanylsäure', *Biochem Z.*, 26, pp. 293–231.

Banumathy, G. *et al.* (2003) 'Heat shock protein 90 function is essential for Plasmodium falciparum growth in human erythrocytes', *Journal of Biological Chemistry*, 278(20), pp. 18336–18345. doi: 10.1074/jbc.M211309200.

Bartoloni, A. and Zammarchi, L. (2012) 'Clinical aspects of uncomplicated and severe malaria', *Mediterranean Journal of Hematology and Infectious Diseases*, 4(1). doi: 10.4084/MJHID.2012.026.

Beaume, N. *et al.* (2013) 'Genome-wide study predicts promoter-G4 DNA motifs regulate selective functions in bacteria: Radioresistance of D. radiodurans involves G4 DNA-mediated regulation', *Nucleic Acids Research*, 41(1), pp. 76–89. doi: 10.1093/nar/gks1071.

Bedrat, A., Lacroix, L. and Mergny, J.-L. (2016) 'Re-evaluation of G-quadruplex propensity with G4Hunter', *Nucleic Acids Research*, 44(4), pp. 1746–1759. doi: 10.1093/nar/gkw006.

Belmonte-Reche, E. *et al.* (2018) 'G-Quadruplex Identification in the Genome of Protozoan Parasites Points to Naphthalene Diimide Ligands as New Antiparasitic Agents', *Journal of Medicinal Chemistry*, 61(3), pp. 1231–1240. doi: 10.1021/acs.jmedchem.7b01672.

Belotserkovskii, B. P. *et al.* (2010) 'Mechanisms and implications of transcription blockage by guanine-rich DNA sequences', *Proceedings of the National Academy of Sciences of the United States of America*, 107(29), pp. 12816–12821. doi: 10.1073/pnas.1007580107.

Belotserkovskii, B. P. *et al.* (2013) 'Transcription blockage by homopurine DNA sequences: Role of sequence composition and single-strand breaks', *Nucleic Acids Research*, 41(3), pp. 1817–1828. doi: 10.1093/nar/gks1333.

Bertschi, N. L. *et al.* (2017) 'Malaria parasites possess a telomere repeat-binding protein that shares ancestry with transcription factor IIIA', *Nature Microbiology*, 2(March), pp. 1–12. doi: 10.1038/nmicrobiol.2017.33.

Bhartiya, D. *et al.* (2016) 'Genome-wide regulatory dynamics of G-quadruplexes in human malaria parasite Plasmodium falciparum', *Genomics*, 108(5–6), pp. 224–231. doi: 10.1016/j.ygeno.2016.10.004.

Bhattacharyya, D., Arachchilage, G. M. and Basu, S. (2016) 'Metal cations in G-quadruplex folding and stability', *Frontiers in Chemistry*, p. 38. doi: 10.3389/fchem.2016.00038.

Biffi, G. *et al.* (2013) 'Quantitative visualization of DNA G-quadruplex structures in human

cells', *Nature Chemistry*. Nature Publishing Group, 5(3), pp. 182–186. doi: 10.1038/nchem.1548.

Bird, J. G. *et al.* (2006) 'Functional Analysis of CbpA, a DnaJ Homolog and Nucleoid-associated DNA-binding Protein', *jbc*, 281(45), pp. 34349–34356. doi: 10.1074/jbc.M603365200.

Biswas, B. *et al.* (2016) 'Genome-wide analysis of G-quadruplexes in herpesvirus genomes', *BMC Genomics*, 17(1), p. 949. doi: 10.1186/s12864-016-3282-1.

Biswas, S. and Sharma, Y. D. (1994) 'Enhanced expression of Plasmodium falciparum heat shock protein PFHSP70-I at higher temperatures and parasite survival', *FEMS Microbiology Letters*, 124(3), pp. 425–429. doi: 10.1111/j.1574-6968.1994.tb07319.x.

Bonnal, S. *et al.* (2003) 'A single internal ribosome entry site containing a G quartet RNA structure drives fibroblast growth factor 2 gene expression at four alternative translation initiation codons', *Journal of Biological Chemistry*, 278(41), pp. 39330–39336. doi: 10.1074/jbc.M305580200.

Bonnefoy, S. *et al.* (1994) 'Molecular characterization of the heat shock protein 90 gene of the human malaria parasite Plasmodium falciparum', *Molecular and Biochemical Parasitology*, 67(1), pp. 157–170. doi: 10.1016/0166-6851(94)90105-8.

Borgognone, M., Armas, P. and Calcaterra, N. B. (2010) 'Cellular nucleic-acid-binding protein, a transcriptional enhancer of c-Myc, promotes the formation of parallel G-quadruplexes', *Biochemical Journal*, 428(3), pp. 491–498. doi: 10.1042/BJ20100038.

Botha, M. *et al.* (2011) 'Plasmodium falciparum encodes a single cytosolic type i Hsp40 that functionally interacts with Hsp70 and is upregulated by heat shock', *Cell Stress and Chaperones*, 16(4), pp. 389–401. doi: 10.1007/s12192-010-0250-6.

Botha, M., Pesce, E. R. and Blatch, G. L. (2007) 'The Hsp40 proteins of Plasmodium falciparum and other apicomplexa: Regulating chaperone power in the parasite and the host', *International Journal of Biochemistry and Cell Biology*, 39(10), pp. 1781–1803. doi: 10.1016/j.biocel.2007.02.011.

Bozdech, Z. *et al.* (2003) 'The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum', *PLoS Biology*, 1(1), pp. 85–100. doi: 10.1371/journal.pbio.0000005.

Brancucci, N. M. B. *et al.* (2014) 'Heterochromatin protein 1 secures survival and transmission of malaria parasites', *Cell Host and Microbe*, 16(2), pp. 165–176. doi: 10.1016/j.chom.2014.07.004.

Brázda, V. *et al.* (2014) 'DNA and RNA quadruplex-binding proteins', *International Journal of Molecular Sciences*, 15(10), pp. 17493–17517. doi: 10.3390/ijms151017493.

Brázda, V. *et al.* (2018) 'The amino acid composition of quadruplex binding proteins reveals a shared motif and predicts new potential quadruplex interactors', *Molecules*, 23(9). doi: 10.3390/molecules23092341.

Brázda, V. *et al.* (2019) 'G4Hunter web application: A web server for G-quadruplex prediction', *Bioinformatics*, 35(18), pp. 3493–3495. doi: 10.1093/bioinformatics/btz087.

Brosh Jr, R. M. (2013) 'DNA helicases involved in DNA repair and their roles in cancer', *Nature Reviews Cancer*, 13(8), pp. 542–558. doi: 10.1038/nrc3560.DNA.

Broxson, C., Beckett, J. and Tornaletti, S. (2011) 'Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human c-myb gene', *Biochemistry*, 50(19), pp. 4162–4172. doi: 10.1021/bi2002136.

Bugaut, A. and Balasubramanian, S. (2012) '5′-UTR RNA G-quadruplexes: Translation regulation and targeting', *Nucleic Acids Research*, 40(11), pp. 4727–4741. doi: 10.1093/nar/gks068.

Buket, O. *et al.* (2019) 'Electrophoretic Mobility Shift Assay and Dimethyl Sulfate Footprinting for Characterization of G-Quadruplexes and G-Quadruplex-Protein Complexes', in *G-Quadruplex Nucleic Acids*, pp. 201–222.

Bunnik, E. M. *et al.* (2013) 'Polysome profiling reveals translational control of gene expression in the human malaria parasite Plasmodium falciparum', *Genome Biology*, 14(11),

pp. 1–18. doi: 10.1186/gb-2013-14-11-r128.

Burge, S. *et al.* (2006) 'Quadruplex DNA: Sequence, topology and structure', *Nucleic Acids Research*, 34(19), pp. 5402–5415. doi: 10.1093/nar/gkl655.

Cahoon, L. A. *et al.* (2013) 'Neisseria gonorrhoeae RecQ helicase HRDC domains are essential for efficient binding and unwinding of the pilE guanine quartet structure required for pilin antigenic variation', *Journal of Bacteriology*, 195(10), pp. 2255–2261. doi: 10.1128/JB.02217-12.

Cahoon, L. A. and Seifert, H. S. (2009) 'An alternative DNA structure is necessary for pilin antigenic variation in Neisseria gonorrhoeae', *Science*, 325(5941), pp. 764–767. doi: 10.1126/science.1175653.

Cairns, M. *et al.* (2012) 'Estimating the potential public health impact of seasonal malaria chemoprevention in African children', *Nature Communications*. Nature Publishing Group, 3, pp. 1–9. doi: 10.1038/ncomms1879.

Callebaut, I. *et al.* (2005) 'Prediction of the general transcription factors associated with RNA polymerase II in Plasmodium falciparum: Conserved features and differences relative to other eukaryotes', *BMC Genomics*, 6, pp. 1–20. doi: 10.1186/1471-2164-6-100.

Calvo, E. P. and Wasserman, M. (2015) 'PfGBP : A Plasmodium falciparum telomere binding protein', *Quimica Organica y Bioq*.

Calvo, E. P. and Wasserman, M. (2016) 'G-Quadruplex ligands: Potent inhibitors of telomerase activity and cell proliferation in Plasmodium falciparum', *Molecular & Biochemical Parasitology*, 207(1), pp. 33–38. doi: 10.1016/j.molbiopara.2016.05.009.

Caro, F. *et al.* (2014) 'Genome-wide regulatory dynamics of translation in the Plasmodium falciparum asexual blood stages', *eLife*, 3, pp. 1–24. doi: 10.7554/elife.04106.

Cartegni, L. *et al.* (1996) 'hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins', *Journal of Molecular Biology*, 259(3), pp. 337–348. doi: 10.1006/jmbi.1996.0324.

Casta, L. J. *et al.* (2008) 'Expression and Biochemical Characterization of the Plasmodium falciparum DNA repair enzyme, FLAP endonuclease-1 (PfFEN-1)', *Mol Biochem Parasitol*, 157(1), pp. 1–12. doi: 10.1016/j.molbiopara.2007.08.008.EXPRESSION.

Castillo Bosch, P. *et al.* (2014) ' FANCJ promotes DNA synthesis through G-quadruplex structures ', *The EMBO Journal*, 33(21), pp. 2521–2533. doi: 10.15252/embj.201488663.

de Castro, E. *et al.* (2006) 'ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins', *Nucleic Acids Research*, 34(WEB. SERV. ISS.), pp. 362–365. doi: 10.1093/nar/gkl124.

Chêne, A. *et al.* (2012) 'PfAlbas constitute a new eukaryotic DNA/RNA-binding protein family in malaria parasites', *Nucleic Acids Research*, 40(7), pp. 3066–3077. doi: 10.1093/nar/gkr1215.

Cheung, I. *et al.* (2002) 'Disruption of dog-1 in Caenorhabditis elegans triggers deletions upstream of guanine-rich DNA', *Nature Genetics*, 31(4), pp. 405–409. doi: 10.1038/ng928.

De Cian, A. *et al.* (2008) 'Plasmodium telomeric sequences: structure, stability and quadruplex targeting by small compounds.', *Chembiochem : a European journal of chemical biology*, 9(16), pp. 2730–2739. doi: 10.1002/cbic.200800330.

Ciesielski, G. L. *et al.* (2013) 'Nucleoid localization of Hsp40 Mdj1 is important for its function in maintenance of mitochondrial DNA', *Biochimica et Biophysica Acta - Molecular Cell Research*. The Authors, 1833(10), pp. 2233–2243. doi: 10.1016/j.bbamcr.2013.05.012.

Claessens, A. *et al.* (2014) 'Generation of Antigenic Diversity in Plasmodium falciparum by Structured Rearrangement of Var Genes During Mitosis', *PLoS Genetics*, 10(12), p. e1004812. doi: 10.1371/journal.pgen.1004812.

Claessens, A. *et al.* (2018) 'RecQ helicases in the malaria parasite Plasmodium falciparum affect genome stability, gene expression patterns and DNA replication dynamics', *PLoS Genetics*, 14(7). doi: 10.1371/journal.pgen.1007490.

Clark, D. W. *et al.* (2012) 'Promoter G-quadruplex sequences are targets for base oxidation

and strand cleavage during hypoxia-induced transcription', *Free Radical Biology and Medicine*, 53(1), pp. 51–59. doi: 10.1016/j.freeradbiomed.2012.04.024.Promoter.

Clontech (2010) 'Matchmaker ™ Gold Yeast One-Hybrid Library Screening System', *Matchmaker ™ Gold Yeast One-Hybrid Library Screening System*, (November 2012), pp. 1–36.

Coleman, B. I. *et al.* (2014) 'A Plasmodium falciparum Histone Deacetylase Regulates Antigenic Variation and Gametocyte Conversion', *Cell Host and Microbe*, 16(2), pp. 177–186. doi: 10.1016/j.chom.2014.06.014.A.

Copeland, W. C. and Wang, T. S. F. (1993) 'Mutational Analysis of the Human DNA Polymerase a', *Journal of Biological Chemistry*, 268(15), pp. 11041–11049.

Cortés, A. (2008) 'Switching Plasmodium falciparum genes on and off for erythrocyte invasion', *Trends in Parasitology*, 24(11), pp. 517–524. doi: 10.1016/j.pt.2008.08.005.

Coulson, R. M. R., Hall, N. and Ouzounis, C. A. (2004) 'Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum', *Genome Research*, 14(8), pp. 1548–1554. doi: 10.1101/gr.2218604.

Cowman, A. F. *et al.* (2016) 'Malaria: Biology and Disease', *Cell*. Elsevier Inc., 167(3), pp. 610–624. doi: 10.1016/j.cell.2016.07.055.

Cowman, A. F. *et al.* (2017) 'The Molecular Basis of Erythrocyte Invasion by Malaria Parasites', *Cell Host and Microbe*. Elsevier Inc., 22(2), pp. 232–245. doi: 10.1016/j.chom.2017.07.003.

Cox, J. *et al.* (2011) 'Andromeda: A peptide search engine integrated into the MaxQuant environment', *Journal of Proteome Research*, 10(4), pp. 1794–1805. doi: 10.1021/pr101065j.

Cox, J. and Mann, M. (2008) 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nature Biotechnology*, 26(12), pp. 1367–1372. doi: 10.1038/nbt.1511.

Cunnea, P. M. *et al.* (2003) 'ERdj5, an endoplasmic reticulum (ER)-resident protein containing DnaJ and thioredoxin domains, is expressed in secretory cells or following ER stress', *Journal of Biological Chemistry*, 278(2), pp. 1059–1066. doi: 10.1074/jbc.M206995200.

Das, A. *et al.* (1997) 'Molecular characterization and ultrastructural localization of Plasmodium falciparum Hsp 60', *Molecular and Biochemical Parasitology*, 88(1–2), pp. 95–104. doi: 10.1016/S0166-6851(97)00081-9.

Demkovičová, E. *et al.* (2017) 'Telomeric G-Quadruplexes: From Human to Tetrahymena Repeats', *Journal of Nucleic Acids*, 2017. doi: 10.1155/2017/9170371.

Doerig, C. *et al.* (2015) 'Post-translational protein modifications in malaria parasites', *Nature Reviews Microbiology*, 13(3), pp. 160–172. doi: 10.1038/nrmicro3402.

van Dooren, G. G. and Striepen, B. (2013) 'The Algal Past and Parasite Present of the Apicoplast', *Annual Review of Microbiology*, 67(1), pp. 271–289. doi: 10.1146/annurev-micro-092412-155741.

Drygin, D. *et al.* (2009) 'Anticancer activity of CX-3543: A direct inhibitor of rRNA biogenesis', *Cancer Research*, 69(19), pp. 7653–7661. doi: 10.1158/0008-5472.CAN-09-1304.

Duraisingh, M. T. *et al.* (2005) 'Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum', *Cell*, 121(1), pp. 13–24. doi: 10.1016/j.cell.2005.01.036.

Duraisingh, M. T. and Skillman, K. M. (2018) 'Epigenetic Variation and Regulation in Malaria Parasites', *Annual Review of Microbiology*, 72(1), pp. 355–375. doi: 10.1146/annurev-micro-090817-062722.

Dzikowski, R., Frank, M. and Deitsch, K. (2006) 'Mutually exclusive expression of virulence genes by malaria parasites is regulated independently of antigen production', *PLoS Pathogens*, 2(3), pp. 0184–0194. doi: 10.1371/journal.ppat.0020022.

Etzioni, S. *et al.* (2005) 'Homodimeric MyoD preferentially binds tetraplex structures of regulatory sequences of muscle-specific genes', *Journal of Biological Chemistry*, 280(29), pp. 26805–26812. doi: 10.1074/jbc.M500820200.

Fan, X. C. and Steitz, J. A. (1998) 'HNS, a nuclear-cytoplasmic shuttling sequence in HuR (nuclear localization‒RNA degradation‒nuclear export)', *Biochemistry*, 95(December), pp. 15293–15298. doi: 10.1073/pnas.95.26.15293.

Figueiredo, L. M. *et al.* (2002) 'A central role for Plasmodium falciparum subtelomeric regions in spatial positioning and telomere length regulation', *EMBO Journal*, 21(4), pp. 815–824. doi: 10.1093/emboj/21.4.815.

Figueiredo, L. and Scherf, A. (2005) 'Plasmodium telomeres and telomerase: The usual actors in an unusual scenario', *Chromosome Research*, 13(5), pp. 517–524. doi: 10.1007/s10577-005-0996-3.

Flueck, C. *et al.* (2009) 'Plasmodium falciparum heterochromatin protein 1 marks genomic loci linked to phenotypic variation of exported virulence factors', *PLoS Pathogens*, 5(9). doi: 10.1371/journal.ppat.1000569.

Frasson, I., Nadai, M. and Richter, S. N. (2019) 'Conserved G-quadruplexes regulate the immediate early promoters of human alphaherpesviruses', *Molecules*, 24(13). doi: 10.3390/molecules24132375.

Freitas, L. H. *et al.* (2005) 'Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites', *Cell*, 121(1), pp. 25–36. doi: 10.1016/j.cell.2005.01.037.

Gao, J. *et al.* (2019) 'DEAD-box RNA helicases Dbp2, Ded1 and Mss116 bind to G-quadruplex nucleic acids and destabilize G-quadruplex RNA', *Chemical Communications*. Royal Society of Chemistry, 55(31), pp. 4467–4470. doi: 10.1039/c8cc10091h.

Gardner, M. J. *et al.* (2002) 'Genome sequence of the human malaria parasite Plasmodium falciparum', *Nature*, 419(6906), pp. 3–9. doi: 10.1038/nature01097.Genome.

Gaur, D. *et al.* (2006) 'Upregulation of expression of the reticulocyte homology gene 4 in the Plasmodium falciparum clone Dd2 is associated with a switch in the erythrocyte invasion pathway', *Molecular and Biochemical Parasitology*, 145(2), pp. 205–215. doi: 10.1016/j.molbiopara.2005.10.004.

Gazanion, E. *et al.* (2020) 'Genome wide distribution of G-quadruplexes and their impact on gene expression in malaria parasites', *Unpublished*.

Gellert, M., Lipsett, M. N. and Davies, D. R. (1962) 'Helix formation by guanylic acid', *PNAS*, 48, pp. 2013–2018. doi: 10.1073/pnas.48.12.2013.

Genevaux, P., Georgopoulos, C. and Kelley, W. L. (2002) 'Scanning Mutagenesis Identifies Amino Acid Residues Essential for the in Vivo activity of th Escherichia coli DnaJ (Hsp40) J-domain.pdf', *Genetics*, 1053(November), pp. 1045–1053.

Giraldo, R. and Rhodes, D. (1994) 'The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA.', *The EMBO Journal*, 13(10), pp. 2411–2420. doi: 10.1002/j.1460-2075.1994.tb06526.x.

González, V. *et al.* (2009a) 'Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein', *Journal of Biological Chemistry*, 284(35), pp. 23622–23635. doi: 10.1074/jbc.M109.018028.

González, V. *et al.* (2009b) 'Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein', *Journal of Biological Chemistry*, 284(35), pp. 23622–23635. doi: 10.1074/jbc.M109.018028.

Gordon, H. *et al.* (2013) 'Depletion of hnRNP A2/B1 overrides the nuclear retention of the HIV-1 genomic RNA', *RNA Biology*, 10(11), pp. 1714–1725. doi: 10.4161/rna.26542.

Götz, S. *et al.* (2019) 'A novel G-quadruplex binding protein in yeast-Slx9', *Molecules*, 24(9), pp. 14–18. doi: 10.3390/molecules24091774.

Gray, N. K. and Hentze, M. W. (1994) 'Regulation of protein synthesis by mRNA structure', *Molecular Biology Reports*, 19(3), pp. 195–200. doi: 10.1007/BF00986961.

Guédin, A. *et al.* (2010) 'How long is too long? Effects of loop size on G-quadruplex stability', *Nucleic Acids Research*, 38(21), pp. 7858–7868. doi: 10.1093/nar/gkq639.

Von Hacht, A. *et al.* (2014) 'Identification and characterization of RNA guanine-quadruplex binding proteins', *Nucleic Acids Research*, 42(10), pp. 6630–6644. doi: 10.1093/nar/gku290.

Halder, K., Halder, R. and Chowdhury, S. (2009) 'Genome-wide analysis predicts DNA structural motifs as nucleosome exclusion signals', *Molecular BioSystems*, 5(12), pp. 1703–1712. doi: 10.1039/b905132e.

Halder, R. *et al.* (2012) 'Bisquinolinium compounds induce quadruplex-specific transcriptome changes in HeLa S3 cell lines', *BMC Res Notes*, 5, p. 138. doi: 10.1186/1756-0500-5-138.

Han, W. and Christen, P. (2003) 'Mechanism of the targeting action of DnaJ in the DnaK molecular chaperone system', *Journal of Biological Chemistry*, 278(21), pp. 19038–19043. doi: 10.1074/jbc.M300756200.

Hänsel-Hertsch, R. *et al.* (2016) 'G-quadruplex structures mark human regulatory chromatin', *Nature Genetics*, 48(10), pp. 1267–1272. doi: 10.1038/ng.3662.

Harris, L. M. *et al.* (2018) 'G-quadruplex DNA motifs in the malaria parasite plasmodium falciparum and their potential as novel antimalarial drug targets', *Antimicrobial Agents and Chemotherapy*, 62(3). doi: 10.1128/AAC.01828-17.

Harris, L. M. and Merrick, C. J. (2015) 'G-Quadruplexes in Pathogens: A Common Route to Virulence Control?', *PLoS Pathogens*. doi: 10.1371/journal.ppat.1004562.

Hazel, P. *et al.* (2004) 'Loop-length-dependent folding of G-quadruplexes', *Journal of the American Chemical Society*, 126(50), pp. 16405–16415. doi: 10.1021/ja045154j.

Henderson, A. *et al.* (2017) 'Detection of G-quadruplex DNA in mammalian cells', *Nucleic Acids Research*, 45(10), p. 6252. doi: 10.1093/nar/gkx300.

Hennessy, F. *et al.* (2000) 'Analysis of the levels of conservation of the J domain among the

various types of DnaJ-like proteins', *Cell Stress and Chaperones*, 5(4), pp. 347–358. doi: 10.1379/1466-1268(2000)005<0347:AOTLOC>2.0.CO;2.

Hennessy, F. *et al.* (2005) 'Not all J domains are created equal: Implications for the specificity of Hsp40-Hsp70 interactions', *Protein Science*, 14(7), pp. 1697–1709. doi: 10.1110/ps.051406805.

Hershman, S. G. *et al.* (2008) 'Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae', *Nucleic Acids Research*, 36(1), pp. 144–156. doi: 10.1093/nar/gkm986.

Hiller, N. L. *et al.* (2004) 'A host-targeting signal in virulence proteins reveals a secretome in malarial infection', *Science*, 306(5703), pp. 1934–1937. doi: 10.1126/science.1102737.

Hon, J. *et al.* (2017) 'pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R', *Bioinformatics (Oxford, England)*, 33(21), pp. 3373–3379. doi: 10.1093/bioinformatics/btx413.

Hosoda, A. *et al.* (2003) 'JPDI, a novel endoplasmic reticulum-resident protein containing both a BiP-interacting J-domain and thioredoxin-like motifs', *Journal of Biological Chemistry*, 278(4), pp. 2669–2676. doi: 10.1074/jbc.M208346200.

Huang, Z. L. *et al.* (2018) 'Identification of G-Quadruplex-Binding Protein from the Exploration of RGG Motif/G-Quadruplex Interactions', *Journal of the American Chemical Society*. American Chemical Society, 140(51), pp. 17945–17955. doi: 10.1021/jacs.8b09329.

Huppert, J. L. *et al.* (2008) 'G-quadruplexes: The beginning and end of UTRs', *Nucleic Acids Research*, 36(19), pp. 6260–6268. doi: 10.1093/nar/gkn511.

Huppert, J. L. (2008) 'Hunting G-quadruplexes', *Biochimie*, 90(8), pp. 1140–1148. doi: 10.1016/j.biochi.2008.01.014.

Huppert, J. L. and Balasubramanian, S. (2005) 'Prevalence of quadruplexes in the human genome', *Nucleic Acids Research*, 33(9), pp. 2908–2916. doi: 10.1093/nar/gki609.

Iwanaga, S. *et al.* (2012) 'Identification of an AP2-family Protein That Is Critical for Malaria Liver Stage Development', *PLoS ONE*, 7(11). doi: 10.1371/journal.pone.0047557.

Jiang, L. *et al.* (2013) 'PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum', *Nature*, 499(7457), pp. 223–227. doi: 10.1038/nature12361.

Jutras, B. L., Verma, A. and Stevenson, B. (2012) 'Identification of Novel DNA-Binding proteins Using DNA-Affinity Chromatography/Pull Down', *Current Protocols in Microbiology*, (SUPPL.24), pp. 1–13. doi: 10.1002/9780471729259.mc01f01s24.

Kafsack, B. F. C. *et al.* (2014) 'A transcriptional switch underlies commitment to sexual development in human malaria parasites', *Nature*, 507(7491), pp. 248–252. doi: 10.1038/nature12920.A.

Keene, D. and Query, C. (1991) 'Nuclear RNA-binding Proteins', *PNAS*, 41, pp. 179–202.

Kelley, W. L. (1998) 'The J-domain family and the recruitment of chaperone power', *Trends in Biochemical Sciences*, 23(6), pp. 222–227. doi: 10.1016/S0968-0004(98)01215-8.

Kerry, L. E. *et al.* (2017) 'Selective inhibition of RNA polymerase I transcription as a potential approach to treat African trypanosomiasis', *PLoS Neglected Tropical Diseases*, 11(3). doi: 10.1371/journal.pntd.0005432.

Kikin, O., D'Antonio, L. and Bagga, P. S. (2006) 'QGRS Mapper: A web-based server for predicting G-quadruplexes in nucleotide sequences', *Nucleic Acids Research*, 34(WEB. SERV. ISS.), pp. 676–682. doi: 10.1093/nar/gkl253.

Kim, N. (2019) 'The Interplay between G-quadruplex and Transcription', *Current Medicinal Chemistry*, 26(16), pp. 2898–2917. doi: 10.2174/0929867325666171229132619.

Knuepfer, E. *et al.* (2017) 'Generating conditional gene knockouts in Plasmodium - A toolkit to produce stable DiCre recombinase-expressing parasite lines using CRISPR/Cas9', *Scientific Reports*, 7(1). doi: 10.1038/s41598-017-03984-3.

König, S. L. B. *et al.* (2013) 'Distance-dependent duplex DNA destabilization proximal to G-

quadruplex/i-motif sequences', *Nucleic Acids Research*, 41(15), pp. 7453–7461. doi: 10.1093/nar/gkt476.

Kosugi, S. *et al.* (2009) 'Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs', *Proceedings of the National Academy of Sciences of the United States of America*, 106(25), pp. 10171–10176. doi: 10.1073/pnas.0900604106.

Kshirsagar, R. *et al.* (2017) 'Probing the Potential Role of Non-B DNA Structures at Yeast Meiosis-Specific DNA Double-Strand Breaks', *Biophysical Journal*. Biophysical Society, 112(10), pp. 2056–2074. doi: 10.1016/j.bpj.2017.04.028.

Kudyba, H. M. *et al.* (2019) 'The endoplasmic reticulum chaperone PfGRP170 is essential for asexual development and is linked to stress response in malaria parasites', *Cellular Microbiology*, 21(9), pp. 1–18. doi: 10.1111/cmi.13042.

Kumar, N. *et al.* (1991) 'Induction and localization of Plasmodium falciparum stress proteins related to the heat shock protein 70 family', *Molecular and Biochemical Parasitology*, 48(1), pp. 47–58. doi: 10.1016/0166-6851(91)90163-Z.

Kuryavyi, V. *et al.* (2012) 'The RecA-binding pilE G4 sequence essential for pilin antigenic variation forms parallel-stranded monomeric and 5'-end stacked dimeric G-quadruplexes', *Structure*, 20(12). doi: 10.1016/j.str.2012.09.013.The.

Kwok, Chun Kit *et al.* (2017) 'G-Quadruplexes: Prediction, Characterization, and Biological Application', *Trends in Biotechnology*, 35(10), pp. 997–1013. doi: 10.1016/j.tibtech.2017.06.012.

LaCount, D. J. *et al.* (2005) 'A protein interaction network of the malaria parasite Plasmodium falciparum', *Nature*, 438(7064), pp. 103–107. doi: 10.1038/nature04104.

Lambros, C. and Vanderberg, J. P. (1979) 'Synchronization of Plasmodium falciparum Erythrocytic Stages in Culture', *The Journal of Parasitology*, 65(3), p. 418. doi: 10.2307/3280287.

Larson, E. D. *et al.* (2005) 'MutS␣ Binds to and Promotes Synapsis of Transcriptionally Activated Immunoglobulin Switch Regions', *Current Biology*, 15, p. 1118. doi: 10.1016/j.

Lasonder, E. *et al.* (2012) 'The plasmodium falciparum schizont phosphoproteome reveals extensive phosphatidylinositol and cAMP-protein kinase A signaling', *Journal of Proteome Research*, 11(11), pp. 5323–5337. doi: 10.1021/pr300557m.

Lavezzo, E. *et al.* (2018) 'G-quadruplex forming sequences in the genome of all known human viruses: A comprehensive guide', *PLoS Computational Biology*, 14(12), pp. 1–20. doi: 10.1371/journal.pcbi.1006675.

Lee, M. C. S. *et al.* (2008) 'Plasmodium falciparum Sec24 marks transitional ER that exports a model cargo via a diacidic motif', *Molecular Microbiology*, 68(6), pp. 1535–1546. doi: 10.1111/j.1365-2958.2008.06250.x.

Leeder, W. M. *et al.* (2016) 'The RNA chaperone activity of the Trypanosoma brucei editosome raises the dynamic of bound pre-mRNAs', *Scientific Reports*. Nature Publishing Group, 6(January), pp. 1–11. doi: 10.1038/srep19309.

Leeder, W. M., Hummel, N. F. C. and Göringer, H. U. (2016) 'Multiple G-quartet structures in pre-edited mRNAs suggest evolutionary driving force for RNA editing in trypanosomes', *Scientific Reports*, 6. doi: 10.1038/srep29810.

Li, Q. *et al.* (2013) 'G4LDB: A database for discovering and studying G-quadruplex ligands', *Nucleic Acids Research*, 41(D1), pp. 1115–1123. doi: 10.1093/nar/gks1101.

Lim, L. and McFadden, G. I. (2010) 'The evolution, metabolism and functions of the apicoplast', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1541), pp. 749–763. doi: 10.1098/rstb.2009.0273.

Lopes, J. *et al.* (2011) 'G-quadruplex-induced instability during leading-strand replication', *EMBO Journal*, 30(19), pp. 4033–4046. doi: 10.1038/emboj.2011.316.

López-Barragán, M. J. *et al.* (2011) 'Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum', *BMC Genomics*. BioMed Central Ltd,

12(1), p. 587. doi: 10.1186/1471-2164-12-587.

Lopez-Rubio, J. J. *et al.* (2007) '5′ Flanking Region of Var Genes Nucleate Histone Modification Patterns Linked To Phenotypic Inheritance of Virulence Traits in Malaria Parasites', *Molecular Microbiology*, 66(6), pp. 1296–1305. doi: 10.1111/j.1365-2958.2007.06009.x.

Lopez-rubio, J., Siegel, T. N. and Artur (2013) *Genome-wide Chromatin Immunoprecipitation-Sequencing in Plasmodium*. doi: 10.1007/978-1-62703-026-7.

Lu, W. *et al.* (2013) 'Telomere- structure, function, and regulation', *Exp Cell Res.*, 319(2), pp. 133–141. doi: 10.1126/scisignal.2001449.Engineering.

Lv, B. *et al.* (2013) 'DNA gyrase-driven generation of a G-quadruplex from plasmid DNA', *Chemical Communications*, 49(75), pp. 8317–8319. doi: 10.1039/c3cc44675a.

M. Njunge, J. *et al.* (2013) 'Hsp70s and J Proteins of Plasmodium Parasites Infecting Rodents and Primates: Structure, Function, Clinical Relevance, and Drug Targets', *Current Pharmaceutical Design*, 19(3), pp. 387–403. doi: 10.2174/1381612811306030387.

Mahmoudi, S. and Keshavarz, H. (2017) 'Efficacy of phase 3 trial of RTS, S/AS01 malaria vaccine: The need for an alternative development plan', *Human Vaccines and Immunotherapeutics*. Taylor & Francis, 13(9), pp. 2098–2101. doi: 10.1080/21645515.2017.1295906.

Maier, A. G. *et al.* (2008) 'Exported Proteins Required for Virulence and Rigidity of Plasmodium falciparum-Infected Human Erythrocytes', *Cell*. Elsevier Inc., 134(1), pp. 48–61. doi: 10.1016/j.cell.2008.04.051.

Mani, P. *et al.* (2009) 'Genome-wide analyses of recombination prone regions predict role of DNA structural motif in recombination', *PLoS ONE*, 4(2), pp. 1–9. doi: 10.1371/journal.pone.0004399.

Maris, C., Dominguez, C. and Allain, F. H. T. (2005) 'The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression', *FEBS Journal*,

272(9), pp. 2118–2131. doi: 10.1111/j.1742-4658.2005.04653.x.

Marsico, G. *et al.* (2019) 'Whole genome experimental maps of DNA G-quadruplexes in multiple species', *Nucleic acids research*, 47(8), pp. 3862–3874. doi: 10.1093/nar/gkz179.

Mcrae, E. K. S., Booy, E. P., Padilla-Meier, G. P., *et al.* (2017) 'On Characterizing the Interactions between Proteins and Guanine Quadruplex Structures of Nucleic Acids'. doi: 10.1155/2017/9675348.

Mcrae, E. K. S., Booy, E. P., Padilla-meier, G. P., *et al.* (2017) 'On characterizing the interactions between proteins and quadruplex structures of nucleic acids.', *Journal of Nucleic Acids*, 2017.

Mendoza, O. *et al.* (2016) 'G-quadruplexes and helicases', *Nucleic Acids Research*, 44(5), pp. 1989–2006. doi: 10.1093/nar/gkw079.

Mergny, J.-L., Phan, A.-T. and Lacroix, L. (1998) 'Following G-quartet formation by UV-spectroscopy', *FEBS Journal*, 435, pp. 74–78. doi: 10.1016/S0014-5793(98)01043-6.

Mergny, J. L. *et al.* (2005) 'Thermal difference spectra: A specific signature for nucleic acid structures', *Nucleic Acids Research*, 33(16), pp. 1–6. doi: 10.1093/nar/gni134.

Metifiot, M. *et al.* (2014) 'G-quadruplexes in viruses: function and potential therapeutic applications', *Nucleic Acids Research*, 42(20), pp. 12352–12366. doi: 10.1093/nar/gku999.

Meyne, J., Ratliff, R. L. and Moyzis, R. K. (1989) 'Conservation of the human telomere sequence (TTAGGG)(n) among vertebrates', *Proceedings of the National Academy of Sciences of the United States of America*, 86(18), pp. 7049–7053. doi: 10.1073/pnas.86.18.7049.

Millevoi, S., Moine, H. and Vagner, S. (2012) 'G-quadruplexes in RNA biology', *Wiley Interdisciplinary Reviews: RNA*, 3(4), pp. 495–507. doi: 10.1002/wrna.1113.

Miroux, B. and Walker, J. E. (1996) *Over-production of Proteins in Escherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High*

Levels, *J. Mol. Biol*. Available at: https://ac.els-cdn.com/S002228369690399X/1-s2.0-S002228369690399X-main.pdf?_tid=41918686-36ba-4529-9f52-7ababf3782f0&acdnat=1552592984_b142d3553dc618ba2e45d8aa36bf4bf5 (Accessed: 14 March 2019).

Mishra, S. K. *et al.* (2016) 'G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins', *Scientific Reports*. Nature Publishing Group, 6(December), pp. 2–10. doi: 10.1038/srep38144.

Mohaghegh, P. (2001) 'The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases', *Nucleic Acids Research*, 29(13), pp. 2843–2849. doi: 10.1093/nar/29.13.2843.

Moll, K. *et al.* (2013) *Methods in malaria reserach_6th edition*, *BEI Resources*. doi: 10.1007/s00436-008-0981-9.

Monchaud, D. and Teulade-Fichou, M. P. (2008) 'A hitchhiker's guide to G-quadruplex ligands', *Organic and Biomolecular Chemistry*, 6(4), pp. 627–636. doi: 10.1039/b714772b.

Morris, M. J. *et al.* (2010) 'An RNA G-quadruplex is essential for cap-independent translation initiation in human VEGF IRES', *Journal of the American Chemical Society*, 132(50), pp. 17831–17839. doi: 10.1021/ja106287x.

Muniyappa, K., Anuradha, S. and Byers, B. (2000) 'Yeast meiosis-specific protein Hop1 binds to G4 DNA and promotes its formation.', *Molecular and cellular biology*, 20(4), pp. 1361–9. doi: 10.1080/1461670X.2016.1265896.

Murat, P. *et al.* (2014) 'G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation', *Nature Chemical Biology*, 10(5), pp. 358–364. doi: 10.1038/nchembio.1479.

Nadolski, M. J. and Linder, M. E. (2007) 'Protein lipidation', *FEBS Journal*, 274(20), pp. 5202–5210. doi: 10.1111/j.1742-4658.2007.06056.x.

Nguyen Ba, A. N. *et al.* (2009) 'NLStradamus: A simple Hidden Markov Model for nuclear

localization signal prediction', *BMC Bioinformatics*, 10, pp. 1–12. doi: 10.1186/1471-2105-10-202.

De Nicola, B. *et al.* (2016) 'Structure and possible function of a G-quadruplex in the long terminal repeat of the proviral HIV-1 genome', *Nucleic Acids Research*, 44(13), pp. 6442–6451. doi: 10.1093/nar/gkw432.

Nicoll, W. S. *et al.* (2007) 'Cytosolic and ER J-domains of mammalian and parasitic origin can functionally interact with DnaK', *International Journal of Biochemistry and Cell Biology*, 39(4), pp. 736–751. doi: 10.1016/j.biocel.2006.11.006.

Njunge, J. M. *et al.* (2015) 'PFB0595w is a Plasmodium falciparum J protein that co-localizes with PfHsp70-1 and can stimulate its in vitro ATP hydrolysis activity', *International Journal of Biochemistry and Cell Biology*. Elsevier Ltd, 62, pp. 47–53. doi: 10.1016/j.biocel.2015.02.008.

Norseen, J., Johnson, F. B. and Lieberman, P. M. (2009) 'Role for G-Quadruplex RNA Binding by Epstein-Barr Virus Nuclear Antigen 1 in DNA Replication and Metaphase Chromosome Attachment', *Journal of Virology*, 83(20), pp. 10336–10346. doi: 10.1128/jvi.00747-09.

Paeschke, K. *et al.* (2005) 'Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo', *Nature Structural and Molecular Biology*, 12(10), pp. 847–854. doi: 10.1038/nsmb982.

Paeschke, K., Capra, John A. and Zakian, V. A. (2011) 'DNA Replication through G-Quadruplex Motifs Is Promoted by the Saccharomyces cerevisiae Pif1 DNA Helicase', *Cell*. Elsevier Inc., 145(5), pp. 678–691. doi: 10.1016/j.cell.2011.04.015.

Paeschke, K., Capra, John A and Zakian, V. A. (2011) 'DNA Replication through G-Quadruplex Motifs Is Promoted by the Saccharomyces cerevisiae Pif1 DNA Helicase', *Cell*, 145(5), pp. 678–691. doi: 10.1016/j.cell.2011.04.015.

Partnership, S. C. T. (2015) 'Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: Final results of a phase 3,

individually randomised, controlled trial', *The Lancet*. Elsevier Ltd, 386(9988), pp. 31–45. doi: 10.1016/S0140-6736(15)60721-8.

Perrone, R., Nadai, M., Frasson, I., *et al.* (2013) 'A dynamic G-quadruplex region regulates the HIV-1 long terminal repeat promoter', *J Med Chem*, 56(16), pp. 6521–6530. doi: 10.1038/jid.2014.371.

Perrone, R., Nadai, M., Poe, J. A., *et al.* (2013) 'Formation of a Unique Cluster of G-Quadruplex Structures in the HIV-1 nef Coding Region: Implications for Antiviral Activity', *PLoS ONE*, 8(8), pp. 1–14. doi: 10.1371/journal.pone.0073121.

Perrone, R. *et al.* (2014) 'Anti-HIV-1 activity of the G-quadruplex ligand BRACO-19', *Journal of Antimicrobial Chemotherapy*, 69(12), pp. 3248–3258. doi: 10.1093/jac/dku280.

Perrone, R. *et al.* (2017) 'Mapping and characterization of G-quadruplexes in Mycobacterium tuberculosis gene promoter regions', *Sci Rep*, 7(1), p. 5743. doi: 10.1038/s41598-017-05867-z.

Pesce, E. R. and Blatch, G. L. (2014) 'Plasmodial Hsp40 and Hsp70 chaperones: Current and future perspectives', *Parasitology*, 141(9), pp. 1167–1176. doi: 10.1017/S003118201300228X.

Phan, A. T. (2010) 'Human telomeric G-quadruplex: Structures of DNA and RNA sequences', *FEBS Journal*, 277(5), pp. 1107–1117. doi: 10.1111/j.1742-4658.2009.07464.x.

Phillips, M. A. *et al.* (2017) 'Malaria', *Nature Reviews Disease Primers*, 3. doi: 10.1038/nrdp.2017.50.

Piazza, A. *et al.* (2010) 'Genetic instability triggered by G-quadruplex interacting Phen-DC compounds in Saccharomyces cerevisiae', *Nucleic Acids Research*, 38(13), pp. 4337–4348. doi: 10.1093/nar/gkq136.

Picard, F. *et al.* (2014) 'The Spatiotemporal Program of DNA Replication Is Associated with Specific Combinations of Chromatin Marks in Human Cells', *PLoS Genetics*, 10(5). doi: 10.1371/journal.pgen.1004282.

Prorok, P. *et al.* (2019) 'Involvement of G-quadruplex regions in mammalian replication origin activity', *Nature Communications*. Springer US, 10(1), pp. 1–16. doi: 10.1038/s41467-019-11104-0.

Qiu, J. *et al.* (2015) 'Biological Function and Medicinal Research Significance of G-Quadruplex Interactive Proteins', *Current Topics in Medicinal Chemistry*, 15(19), pp. 1971–1987. doi: 10.2174/1568026615666150515150803.

Qureshi, M. H. *et al.* (2012) 'Replication protein A unfolds G-quadruplex structures with varying degrees of efficiency', *Journal of Physical Chemistry B*, 116(19), pp. 5588–5594. doi: 10.1021/jp300546u.

Rachwal, P. A., Brown, T. and Fox, K. R. (2007) 'Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes', *Biochemistry*, 46(11), pp. 3036–3044. doi: 10.1021/bi062118j.

Radfar, A. *et al.* (2009) 'Synchronous culture of Plasmodium falciparum at high parasitemia levels', *Nature Protocols*, 4(12), pp. 1899–1915. doi: 10.1038/nprot.2009.198.

Raiber, E. A. *et al.* (2012) 'A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro', *Nucleic Acids Research*, 40(4), pp. 1499–1508. doi: 10.1093/nar/gkr882.

Ralph, R. K., Connors, W. J. and Khorana, H. . . (1962) 'Secondary Structure and aggregation in deoxyguanosine oligonucleotides', *Journal of the American Chemical Society*, 84(11), pp. 2265-2266. Available at: https://pubs.acs.org/sharingguidelines (Accessed: 22 October 2019).

Ravichandran, S., Ahn, J. H. and Kim, K. K. (2019) 'Unraveling the Regulatory G-Quadruplex Puzzle: Lessons From Genome and Transcriptome-Wide Studies', *Frontiers in Genetics*, 10(October). doi: 10.3389/fgene.2019.01002.

Rawal, P. *et al.* (2006) 'Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation', *Genome Res*, 16(5), pp. 644–655. doi: 10.1101/gr.4508806.

Rehm, C. *et al.* (2015) 'Investigation of a quadruplex-forming repeat sequence highly enriched in Xanthomonas and Nostoc sp.', *PLoS ONE*, 10(12), pp. 1–21. doi: 10.1371/journal.pone.0144275.

Rezler, E. M., Bearss, D. J. and Hurley, L. H. (2002) 'Telomeres and telomerases as drug targets', *Current Opinion in Pharmacology*, 2(4), pp. 415–423. doi: 10.1016/S1471-4892(02)00182-0.

Rhodes, D. and Lipps, H. J. (2015) 'Survey and summary G-quadruplexes and their regulatory roles in biology', *Nucleic Acids Research*, 43(18), pp. 8627–8637. doi: 10.1093/nar/gkv862.

Roch, K. G. Le *et al.* (2004) 'Global analysis of transcript and protein levels across the Plasmodium falciparum life cycle', *Genome Research*, pp. 2308–2318. doi: 10.1101/gr.2523904.7.

Rug, M. and Maier, A. G. (2011) 'The heat shock protein 40 family of the malaria parasite Plasmodium falciparum.', *IUBMB life*, 63(12), pp. 1081–1086. doi: 10.1002/iub.525.

Ruggiero, E. and Richter, S. N. (2018) 'Survey and summary G-quadruplexes and G-quadruplex ligands: Targets and tools in antiviral therapy', *Nucleic Acids Research*. Oxford University Press, 46(7), pp. 3270–3283. doi: 10.1093/nar/gky187.

Salanti, A. *et al.* (2003) 'Selective upregulation of a single distinctly structured var gene in chondroitin sulphate A-adhering Plasmodium falciparum involved in pregnancy-associated malaria.', *Molecular microbiology*, 49(1), pp. 179–191.

Sanders, C. M. (2010) 'Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity', *Biochem. J*, 430, pp. 119–128. doi: 10.1042/BJ20100612.

Sargeant, T. J. *et al.* (2006) 'Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites', *Genome Biology*, 7(2). doi: 10.1186/gb-2006-7-2-r12.

Saul, A. *et al.* (1982) 'Purification of mature schizonts of Plasmodium falciparum on

colloidal silica gradients', *Bulletin of the World Health Organization*, 60(5), pp. 755–759.

Scalabrin, M. *et al.* (2017) 'The cellular protein hnRNP A2/B1 enhances HIV-1 transcription by unfolding LTR promoter G-quadruplexes', *Scientific Reports*, 7. doi: 10.1038/srep45244.

Schaffitzel, C. *et al.* (2001) 'In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei', *Proceedings of the National Academy of Sciences of the United States of America*, 98(15), pp. 8572–8577. doi: 10.1073/pnas.141229498.

Scherf, A. (1996) 'Plasmodium telomeres and telomere proximal gene expression', *Seminars in Cell and Developmental Biology*, 7(1), pp. 49–57. doi: 10.1006/scdb.1996.0008.

Scherf, A., Lopez-Rubio, J. J. and Riviere, L. (2008) 'Antigenic Variation in Plasmodium falciparum', *Annual Review of Microbiology*, 62(1), pp. 445–470. doi: 10.1146/annurev.micro.61.080706.093134.

Schiavone, D. *et al.* (2014) 'Determinants of G quadruplex-induced epigenetic instability in REV 1-deficient cells', *The EMBO Journal*, 33(21), pp. 2507–2520. doi: 10.15252/embj.201488398.

Schroder, H. *et al.* (1993) 'DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage', *EMBO Journal*, 12(1), pp. 4137–4144.

Sen, D. and Gilbert, W. (1988) 'Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis', *Nature*, 334, pp. 364–366. doi: 10.1093/nq/s8-VII.175.348-b.

Sen, D. and Gilbert, W. (1990) 'A sodium-potassium switch in the formation of four-stranded G4-DNA', *Nature*, 344(6265), pp. 410–414. doi: 10.1038/344410a0.

Sequeira-Mendes, J. *et al.* (2009) 'Transcription Initiation Activity Sets Replication Origin Efficiency in Mammalian Cells', *PLoS Genet*, 5(4), p. 1000446. doi: 10.1371/journal.pgen.1000446.

Sharma, S. K. *et al.* (2010) 'The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase', *Nature Chemical Biology*. Nature Publishing Group, 6(12), pp. 914–920. doi: 10.1038/nchembio.455.

Shorter, J. (2011) 'The mammalian disaggregase machinery: Hsp110 synergizes with Hsp70 and Hsp40 to catalyze protein disaggregation and reactivation in a cell-free system', *PLoS ONE*, 6(10). doi: 10.1371/journal.pone.0026319.

Shum, K. T. *et al.* (2011) 'Aptamer-mediated inhibition of mycobacterium tuberculosis polyphosphate kinase 2', *Biochemistry*, 50(15), pp. 3261–3271. doi: 10.1021/bi2001455.

Siddiqui-Jain, A. *et al.* (2002) 'Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription', *Proceedings of the National Academy of Sciences of the United States of America*, 99(18), pp. 11593–11598. doi: 10.1073/pnas.182256799.

Sierra-Miranda, M. *et al.* (2017) 'PfAP2Tel, harbouring a non-canonical DNA-binding AP2 domain, binds to Plasmodium falciparum telomeres', *Cellular Microbiology*, 19(9). doi: 10.1111/cmi.12742.

Silvia Götz (2017) *Zuo1-ein neues G - Quadruplex - bindendes Protein in Saccharomyces cerevisiae*.

Singh, B. and Daneshvar, C. (2013) 'Human infections and detection of plasmodium knowlesi', *Clinical Microbiology Reviews*, 26(2), pp. 165–184. doi: 10.1128/CMR.00079-12.

Sissi, C., Gatto, B. and Palumbo, M. (2011) 'The evolving world of protein-G-quadruplex recognition: A medicinal chemist's perspective', *Biochimie*. Elsevier Masson SAS, 93(8), pp. 1219–1230. doi: 10.1016/j.biochi.2011.04.018.

Smargiasso, N. *et al.* (2009) 'Putative DNA G-quadruplex formation within the promoters of Plasmodium falciparum var genes', *BMC Genomics*, 10. doi: 10.1186/1471-2164-10-362.

Spiegel, J., Adhikari, S. and Balasubramanian, S. (2019) 'The Structure and Function of DNA G-Quadruplexes', *Trends in Chemistry*. The Author(s), xx(xx), pp. 1–14. doi:

10.1016/j.trechm.2019.07.002.

Stanton, A. *et al.* (2016) 'Recombination events among virulence genes in malaria parasites are associated with G-quadruplex-forming DNA motifs', *BMC Genomics*, 17(1). doi: 10.1186/s12864-016-3183-3.

Stubbs, J. *et al.* (2005) 'Microbiology: Molecular mechanism for switching of P. falciparum invasion pathways into human erythrocytes', *Science*, 309(5739), pp. 1384–1387. doi: 10.1126/science.1115257.

Sturm, A. *et al.* (2006) 'Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids', *Science*, 313(5791), pp. 1287–1290. doi: 10.1126/science.1129720.

Sun, H. *et al.* (1998) *The Bloom's Syndrome Helicase Unwinds G4 DNA**. Available at: http://www.jbc.org/ (Accessed: 15 November 2018).

Sun, H., Bennett, R. J. and Maizels, N. (1999) 'The Saccharomyces cerevisiae Sgs1 helicase efficiently unwinds G-G paired DNAs', *Nucleic Acids Research*, 27(9), pp. 1978–1984. doi: 10.1093/nar/27.9.1978.

Sundquist, W. I. and Heaphy, S. (1993) 'Evidence for interstrand quadruplex formation in the dimerization of human immunodeficiency virus 1 genomic RNA', *Proceedings of the National Academy of Sciences of the United States of America*, 90(8), pp. 3393–3397. doi: 10.1073/pnas.90.8.3393.

Takahama, K. *et al.* (2011) 'Identification of Ewing's sarcoma protein as a G-quadruplex DNA- and RNA-binding protein', *FEBS Journal*, 278(6), pp. 988–998. doi: 10.1111/j.1742-4658.2011.08020.x.

Tavares, J. *et al.* (2013) 'Role of host cell traversal by the malaria sporozoite during liver infection', *Journal of Experimental Medicine*, 210(5), pp. 905–915. doi: 10.1084/jem.20121130.

Thakur, R. S. *et al.* (2014) 'Mycobacterium tuberculosis DinG is a structure-specific helicase that unwinds G4 DNA: Implications for targeting G4 DNA as a novel therapeutic approach',

Journal of Biological Chemistry, 289(36), pp. 25112–25136. doi: 10.1074/jbc.M114.563569.

Thandapani, P. *et al.* (2013) 'Defining the RGG/RG Motif', *Molecular Cell*, 50(5), pp. 613–623. doi: 10.1016/j.molcel.2013.05.021.

Tonkin, C. J. *et al.* (2009) 'Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in Plasmodium falciparum', *PLoS Biology*, 7(4), pp. 0771–0788. doi: 10.1371/journal.pbio.1000084.

Tosoni, E. *et al.* (2015) 'Nucleolin stabilizes G-quadruplex structures folded by the LTR promoter and silences HIV-1 viral transcription', *Nucleic Acids Research*, 43(18), pp. 8884–8897. doi: 10.1093/nar/gkv897.

Treeck, M. *et al.* (2011) 'The Phosphoproteomes of Plasmodium falciparum and Toxoplasma gondii reveal unusual adaptations within and beyond the parasites' boundaries', *Cell Host & Microbe*, 10(4), pp. 410–419. doi: 10.1016/j.chom.2011.09.004.The.

Tsai, J. and Douglas, M. G. (1996) 'A conserved HPD sequence of the J-domain is necessary for YDJ1 stimulation of Hsp70 ATPase activity at a site distinct from substrate binding', *Journal of Biological Chemistry*, 271(16), pp. 9347–9354. doi: 10.1074/jbc.271.16.9347.

Ukaegbu, U. E. *et al.* (2014) 'Recruitment of PfSET2 by RNA Polymerase II to Variant Antigen Encoding Loci Contributes to Antigenic Variation in P. falciparum', *PLoS Pathogens*, 10(1). doi: 10.1371/journal.ppat.1003854.

Uribe, D. J. *et al.* (2011) 'Heterogeneous nuclear ribonucleoprotein K and nucleolin as transcriptional activators of the vascular endothelial growth factor promoter through interaction with secondary DNA structures', *Biochemistry*, 50(18), pp. 3796–3806. doi: 10.1021/bi101633b.

Ushioda, R. *et al.* (2008) 'ERdj5 is required as a disulfide reductase for degradation of misfolded proteins in the ER', *Science*, 321(5888), pp. 569–572. doi: 10.1126/science.1159293.

Vallur, A. C. and Maizels, N. (2010) 'Distinct activities of exonuclease 1 and flap

endonuclease 1 at telomeric G4 DNA', *PLoS ONE*, 5(1). doi: 10.1371/journal.pone.0008908.

Vazquez de Aldana, C. R., Marton, M. J. and Hinnebusch, A. G. (1995) 'GCN20, a novel ATP binding cassette protein, and GCN1 reside in a complex that mediates activation of the eIF-2 alpha kinase GCN2 in amino acid-starved cells.', *The EMBO Journal*, 14(13), pp. 3184–3199. doi: 10.1002/j.1460-2075.1995.tb07321.x.

Verma, A. *et al.* (2008) 'Genome-wide computational and expression analyses reveal G-quadruplex DNA motifs as conserved cis-regulatory elements in human and related species', *Journal of Medicinal Chemistry*, 51(18), pp. 5641–5649. doi: 10.1021/jm800448a.

Vernick, K. D. and McCutchan, T. F. (1988) 'Sequence and structure of a Plasmodium falciparum telomere', *Molecular and Biochemical Parasitology*, 28(2), pp. 85–94. doi: 10.1016/0166-6851(88)90055-2.

del Villar-Guerra, R., Trent, J. O. and Chaires, J. B. (2018) 'G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy', *Angewandte Chemie - International Edition*, 57(24), pp. 7171–7175. doi: 10.1002/anie.201709184.

Vlasenok, M. *et al.* (2018) 'Data set on G4 DNA interactions with human proteins', *Data in Brief*. Elsevier Inc., 18, pp. 348–359. doi: 10.1016/j.dib.2018.02.081.

Wang, F. *et al.* (2012a) 'Telomere- and telomerase-interacting protein that unfolds telomere G-quadruplex and promotes telomere extension in mammalian cells', *Proceedings of the National Academy of Sciences of the United States of America*, 109(50), pp. 20413–20418. doi: 10.1073/pnas.1200232109.

Wang, F. *et al.* (2012b) 'Telomere- and telomerase-interacting protein that unfolds telomere G-quadruplex and promotes telomere extension in mammalian cells', *Proceedings of the National Academy of Sciences of the United States of America*, 109(50), pp. 20413–20418. doi: 10.1073/pnas.1200232109.

Watanabe, J. (1997) 'Cloning and characterization of heat shock protein DnaJ homologues from Plasmodium falciparum and comparison with ring infected erythrocyte surface antigen', *Molecular and Biochemical Parasitology*, 88(1–2), pp. 253–258. doi: 10.1016/S0166-

6851(97)00073-X.

Wellinger, R. J. and Sen, D. (1997) 'The DNA structures at the ends of eukaryotic chromosomes', *European Journal of Cancer Part A*, 33(5), pp. 735–749. doi: 10.1016/S0959-8049(97)00067-1.

Westmark, C. J. and Malter, J. S. (2007) 'FMRP mediates mGluR5-dependent translation of amyloid precursor protein', *PLoS Biology*, 5(3), pp. 0629–0639. doi: 10.1371/journal.pbio.0050052.

WHO (2019) *World Malaria Report. Colombia*.

Wickramasinghe, C. M. *et al.* (2015) 'Contributions of the specialised DNA polymerases to replication of structured DNA', *DNA Repair*. Elsevier B.V., 29, pp. 83–90. doi: 10.1016/j.dnarep.2015.01.004.

Williams, P. *et al.* (2017) 'Identification of SLIRP as a G Quadruplex-Binding Protein Computational analysis using the consensus G4 sequence motif of G 3+ N 1-7 G 3+ N 1-7 G 3+ N 1-7 G 3+ revealed >300 000 motifs in the human genome with potential HHS Public Access', *J Am Chem Soc*, 139(36), pp. 12426–12429. doi: 10.1021/jacs.7b07563.

Wong, H. M. and Huppert, J. L. (2009) 'Stable G-quadruplexes are found outside nucleosome-bound regions', *Molecular BioSystems*, 5(12), pp. 1713–1719. doi: 10.1039/b905848f.

Wu, G. and Han, H. (2019) 'A DNA Polymerase Stop Assay for Characterization of G-Quadruplex Formation and Identification of G-Quadruplex-Interactive Compounds', in *G-Quadruplex Nucleic Acids*, pp. 223–231.

Wu, Y. *et al.* (2005) 'The crystal structure of the C-terminal fragment of yeast Hsp40 Ydj1 reveals novel dimerization motif for Hsp40', *Journal of Molecular Biology*, 346(4), pp. 1005–1011. doi: 10.1016/j.jmb.2004.12.040.

Young, C. L., Britton, Z. T. and Robinson, A. S. (2012) 'Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications',

Biotechnology Journal, 7(5), pp. 620–634. doi: 10.1002/biot.201100155.

Young, J. C. *et al.* (2004) 'Pathways of chaperone-mediated protein folding in the cytosol', *Nature Reviews Molecular Cell Biology*, 5(10), pp. 781–791. doi: 10.1038/nrm1492.

Yuda, M. *et al.* (2009) 'Identification of a transcription factor in the mosquito-invasive stage of malaria parasites', *Molecular Microbiology*, 71(6), pp. 1402–1414. doi: 10.1111/j.1365-2958.2009.06609.x.

Yuda, M. *et al.* (2010) 'Transcription factor AP2-Sp and its target genes in malarial sporozoites', *Molecular Microbiology*, 75(4), pp. 854–863. doi: 10.1111/j.1365-2958.2009.07005.x.

Zakian, V. A. (1995) 'Telomeres: Beginning to understand the end', *Science*, 270(5242), pp. 1601–1607. doi: 10.1126/science.270.5242.1601.

Zakian, V. A. (2012) 'Telomeres: The beginnings and ends of eukaryotic chromosomes', *Experimental cell research*, 318(12), pp. 1456–1460. doi: 10.1016/j.yexcr.2012.02.015.Telomeres.

Zaug, A. J., Podell, E. R. and Cech, T. R. (2005) 'Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro', *Proceedings of the National Academy of Sciences*, 102(10), pp. 10864–10869. doi: 10.1073/pnas.0504744102.

Zhang, M. *et al.* (2018) 'Uncovering the essential genome of the human malaria parasite Plasmodium falciparum by saturation mutagenesis', *Science*, 360(6388), pp. 139–148. doi: 10.1016/j.physbeh.2017.03.040.

Zhang, S. *et al.* (1992) 'Zuotin, a putative Z-DNA binding protein in Saccharomyces cerevisiae.', *The EMBO Journal*, 11(10), pp. 3787–3796. doi: 10.1002/j.1460-2075.1992.tb05464.x.

Zhou, Z. *et al.* (2013) 'G-quadruplex-based fluorescent assay of S1 nuclease activity and K +', *Analytical Chemistry*, 85(4), pp. 2431–2435. doi: 10.1021/ac303440d.

Appendices

Appendices

3. Chapter 3

Table 3-1.List of primers or oligonucleotides used in Y1H assay and DNA pull down assay

Primer ID	**Purpose, F-Forward and R-Reverse**	**Sequence (5'- 3')**
PG1	Cloning in pAbAi_G4 (F)	ATGAATTGAAAAGCTTATAATTTAAAAATTTTGGGATTTGGGAGGGGGGGA
PG2	Cloning in pAbAi_G4 (R)	GAGCACATGCCTCGAGTATATAATTTTTTGTTCCCCCCCTCCCAAATCCC
PG15	Cloning in pAbAi_mutG4 (F)	ATGAATTGAAAAGCTTATAATTTAAAAATTTTGAGATTTGAGAGGGAGGGA
PG16	Cloning in pAbAi_mutG4 (R)	GAGCACATGCCTCGAGTATATAATTTTTTGTTCCCTCCCTCTCAAATCTC
PG19	Sequencing primer pAbAi_URA3_P_F	TATGCGTATATATACCAATC
PG33	DNA-pull down assay_ Biotin_PDA_WT_G4	AAAACAAGGGGGGGAGGGTTTAGGGTTTTA
PG34	DNA pull-down assay Biotin_G4_PG_PDA_mut_G4	AAAACAAGGGAGGGAGAGTTTAGAGTTTTA

Table 3-2.List of candidates obtained from Yeast one-Hybrid system

PlasmoDB ID	Name	Growth phenotype of the yeast
Candidates binding to the WT_G4		
PF3D7_1407400	Conserved Plasmodium protein, unknown protein	normal
PF3D7_1229400	Macrophage migration inhibitory factor	normal
PF3D7_1019100	Conserved Plasmodium protein, unknown protein	normal
PF3D7_0309600	60S acidic ribosomal protein P2	normal
PF3D7_1423500	Conserved Plasmodium protein, unknown protein	normal
PF3D7_1136500	Casein kinase I	normal
PF3D7_0919900	Regulator of chromosome condensation-PP1-binding protein	normal
PF3D7_0504800	Conserved Plasmodium protein, unknown protien	normal
PF3D7_1201000	Conserved Plasmodium protein, unknown protein	normal
PF3D7_1020500	Ribosomal silencing factor RsfS, putative	normal
PF3D7_0408500	Flap endonuclease 1	normal
PF3D7_1035200	S-antigen	normal
PF3D7_0823800	DnaJ protein, putative	normal
XM_002808987.1	Conserved Plasmodium protein, unknown protein	normal
PF3D7_1402800	Conserved Plasmodium protein, unknown protein	normal
PF3D7_1408700	Conserved Plasmodium protein, unknown protein	slow
PF3D7_1223300	DNA gyrase subunit A	slow
PF3D7_0514000	Tubulin--tyrosine ligase, putative	slow

PF3D7_0518500	ATP-dependent RNA helicase DDX23, putative	slow
PF3D7_1145300	Cysteine-rich PDZ-binding protein, putative	slow
PF3D7_0706000	Importin-7, putative	slow
Candidates binding to the mut_G4		
PF3D7_1128400	Bifunctional farnesyl/geranylgeranyl diphosphate synthase	normal
PF3D7_0615600	Zinc finger protein, putative	normal
PF3D7_0721100	Uncharacterized protein	normal
PF3D7_0202500	Early transcribed membrane protein 2	normal
PF3D7_1329000	DNA-directed RNA polymerase subunit	normal
PF3D7_0801800	Mannose-6-phosphate isomerase, putative	normal
PF3D7_1105100	Histone H2B	normal
PF3D7_1201000	Uncharacterized protein	normal
PF3D7_1429400	rRNA (Adenosine-2'-O-)-methyltransferase, putative	normal
PF3D7_1128400	Bifunctional farnesyl/geranylgeranyl diphosphate synthase	normal
PF3D7_0316300	Probable inorganic pyrophosphatase	normal
PF3D7_1325900	Uncharacterized protein	normal
PF3D7_1345200	Rhomboid protease ROM6, putative	normal
PF3D7_0422400	40S ribosomal protein S19	normal
PF3D7_0510500	Topoisomerase I	normal
PF3D7_0825500	Protein KRI1, putative	normal
PF3D7_112300	Uncharacterized protein	normal
PF3D7_1002600	Uncharacterized protein	normal

PF3D7_1136500	casein kinase 1	normal
PF3D7_1414300	60S ribosomal protein L10, putative	normal
PF3D7_0921000	Ubiquitin-conjugating enzyme E2, putative	normal
PF3D7_1024800	Exported protein 3	normal
PF3D7_1115600	Peptidyl-prolyl cis-trans isomerase	normal
PF3D7_1475600	Bromodomain protein, putative	normal
PF3D7_1448300	Uncharacterized protein	normal
PF3D7_0524000	Karyopherin beta	normal
PF3D7_1241100	Polyadenylation factor subunit 2, putative	normal
PF3D7_1029300	Uncharacterized protein	normal
PF3D7_0532100	Early transcribed membrane protein 5	normal
PF3D7_0304500	1-cys-glutaredoxin-like protein-1	normal
PF3D7_0404900	Merozoite surface protein P41	normal
PF3D7_1434500	Dynein-related AAA-type ATPase, putative	normal
PF3D7_1317300	Uncharacterized protein	normal
U15994.1	Histone H3	slow
PF3D7_1227000	Uncharacterized protein	normal
PF3D7_0719900	Uncharacterized protein	normal
PF3D7_1439300	Uncharacterized protein	slow
PF3D7_0323700	U4/U6.U5 tri-snRNP-associated protein 1, putative	normal
PF3D7_0322500	Iron-sulfur assembly protein, putative	normal
PF3D7_1027300	Peroxiredoxin	normal
PF3D7_1447100	Uncharacterized protein	normal

PF3D7_1133900	Uncharacterized protein	slow
PF3D7_1133200	Uncharacterized protein	normal
PF3D7_0422400	40S ribosomal protein S19	normal
PF3D7_0527000	DNA replication licensing factor MCM3, putative	normal
PF3D7_0212900	Arginyl-tRNA--protein transferase	normal
PF3D7_0522400	Uncharacterized protein	normal
FJ406826.1	Merozoite surface protein 9	slow
PF3D7_1414000	26S proteasome regulatory subunit RPN13, putative	slow
PF3D7_0726400	Uncharacterized protein	slow
CP017005.1		slow
PF3D7_0807800	26S proteasome regulatory subunit RPN10, putative	slow
PF3D7_1033700	Bromodomain protein 1	slow
PF3D7_0207800	Serine repeat antigen 3	slow
CP016996.1		slow
PF3D7_0312800	60S ribosomal protein L26, putative	slow
PF3D7_1117700	GTP-binding nuclear protein	slow
PF3D7_1346300	DNA/RNA-binding protein Alba 2	slow
PF3D7_1027800	60S ribosomal protein L3	slow
PF3D7_0607900	Uncharacterized protein	slow
PF3D7_0803800	Proteasome subunit beta	slow
PF3D7_1335500	DNA replication complex GINS protein, putative	slow
PF3D7_1039000	Serine/threonine protein kinase, FIKK family	slow
PF3D7_1407400	Conserved Plasmodium protein, unknown protein	normal

Table 3-3.List of candidates obtained from DNA pull down assay followed by MS

		Unique peptide						MS/MS count						
		WT_G4			Mut_G4			WT_G4			Mut_G4			
PlasmoDB ID	Name	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3	Mean_ MS/Ms count of WT/mut
PF3D7_0501500	Rhoptry-associated protein 3	1	19	3	2	9	5	0	35	2	1	1	4	11.8
PF3D7_0904700	Bacterial histone-like protein	5	6	4	5	3	4	8	10	3	5	0	6	4.03
PF3D7_1022400	Serine/arginine-rich splicing factor 4	0	8	1	2	4	2	0	11	1	3	0	2	3.83
PF3D7_0207500	Serine repeat antigen 6	5	6	5	4	1	6	6	7	3	2	0	10	3.43
PF3D7_1346300	DNA/RNA-binding protein Alba 2	10	8	7	12	4	7	32	8	11	25	1	19	3.28
PF3D7_1434800	Mitochondrial acidic protein MAM33, putative	2	4	1	2	3	1	2	9	0	3	1	1	3.22
PF3D7_1129000	Spermidine synthase	3	3	4	5	2	4	0	3	5	8	0	1	2.66
PF3D7_1462800	Glyceraldehyde-3-phosphate dehydrogenase	5	4	4	5	1	4	4	3	4	8	1	1	2.5
PF3D7_1224300	Polyadenylate-binding protein	15	15	17	23	12	13	11	17	24	41	7	5	2.49
PF3D7_0918000	Glideosome-associated protein 50	2	5	4	3	3	3	0	5	2	4	0	1	2.33
PF3D7_1136300	Tudor staphylococcal nuclease	15	2	5	12	1	9	29	1	0	5	0	9	2.26
PF3D7_1325100	Phosphoribosylpyrophosphate synthetase	2	5	1	3	3	1	3	5	0	2	1	1	2.16

PF3D7_1473200	DnaJ protein, putative	0	14	0	1	10	0	0	25	0	0	4	0	2.08
PF3D7_0619400	Cell division cycle protein 48 homologue, putative	1	8	2	1	1	7	0	6	1	1	0	8	2.0
PF3D7_1002400.1	Transformer-2 protein homolog beta, putative	2	3	1	3	3	1	0	6	0	2	1	1	2
PF3D7_0517300	Serine/arginine-rich splicing factor 1	2	4	0	4	2	0	0	6	0	7	1	0	2
CK1	Casein kinase I	4	0	0	3	0	0	6	0	0	0	0	0	2
PF3D7_1223300	DNA gyrase subunit A	0	6	0	1	1	0	0	6	0	0	0	0	2
PF3D7_0818900	Heat shock protein 70	16	14	16	20	11	13	15	14	23	49	6	8	1.83
PF3D7_1230400	ATP-dependent protease subunit ClpQ	2	0	3	2	0	2	2	0	5	6	0	1	1.77
PF3D7_1121600	Exported protein 1	3	2	2	3	1	2	1	3	2	6	0	1	1.72
PF3D7_1357000	Elongation factor 1-alpha	24	14	12	30	12	12	59	17	17	100	5	16	1.68
PF3D7_1206000	Shewanella-like protein phosphatase 2	6	16	4	6	11	5	12	33	1	8	10	4	1.68
PF3D7_1220900	Heterochromatin protein 1	2	2	1	2	1	1	0	4	1	6	0	1	1.66
PF3D7_1320000	Golgi protein 1	4	1	0	1	1	4	4	1	0	0	0	4	1.66
PF3D7_1143400	Translation initiation factor eIF-1A, putative	3	4	0	7	2	0	0	5	0	13	0	0	1.66
PF3D7_0930300	Merozoite surface protein 1	33	14	22	44	14	22	27	12	25	78	5	14	1.51
PF3D7_0823200	RNA-binding protein, putative	6	4	1	5	4	4	9	6	0	7	2	5	1.42
PF3D7_0929	High molecular weight rhoptry	2	36	7	2	2	7	0	51	1	1	13	9	1.34

400	protein 2					7								
PF3D7_0424600	PRESAN domain-containing protein	1	2	2	1	0	2	0	2	2	1	0	0	1.33
PF3D7_1352500	Thioredoxin-related protein, putative	2	1	2	3	1	2	0	1	3	5	0	1	1.33
PF13_0233	Myosin-A	0	1	3	0	0	3	0	1	3	0	0	0	1.33
PF3D7_0501600	Rhoptry-associated protein 2	2	20	14	3	15	15	1	36	18	1	16	26	1.31
PF3D7_1410400	Rhoptry-associated protein 1	7	32	30	8	27	27	3	82	42	10	28	60	1.30
PF3D7_0905400	High molecular weight rhoptry protein 3	10	17	12	12	11	11	2	18	10	27	7	8	1.29
PF3D7_1347500	DNA/RNA-binding protein Alba 4	17	11	11	16	11	12	81	13	16	65	9	16	1.23
PF3D7_0731600	Acyl-CoA synthetase	4	1	2	3	0	2	2	1	1	1	0	2	1.16
PF3D7_1006800	Single-strand telomeric DNA-binding protein GBP2, putative	2	8	3	2	7	4	4	10	1	2	8	5	1.15
PF3D7_0520400	Single-stranded DNA-binding protein, putative	2	4	1	0	5	2	3	3	0	0	7	2	1.14
PF3D7_1027300	Peroxiredoxin	6	15	9	5	18	10	4	21	5	2	28	9	1.10
PF3D7_1011800	PRE-binding protein	16	18	8	14	18	11	8	18	1	15	7	11	1.06
PF3D7_0912400	Alkaline phosphatase, putative	3	3	0	5	1	0	1	3	0	6	0	0	1.05
PF3D7_0102200	Ring-infected erythrocyte surface antigen	5	3	1	7	1	1	1	3	0	13	0	1	1.02
PF3D7_0936800	PRESAN domain-containing protein	1	2	0	1	2	0	2	1	0	1	1	0	1
PF3D7_1134	Protein disulfide	2	4	2	3	2	3	0	3	0	2	0	3	1

100	isomerase													
PF3D7_1459400	Uncharacterized protein	0	3	1	2	0	1	0	3	0	2	0	1	1
PF3D7_1411900	p1/s1 nuclease, putative	2	1	0	1	4	1	3	0	0	0	6	1	1
PF3D7_0728000	Eukaryotic translation initiation factor 2 subunit alpha	3	0	0	1	0	0	3	0	0	0	0	0	1
PF3D7_0413500	Phosphoglucomutase-2	0	2	0	0	0	0	0	3	0	0	0	0	1
PF3D7_0214900	Rhoptry neck protein 6	0	3	0	0	0	0	0	3	0	0	0	0	1
PF3D7_1330800	RNA-binding protein, putative	1	5	2	3	3	4	0	5	1	1	2	4	0.91
PF3D7_0500800	Mature parasite-infected erythrocyte surface antigen	17	20	17	32	19	19	7	24	12	55	13	19	0.86
PF3D7_1149000	Antigen 332, DBL-like protein	25	33	31	25	32	34	28	29	18	24	29	42	0.86
PF3D7_0708400	Heat shock protein 90	12	0	2	13	0	2	10	0	2	18	0	0	0.85
PF3D7_0904800	Replication protein A1, small	12	1	12	12	2	12	23	1	7	15	2	16	0.823
PF3D7_1301600	Erythrocyte binding antigen-140	1	2	4	2	0	4	0	1	4	3	0	3	0.777
PF3D7_0721100	Uncharacterized protein	10	1	4	10	0	3	10	1	2	30	0	2	0.777
PF3D7_0917900	Heat shock protein 70	5	6	4	7	7	4	2	5	4	9	4	5	0.757
PF3D7_1451100	Elongation factor 2	8	1	0	9	0	1	12	1	0	11	0	0	0.6967
PF3D7_0817900	High mobility group protein B2	4	1	1	5	1	1	27	2	0	25	2	2	0.693
PF3D7_1310700	RNA-binding protein, putative	1	1	2	1	1	2	1	1	0	0	0	2	0.666

PF3D7_0525800	Inner membrane complex protein 1g, putative	0	2	1	0	2	1	0	2	1	0	2	1	0.6666667
PF3D7_0501300	Protein phosphatase	0	1	1	0	1	0	0	1	1	0	0	0	0.6666667
PF3D7_1246400	Myosin A tail domain binding protein	0	1	1	0	0	0	0	1	1	0	0	0	0.6666667
PF3D7_1476300	PRESAN domain-containing protein	2	3	1	2	0	1	0	2	0	1	0	1	0.6666667
PF3D7_0919000	Nucleosome assembly protein	3	0	6	4	0	6	4	0	0	2	0	8	0.6666667
PF3D7_1471100	Exported protein 2	1	0	2	3	0	2	0	0	2	5	0	1	0.6666667
vapA	V-type proton ATPase catalytic subunit A	0	0	1	2	0	1	0	0	2	3	0	0	0.6666667
PF3D7_0629200	DnaJ protein, putative	0	2	0	1	1	0	0	2	0	1	0	0	0.6666667
PF3D7_0919900	Regulator of chromosome condensation-PP1-binding protein	0	1	0	0	1	0	0	2	0	0	1	0	0.6666667
PF3D7_1108700	Heat shock protein J2	0	2	0	1	1	0	0	2	0	0	0	0	0.6666667
PF3D7_1133800	RNA (Uracil-5-)methyltransferase, putative	0	2	0	0	0	0	0	2	0	0	0	0	0.6666667
PF3D7_0102900	Aspartate--tRNA ligase	2	0	0	2	0	0	2	0	0	0	0	0	0.6666667
PF3D7_1401600	PRESAN domain-containing protein	0	2	0	1	1	0	0	2	0	0	0	0	0.6666667
PF3D7_0823800	DnaJ protein, putative	0	2	0	0	0	0	0	2	0	0	0	0	0.6666667
PF3D7_1007900	Eukaryotic translation initiation factor 3 subunit D	0	0	2	0	0	1	0	0	2	0	0	0	0.6666667
PF3D7_1464	Serine/threonine protein phosphatase	0	2	0	0	0	0	0	2	0	0	0	0	0.66666

600	UIS2, putative													67
PF3D7_1410600	Eukaryotic translation initiation factor 2 subunit gamma, putative	5	1	2	6	1	2	6	1	0	7	1	2	0.6190476
PF3D7_1006200	DNA/RNA-binding protein Alba 3	7	6	5	10	7	5	8	8	3	13	11	6	0.6142191
PF3D7_0916700	RNA-binding protein musashi, putative	0	4	0	0	4	2	0	5	0	0	3	2	0.5555556
PF3D7_0818200	14-3-3 protein	7	7	8	8	7	8	5	4	3	9	5	10	0.5518519
PF3D7_1232100	60 kDa chaperonin	5	12	0	4	14	0	2	9	0	2	14	0	0.547619
PF3D7_1202900	High mobility group protein B1	7	2	2	8	1	2	25	1	0	41	0	2	0.5365854
PF3D7_0814200	DNA/RNA-binding protein Alba 1	5	9	5	8	9	6	6	6	3	14	7	10	0.5285714
PF3D7_0831400	Uncharacterized protein	2	1	0	2	0	0	1	1	0	2	0	0	0.5
PF3D7_0402000	PRESAN domain-containing protein	2	2	3	3	1	4	3	1	0	3	2	4	0.5
PF3D7_1126200	40S ribosomal protein S18, putative	2	7	4	3	3	2	1	10	3	4	0	0	4.4166667
PF3D7_0312800	60S ribosomal protein L26, putative	0	5	3	1	3	3	0	8	5	1	0	1	4.3333333
PF3D7_1358800	40S ribosomal protein S15	4	7	3	4	5	3	0	11	4	6	2	1	3.1666667
PF3D7_1130100	60S ribosomal protein L38	4	1	3	4	0	3	9	1	5	14	0	0	2.2142857
PF3D7_0422400	40S ribosomal protein S19	5	3	5	6	1	5	7	4	5	10	0	3	2.1222222
PF3D7_1302800	40S ribosomal protein S7	7	5	4	7	3	5	13	8	2	6	2	11	2.1161616
PF3D7_1408	40S ribosomal	4	12	5	6	9	5	1	24	7	6	9	2	2.11111

600	protein S8													11
PF3D7_0217800	40S ribosomal protein S26	1	3	3	2	2	3	0	3	3	2	0	0	2
PF3D7_0813900	40S ribosomal protein S16, putative	6	3	2	7	2	2	0	5	2	8	0	2	2
PF3D7_1365900	Ubiquitin-60S ribosomal protein L40	3	5	3	3	5	2	5	19	1	13	4	3	1.8226496
PF3D7_1431700	60S ribosomal protein L14, putative	4	4	6	5	2	7	0	5	3	6	1	9	1.7777778
PF3D7_1460300	60S ribosomal protein L29	1	2	2	1	2	2	1	2	2	2	1	1	1.5
PF3D7_0516200	40S ribosomal protein S11	4	3	2	6	2	2	4	3	2	12	1	2	1.4444444
PF3D7_1130200	60S acidic ribosomal protein P0	7	2	2	8	2	2	2	2	2	7	1	1	1.42
PF3D7_0507100	60S ribosomal protein L4	3	4	1	3	4	1	1	4	0	4	1	1	1.41
PF3D7_1460700	60S ribosomal protein L27	1	2	2	4	1	2	1	2	2	5	0	1	1.4
PF3D7_0316800	40S ribosomal protein S15A, putative	2	2	1	3	2	1	1	3	1	7	1	1	1.38
PF3D7_1424400	60S ribosomal protein L7-3, putative	3	14	7	3	11	7	1	23	6	2	9	6	1.35
PF3D7_0306900	40S ribosomal protein S23, putative	0	2	1	0	1	1	0	2	2	0	1	0	1.33
MAL3P7.35	40S ribosomal protein S3a	1	2	3	1	1	3	0	3	2	2	0	2	1.33
PF3D7_1124900	60S ribosomal protein L35, putative	1	1	2	1	0	1	0	1	3	1	0	1	1.33
MAL3P3.19	60S acidic ribosomal protein	1	4	0	1	2	0	0	4	0	1	0	0	1.33

	P2													
PF3D7_1317800	40S ribosomal protein S19	0	3	0	0	1	0	0	4	0	0	1	0	1.33
PF3D7_1105400	40S ribosomal protein S4	1	5	3	3	3	3	1	6	2	4	2	3	1.30
PF3D7_0719600	60S ribosomal protein L11a, putative	2	3	4	2	3	4	7	4	1	5	2	5	1.2
PF3D7_0317600	40S ribosomal protein S11, putative	1	1	1	1	0	1	1	1	2	2	0	1	1.16
PF3D7_0519400	40S ribosomal protein S24	1	3	1	2	1	1	0	3	1	3	1	2	1.166
PF3D7_0814000	60S ribosomal protein L13	1	5	2	1	5	2	0	4	2	1	4	0	1
RPL44	60S ribosomal protein L44	1	1	1	1	1	1	0	2	1	2	1	1	1
PF3D7_0710600	60S ribosomal protein L34	0	2	1	0	2	1	0	2	1	0	0	1	1
PF3D7_1109900	60S ribosomal protein L36	0	2	1	0	1	1	0	2	1	0	0	0	1
PF3D7_0520000	40S ribosomal protein S9, putative	0	3	1	0	3	1	0	3	0	0	0	0	1
PF3D7_1142600	60S ribosomal protein L35ae, putative	1	1	3	1	1	3	0	2	2	1	1	3	0.888
PF3D7_1338200	60S ribosomal protein L6, putative	3	5	3	4	5	3	1	7	0	5	3	2	0.844
PF3D7_1341200	60S ribosomal protein L18a	1	7	2	1	6	2	0	8	2	2	6	2	0.77
PF10_0264	40S ribosomal protein SA	1	1	1	1	1	1	0	1	1	1	1	1	0.66
PF3D7_0903900	60S ribosomal protein L32	1	1	2	1	1	2	0	1	1	1	0	0	0.66
PF3D7_0721600	40S ribosomal protein S5, putative	1	0	4	1	0	4	1	0	4	1	0	4	0.66

PF3D7_1424100	60S ribosomal protein L5, putative	1	2	6	3	2	6	0	2	0	2	0	6	0.66
PF3D7_1323100	60S ribosomal protein L6, putative	2	2	0	2	1	2	0	2	0	2	0	2	0.66
PF3D7_0614500	60S ribosomal protein L19	0	2	1	0	2	1	0	2	0	0	1	1	0.66
PF3D7_0307200	60S ribosomal protein L7, putative	0	2	0	1	2	0	0	2	0	1	1	0	0.66
PF3D7_1242700	40S ribosomal protein S17, putative	0	2	0	1	1	0	0	2	0	1	0	0	0.66
RPL37A	60S ribosomal protein L37a	0	0	1	0	0	1	0	0	2	0	0	0	0.66
PF3D7_0618300	60S ribosomal protein L27a, putative	0	2	0	0	2	0	0	2	0	0	0	0	0.66
PF3D7_1342000	40S ribosomal protein S6	1	6	2	1	5	2	0	7	0	1	4	2	0.58
PF3D7_1421200	40S ribosomal protein S25	2	1	1	3	1	1	3	0	1	5	1	0	0.533

4. Chapter 4

Table 4-1.List of primers and gblock used in chapter 4, characterization of PfGBP2

Primer name	**Purpose,Amplicon , F-forward and R- reverse**	**5' – 3' sequence**
Amplification of Homologous region1 and 2		
VG1	HR1_F	TTAGCTAAGCATGCGGGCCCATGTCGATGGAAAATA ATGTAAG
VG2	HR1_R	TAAATTTCCAACATACACTCTACAACCTTTGCTTG
VG3	HR2_F	ACGAAGTTATACTAGTGTTAATTATGAAAATAAC
VG4	HR2_R	TTCTATTTATCTCGAGACATAACATAAAAATATATG
PG101	gRNA_F	AAGTATATAATATTGATTGGCCACGTAACCTTCCAG TTTTAGAGCTAGAA
PG102	gRNA_R	TTCTAGCTCTAAAACTGGAAGGTTACGTGGCCAATC AATATTATATACTT
PG106	Sequencing primer for gRNA verification	GATTGGCCACGTAACCTTCCA
Verification of Transgenic line		
PG120	Box1_F (Whole_locus _F)	TTATCTCTTCTCTTCTCTTCC
PG121	Box1_R	GAATCGTTAAGAGAAGATATG
PG122	Box2_F	GAAGCTAAGAACGCAATAGAC
PG123	Box2_R (Whole locus_R)	GCTAGCTGATTAATAACAATCC
Cloning for recombinant protein		
PG111	pET15b_F	CGCGCGGCAGCCATATGTCGATGGAAAAT
PG112	pET15b_R	CAGCCGGATCCTCGAGTTATTCGTTGTTTTCATAATT G

Gblock synthesized for PfGBP2 to generate floxed recodononized gene

5’tatgttggaaatttaCCCTGGAAAGTAACGTGGCCAATATTAAAAAATCATATGAAAAAGGCTGGAGACG
TAGTGAGGGTGGATATTTTCGAAGACACTCAAGGAAGATCTAAGGTAAATAAAAAAAATAATATA
CAATAACTTCGTATAGCATACATTATACGAAGTTATTATATATGTATATATATATATATTTATATAT
TTTATATTCTTTTAGGGTTGCGGAATTGTTGAATATGCCACTTATGAAGAGGCTCAGGAAGCCATA
TCTTCTCTTAACGATTCAAAACTTGAGGATAGGTTAATTTTTGTTCGTGAGGATAGGGAAGAGAAT
TCAGGAAACTTTGAAAAGAGGAAATTCAACAATGTGAGAAAGGACAAGTTTTATGAGTCTCGTCG
TCGTAGGGACTATGACTACCGTAAAGAATACAGAAGGGATGACTACCGTAGGGACTTTCGTAGGG
ACGAGTTCAGACGTGGAGGTGGAGAATTTAGAAGAGATTATCGTAGAGATGAGTTTAGAAGGGGT
GGTGGTGAATTTCGTAGAAGTTCTAAAAGGAATTGCACACTTATAGTATACAATTTACCACCTCAA
GTAACTTGGAAGGAACTTAAAGACTTATTCCGTAAGCACGGACGTGTTGTTCGTGCCGATTTAAAA
AATGAGGATAATAGTAGTAAGGAGTTAATAGGAGTTGTAATAATGGAAAACGAGTATGAAGCTAA
GAACGCAATAGACGCATTGAACTTTTGCAATTTTGACGGATATATACTTAAGGTAAACTACGAAAA
CAACGAAacgcgttacccgtacg 3’

Table 4-2.MS analysis of recombinant protein PfGBP2

UniProt ID	Protein name	Origin	Molecular weight (kDa)	MS/MS count
Q8IJX3	Single-strand telomeric DNA-binding protein GBP2, putative	*Plasmodium falciparum* (isolate 3D7)	29.53	96
P60422	50S ribosomal protein L2	*Escherichia coli* (strain K12)	29.86	75
P06992	Ribosomal RNA small subunit methyltransferase A	*Escherichia coli* (strain K12)	30.42	1
P32684	Dual-specificity RNA pseudouridine synthase RluF	*Escherichia coli* (strain K12)	32.476	1
P37765	Ribosomal large subunit pseudouridine synthase B	*Escherichia coli* (strain K12)	32.711	1
P0A910	Outer membrane protein A	*Escherichia coli* (strain K12)	37.2	2

Table 4-3.List of primers used for qRT-PCR to verify the expression of var gene

Target_Gene_ID	Primer_set_Salanti*	Forward_primer	Reverse_primer
PF3D7_1240600	15	CATCCATTACGCAGGATACG	AAATAGGGTGGGCGTAACAC
PF3D7_1300300	20	CACAGGTATGGGAAGCAATG	CCATACAGCCGTGACTGTTC
PF3D7_0412700	27	TAAAAGACGCCAACAGATGC	TCATCGTCTTCGTCTTCGTC
PF3D7_0412900/PF3D7_0413100	28	ACTTTCTGGTGGGGAATCAG	TTCACCGCCACTTACTTCAG
PF3D7_0425800	35	AAACACGTTGAATGGCGATA	GACGCCGAGGAGGTAAATAG
PF3D7_0500100	38	GAAGCTGGTGGTACTGACGA	TATTTTCCCACCAGGAGGAG
PF3D7_0632500	44	ATGTGTGCGAGAAGGTGAAG	TGCCTTCTAGGTGGCATACA
PF3D7_0421100	54	ACGATTGGTGGGAAACAAAT	CCCCATTCTTTTATCCATCG
PF3D7_0421300	34	AAAGGAATTGAGGGGGAAAT	TAAACCACGAAACGGACTGA
PF3D7_0617400	39	ATTTGTCGCACATGAAGGAA	AACTTCGTGCCAATGCTGTA
PF3D7_0412400	26	ACCGCCCCATCTAGTGATAG	CACTTGGTGATGTGGTGTCA
PF3D7_0420700/PF3D7_0420900	31	AGAGGGTTATGGGAATGCAG	GCATTCTTTGGCAATTCCTT
PF3D7_0533100	37	AAGAAAGTGCCACAACATGC	GTTCGTACGCCTGTCGTTTA
PF3D7_0937800	58	CACACGTGGACCTCAAGAAC	AAAACCGATGCCAATACTCC
PF3D7_0600200	94	TGGAAAGAACATGGACCTGA	TTCCTCGAGGGAAGAATCAC
PF3D7_0712800	51	ACGTGGTGGAGACGTAAACA	CCTTTGTTGTTGCCACTTTG
PF3D7_0712400	46	GCGACGCTCAAAAACATTTA	TCATCCAACGCAATCTTTGT
PF3D7_1200400	9	TCGATTATGTGCCGCAGTAT	TTCCCGTACAATCGTATCCA
PF3D7_0100300	97	TCATTATGGGAAGCACGATT	TGATTTCTACCATCGCAAGG

PF3D7_0712000	49	GTTGAGTCTGCGGCAATAGA	CTGGGGTTTGTTCAACACTG
PF3D7_0712600	52	CGTGGTAGTGAAGCACCATC	CCCACCTTCTTGTGGTTTCT
PF3D7_0711700	45	CAATTTTTCCGACGCTTGTA	CACATATAGCGCCGTCCTTA
PF3D7_0712900	50	CACACATGTCCACCACAAGA	ACCCTTCTGTGGTGTCTTCC
PF3D7_0808600	59	CCTAAAAAGGACGCAGAAGG	CCAGCAACACTACCACCAGT
PF3D7_0712300	47	GGTGGAGGTAGTCCACAGGA	CAGCTATTTCCCCACCAGAA
PF3D7_0809100	92	TGCAAGGGTGCTAATGGTAA	CCTGCATTTTGACATTCGTC
PF3D7_0808700	55	TTTGTCCGGAAGACGATACA	ATCTGGGGCAGAATTACCAC
PF3D7_1240300	12	AGCAAAATCCGAAGCAGAAT	CCCACAGATCTTTTCCTCGT
PF3D7_0800200	41	GGTGTCAAGGCAGCTAATGA	TATGTCCTGCGCTATTTTGC
PF3D7_0400400	25	ATATGGGAAGGGATGCTCTG	TGAACCATCGAAGGAATTGA
PF3D7_1100200	7	GACGGCTACCACAGAGACAA	CGTCATCATCGTCTTCGTTT
PF3D7_0100300/PF3D7_0600400/PF3D7_0937600	57	CGTAAAACATGGTGGGATGA	GGCCCATTCAGTTAACCATC
PF3D7_1150400	8	TGCTGAAGACCAAATTGAGC	TTGTTGTGGTGGTTGTTGTG
PF3D7_0632800	93	GACAAATACGGCGACTACGA	TGTTTCACCCCATTCTTCAA
PF3D7_0733000	53	TGACGACGATAAATGGGAAA	TTCTTTTGGAGCAGGGAGTT
PF3D7_0800100	43	GTCGTGGAAAAACGAAAGGT	TATCTATCCAGGGCCCAAAG
PF3D7_1000100	4	GCGTTCTGTGCTCAAACAAA	GTCCTCGTCATCATCGTCCT
PF3D7_1041300	5	GTGCACCAAAAGAAGCTCAA	ACAAAACTCCTCTGCCCATT
PF3D7_1100100	6	GAGGCTTATGGGAAACCAGA	AGGCAGTCTTTGGCATCTTT
PF3D7_1300100	91	ACAAAGGAACGTCCATCTCC	GCCAATACTCCACATGATCG
PF3D7_1373500	19	CGGAATTAGTTGCCTTCACA	CATTGGCCACCAAGTGTATC

PF3D7_0100100	1	TGCGCTGATAACTCACAACA	AGGGGTTCATCGTCATCTTC
PF3D7_0115700	3	AACCCCCAATACCATTACGA	TTCCCCACTCATGTAACCAA
PF3D7_0200100	22	ATGTGCGCTACAAGAAGCTG	TTGATCTCCCCATTCAGTCA
PF3D7_0223500	21	CAATTTTGGGTGTGGAATCA	CACTGGCCACCAAGTGTATC
PF3D7_0300100/PF3D7_0324900	23	CAATCTGCGGCAATAGAGAC	CCACTGTTGAGGGGTTTTCT
PF3D7_0400100	30	GACGACGATGAAGACGAAGA	AGATCTCCGCATTTCCAATC
PF3D7_0426000	36	TGACGACTCCTCAGACGAAG	CTCCACTGACGGATCTGTTG
PF3D7_0900100	56	TGCAAACCACCAGAAGAAAG	GTTCTCCGTGTTGTCCTCCT
PF3D7_1200100	18	CGGAGGAGGAAAAACAAGAG	TGCCGTATTTGAGACCACAT
PF3D7_1219300	11	GACGCCTGCACTCTCAAATA	TTGGAGAGCACCACCATTTA
PF3D7_1255200	17	GCGAGGTCTTCTCGTTCTTG	ATGACGAAGAAGCAGCAGGT
PF3D7_0800300	40	GGAGGAGGAAGAGGAAAACG	CCACCTCCTCTTGTTGTGGT
PF3D7_1240400/PF3D7_1240900	13	AAAGCCACTAGCGAGGGTAA	TGTTTTTGCCCACTCCTGTA
PF3D7_0420700	95	TCACAACCTGACCCCCTACT	TCTTCGTCGTTGTCATCCTC
PF3D7_0937600	96	TGACCAAGACGAAGTATGGAA	TTGATCTCTGTTCGCTGTCC
PF3D7_1200600	10	TGGTGATGGTACTGCTGGAT	TTTATTTTCGGCAGCATTTG

Table 4-4.qRT-PCR results for the expression of var genes expression

***var* gene ID**	**ups**	**DMSO**		**Rapa**	
		Normalized Relative absorbance	**SD**	**Normalized Relative absorbance**	**SD**
PF3D7_1150400	A	0.006846	0.00072	0.069336	0.003908
PF3D7_1300300	A	0.38297	0.620579	0.002306	0.003262
PF3D7_0400400	A	3.81E-05	1.24E-06	0.000101	2.75E-05
PF3D7_0425800/PF3D7_0425800	A	0.007424	0.000304	0.062979	0.006468
PF3D7_0800200	A	0.001688	0.001	0.01041	0.000214
PF3D7_0100300/PF3D7_0937600/ PF3D7_0600400	A	0.004978	0.001217	0.018969	0.001542
PF3D7_0100300	A	8.83E-04	7.11E-05	0.001854	5.09E-05
PF3D7_1200400	BA	0.00987	0.000508	0.0368	0.00242
PF3D7_0632500	BA	0.039093	0.002762	0.186354	0.011394
PF3D7_0100100	B	0.02875	3.22E-03	0.039176	0.003762
PF3D7_0115700	B	0.002214	0.000309	0.007912	0.000282
PF3D7_1000100	B	0.00365	0.00085	0.0122	0.000953
PF3D7_1041300	B	0.013017	0.005042	0.038283	0.003588
PF3D7_1100100	B	0.014201	0.002543	0.048747	0.002898
PF3D7_1219300	B	0.007328	0.000855	0.019518	0.000205
PF3D7_1255200	B	0.001759	0.000269	0.005592	0.000375
PF3D7_1200100	B	0.00139	0.000307	0.00654	0.000408
PF3D7_1373500	B	0.031545	0.001134	0.118664	0.009615

PF3D7_0223500	B	0.00681	0.000712	0.0112	0.00078
PF3D7_0324900/PF3D7_0300100	B	0.012338	0.002936	0.042073	0.004491
PF3D7_0400100	B	0.01194	0.000842	0.052572	0.002263
PF3D7_0426000	B	0.00248	0.000646	0.00378	0.000442
PF3D7_0500100	B	5.47E-02	0.002063	0.122232	0.001836
PF3D7_0800100	B	3.70E-02	0.003017	0.15398	0.005771
PF3D7_0733000	B	0.005742	0.001668	0.046329	0.007527
PF3D7_0900100	B	0.014378	0.001065	0.029281	0.002015
PF3D7_0937800	B	0.032929	0.003711	0.055573	0.004793
PF3D7_1300100	B	0.004828	0.002193	0.004808	0.001014
PF3D7_0632800	B	0.019227	0.002608	0.086262	0.003993
PF3D7_1240300	BC	0.033151	0.003705	0.123187	0.005008
PF3D7_1240400/PF3D7_1240900	BC	0.054844	0.004551	0.197959	0.008731
PF3D7_0712400	BC	0.057291	0.00515	0.195975	0.003605
PF3D7_0712800	BC	0.00723	0.002324	0.042265	0.004461
PF3D7_0421100/PF3D7_0421300	BC C	0.026811	0.002128	0.081456	0.001518
PF3D7_0808700	BC	0.025677	0.000702	0.080158	0.003665
PF3D7_0809100	BC	0.018679	0.002798	0.081226	0.009153
PF3D7_1240600	C	0.155639	0.01334	0.566558	0.002243
PF3D7_0412400	C	0.043421	0.006987	0.10741	0.015462
PF3D7_0412700	C	0.067255	0.008141	0.128199	0.008657
PF3D7_0412900/PF3D7_0413100	C	0.002609	0.001417	0.015151	0.001219
PF3D7_0420700/PF3D7_0420900	C	0.027183	0.000229	0.106235	0.002703

PF3D7_0421300	C	0.053325	0.004031	0.140831	0.010451
PF3D7_0617400	C	9.28E-03	0.004127	0.021886	0.001232
PF3D7_0711700	C	13.8612	1.336397	15.63989	0.783099
PF3D7_0712000	C	0.048359	0.005014	0.077751	0.004082
PF3D7_0712900	C	0.012257	0.001425	0.048495	0.003321
PF3D7_0712600	C	0.066196	0.012882	0.209804	0.028811
PF3D7_0420700	C	0.010512	0.002422	0.070883	0.003509
PF3D7_0533100	D	0.011973	0.003582	2.27E-02	0.00189
PF3D7_1200600	E	0.177737	0.01111	0.187196	0.004976

5. Chapter 5

Table 5-1.List of primers and gblock used in the chapter 5, characterization of PfDNAJb

Primer name	**Purpose,Amplicon , F-forward and R- reverse**	**5' – 3' sequence**
Amplification of Homologous region1 and 2		
PG134	HR1_F	TTAGCTAAGCATGCGGGCCCccttatcatttcatttcttttg
PG135	HR1_R	tgaggagtagcaGAT ccgttcacatctgaattc

PG136	HR2_F	ACGAAGTTATACTAGT attggataaaccaattgg
PG137	HR2_R	TTCTATTTATCTCGAG tcattaaagtgtatcac
PG145	gRNA1_F	AAGTATATAATATTGTAAAAGGAAATCTAAAGGGAg ttttagagctaGAA
PG146	gRNA1_R	TTCTAGCTCTAAAACTCCCTTTAGATTTCCTTTTACA ATATTATATACTT
PG147	gRNA2_F	AAGTATATAATATTGGTTTTGTTTGCTTTATCCGTgtttt agagctaGAA
PG148	gRNA2_R	TTCTAGCTCTAAAACACGGATAAAGCAAACAAAACC AATATTATATACTT
PG149	Sequencing primer for gRNA verification	TATTG TAAAAGGAAATCTAAAGGGA
Verification of Transgenic line		
PG174	Box1_F (Whole_locus _F)	cataacaaaaaaatgtaccttg
PG175	Box1_R	CATATTACCACTGAAATTACC
PG176	Box2_F	CTCTTTCTGTGGGATTTAG
PG182	Box2_R (Whole locus_R)	CTAATAAGAACCACGAAACTAAG
Cloning for recombinant protein		
PG132	pET15b_F	CATCAGTTTCGGGCGCATC
PG133	pET15b_R	GTTTGTTGGGCTAATGCGGC

Gblock synthesized for PfDNAJb to generate floxed recodononized gene

Gblock1

5’
AGTAGACTTAAAAAGGCTTACGATAACAGTAAGGGTAAGGTACTTAAGTTGATTGAATCTAACTCT
ATTAACTTAAATATGAAAAACTATGTAGACTTGGTGGAGAATTCTAACGATTACTGGATTATTCAG
ATTTATTCAGATACAGACTCATTGTGCTTGAGTTTTTCAAATATTTGGGAGGAGTCATTCGACAAAT
ATCACGAGTACATTTCTTTTGGACGTATTAACATTTCTACCGACAAAAAGTTAATTAAGCAGAAGG
TGCCATTCAACGTGAAAATATACCCAACTATTTTTATACTTTCTCCTGATGGTACTTATCAGTTGTA

CTCAAATATATTCAATGCTACATCTAAAGATTTTCAAAATTTCATAACCAGTAATTATCCAAACAA
TATATATGACTTGAAAATGCTTAATGATGCCTTATACAAAATGTCAAGTTTAAACGTAAAGAAGAG
TGGTTACATAGATAAGAACCATCATGTGATACTTTTGACAAACAAGAAGAAATTATCTCTTCAGGC
AAAGCAGATAACATATAAGTTTAATAACATTTATAAAACTTATAGTATTAAATATAACGAGATAG
ATAACATTAGTAACGAATCAATAAAAAAGTCTATAATTGACTCATTGAAGGTTCTTTCAATTAAAA
AAGATGAGTATTTGAAAGAGAACCAGAATATAGATTACTTTATTTTAGTTAATAACAACCAGGTAA
AAATAATTAGGCGTATATCACCTACCAACATTAAGAACGTATATAATGACGCACTTAAAAAGAAC
ATAGTTGAGATAAACTCTATTAATGTGGACAGTGTGTGTTCAACTATGGGAAGTAGGAAGACATAT
TGCTACGTGATTTTCGTAGACAATGTGGACGATGAGAAAAATGTGAACTACATGCGTAAGACCTA
CCAGAACATTAACAGTTCTTACAACAAGTTCGTTTCAAAAAATGGTTCTGAAGAGACCGAGGAAC
AGTTGTTCATTCAACCAGTGTACGTGTTAAAGAAGAACTTAACCAAAAAGTTTAACAAATTCATTA
ATGACAAAACAATAAATAAATATGATTCTTTTTTGTTAGATTATAGTTCTAATACATTTTCTTCTAT
TAACGAAATTAAGAACTTATCAAACTATTCACAGAATAAGGATGACTTGTCATTCTTGCACAACAT
ATACAAAGACATAGAGATATTGAATTTCCAAAAAATACCTAAATACTGCTTACCATTCAATATTAA
TTGCTTATTTAATATAAAGACCACCACTCCATACAAATTGTACAACATTATAAATAGAACAAGTAA
ACTTCAAATTATTTTCTCTTTGATAGTGGGTTTCATGTTGTTCCCTACTTTTAAGGAATACGGTTTCT
TAAAATACGTTTTGTTCGCTCTTTCTGTGGGATTTAGTGTGATTGTTATAAACATAAAAGACTTTAT
TCTTTTATTGGCCAGCacgcgttacccgtacg 3’

Gblock 2

5’
ttcagatgtgaacggtgctgttTCTAAAAGAAAGAGTAAGGGTAAGGACGACATAAACGATATTATGAACGACT
TGAACAagGTAAATAAAAAAAATAATATACAATAACTTCGTATAGCATACATTATACGAAGTTATT
ATATATGTATATATATATATATTTATATATTTTATATTCTTTTAGaaAAAGATGAACGAAAAAAAGA
ATATGAAGGATAAAAAAGACGGTAATTTCAGTGGTAATATGATAGGAGGTATGTATAAAAACAAA
TTACGTTACATTCTTGACAATATTATTATAATTAGTATAATACTTAGTTTCGTGGTTCTTATTAGTTT
TAAGTACCTTGAGGAGAAGTATTCAATTGAGAGTTATTTTGAGGACGGAGACTCTTTTGACTACTA
CGAGGTATTAAAATGCAAAAGGGGTGACTCTATACAAAAAATAAAGAAAAACTATAGGGATTTAT
CAAAACAATATCATCCTGATTCAAATAAAAATTGCAAAGACTGCGACAAAAAATTCCAAGAAATT
ACAAAAGCCTATAAGACATTGAGTGATagtagacttaaaaag 3’

6. Chapter 6

Table 6-1.List of candidates obtained in silico search of RGG motif in P. falciparum

Gene ID	**Product Description**	**Common in Y1H/ DNA pull down assay**
PF3D7_0100100	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0104300	ubiquitin carboxyl-terminal hydrolase 1, putative	
PF3D7_0113600	surface-associated interspersed protein 1.2	
PF3D7_0113800	DBL containing protein, unknown function	
PF3D7_0200100	erythrocyte membrane protein 1, PfEMP1	

PF3D7_0201600	PHISTb domain-containing RESA-like protein 1	
PF3D7_0211700	tyrosine kinase-like protein, putative	
PF3D7_0212300	peptide chain release factor subunit 1, putative	
PF3D7_0216500	conserved Plasmodium protein, unknown function	
PF3D7_0223500	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0305500	conserved Plasmodium protein, unknown function	
PF3D7_0308200	T-complex protein 1 subunit eta	
PF3D7_0310500	ATP-dependent RNA helicase DHX57, putative	
PF3D7_0311100	pre-mRNA splicing factor, putative	
PF3D7_0311400	protein kinase, putative	
PF3D7_0318100	stomatin-like protein	
PF3D7_0318200	DNA-directed RNA polymerase II subunit RPB1	
PF3D7_0320300	T-complex protein 1 subunit epsilon	
PF3D7_0321800	WD repeat-containing protein, putative	
PF3D7_0322400	regulator of initiation factor 2 (eIF2)	
PF3D7_0323100	conserved Plasmodium protein, unknown function	
PF3D7_0324900	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0400100	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0400400	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0401400	rifin	
PF3D7_0405300	liver specific protein 2, putative	
PF3D7_0407400	conserved Plasmodium protein, unknown function	

PF3D7_0409100	U4/U6 small nuclear ribonucleoprotein PRP31, putative	
PF3D7_0409400	chaperone protein DnaJ	
PF3D7_0412100	30S ribosomal protein S12, mitochondrial, putative	
PF3D7_0412700	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0415300	cdc2-related protein kinase 3	
PF3D7_0418600	regulator of chromosome condensation, putative	
PF3D7_0420300	AP2 domain transcription factor, putative	
PF3D7_0420700	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0420900	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0421100	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0424300	erythrocyte binding antigen-165, pseudogene	
PF3D7_0424400	surface-associated interspersed protein 4.2 (SURFIN 4.2)	
PF3D7_0425800	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0426000	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0500100	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0504800	conserved Plasmodium protein, unknown function	Y1H assay
PF3D7_0507100	60S ribosomal protein L4	DNA pull-down assay
PF3D7_0514600	ribose-5-phosphate isomerase, putative	
PF3D7_0517300	serine/arginine-rich splicing factor 1	DNA pull-down assay
PF3D7_0519500	carbon catabolite repressor protein 4, putative	
PF3D7_0519700	FoP domain-containing protein, putative	

PF3D7_0520800	conserved Plasmodium protein, unknown function	
PF3D7_0522400	conserved Plasmodium protein, unknown function	
PF3D7_0524900	S-adenosyl-L-methionine-dependent tRNA 4-demethylwyosine synthase, putative	
PF3D7_0530900	formin 1	
PF3D7_0533100	erythrocyte membrane protein 1 (PfEMP1), pseudogene	
PF3D7_0600300	rifin	
PF3D7_0603000	conserved Plasmodium protein, unknown function	
PF3D7_0609800	palmitoyltransferase DHHC2, putative	
PF3D7_0619400	cell division cycle protein 48 homologue, putative	DNA pull-down assay
PF3D7_0619500	acyl-CoA synthetase	
PF3D7_0621700	conserved Plasmodium protein, unknown function	
PF3D7_0624100	conserved Plasmodium protein, unknown function	
PF3D7_0627800	acetyl-CoA synthetase, putative	
PF3D7_0628100	HECT-domain (ubiquitin-transferase), putative	
PF3D7_0629700	SET domain protein, putative	
PF3D7_0632800	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0700100	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0704000	conserved Plasmodium membrane protein, unknown function	
PF3D7_0705600	RNA helicase, putative	
PF3D7_0706500	conserved Plasmodium protein, unknown function	
PF3D7_0712300	erythrocyte membrane protein 1, PfEMP1	

PF3D7_0712400	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0712800	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0712900	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0723700	metallo-hydrolase/oxidoreductase, putative	
PF3D7_0728600	RING zinc finger protein, putative	
PF3D7_0728900	RNA-binding protein, putative	
PF3D7_0731500	erythrocyte binding antigen-175	
PF3D7_0800300	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0801000	Plasmodium exported protein (PHISTc), unknown function	
PF3D7_0801800	mannose-6-phosphate isomerase, putative	
PF3D7_0802000	glutamate dehydrogenase, putative	
PF3D7_0806800	V-type proton ATPase subunit a, putative	
PF3D7_0807600	conserved Plasmodium protein, unknown function	
PF3D7_0808600	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0809100	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0814200	DNA/RNA-binding protein Alba 1	DNA pull down assay
PF3D7_0818000	U4/U6.U5 small nuclear ribonucleoprotein, putative	
PF3D7_0819400	perforin-like protein 4	
PF3D7_0820500	protein transport protein YIF1, putative	
PF3D7_0823600	lipoate-protein ligase B	
PF3D7_0824300	GTP-binding protein, putative	

PF3D7_0833500	erythrocyte membrane protein 1, PfEMP1	
PF3D7_0907100	conserved Plasmodium protein, unknown function	
PF3D7_0909800	small nuclear ribonucleoprotein Sm D3, putative	
PF3D7_0912200	conserved Plasmodium membrane protein, unknown function	
PF3D7_0928200	conserved Plasmodium protein, unknown function	
PF3D7_0937300	rifin	
PF3D7_0937800	erythrocyte membrane protein 1, PfEMP1	
PF3D7_1000100	erythrocyte membrane protein 1, PfEMP1	
PF3D7_1000800	stevor, pseudogene	
PF3D7_1006800	single-strand telomeric DNA-binding protein GBP2, putative	DNA pull down assay
PF3D7_1016800	Plasmodium exported protein (PHISTc), unknown function	
PF3D7_1018300	conserved Plasmodium protein, unknown function	
PF3D7_1021300	apicoplast integral membrane protein, putative	
PF3D7_1022400	serine/arginine-rich splicing factor 4	DNA pull-down assay
PF3D7_1023100	dynein heavy chain, putative	
PF3D7_1023900	chromodomain-helicase-DNA-binding protein 1 homolog, putative	
PF3D7_1028700	merozoite TRAP-like protein	
PF3D7_1031400	OTU-like cysteine protease	
PF3D7_1033000	conserved Plasmodium protein, unknown function	
PF3D7_1038400	gametocyte-specific protein	

PF3D7_1040200	stevor	
PF3D7_1041300	erythrocyte membrane protein 1, PfEMP1	
PF3D7_1100100	erythrocyte membrane protein 1, PfEMP1	
PF3D7_1100200	erythrocyte membrane protein 1, PfEMP1	
PF3D7_1104200	chromatin remodeling protein	
PF3D7_1105000	histone H4	
PF3D7_1107000	U6 snRNA-associated Sm-like protein LSm4, putative	
PF3D7_1110400	RNA-binding protein, putative	
PF3D7_1111400	conserved Plasmodium protein, unknown function	
PF3D7_1115400	cysteine proteinase falcipain 3	
PF3D7_1118300	insulinase, putative	
PF3D7_1119300	splicing factor U2AF small subunit, putative	
PF3D7_1123400	translation elongation factor EF-1, subunit alpha, putative	
PF3D7_1126600	steryl ester hydrolase, putative	
PF3D7_1132700	50S ribosomal protein L2, putative	
PF3D7_1135100	protein phosphatase PPM8, putative	
PF3D7_1141500	conserved Plasmodium protein, unknown function	
PF3D7_1142300	conserved Plasmodium membrane protein, unknown function	
PF3D7_1148000	serine/threonine protein kinase, putative	
PF3D7_1150400	erythrocyte membrane protein 1, PfEMP1	
PF3D7_1211000	kinesin-7, putative	

PF3D7_1211800	polyubiquitin	
PF3D7_1219300	erythrocyte membrane protein 1, PfEMP1	
PF3D7_1219400	erythrocyte membrane protein 1 (PfEMP1), pseudogene	
PF3D7_1225000	conserved Plasmodium protein, unknown function	
PF3D7_1225900	conserved Plasmodium protein, unknown function	
PF3D7_1229800	myosin J, putative	
PF3D7_1230800	pre-mRNA-splicing regulator, putative	
PF3D7_1232500	CG2-related protein, putative	
PF3D7_1233600	asparagine and aspartate rich protein 1	
PF3D7_1234100	bromodomain protein, putative	
PF3D7_1234500	conserved Plasmodium protein, unknown function	
PF3D7_1240900	erythrocyte membrane protein 1, PfEMP1	
PF3D7_1254000	rifin	
PF3D7_1254400	rifin	
PF3D7_1255200	erythrocyte membrane protein 1, PfEMP1	
PF3D7_1303500	sodium/hydrogen exchanger, Na+, H+ antiporter	
PF3D7_1303800	conserved Plasmodium protein, unknown function	
PF3D7_1307600	DNA-directed RNA polymerase subunit alpha, putative	
PF3D7_1308000	conserved Plasmodium membrane protein, unknown function	
PF3D7_1309700	vacuolar protein sorting-associated protein 18, putative	

PF3D7_1313000	ubiquitin-like protein nedd8 homologue, putative	
PF3D7_1318100	ferredoxin, putative	
PF3D7_1322900	conserved Plasmodium protein, unknown function	
PF3D7_1330800	RNA-binding protein, putative	DNA pull down assay
PF3D7_1333700	histone H3-like centromeric protein CSE4	
PF3D7_1343400	DNA repair protein RAD5, putative	
PF3D7_1344000	aminomethyltransferase, putative	
PF3D7_1344300	zinc finger protein, putative	
PF3D7_1346300	DNA/RNA-binding protein Alba 2	DNA pull down assay
PF3D7_1349500	conserved Plasmodium protein, unknown function	
PF3D7_1352400	conserved Plasmodium protein, unknown function	
PF3D7_1353900	proteasome subunit alpha type-7, putative	
PF3D7_1354400	V-type proton ATPase 21 kDa proteolipid subunit, putative	
PF3D7_1361800	glideosome-associated connector	
PF3D7_1362400	calpain	
PF3D7_1365900	ubiquitin-60S ribosomal protein L40	DNA pull down assay
PF3D7_1366300	conserved Plasmodium protein, unknown function	
PF3D7_1366900	conserved Plasmodium protein, unknown function	
PF3D7_1368700	mitochondrial carrier protein, putative	
PF3D7_1373500	erythrocyte membrane protein 1, PfEMP1	
PF3D7_1407100	rRNA 2'-O-methyltransferase fibrillarin, putative	

PF3D7_1410800	ankyrin-repeat protein, putative	
PF3D7_1411600	GTP-binding protein, putative	
PF3D7_1411900	p1/s1 nuclease, putative	DNA pull-down assay
PF3D7_1412000	p1/s1 nuclease, putative	
PF3D7_1414700	ubiquitin carboxyl-terminal hydrolase, putative	
PF3D7_1417200	NOT family protein, putative	
PF3D7_1417800	DNA replication licensing factor MCM2	
PF3D7_1426700	phosphoenolpyruvate carboxylase	
PF3D7_1429100	apicoplast ribosomal protein L15 precursor, putative	
PF3D7_1433400	zinc finger protein, putative	
PF3D7_1434700	mitochondrial import inner membrane translocase subunit TIM17, putative	
PF3D7_1435300	NAD(P)H-dependent glutamate synthase, putative	
PF3D7_1439100	DEAD/DEAH box helicase, putative	
PF3D7_1445700	conserved Plasmodium protein, unknown function	
PF3D7_1447000	40S ribosomal protein S5	
PF3D7_1450900	acetyl-CoA acetyltransferase, putative	
PF3D7_1454700	6-phosphogluconate dehydrogenase, decarboxylating, putative	
PF3D7_1456600	mitochondrial ribosomal protein L28 precursor, putative	
PF3D7_1456800	V-type H(+)-translocating pyrophosphatase, putative	
PF3D7_1461300	40S ribosomal protein S28e, putative	

PF3D7_1468400	zinc finger protein, putative	
PF3D7_1469600	acetyl-CoA carboxylase, putative	
PF3D7_1472900	dihydroorotase, putative	
PF3D7_1474000	probable protein, unknown function	