Characterisation of the mediator complex in the malarial parasite *Plasmodium falciparum*

Uthra Balasubramaniyan Iyer

2018

Uthra Balasubramaniyan Iyer. (2018). Characterisation of the mediator complex in the malarial parasite *Plasmodium falciparum*. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/73388>

<https://doi.org/10.32657/10356/73388>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**CHARACTERISATION OF THE MEDIATOR COMPLEX IN
THE MALARIAL PARASITE *PLASMODIUM FALCIPARUM***

**UTHRA BALASUBRAMANIYAN IYER
SCHOOL OF BIOLOGICAL SCIENCES**

2018

**CHARACTERISATION OF THE MEDIATOR COMPLEX IN
THE MALARIAL PARASITE *PLASMODIUM FALCIPARUM***

UTHRA BALASUBRAMANIYAN IYER

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2018

ACKNOWLEDGEMENT

First and foremost, I would like to express my gratitude to my supervisor, Dr. Mark Featherstone, for being an excellent mentor to me. He has been encouraging when I was successful, patient when I struggled, and insightful when I needed help. He has been an inspiration to me due to his dedication to research and education, and especially because of his outstanding integrity that is observable in everything he does.

I am also profoundly thankful to Associate Prof. Zbynek Bozdech for his guidance and critical insights over the years as my thesis advisory committee member and for taking over as my supervisor during the final stages of my dissertation.

I am extremely grateful to Asst. Prof. Xavier Roca, for being a part of my thesis advisory committee and for providing valuable suggestions and reviewing my progress. I am greatly indebted to Prof. Gerhard Gruber and Dr. Ardina Gruber for their expert guidance for protein purification experiments carried out in their laboratory. I would also like to convey my gratitude to Associate Prof. Newman Sze and members of the NTU mass spectrometric facility who helped me immensely with my mass spectrometry experiments.

This work would not have been possible without financial support from NTU, in the form of the NTU research scholarship, for which I will be forever grateful.

I would like to thank all the members of the Featherstone lab for help, collaboration, advice, and specially for making these last four years an amazing and rewarding life experience. I want to thank Dr. Ragini Rai and Dr. Sandeep Srivastava for their advice and help with experimental protocols. I extend my warmest thanks to my fellow Phd mates, Changqing and Ying, for their cheerful presence and friendship, and I shall always cherish the camaraderie I have shared with them. I would also like to express my gratitude to Lawrence, for his kind words and constant encouragement. My sincere thanks to Aarthi for helping me with the data analysis of ChIP-Seq results and for being a great friend throughout all these years. I would also like to thank my other wonderful

labmates Reshma, Anusha and Ana for their cheerful banter in the lab and support over the years.

I would like to thank all my friends who have made my life in Singapore truly memorable. I extend my warmest thanks especially to Areetha and Oviya, for being my writing buddies and for always being there for me whenever I needed help.

There are no words to describe the support that I have had from my husband Rajesh. He has been my pillar of strength and has always believed in me. He has been encouraging, understanding and caring throughout all these years and has always managed to bring a smile on my face during tough times. I am also blessed to have wonderful in-laws who have always supported me. Last but not the least, I am indeed greatly indebted to my parents and my sister for all their love and unwavering faith in me. I extend my heartfelt gratitude to them for their selfless support, understanding and encouragement for all decisions in my life. My family means the most to me and I dedicate this thesis to them.

TABLE OF CONTENTS

Acknowledgement	i
List of Figures	ix
List of Tables	xii
Abbreviations	xiii
Abstract	xviii
1 Introduction	1
1.1 Malaria	1
1.2 The life cycle of the malarial parasite	3
1.3 Control and treatment of malaria	6
1.3.1 Preventive measures against malaria	6
1.3.2 Anti-malarial drugs	6
1.3.3 Anti-malarial vaccines	8
1.4 An overview of eukaryotic transcription	9
1.5 Regulation of gene expression in <i>P. falciparum</i>	12
1.5.1 Chromatin landscape	13
1.5.2 Transcriptional control	15
1.5.3 Post-transcriptional regulatory mechanisms	17
1.6 The Mediator complex and its role in eukaryotic gene expression	20
1.6.1 Discovery of the Mediator complex	21
1.6.2 Composition and structural organization of the Mediator complex	23
1.6.3 Functions of the Mediator complex in transcription	27
1.6.3.1 Role of Mediator in transcription initiation	27
1.6.3.2 Mediator as a complex with RNAPII – the holoenzyme hypothesis	29
1.6.3.3 Genome-wide occupancy of the Mediator complex	30

1.6.3.4	Protein-protein interactions of the Mediator complex	33
1.7	Evolutionary conservation of the Mediator complex	36
1.8	The Mediator complex in <i>P. falciparum</i> : What do we know?	38
1.9	Project rationale & research objectives	41
2	Materials & Methods	43
2.1	<i>P. falciparum</i> culture and maintenance	43
2.2	Genomic DNA extraction from <i>P. falciparum</i>	43
2.3	Generation of plasmid constructs	43
2.3.1	Amplification of insert sequences	43
2.3.2	Restriction digestion of plasmid vector	45
2.3.3	Ligation-independent cloning (LIC)	45
2.3.4	Transformation of LIC product	46
2.3.5	Screening of recombinant clones	46
2.3.6	Preparation of recombinant plasmid constructs	47
2.4	Recombinant expression of <i>his</i> -tagged Med31 in <i>E. coli</i>	47
2.4.1	Induction of recombinant protein expression	47
2.4.2	Solubility assay of recombinant protein	48
2.4.3	Preparation of <i>E. coli</i> lysate under denaturing conditions	48
2.4.4	Ni-NTA purification under denaturing conditions	48
2.5	Generation of antibodies against PfMED31	49
2.5.1	Epitope design	49
2.5.2	Conjugation of peptide antigen with an immunogenic carrier	49
2.5.3	Polyclonal antibody production	49
2.6	Synchronization of ring-stage parasites	50
2.7	Transfection of <i>P. falciparum</i>	50

2.8	Total RNA extraction from <i>P. falciparum</i>	50
2.9	Reverse-transcription PCR.....	51
2.10	Genetic analysis of parasite transgenic lines	51
2.11	Cloning by limiting dilution.....	54
2.12	Western blot analysis	55
2.12.1	Preparation of <i>P. falciparum</i> total protein lysate	55
2.12.2	Estimation of protein concentration by Bradford assay.....	55
2.12.3	SDS-PAGE and western blot	56
2.13	Preparation of cytoplasmic and nuclear protein fractions.....	57
2.14	Immunofluorescence analysis	58
2.15	Co-immunoprecipitation	58
2.15.1	Preparation of protein lysate	58
2.15.2	Immunoprecipitation.....	59
2.15.3	Elution of HA-tagged protein	59
2.16	Mass spectrometric analysis	60
2.17	Chromatin immunoprecipitation-Sequencing (ChIP-Seq).....	61
2.17.1	Saponin lysis and formaldehyde cross-linking	61
2.17.2	Chromatin immunoprecipitation.....	61
2.17.3	ChIP-Seq library preparation	63
2.17.4	ChIP-Seq data analysis	63
2.17.4.1	Quality check and data pre-processing.....	63
2.17.4.2	Mapping and peak calling.....	64
2.17.4.3	Features.....	65
2.17.4.4	Extending reads to fragments	65
2.17.4.5	Normalization	65

2.18	Real-time PCR.....	66
3	Generation of tools to study Mediator subunits.....	68
3.1	Introduction	68
3.2	Generation of antibodies against Med31 and Med7.....	68
3.2.1	Recombinant expression of <i>Pf</i> Med31 protein in <i>E. coli</i>	68
3.2.1.1	Induction of protein expression using BL21 cells	69
3.2.1.2	Assay for solubility of <i>his</i> -Med31 protein in BL21 cells	71
3.2.1.3	Induction of protein expression using alternate BL21-derived cells	73
3.2.1.4	Assay for solubility of <i>his</i> -Med31 protein in alternate BL21-derived cells	74
3.2.1.5	Purification of recombinant <i>his</i> -Med31 under denaturing conditions ..	77
3.2.1.6	Ni-NTA Purification of <i>his</i> -tagged MED31 protein.....	78
3.2.2	Polyclonal antibody against Med31	81
3.2.3	Monoclonal antibody against Med7.....	84
3.3	Conditional Knockdown of Med31 protein	85
3.3.1	Endogenous epitope tagging approach.....	85
3.3.2	Genomic loci analysis	87
3.4	Episomal expression of epitope-tagged Med31	92
3.4.1	Episomal DNA and transcript analysis	93
3.5	Episomal expression of epitope-tagged putative Mediator subunits.....	95
4	Functional validation of the Mediator complex in <i>P. falciparum</i>	99
4.1	Conditional knockdown of endogenous Med31 protein	99
4.1.1	Expression of Med31 ^{HA-DDD} protein.....	99
4.1.2	Analysis of protein expression in the presence and absence of stabilizing ligand	101

4.1.3	Growth analysis of parasites in the presence and absence of stabilizing ligand	102
4.2	Analysis of the Mediator complex subunit Med7	103
4.3	Analysis of the Mediator complex subunit Med31	105
4.3.1	Molecular characterization of episomally expressed Med31-HA protein ..	105
4.3.1.1	Protein expression of Med31	105
4.3.1.2	Intracellular localization of Med31	108
4.3.1.3	Immunoprecipitation of Med31 from <i>P. falciparum</i> total protein lysate	112
4.3.2	Identification of protein interaction partners of epitope tagged Med31	113
4.3.2.1	Co-immunoprecipitation of Med31 and mass spectrometric analysis.	113
4.3.2.2	Gene ontology analysis of interacting partner proteins	116
4.3.2.3	Bioinformatic analysis of the putative Mediator subunits.....	119
4.3.2.4	Episomal expression of putative Mediator subunits.....	121
4.3.3	Mediator occupancy across the <i>P. falciparum</i> genome	122
4.3.3.1	Enrichment-based analysis	122
4.3.3.2	Peak-based analysis	137
4.3.3.3	Expression pattern of genes that display exclusive binding by Mediator	143
4.3.3.4	Motif analysis in regions associated with Mediator	145
4.3.3.5	Validation of ChIP-Seq targets by real-time PCR.....	147
5	Discussion.....	150
5.1	Production of antibodies against Mediator subunits	150
5.2	Conditional knockdown of Med31 protein expression	153
5.3	Genome-wide occupancy of Mediator	156
5.3.1	Mediator binds predominantly to coding regions	156

5.3.2	Mediator binds to genes involved in metabolic pathways	160
5.4	Conserved core Mediator complex is present in <i>P. falciparum</i>	163
6	Future Direction	172
7	Conclusion	174
8	References	176

LIST OF FIGURES

Figure 1.1 - Life cycle of the <i>Plasmodium</i> parasite	3
Figure 1.2 - A model of the yeast Mediator complex	21
Figure 1.3 - Cryo-EM map of yeast Mediator complex.....	25
Figure 1.4 - Conservation of Mediator subunits in <i>P. falciparum</i> compared to yeast and human Mediator	39
Figure 3.1 - pNIC28-Bsa4 vector used for recombinant protein expression	69
Figure 3.2 - <i>his</i> -Med31 protein expression in BL21 cells cultured at 18°C (A) and 37°C (B)	71
Figure 3.3 - Solubility of <i>his</i> -Med31 protein in BL21 cells at 18°C (A) and 37°C (B)	72
Figure 3.4 – <i>his</i> -Med31 protein expression in BL21-CodonPlus-RIL strain (RIL) and Rosetta-gami™ 2 (DE3) cells cultured at 37°C and 18°C.....	74
Figure 3.5 - Solubility of <i>his</i> -Med31 protein in BL21-CodonPlus-RIL cells at 18°C (A) and 37°C (B)	75
Figure 3.6 - Solubility of <i>his</i> -Med31 protein in Rosetta-gami™ 2 cells at 18°C (A) and 37°C (B).....	76
Figure 3.7 - Protein solubilisation from inclusion bodies using urea (A) and guanidinium chloride (B)	78
Figure 3.8 - Ni-NTA purification of <i>his</i> -Med31 under denaturing conditions (A & B).....	80
Figure 3.9 - Concentration of imidazole eluted protein fractions	81
Figure 3.10 - Western blot analysis of polyclonal serum against Med31, obtained from two rabbits (A & B)	83
Figure 3.11 - Western blot analysis of Med7 monoclonal cell line	84
Figure 3.12 - Schematic diagram of Med31-pHADB plasmid	86
Figure 3.13 - Endogenous tagging of Med31 by homologous recombination	87
Figure 3.14 - Diagnostic PCR shows integration into the endogenous locus	88
Figure 3.15 - Genomic DNA analysis of clones obtained from limiting dilution	90
Figure 3.16 - Transcript analysis by reverse transcription-PCR.....	91
Figure 3.17 - Plasmid vector map of pBcamHA.....	92

Figure 3.18 - Schematic representation of the HA-tagged Med31 gene present on the pBcamHA plasmid	93
Figure 3.19 - PCR analysis of DNA extracted from wild-type (T996) and drug resistant parasites (Med31-T996)	94
Figure 3.20 - RT-PCR analysis for detection of endogenous Med31 & episomally expressed HA-tagged Med31 transcript.....	95
Figure 3.21 - Schematic representation of the PF3D7_0526800 and PF3D7_1363600 expression constructs.....	96
Figure 3.22 - Restriction digestion and PCR analysis of recombinant PF3D7_0526800-HA and P3D7_1363600-HA clones.....	97
Figure 3.23 - PCR analysis of DNA extracted from parental T996 strain (wild-type) and drug resistant parasites (transfectants)	98
Figure 4.1 - Western blot showing endogenous Med31 ^{HA-DDD} protein expression (A) & Med31 ^{HA-DDD} protein expression across the IDC (B)	100
Figure 4.2 - Med31 ^{HA-DDD} protein stability in the presence or absence of TMP (A) & Immunoprecipitation of endogenous Med31 ^{HA-DDD} protein from asynchronous cultures grown in the presence or absence of TMP (B).....	102
Figure 4.3 - Growth analysis of asynchronous Med31 ^{HA-DDD} parasites in the presence and absence of TMP.....	103
Figure 4.4 - Evaluation of Med7 monoclonal antibody by western blot analysis (A) & .	104
Figure 4.5 - Episomal Med31-HA protein expression (A) & Expression of episomal Med31-HA protein with increasing blasticidin concentration (B).....	106
Figure 4.6 - Episomal Med31-HA protein expression across the IDC	107
Figure 4.7 - Intracellular localization of Med31-HA protein in Med31-HA cell lines cultured with BSD 2.5 µg/ml (A) and 7.5 µg/ml (B).	109
Figure 4.8 - Co-localisation of Med31-HA with a cytoplasmic marker	110
Figure 4.9 - Episomal Med31-HA protein is present in both nucleus and cytoplasm	111
Figure 4.10 - Immunoprecipitation of Episomal Med31-HA protein	113
Figure 4.11 - Co-immunoprecipitation of Episomal Med31-HA protein	114
Figure 4.12 - Immunoprecipitation of episomal PF3D7_0526800-HA parasite line.....	121

Figure 4.13 - Mediator occupancy across the transcription unit for Med31-HA parasites (red) and wild-type T996 parasites (green) during schizont stage.....	124
Figure 4.14 - Profile plot showing mediator occupancy for Med31-HA parasites (red) and wild-type T996 parasites (green) during schizont stage for special groups of genes	127
Figure 4.15 - Overall fragment distribution in gene regions for Med31 cell line and the T996 wild-type control	129
Figure 4.16 - Pattern of fragment distribution in regions 1 kb upstream and downstream of ATG	133
Figure 4.17 - Pattern of fragment distribution in regions 1 kb upstream and downstream of STOP	136
Figure 4.18 - Venn diagram of gene names overlapping between Med31 and T996 ChIP	139
Figure 4.19 - Med31 and T996 ChIP fragment distribution on 407 genes specific to Med31	140
Figure 4.20 - Genes common between Med31 and T996	141
Figure 4.21 - Peak frequency in bins from Med31 ChIP (Bar plot); Expression pattern of genes with peaks in the three genomic locations (Box plot)	143
Figure 4.22 - Expression of genes that exclusively bind Med31 in the Schizont stage....	145
Figure 4.23 - Gene ontology and DNA motif analysis for 3 clusters specific to Med31 .	146
Figure 4.24 - Validation of ChIP-Seq data by real-time PCR for Cluster 1 genes centered around the ATG and STOP codon	149
Figure 5.1 - Schematic diagram of the Mediator complex in <i>P. falciparum</i>	170

LIST OF TABLES

Table 1.1 - Predicted Mediator subunits in <i>P. falciparum</i>	39
Table 2.1 - Primers used for amplifying insert sequences for cloning.....	44
Table 2.2 - Restriction enzymes used for cloning into plasmid constructs.....	45
Table 2.3 - Sequencing primers used for confirming recombinant constructs.....	47
Table 2.4 - List of primers used for genetic analysis of transfectants for episomal expression of epitope-tagged Med31	53
Table 2.5 - List of primers used for genetic analysis of transfectants for conditional knockdown of Med31 expression	53
Table 2.6 – List of primers used for genetic analysis of transfectants for episomal expression of putative HA-tagged Mediator subunits.....	54
Table 2.7 - Reagent composition for preparing SDS-polyacrylamide resolving and stacking gels	57
Table 2.8 - Counts of raw and trimmed reads obtained from Mi-Seq paired-end sequencing.....	64
Table 2.9 - List of primers used for quantitative real-time PCR for validating ChIP-Seq results	67
Table 4.1 - Mediator subunits of <i>P. falciparum</i> identified by Co-IP and LC-MS/MS analysis of Med31	116
Table 4.2 - Gene ontology (GO) analysis of proteins identified by LC-MS/MS analysis	119
Table 4.3 - Mapping statistics using pre-processed reads for ChIP and input DNA	123
Table 4.4 - p-values from Mann-Whitney test for clusters with fragment distribution in regions 1 kb upstream and downstream of ATG.....	134
Table 4.5 - p-values from Mann-Whitney test for clusters with fragment distribution in regions 1 kb upstream and downstream of STOP	137
Table 5.1 - List of Mediator subunits identified by Co-IP and LC-MS/MS analysis	165

ABBREVIATIONS

ACN	Acetonitrile
ACT	Artemisinin combination therapy
Alba	Acetylation lowers binding affinity
AMA-1	Apical membrane antigen 1
ampR	Ampicillin resistance
AP2	Apetala-2
AP2/ERF	Apetala2/ethylene response factor
BLASTp	Basic Local Alignment Search Tool-protein
BRE	TFIIB-recognition element
BSA	Bovine serum albumin
BSD	Blasticidin
C/EBP β	CCAAT enhancer-binding protein β
Caf1	CCR4-associated factor 1
cam	Calmodulin
cDNA	Complementary DNA
CF1	Cleavage factor 1
CHD	Chromodomain helicase DNA-binding
ChEC-Seq	Chromatin endogenous cleavage and high-throughput sequencing
ChIP	Chromatin immunoprecipitation
ChIP-seq	Chromatin immunoprecipitation-sequencing
CKM	Cdk8 kinase module
Co-IP	Co-immunoprecipitation
CPE	Core promoter element
CPSF	Cleavage and polyadenylation specificity factor
CstF	Cleavage stimulatory factor
Ct	Threshold cycle
CTD	C-terminal domain
DAPI	4',6-Diamidino-2-Phenylindole, Dihydrochloride
DDT	Dichlorodiphenyltrichloroethane
DEPC	Diethyl pyrocarbonate

DHFR	Dihydrofolate reductase
DHFR-TS	Dihydrofolate reductase–thymidylate synthase
DPE	Downstream promoter element
DSIF	DRB-sensitivity-inducing factor
dsRNA	Double-stranded RNA
DTT	Dithiothreitol
E-value	Expect value
EAF	ELL-associated factor
ecDDD	<i>E. coli</i> dihydrofolate reductase-based destabilisation domain
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
ELL	Eleven-nineteen Lys-rich leukaemia
EM	Electron microscopy
emPAI	Exponentially modified protein abundance score
EMSA	Electrophoretic mobility shift assay
GFP	Green fluorescent protein
GO	Gene Ontology
GST	Glutathione S-transferase
GTFs	General transcription factors
GuHCl	Guanidinium hydrochloride
HA	Hemagglutinin
HAT	Histone acetyltransferase
HCD	Higher energy collisional dissociation
HCl	Hydrochloric acid
hDHFR	Human dihydrofolate reductase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
<i>his</i>	Histidine
HMG	High-mobility-group
HMGB	HMG box domain

HMM	Hidden Markov model
hnRNP	Heterogeneous nuclear ribonucleoprotein
hp1	Heterochromatin protein 1
HRP	Horse radish peroxidase
HSP	Heat shock protein
HSPGs	Heparan sulphate proteoglycans
IDC	Intraerythrocytic developmental cycle
IDRs	Intrinsically disordered regions
IEDB	Immune Epitope Database
IgG	Immunoglobulin G
Inr	Initiator
IP	Immunoprecipitation
IPTG	Isopropyl β -D-1-thiogalactopyranoside
IPTp	Intermittent preventive treatment in pregnancy
IRS	Indoor residual spraying
ISWI	Imitation switch
K13	Kelch 13
KAHRP	Knob-associated histidine rich proteins
KCl	Potassium chloride
KEGG	Kyoto Encyclopedia of Genes and Genomes
KLH	Keyhole limpet hemocyanin
LB	Luria-Bertani
LC-MS/MS	Liquid chromatography-tandem mass spectrometry
LIC	Ligation-independent cloning
LiCl	Lithium chloride
LLINs	Long-lasting insecticidal nets
LPS	Lipopolysaccharide
MAb	Monoclonal antibody
MBP	Maltose-binding protein
MCS	Multiple cloning site
MES	2-(N-morpholino)ethanesulfonic acid
MPM	Malaria Parasite Metabolic Pathway
MS	Mass spectrometry
MSP1	Merozoite surface protein 1

NaCl	Sodium chloride
ncRNA	Non-coding RNA
NELF	Negative elongation factor
NHE	Na ⁺ /H ⁺ exchanger
Ni	Nickel
NP-40	Nonylphenoxypolyethoxylethanol
NPC	Nuclear pore complex
NTA	Nitrilotriacetic acid
ORF	Open reading frame
P-TEFb	Positive transcription elongation factor-b
p-value	Probability value
PAGE	Polyacrylamide gel electrophoresis
PAI	Protein abundance score
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PE	Paired end
PfBDP1	<i>P. falciparum</i> bromodomain protein 1
PfCRT	<i>P. falciparum</i> chloroquine resistance transporter
PfEMP1	<i>P. falciparum</i> erythrocyte membrane protein-1
PfSR1	Serine/Arginine rich protein
PIC	Pre-initiation complex
PMSF	Phenylmethanesulfonyl fluoride
PSI-BLAST	Position-Specific Iterated Basic Local Alignment Search Tool
PTM	Post-translational modification
PV	Parasitophorous vacuole
PVDF	Polyvinyl difluoride
qPCR	Quantitative polymerase chain reaction
RBC	Red blood cell
REST	RE1 silencing transcription factor
RNA-Seq	RNA sequencing
RNAPII	RNA Polymerase II
RPM	Revolutions per minute
RT-PCR	Reverse transcriptase polymerase chain reaction
SAGA	Spt-Ada-Gcn5-Acetyltransferase

SDS	Sodium dodecyl sulfate
SEC	Super elongation complex
Ser	Serine
SMCC	SRB/MED-containing cofactor complex
SRB	Suppressors of RNA polymerase B
SSM	Signature sequence motif
SWI/SNF	Switch/sucrose non-fermentable
TAFs	TBP-associated factors
TAPs	Transcription-associated proteins
TBP	TATA-binding protein
TBS	Tris-buffered saline
TEMED	Tetramethylethylenediamine
TMP	Trimethoprim
TRAP	Thrombospondin-related anonymous protein
TRAP	Thyroid hormone receptor-associated protein
TREX-2	Transport & Export complex-2
Tris	Tris(hydroxymethyl)aminomethane
TSS	Transcription start site
UAS	Upstream activation sequence
UPR	Unfolded protein response
UTR	Untranslated region
VP16	Herpes viral protein 16
WHO	World health organisation

ABSTRACT

The deadly manifestations of malaria are caused by the apicomplexan parasite, *Plasmodium falciparum*, as it multiplies within red blood cells in the human body. Progression through the intraerythrocytic developmental cycle is characterized by an ordered and reproducible pattern of gene expression. Such a unique expression pattern demands a highly specialized regulatory system involving a synchrony between chromatin dynamics, transcription and post-transcriptional processes. In order to understand how the parasite orchestrates such a tight control over gene expression, it is essential to investigate the role of the Mediator complex which is known to interact with components involved in each of these processes.

The Mediator is a highly conserved multi-protein complex which provides a dynamic interface for protein-protein interactions and can be envisaged as the “integrative hub” of transcription. In other eukaryotes, it functions as a bridge between activators bound to enhancer regions and the general transcription machinery assembled at the promoter. While a Mediator complex of 25 to 30 subunits is conserved from yeast to human, only 14 Mediator subunits are encoded by the *P. falciparum* genome based on computational analyses. It is unknown whether additional Mediator subunits exist and whether the Mediator complex interacts with novel or previously identified transcriptional components in the parasite.

In this study, we have characterized the Mediator complex in *P. falciparum* by studying the protein-protein interactions and genome-wide occupancy of the highly conserved Mediator subunit, Med31. Genome-wide ChIP-Seq analysis of episomally expressed Med31 showed that it occupies the coding region of most genes in the schizont stage of the parasite. Our results show that 407 genes are exclusively associated with the Mediator. Gene ontology analysis revealed that these genes were particularly enriched in metabolic processes such as selenocysteine biosynthesis, tRNA modifications, inorganic cation transport and copper ion transport. Therefore, Mediator seems to be involved in regulation of specific metabolic pathways in the parasite.

Using co-immunoprecipitation and mass spectrometric analysis of episomally expressed Med31, we have shown the physical presence of the Mediator complex in *P. falciparum* for the first time. We have identified 11 out of 14 predicted *P. falciparum* Mediator subunits. Interestingly, two uncharacterized proteins; PF3D7_0526800 and PF3D7_1363600 were also identified which seem to be strongly associated with the Mediator complex and therefore, could be presumptive Mediator subunits that might have evolved for specialized functions in the parasite. Co-IP and mass spectrometric analysis of these proteins may further help in substantiating their identity as *bona fide* components of the Mediator complex. *Plasmodium*-specific Mediator subunits could have a significant role to play in the unique control of gene expression programs and could be explored as potential drug targets.

1 INTRODUCTION

Malaria is a severe public health concern in the developing world. It is an infectious disease caused by *Plasmodium* parasites that are transmitted by infected female *Anopheles* mosquitoes. There are five parasite species that are known to infect humans, of which, *Plasmodium falciparum* poses the maximum threat to human mortality. Efforts for treatment and prevention of malaria are being threatened by development of drug and insecticide resistance. The entire genome of *P. falciparum* has been sequenced and is the subject of research efforts worldwide in the hope of identifying novel therapeutic targets. An understanding of drug resistance, pathogenesis and evasion of host immune responses at the molecular level may provide better tools to tackle this age-old parasite. The clinical symptoms of malaria are due to the blood stage of the disease and genome-wide studies have shown that the parasite has a co-ordinated pattern of gene expression. Analysis of regulators of gene expression may reveal as yet unknown features of its unique gene regulation program. One such regulatory component is the Mediator complex that is known to control gene expression in other organisms. The *Plasmodium spp.* Mediator has not been studied in detail and may shed some light on how the parasite achieves a tight control over its genes. This thesis therefore focuses on the characterisation of the Mediator complex in the human malarial parasite, *P. falciparum*.

The following sections provide an overview of malaria, the *Plasmodium* parasite, gene regulation in *Plasmodium*, functions of the Mediator in eukaryotic transcription and the importance of studying the Mediator complex in *P. falciparum*.

1.1 Malaria

Malaria is a debilitating mosquito-borne disease that has afflicted mankind for millennia. It is caused by unicellular parasites belonging to the *Plasmodium* genus and the phylum, apicomplexa. Malaria has ancient origins with references in Chinese documents dating back to 2700 BC. The etymology of the word, malaria, has been connected to the Italian word meaning ‘bad air’, as malarial fevers were thought to be caused by breathing in air rising from marshy areas. The *Plasmodium* parasite was later identified as the

causative agent of malaria by Charles Louis Alphonse Laveran in 1880. In 1897, Ronald Ross discovered that culicine mosquitoes transmit avian malaria parasites. Research over the years showed that female *Anopheles* mosquitoes transmitted human malaria parasites [1].

Malaria ranks among the most infectious diseases in the world and according to the latest estimates, 212 million cases were reported in the year 2015 leading to 429,000 deaths globally [2]. Two-thirds of the global malaria deaths occur in children who are particularly vulnerable to this disease. Malaria attacks the poorest and most vulnerable populations in the world such as sub-Saharan Africa and parts of Asia. *Plasmodium* species known to infect human beings are *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale* and, more recently, *P. knowlesi*.

Malaria caused by *P. vivax* and *P. falciparum* is the greatest threat to global health. Although *P. vivax* malaria is rarely fatal, it has a wider geographical reach due to its ability to thrive in cooler climates. It was previously thought that *P. vivax* can only invade reticulocytes expressing the Duffy blood group antigen [3]. However, recent studies have shown the existence of Duffy-independent invasion pathways [4]. *P. vivax* parasites can also remain in the liver cells in a dormant form termed as a hypnozoite and cause a relapse of infection after several years. Since *P. vivax* differentiates into gametocytes soon after the merozoites are released from liver, transmission can occur even before the symptomatic blood stage begins. In contrast, *P. falciparum* differentiates much later in the blood stage and therefore disease transmission in *P. falciparum* can be successfully targeted by blood-stage chemotherapy [5].

Patients suffering from malaria develop symptoms such as recurrent episodes of fever, chills, headache, vomiting, diarrhea and anemia. Highest levels of mortality is associated with *P. falciparum* malaria [2]. Infection with *P. falciparum*, if left untreated, can lead to fatal complications such as cerebral malaria, metabolic acidosis and severe anemia, leading to death. Red blood cells infected by *P. falciparum* adhere to the linings of small blood vessels. Attachment to the endothelium protects the parasite from clearance by the spleen. The parasites are capable of sequestering themselves in various organs such

as brain, heart, lung, kidney, placenta and subcutaneous tissues. Such a sequestration of parasites causes an obstruction in blood circulation through tissues. Sequestration of parasites in the brain may lead to cerebral malaria which is a severe neurological complication caused by *P. falciparum* infection. It is characterized by coma, seizures and retinal abnormalities in patients [6]. Infected parasites also adhere to the placenta of pregnant women, causing placental malaria, which may lead to anemia in the mother and low birth weight and high mortality rates in the infants. Compromised blood flow also leads to dehydration and an overall reduction in circulating blood volume in the body. Lysis of red blood cells causes anemia by lowering the amount of oxygen delivered to various tissues of the body [7].

1.2 The life cycle of the malarial parasite

The life cycle of the *Plasmodium* parasite was elucidated over the course of nearly 70 years [1]. *Plasmodium* has a complex life cycle with distinct morphological stages in the mosquito and human hosts (Fig 1.1).

Figure 1.1 - Life cycle of the *Plasmodium* parasite.

The malaria parasite has developmental stages in the female *Anopheles* mosquito (**mosquito stage**) and the human host (**liver stage** and the **erythrocytic stage**). Reprinted from [8], with permission from Elsevier.

Infection by *Plasmodium* begins as an infected female *Anopheles* mosquito takes a blood meal from a human being. During its blood meal, sporozoites present in the mosquito's saliva are deposited into the subcutaneous tissue. The sporozoites travel through the bloodstream for a short time before entering the liver. Inside the liver, the sporozoites migrate through several hepatocytes, finally invading a hepatocyte and forming a parasitophorous vacuole. Parasite surface proteins such as thrombospondin-related anonymous protein (TRAP) and circumsporozoite protein have been shown to interact with heparan sulphate proteoglycans (HSPGs) of liver cells for invasion of hepatocytes. Inside the hepatocyte, the sporozoites multiply and differentiate, ultimately releasing several thousands of merozoites. The merozoites then enter the bloodstream and invade red blood cells in the pathologically severe phase of the infection, the erythrocyte (blood) stage [9].

The development of *P. falciparum* inside an RBC spans a period of 48 h and is referred to as the intraerythrocytic developmental cycle (IDC). The parasite adopts four different morphological forms during the IDC: merozoite, ring, trophozoite and schizont (Fig 1). *Plasmodium* parasites belong to the family of apicomplexans and are characterized by the presence of an apical complex that is involved in invading the host cell.

The merozoite is the invasive form of the parasite, and is characterized by the presence of three major organelles: rhoptries, micronemes and dense granules. The rhoptries are flask shaped organelles located at the apical end, which contain proteins and lipid components involved in the formation of the parasitophorous vacuole (PV). Micronemes are elongated organelles containing proteins required to adhere to the RBCs. Dense granules are round-shaped vesicles containing proteins required for the development of the parasite after invasion. The merozoite is also equipped with a full set

of organelles required for development including mitochondria, apicoplast and nucleus [10].

The invasion process begins when merozoites adhere to RBCs by binding to surface receptors through low-affinity interactions. The merozoites then re-orient themselves such that their apical end is close to the RBC membrane. Through high-affinity ligand-receptor interactions, a tight junction is then formed between the merozoite and the erythrocyte membrane. The action of actin-myosin molecular motors in the parasite moves the tight junction from the apical to the posterior end. The ligand molecules involved in the invasion process are subsequently removed and the parasite forms a protective membrane called the parasitophorous vacuole (PV) for its existence within the host cell [11].

The ring stage (0-16 h) represents the feeding phase of the parasite, wherein haemoglobin from the host cell cytoplasm is digested inside acidic peripheral vesicles. Haemoglobin is converted to heme, which is further broken down to hematin. Hematin is toxic to parasites and therefore is neutralized in the form of hemazoin, a crystal formed of hematin dimers. The acidic peripheral vesicles fuse to form the digestive vacuole, which is observed during the trophozoite stage [12].

The trophozoite stage (17-32 h) is the most metabolically active stage of the parasite accompanied by several cellular and morphological changes. The host RBC membrane is altered for adherence of infected RBCs to the endothelium [13] and also for the uptake of nutrients, which is facilitated by certain ATP translocons present on the PV [14]. Structures called knobs are formed by an array of knob-associated histidine rich proteins (KAHRP) [15] and they provide a scaffold for the major adhesion protein *P. falciparum* erythrocyte membrane protein-1 (PfEMP1). Another characteristic feature of trophozoites is the presence of golgi-like vesicular compartments called Maurer's clefts, containing virulence proteins that are to be delivered to the surface of the infected RBC [12]. DNA synthesis also begins in this stage from 28 – 31 h to prepare for subsequent cell division in the schizont stage [16].

During the schizont stage (33-48 h), the parasite undergoes multiple rounds of nuclear division to produce 16 to 22 nuclei from a single cell [17, 18]. After nuclear division, the infected erythrocyte bursts open and releases the merozoites. During the process of merozoite egress, the infected RBC membrane curls outwards until it turns inside out and thrusts the merozoites outward [19, 20]. In this way, merozoites are released into the bloodstream to initiate another fresh round of infection.

1.3 Control and treatment of malaria

1.3.1 Preventive measures against malaria

Control of malaria relies on our ability to control mosquito populations that transmit malaria parasites. Indoor residual spraying (IRS) and the use of long-lasting insecticidal nets (LLINs) are widely used vector control methods. Usage of nets treated with an effective insecticide creates a physical barrier and an insecticidal effect thereby reducing human-vector contact. Insecticide treated nets have been shown to reduce death rates by 55% in sub-Saharan Africa [2].

In the 1950s and 1960s, malaria was successfully eliminated from several regions of the world including North America and Europe by the use of a synthetic insecticide, dichlorodiphenyltrichloroethane (DDT). Due to concern over environmental and health hazards, the widespread use of DDT for malaria vector control was later restricted [21]. However, based on extensive research and testing, WHO again recommended the use of DDT for IRS in limited situations as its efficacy lasts for more than 6 months.

Another way to prevent malaria is by administration of drugs that can suppress infection. Sulfadoxine-pyrimethamine has been used in intermittent preventive treatment in pregnancy (IPTp) in sub-Saharan Africa. Reductions in maternal anemia and low birth weight have been reported due to IPTp [2].

1.3.2 Anti-malarial drugs

Several drugs were found to be effective in the treatment of malaria such as sulfadoxine-pyrimethamine and chloroquine. Chloroquine was widely used in the

treatment of malaria. Chloroquine affects hemazoin formation in the digestive vacuole of the parasite. However, chloroquine resistance has been observed in all areas of the world. Resistance to chloroquine has been found to occur due to multiple mutations in the chloroquine resistance transporter gene, PfCRT. Other malarial drugs such as cycloguanil and pyrimethamine act by inhibiting the enzyme, dihydrofolate reductase (DHFR), involved in nucleic acid biosynthesis. Resistance to these drugs is prevalent due to mutations in the enzyme. Most of the anti-malarial drugs act against the blood stages of the parasite. There are two drugs that act in the liver stage. These include a combination of atovaquone/proguanil and primaquine. Primaquine is particularly effective in clearing schizonts and also dormant liver-stage parasites found in *P. vivax* and *P. ovale* infections known as hypnozoites [8]. Atovaquone affects the mitochondrial membrane electropotential and electron transport in malaria parasites and proguanil enhances this inhibitory effect on membrane electropotential [22]. Primaquine is also thought to act by disrupting parasite mitochondrial configuration [8].

Artemisinin combination therapy (ACT) is being prescribed as the first line of treatment for uncomplicated *P. falciparum* malaria and also for *P. vivax* malaria in areas with prevalent resistance to chloroquine. Artemisinin and its derivatives have a quick mode of action but are cleared quickly and are therefore combined with other drugs that are cleared at a slower rate. These partner drugs usually include lumefantrine and piperaquine. However, artemisinin resistance has already been detected in five countries: Cambodia, Thailand, Myanmar, Vietnam and the Lao People's Democratic Republic [23]. Resistance to artemisinin is a major impediment in worldwide efforts to combat malaria and the molecular mechanisms behind artemisinin resistance are being investigated. Specific mutations in the Kelch 13 (K13)- propeller domain was found to be a marker of artemisinin resistance [24]. Transcriptomic analysis of clinical isolates collected from patients with acute malaria showed that artemisinin resistance is associated with upregulation of genes involved in the unfolded protein response pathways (UPR), suggesting that the parasite may use these pathways to repair or replenish its proteins that have been damaged due to the oxidative effects of artemisinin [25]. These studies highlight the importance of studying molecular mechanisms underlying drug resistance in the malaria parasite. Emergence of resistance against currently used drugs is troubling and

it is imperative to develop new anti-malarial drugs.

1.3.3 Anti-malarial vaccines

The *Plasmodium* parasites express thousands of antigens at different points in their life cycle, which could be potential vaccine candidates. However, one of the biggest challenges to vaccine development is the parasite's intracellular existence within red blood cells and its ability to escape human immune responses by antigenic variation.

A few *P. falciparum* antigens have been studied as potential vaccine candidates. Erythrocytic stage vaccines have focused on apical membrane antigen 1 (AMA-1) and merozoite surface protein 1 (MSP1) [26]. However, these vaccine candidates did not proceed to clinical trials. The only vaccine that has completed phase III clinical trials is a pre-erythrocytic stage vaccine known as RTS,S. The RTS,S vaccine is based on the *P. falciparum* circumsporozoite protein (CSP) and is composed of two proteins, RTS and S. RTS refers to the chimeric protein formed by a genetic fusion between carboxy-terminal half of CSP (labelled as RT) and the hepatitis B virus gene encoding the virus surface protein (labelled as S). The RTS and S proteins are simultaneously expressed in genetically engineered yeast cells, where they assemble into virus-like particles displaying CSP and S sequences at their surface [27]. In clinical trials, children aged 5 – 17 months who received four doses of the vaccine showed a reduction in severe malaria by 31.5%. RTS, S will be implemented on a pilot level in a few sub-Saharan African countries [2], before large-scale implementation.

With the availability of genome sequences, transcriptomic profiles, protein expression data and advanced genetic analysis tools, it is now possible to investigate the mechanisms behind immune evasion, drug resistance and pathogenesis at a molecular level. This could pave the way for faster identification of potential vaccine candidates. The following section provides a glimpse of eukaryotic transcription before moving on to gene regulation in *P. falciparum*.

1.4 An overview of eukaryotic transcription

Transcription is a complex process that involves the interplay of a multitude of proteins and regulatory complexes. Eukaryotic RNA polymerases require a cohort of proteins called the general transcription factors (GTFs) to assemble at the core promoter before polymerase can begin transcription. Assembly of the GTFs at the core promoter begins with the recruitment of TATA-binding protein (TBP) to a consensus AT-rich sequence called TATA box located near the transcription start site (TSS). TBP and TBP-associated factors (TAFs) are components of TFIID, a GTF that is the first to bind in the sequence of binding events leading to the formation of the pre-initiation complex (PIC). Other GTFs such as TFIIA, TFIIB, TFIIIE, and TFIIF are then recruited to the core promoter. TFIIB binds to regions known as BREs (TFIIB-recognition elements), located both upstream (BRE^u) and downstream (BRE^d) of the TATA box [28]. TFIIA stabilizes the interaction between TBP and DNA at the TATA box. TFIIIE functions closely with TFIIH and is known to stimulate its enzymatic activities involving the C-terminal domain (CTD) kinase and ATPase of TFIIH. TFIIF is similar to the sigma factor in bacteria and can bind to RNA Polymerase II (RNAPII) and stabilize the PIC [29].

The promoters of many eukaryotic genes are now known to lack canonical TATA box elements. In such TATA-less promoters, other core promoter elements (CPEs) such as Initiator (Inr) and downstream promoter element (DPE) can aid in promoter recognition by the transcriptional apparatus. While both TATA-containing and TATA-less promoters are bound by TBP, the SAGA complex may replace most components of TFIID in the recognition of TATA-containing promoters [30].

Finally, TFIIH is recruited which gives RNAPII access to the template strand because of its DNA helicase activity. TFIIH also contains a protein kinase activity, which phosphorylates the CTD of RNAPII. Once RNAPII begins to transcribe the gene, most of the GTFs are released from the DNA so that they can initiate another round of transcription with new polymerase molecules.

RNAPII along with the GTFs is capable of initiating transcription on naked DNA templates *in vitro*. However, in a cellular milieu, packaging of DNA into compact

structures further complicates transcription. In a eukaryotic cell, DNA is wrapped around an octamer of histone proteins to form structures called nucleosomes, which are further coiled and packed together to form higher-order structures. The various structures consisting of DNA packaged into nucleosomes and potentially associated with a wide range of additional proteins are collectively called chromatin. The histone octamer consists of 2 copies each of core histones H2A, H2B, H3 and H4. Each nucleosome encompasses approximately 147 bp of DNA wrapped twice around the histone octamer [31]. While the organization of eukaryotic DNA into chromatin may have been necessary for its packaging and organization within the confines of the nucleus, such chromatin is an impediment to transcription. This has been exploited by eukaryotes to control gene expression via the control of chromatin conformation. Epigenetics refers to heritable changes in gene expression that are not caused by changes in DNA sequence. Epigenetic regulation can occur by alteration of chromatin structure. Two ways by which this is accomplished is through the post-translational modifications (PTMs) of histone tails and the remodeling of chromatin structure.

Post-translational modifications of histones play a crucial role in recruitment of factors regulating transcription, and the spectrum and presentation of such modifications has been proposed to convey a transcriptional state for associated genes via a so-called histone code. The most common PTMs of histones include acetylation (ac), methylation (me), ubiquitination, ADP-ribosylation, sumoylation of lysine (K) residues and phosphorylation of serine and threonine residues. While histone modifications such as H3K9ac or H3K14ac yield an accessible form of chromatin known as euchromatin, H3K9me or H3K27me modifications result in inaccessible chromatin known as heterochromatin. Histone modifications usually affect interactions between DNA and histones and are also recognized by specific proteins that influence chromatin structure [31, 32].

Chromatin structure can also be regulated by the action of protein complexes that alter the interaction between DNA and histones by means of ATP hydrolysis. Such complexes are referred to as ‘chromatin remodelers’. These include complexes such as imitation switch (ISWI), switch/sucrose non-fermentable (SWI/SNF) and chromodomain

helicase DNA-binding (CHD) [33]. The chromatin environment controls the accessibility of the genome to the general transcriptional machinery.

For transcription in a eukaryotic cell, a regulatory protein called a transcriptional activator binds to specific DNA sequences and leads to the recruitment of RNAPII and other GTFs to the core promoter. Activators interact with several co-activators, which help in communicating activation signals to the general transcriptional machinery. These co-activators do not bind DNA directly and function at specific sets of genes. One class of such co-activators includes protein complexes that can alter histone proteins located in promoter regions or remodel chromatin via an ATP-dependent process. Chromatin remodeling complexes and chromatin-modifying enzymes are required to make the DNA accessible to RNAPII and its accessory proteins. The other class of co-activators includes the Mediator complex that aids in recruitment of GTFs and RNAPII to the core promoter. Mediator also helps the activators to communicate efficiently with the general transcriptional machinery at the core promoter. The SAGA complex, already mentioned above as a co-activator that functions in GTF recruitment, also plays a role in chromatin remodeling. The SAGA (Spt-Ada-Gcn5-Acetyltransferase) complex is composed of 21 proteins, and in one study was shown to activate 10% of yeast genes in response to environmental cues [34]. The Spt3/Spt8 sub-module of the SAGA complex is capable of recruiting TBP and SAGA can also acetylate histones in yeast genes [35]. Genome-wide requirement of SAGA in all RNAPII transcription events has been recently uncovered in a study by Bonnet et al. According to the findings of this study, SAGA is involved in acetylation and deubiquitination of promoter and transcribed regions respectively of all expressed genes and is also important for RNAPII recruitment across the genome [36].

Rpb1 is the largest subunit of RNAPII which contains a C-terminal domain composed of multiple repeats of a heptamer sequence (Y-S-P-T-S-P-S)_n. The number of heptad repeats varies across species. For example, humans contain 52 repeats while there are 26 repeats present in *Saccharomyces cerevisiae*. Phosphorylation of the CTD of RNAPII leads to elongation to produce an RNA transcript. The Mediator complex binds to unphosphorylated RNAPII, but it can also strongly stimulate the kinase activity of TFIIH, which phosphorylates Ser5 of CTD concomitant with the release of RNAPII from the

promoter. Kin28, the kinase subunit of TFIIH phosphorylates RNAPII CTD and leads to dissociation of Mediator from PIC. The phosphorylated Ser5 is recognized by the capping enzyme, which catalyzes the addition of a methylguanosine cap to the 5' end of nascent mRNA. Thus, the CTD of RNAPII aids co-transcriptional processing of the RNA transcript.

After its escape from the promoter, the polymerase does not immediately transcribe the entire gene. It has been observed that the polymerase pauses 30-60 nucleotides downstream of the transcription start site. This promoter-proximal pausing is a key regulatory step in transcription. Proteins such as negative elongation factor (NELF) and DRB-sensitivity-inducing factor (DSIF) are known to stabilize the paused RNAPII in higher eukaryotes. Paused RNAPII is released by the action of positive transcription elongation factor-b (P-TEFb) complex, which phosphorylates the CTD of RNAPII at Ser2 and also phosphorylates NELF and DSIF thereby relieving their inhibitory effects on elongation [37].

Other transcription elongation factors such as TFIIIS and elongin also regulate transcript processing by RNAPII. As the polymerase moves towards the 3' end of the coding region, there's a decrease in the Ser5 phosphorylation and an increase in Ser2 phosphorylation mediated by the CDK9 kinase subunit of p-TEFb. Elongation also involves co-transcriptional splicing of the nascent mRNA transcript. Ser2 phosphorylation of CTD of RNAPII also acts as a scaffold for the recruitment of transcription termination complexes such as cleavage and polyadenylation specificity factor (CPSF), Poly (A) polymerase and cleavage stimulatory factor (CstF). The CPSF complex binds to the Poly-A signal causing RNAPII to pause and causes cleavage and release of RNAPII from the gene. CTD Phosphorylation thus plays a role in recruitment of different polymerase accessory factors that are important for elongation and RNA processing.

1.5 Regulation of gene expression in *P. falciparum*

Development of *P. falciparum* within varied cellular environments such as the mosquito midgut, liver and red blood cells is accompanied by coordinated changes in gene

expression. Genome sequences made available for *P. falciparum* and its vector *Anopheles gambiae* [38] have paved the way for genome-wide transcriptional and proteome analysis of several life-cycle stages of the parasite. *P. falciparum* possesses one of the most AT-rich nuclear genomes sequenced to date, consisting of 14 chromosomes and approximately 6,372 protein-coding genes [39]. The AT composition, overall, is almost 80% and can rise to 90-95% in intergenic and intronic regions. Genome-wide transcriptomic analyses have shown that the parasite has a distinct pattern of gene expression wherein more than 75% of genes have highest mRNA levels at only one time-point during the IDC [40-42]. Furthermore, for 30% of genes, a time interval was observed between mRNA levels and protein expression [43]. Thus, the developmental stages of the parasite are accompanied by a unique co-ordinated pattern of gene expression.

In eukaryotes, tissue and cell-type specific gene expression is orchestrated in part by gene-specific transcription factors, which recruit transcription initiation complexes to gene promoters. Remarkably, very few gene-specific transcription factor families have been identified in the *Plasmodium* genome. Gene expression can also be regulated by controlling access of DNA to transcription-related proteins or at the post-transcriptional level. Several studies have been carried out to understand each of these gene regulatory steps in *P. falciparum*.

1.5.1 Chromatin landscape

The *Plasmodium* parasite is able to evade host immune responses by a process of antigenic variation. The multigene family comprising 60 *var* genes encodes variants of the *P. falciparum* erythrocyte membrane protein 1 (PfEMP1). Most commonly, only a single PfEMP1 variant is expressed on the surface of each infected erythrocyte. Thus, out of 60 *var* genes, in most cases, only one is expressed at any given time. The parasite thus evades host immune responses by switching *var* gene expression. Antigenic variation in *Plasmodium* has been attributed to underlying epigenetic mechanisms. It has been observed that silent *var* genes are clustered to repressive regions at the periphery of nucleus and are marked by H3K9me3 and PfHP1. Histone deacetylases, in particular, PfSIR2A and PfSIR2B, class II HDAC protein PfHDA2 have also been shown to regulate

these repressive *var* gene clusters. The variant-silencing SET gene (PfSETvs) controls H3K36me3 on *var* genes leading to silencing of 59 out of 60 *var* genes [44]. The active *var* gene, on the other hand, contains H3K4me3 and H3K9ac marks and is physically separated from the repressive cluster in the nucleus. Gene families encoding variant antigens such as *var*, *rifin*, *stevor* and invasion specific gene families such as *eba* and *clag* are known to be located in these heterochromatin clusters [39].

The *Plasmodium* epigenome has been studied over the years to further understand epigenetic mechanisms of gene regulation. While most chromatin modifications used by the parasite are common to other eukaryotic organisms, there are some features that are unique to the parasite. A complete set of chromatin remodeling proteins has been identified in *P. falciparum* along with histone proteins including core histones, (H2A, H2B, H3, H4) and variant histones (H3.3, CenH3, H2AZ, H2AX and H2Bv). The variant histones H2A.Z and H2B.Z are preferentially enriched at AT-rich DNA regions [45] [46]. Some parasite species such as trypanosomatids also contain a specialized linker histone, H1. This linker histone H1, however, has not been identified in *Plasmodium* yet [31, 47].

Most of the *Plasmodium* genome has an open chromatin configuration, known as euchromatin. Only a small subset of the genome is in the form of condensed heterochromatin [48]. Majority of the parasite genome is enriched in activating histone marks such as H3K4me3 and H3K9ac compared to repressive marks such as H3K9me3 and H3K36me3. The heterochromatin regions contain H3K9me3 histone modifications and are occupied by heterochromatin protein 1 (hp1) [39]. A study of the dynamics of 12 histone modifications across the *P. falciparum* IDC showed that the presence of H3K9ac and H3K4me3 marks in the 5' ends of ORFs correlated positively with transcription, while H4K8ac was associated with putative promoter regions [49]. Histone modifications have been shown to play a critical role in progression through the IDC. This has been demonstrated by treatment of parasites with the drug apicidin which acts by inhibiting the histone deacetylase enzyme. Treatment with the drug showed that normal parasite growth was disrupted with misregulated gene expression levels [50].

Invasion of erythrocytes requires interactions between host cell receptors and

parasite ligands, which is governed by specific gene expression. A parasite-specific bromodomain protein, PfBDP1, has been identified that binds to chromatin at TSSs of genes linked to invasion. PfBDP1 was found to positively regulate transcription by binding to acetylated histone H3 [51].

Global histone levels have also been found to vary during the parasite's asexual developmental cycle and are particularly lower during trophozoite stage, which is transcriptionally active. Schizont stage involves repacking of the genome to prepare for invasion and histone levels are restored to normal levels [39].

These studies have highlighted the importance of epigenetic mechanisms in key parasitic processes such as antigenic variation and host cell invasion. Detailed molecular studies of epigenetic factors and their interactions with other transcription regulatory factors will help in further understanding of the checkpoints of gene regulation.

1.5.2 Transcriptional control

Since the publication of the complete genome sequence of *P. falciparum*, considerable research efforts have been devoted to studying the transcriptional components encoded by the parasite. Most of the components of the basal transcriptional machinery have been identified including the 12 subunits of RNAPII and several subunits of general transcription factors such as TFIIA, -B, -D, -E, -F, -H. Although the TATA binding protein (TBP) has been identified, very few TBP-associated factors (TAFs) such as TAF1, TAF2, TAF7 and TAF10 have been identified [52]. In yeast, the TFIID complex is composed of TBP and TAFs, several of which contain the histone fold domain that is required for TAFs to assemble into heterodimers. None of the classical TFIID subunits containing the histone fold domain have been identified in the parasite suggesting that the TFIID complex is divergent from other eukaryotes. Given the absence of many TAFs, it may be that the parasite has dispensed with certain mechanisms of transcriptional regulation and that alternative mechanisms have evolved.

A limited number of genes encoding specific transcription factors have been annotated in the *Plasmodium* genome. Bioinformatic analyses based on sequence

similarity to known eukaryotic TFs indicated an apparent absence of gene-specific TFs [52, 53]. One of the first *Plasmodium* transcription factors to be identified was PfMyb1, a myb-related protein belonging to the tryptophan cluster family [54]. When a long double-stranded RNA (dsRNA) was used to interfere with PfMyb1 mRNA expression, a reduction in the growth of parasite cultures was observed and the development from trophozoite to schizont stage was affected. PfMyb1 was found to control the transcript expression of several genes including phosphoglycerate kinase and calcium dependent kinase, demonstrating its role in regulating gene expression in the parasite [55].

Other transcription factors that have been identified in *Plasmodium* are the HMGB proteins, which belong to the high-mobility-group (HMG) superfamily and contain a characteristic HMG box domain (HMGB). Instead of binding to specific DNA sequences, HMGB proteins are known to bind to AT-rich DNA sequences and cause bending, thereby altering the nucleosome positioning on DNA. In an initial study carried out before complete sequencing of the *Plasmodium* genome, a *P. falciparum* gene encoding a typical HMG box was reported [56]. Later, four putative HMG proteins, PfHMGB 1-4 were predicted in the *Plasmodium* genome based on sequence homology. Detailed analysis of PfHMGB1 and PfHMGB2 proteins showed that these proteins were expressed in both asexual and gametocyte stages and were present mainly in the nucleus of parasites. The proteins could interact with distorted DNA structures and bend linear DNA under *in vitro* conditions [57]. HMGB2 protein in *Plasmodium yoelii* has also been studied. *In vivo* analysis in vertebrate and invertebrate hosts showed that although parasites lacking PyHMGB2 were able to develop into gametocytes, they were severely impaired in oocyte formation in the mosquito vector. Thus, PyHMGB2 plays an important role in sexual stage gene expression [58].

A major breakthrough in identifying specific TFs in *Plasmodium* came about with the discovery of conserved proteins containing putative plant *apetala 2* (AP2) DNA-binding domains [59]. These proteins are now known as the apicomplexan AP2 (ApiAP2) protein family and comprise 27 members. The ApiAP2 family is the largest class of transcription factors that has been identified in *P. falciparum*. The ApiAP2 family members harbor domains that are structurally related to the DNA-binding domain found in

the plant Apetala2/ethylene response factor (AP2/ERF) transcription factor family [59]. The parasite ApiAP2 family consists of 27 genes, 21 of which are expressed during the IDC [47]. *In vitro* DNA-binding activity of all the *P. falciparum* AP2 domains has been biochemically and computationally characterized using protein binding microarrays and electrophoretic mobility shift assays (EMSA) [60]. Distinct recognition motifs in DNA have been identified for 24 AP2 proteins. However, these DNA-binding interactions need to be further validated *in vivo*.

The functional role of ApiAP2 proteins as transcription factors has been explored in a few studies. The ApiAP2 protein from *P. berghei*, AP2-O has been shown to regulate gene expression by binding to a *cis* motif 5'-TAGCTA-3' in the 5' upstream regions of ookinete-stage specific genes [61]. AP2-Sp in *P. berghei* is involved in the regulation of sporozoite specific genes by binding to the *cis* motif 5'-GCATGCA-3' [62]. In *P. falciparum*, PfSIP2 has been found to bind to the SPE2 motif located upstream of subtelomeric UpsB *var* genes [63]. Two independent studies in *P. falciparum* and *P. berghei* have identified the same AP2 protein, AP2-G, as an important regulator of sexual differentiation [64], [65]. Such studies have strengthened the view of ApiAP2 proteins as important regulators of gene expression.

1.5.3 Post-transcriptional regulatory mechanisms

The apparently reduced spectrum of gene-specific transcription factors and the euchromatic state of the majority of the *Plasmodium* genome has pointed towards post-transcriptional mechanisms of gene regulation. The core components of the translation machinery are conserved in *Plasmodium* spp. with respect to ribosome components, tRNA molecules and aminoacyl-tRNA synthetases [66]. Translation is known to occur in the parasite cytoplasm as well as the apicoplast and mitochondria organelles. Genes encoding rRNAs, tRNAs, limited ribosomal proteins and the bacterial translation elongation factor EF-Tu are present in the apicoplast genome while the mitochondrial genome codes for a few fragmented rRNAs.

Translational control of the blood stages of the parasite has been demonstrated in several studies. Primary evidence for translational control came from the finding that the dihydrofolate reductase–thymidylate synthase (DHFR-TS) protein of *Plasmodium* binds its own mRNA, thereby inhibiting its own translation [67]. Early transcriptomic studies showed that for most *Plasmodium* genes, mRNA levels were at the highest levels during only one time point in the IDC [40, 41]. However, analysis of genome-wide protein profiles showed that for approximately 30% of the total genes analysed, there was a gap of 11 – 18 h between highest mRNA levels and protein expression [43]. Another study, which used microarray and 2-D differential gel electrophoresis protein gels for assessing the mRNA and protein levels respectively, also observed a similar delay of 11 – 15 h between peak mRNA and protein levels [42]. Ribosome profiling studies have also been carried out to understand translational regulation. Bunnik et al., isolated mRNA associated with polysomes from different time-points during the IDC and analysed protein and RNA composition by MudPIT and RNA-seq [68]. The polysome-associated mRNA levels were compared with steady-state mRNA levels. For more than 30% of genes, there was a marked delay between peak transcript abundance in the polysomal sample compared to steady-state mRNA levels. Work carried out in our laboratory used specific monoclonal antibodies against unphosphorylated, Ser5-P and Ser2/5-P forms of RNAPII CTD in ChIP-on-chip experiments to assess the polymerase occupancy across the IDC. RNAPII displayed a simple binding pattern wherein genes could be divided into two classes based on whether they bind RNAPII early or late during the IDC. Comparison of RNAPII occupancy with mRNA abundance showed that a sub-set of genes displayed a high level of RNAPII occupancy and lower levels of corresponding transcripts and *vice versa*. These findings indicated that gene expression might be controlled by post-transcriptional mechanisms rather than by RNAPII recruitment [69].

The half-life of mRNA also has implications in its translational capacities and can be altered during developmental processes and in response to environmental stimuli. Analysis of mRNA decay in *P. falciparum* by Shock et al., [70] showed that the half-life of mRNAs varied from 1 min to 138 min. Interestingly, genes which were involved in the same pathway had the same decay pattern, suggesting coordinated regulation. Also, for a

large number of mRNAs, the half-life increased during the IDC, peaking at the schizont stage.

mRNA decay typically involves decapping, deadenylation and ribonuclease-mediated degradation. Genes encoding mRNA decapping enzymes and components of the CCR4/NOT complex which is involved in mRNA deadenylation are conserved in *Plasmodium spp.* The CCR4-associated factor 1, Caf1, has been found to regulate the expression of more than 1000 genes during the IDC, as Caf1 knockout caused mistimed expression of proteins, particularly those involved in egress and invasion [71], confirming the importance of mRNA metabolism in the control of gene expression in the parasite.

RNA binding proteins also play a key role in translational regulation. Alba (Acetylation lowers binding affinity) family proteins are important in this regard. Alba family proteins are DNA/RNA binding proteins that have been identified in archaea and several eukaryotes as important factors regulating RNA stability and metabolism [72]. PfAlba1 has been shown to bind to mRNAs coding for RAP1, AMA1 and RhopH3 proteins, which are involved in invasion, thereby regulating the stage-specific expression of these genes [73].

An exoribonuclease, PfRNase II, which is structurally related to yeast core exosome component Rrp44, has been identified in *P. falciparum* that has been found to control silencing of upsA var genes [74, 75]. PfRNase II functions independently of the RNA exosome complex and is recruited to promoters and introns of silenced upsA genes. Nascent transcripts resulting from these loci are therefore produced as cryptic RNAs. This gene silencing mechanism was shown to be widespread, as RNA-seq analysis revealed that more than 200 transcripts are upregulated upon depletion of PfRNaseII.

Plasmodium genes contain fewer introns and appropriate homologs of spliceosome components have been identified in *P. falciparum* and *P. berghei* [76]. Alternative splicing has also been found to occur in *Plasmodium* for several genes indicating that this might be a mechanism for diversifying the proteome repertoire [77]. PfSR1, a Serine/Arginine rich protein, has also been identified as a splicing regulator in *P.*

falciparum [78]. PfSR1 was found to localize to both nucleus and cytoplasm and also adjacent to NPC clusters in early stage parasites. RNA immunoprecipitation and microarray analysis of overexpressed PfSR1 showed that it binds to 64 endogenous transcripts, indicating that these transcripts may be regulated by alternative splicing mechanisms [79].

The parasite's ability to traverse a series of hosts and tissues across asexual and sexual life cycles demands a strict control of gene expression, which is influenced by chromatin dynamics, transcription and post-transcriptional processing. Apart from studying these processes individually, it is also essential to understand the interplay between them in order to develop an integrated view of gene regulation. An important co-factor complex that is known to interact with several constituents of each of these processes is the Mediator. Detailed study of the Mediator complex in the malarial parasite may help us to understand how multiple regulatory elements impinge on each other and bring about specific gene regulation.

1.6 The Mediator complex and its role in eukaryotic gene expression

One of the defining features of eukaryotic transcription is the presence of the Mediator complex, which is a multi-subunit complex involved in both basal and activated transcription. Mediator provides a scaffold within the pre-initiation complex and is also known to regulate the activity of RNAPII. It helps in relaying signals from gene-specific transcription factors that are bound to enhancer sites to the general transcriptional machinery assembled at the promoter. The entire complex has a molecular weight of ~1 MDa and is composed of 25 – 30 different proteins.

Figure 1.2 - A model of the yeast Mediator complex

The figure shows the yeast Mediator complex associated with activators/repressors bound to UAS and the general transcription machinery assembled at the promoter region. Reprinted from [80], by permission of Oxford University Press.

A combination of structural, genetic and biochemical studies have shown that the Mediator is organized into four modules: head, middle, tail and a dissociable cdk8 kinase module (CKM) (Fig 1.2). Mediator is conserved in all eukaryotes and has a universal role in gene regulation. Mediator controls different aspects of transcription such as chromatin architecture, RNAPII initiation, re-initiation, pausing and elongation.

1.6.1 Discovery of the Mediator complex

Discovery of the Mediator complex came about with studies in yeast, wherein overexpression of a transcriptional activator GAL4 was found to inhibit expression of genes lacking GAL4 binding sites. This inhibition was termed as ‘squenching’ and was attributed to a decrease in levels of a common target that was required for activation of transcription. Similar results were observed with another activator, herpes viral protein 16 (VP16), which severely reduced transcription of unrelated genes [81]. The overall view was that activators interact with components of the basal transcriptional machinery and overexpression of specific activators leads to sequestration of a common limiting factor

making it unavailable for other activator proteins. It was thought that the common factor could be one of the GTFs or RNAPII since activators were known to interact with them [82].

To identify this common factor, interference between activator proteins was studied in yeast RNAPII transcription system under both *in vivo* and *in vitro* conditions. A hybrid GAL4-VP16 activator containing the DNA binding domain of GAL4 fused with activation domain of VP16 was used. The hybrid activator was found to inhibit transcription by a thymidine-rich binding factor, implying that these two activators compete for a common target. This interference was relieved by the addition of a partially purified yeast extract and not by purified RNAPII or other purified GTFs. This factor present in the partially purified yeast extract was termed as Mediator of activation and was thought to play an important role in the mechanism of transcription activation [83].

In 1991, studies from Roger Kornberg's lab showed that Mediator is indeed required for stimulating transcription *in vitro* by the activators GAL4-VP16 and GCN4. Yeast extracts were resolved into five fractions and together with purified RNAPII were used in a reconstituted *in vitro* transcription system. Although this *in vitro* system was capable of basal transcription, it was not responsive to GAL4-VP16 activator. Addition of the partially purified yeast extract prepared by Kelleher et al., induced a response to the activator and showed that Mediator is required for activated transcription [84].

Indirect evidence of the Mediator complex and its interaction with the CTD of RNAPII also came about from another study which employed RNAPII CTD truncation mutants. Extragenic suppressors of the CTD truncation mutations were identified through genetic screens in yeast and these genes were termed as suppressors of RNA polymerase B (SRB). However, at that time these SRB proteins were thought to be TAFs since they could be isolated as a complex with TBP and RNAPII [85].

Mediator was finally purified to homogeneity in the year 1994, as a complex consisting of about 20 polypeptides. Most of the SRB proteins identified earlier were confirmed to be Mediator subunits [85] [86]. *In vitro* transcription assays with purified Mediator showed that it is required for both basal and activated transcription. Mediator

was also found to interact with the CTD of RNAPII leading to CTD phosphorylation. In the same study, Mediator was found to form a complex with RNAPII as a holoenzyme [87].

1.6.2 Composition and structural organization of the Mediator complex

The organization and structure of the Mediator complex plays a pivotal role in governing its interactions and molecular functions. Initial electron microscopy studies of yeast Mediator showed that the complex has a compact structure. Various biochemical, structural and genetic studies showed that Mediator is structured into four modules termed as head, middle, tail and the CKM. Each module serves varied cellular functions. The tail module is known to contact activators and repressors to regulate transcription. The head and middle modules contact RNAPII and basal transcriptional factors while the kinase module has an inhibitory role in gene expression.

EM analysis of the Mediator showed that it had an extended conformation upon interaction with RNAPII. Mediator contacted CTD of RNAPII through its middle module and the DNA-binding channel of RNAPII through its head domain. The tail module was not involved in contacting RNAPII. Yeast and murine Mediator appeared similar in conformational changes indicating an overall conservation of Mediator structure [88]. Purification of yeast Mediator and subsequent mass spectrometric analysis showed the existence of Mediator complexes with and without the CDK8 kinase module (CKM) [89].

EM analysis of yeast Mediator complex, murine Mediator and human thyroid hormone receptor-associated protein (TRAP) complex showed that the complex was divided into a head and a middle-tail domain. While the head domain was most similar among the three complexes and found to interact with RNAPII, the middle-tail domain displayed a higher structural divergence [90]. Additional structure-based studies of the yeast Mediator-RNAPII holoenzyme showed that Mediator undergoes extensive conformational changes in order to interact with RNAPII. Mediator was found to specifically contact the RNAPII Rpb3/Rpb11 heterodimer [91]. EM analysis of the human Mediator have also detected that it undergoes changes in structure upon interaction with

activators and nuclear receptors [92]. These studies highlighted the contribution of structural changes to its functions. Yeast Mediator was purified using affinity purification and its structure was deduced by reconstructed cryo-EM images. A comparison of free Mediator and RNAPII associated forms showed that Mediator has an inherent flexibility that contributes to structural redistribution of modules to make contacts with RNAPII. Interestingly, yeast and human Mediator complexes were found to be structurally similar despite low sequence conservation between the two species [92].

Owing to its large size, low abundance and complexity, it has been difficult to study the Mediator complex in its entirety. The complete Mediator complex has been mainly studied by electron microscopy. X-ray crystal structures are available for the yeast head module comprising seven Mediator proteins [93] and several other single subunits and subunit segments [94, 95]. Using lysine-lysine crosslinking data obtained from mass spectrometric analysis and partial X-ray structures of available subunits, the model structure of the middle module has also been reported [96]. The structure of a sub-module composed of N-terminal part of Med7 and Med31 was studied by X-ray crystallography, which revealed a unique fold encompassing the right-handed four-helix bundle of Med31 engulfed by two proline-rich stretches present in Med7N. Phenotyping and transcriptomic analysis further showed a distinct functional requirement of this submodule in activated transcription [94].

Structural analyses have revealed various aspects of the modular organization of Mediator. Initial studies showed that the middle and tail modules interacted with each other and form the upper portion of the Mediator. The head module was shown to be present separately as a dense domain at the base of the complex [88, 97]. This modular organization however has contradicted several functional studies. The modularity of the Mediator has been redefined by several studies since then. Work by Wang et al., showed that the head and middle modules actually fold back on one another and are located on the upper part of the Mediator complex and the tail module forms a dense domain at the base of the complex. Surprisingly, the head module was found to be exposed and readily available for interactions with RNAPII and GTFs [98]. This finding contradicted previous

studies, which argued that Mediator undergoes a large-scale rearrangement in its structure for PIC assembly [97].

A definitive description of the complete subunit architecture of Mediator has been obtained from electron microscopy (EM) of purified yeast and human Mediator complexes [99]. The localization and molecular interactions of the constituent subunits were mapped based on known crystal structures, subunit localization and biochemical studies (Fig 1.3). Several important features of the Mediator architecture came forward from this study. Med17 provides an extensive area for contact between the head and the other modules. The subunits, Med7N and Med31 are very important for the interaction between middle and head modules. Med31 present in the middle module is situated close to the portion of the head module that binds to CTD of RNAPII. Med14 forms a core interface connecting components from head, middle and tail regions. The CKM interacts with core Mediator via contacts with the head and the middle modules.

Figure 1.3 - Cryo-EM map of yeast Mediator complex

Mediator complex is arranged into different modules namely head, middle, tail and the dissociable CKM module. The organization of the 25 yeast Mediator subunits have been

shown in the picture. Known X-ray structures of individual subunits and submodules have also been indicated in the cryo-EM map. Reprinted from [99], with permission from Elsevier.

The boundary between head and middle modules is formed mainly by a link between the head module and Med7N-Med31 subcomplex and was identified to be dynamic in nature. Deletion of Med31 showed that the head's neck detached from the middle module resulting in increased mobility of the head. In contrast, the head-tail interface formed by Med17-Med14 was found to be relatively more stable. Overall, deletion and labeling of specific subunits showed that the structural conformation of individual subunits is relatively constant, whereas the interfaces between modules were responsible for major conformational changes. The module segments from the cryo-EM map in the study were fitted onto a published EM map of the holoenzyme formed by yeast Mediator, RNAPII and TFIIF in order to study changes in Mediator conformation upon binding to RNAPII. To allow Mediator to interact with RNAPII, the middle module was found to shift away, causing a rotation in the head module and relocation of the tail module. RNAPII bound subsequently to Mediator in the original location of the middle module. This indicated that Mediator changes its conformation upon binding to RNAPII by altering the structural interfaces between various modules.

Comparison of the yeast and human mediator showed that the overall subunit architecture was conserved between the two species. The core Mediator does not possess any enzymatic activity and plays a mechanistic role in gene transcription, by acting as a scaffold for interaction with multiple proteins involved in transcription. Interaction of multiple factors with the Mediator may lead to a rearrangement between individual modules ultimately affecting interaction with RNAPII and gene transcription [99].

These structural features of the Mediator have been validated by another study that relied on an integrative modeling approach to determine a 3D model of the Mediator complex [100]. The study highlighted the critical roles of Med17 and Med14 in Mediator complex architecture. While Med17 and Med14 were previously classified as head and tail module subunits respectively, the N-terminal regions of these proteins were found to form

part of the middle module. Med17 was found to connect the head and middle modules while Med14 was found to span almost the entire complex connecting all three modules.

A cryo-EM structure of a 15-subunit core Mediator bound to a core initiation complex has also been determined. The head module was seen to bind to RNAPII near the CTD and TFIIB, thereby stabilizing the initiation complex. The Mediator domains were also found to form a structure that positioned CTD and TFIIH in close proximity for TFIIH-mediated CTD phosphorylation [101]. Structure of the complete Mediator-PIC complex has also been determined recently. The complete Mediator-PIC complex was assembled *in vitro* and this 52-protein, 2.5-million dalton complex was also demonstrated to be transcriptionally active. The molecular basis for dissociation of Mediator from RNAPII upon CTD phosphorylation was shown in this study. Interestingly, the CTD of RNAPII was found to pass through a groove located in the head and then through a channel formed between the head module and Med7N-Med31 located in the middle module. In this way, Mediator brings the CTD in close proximity to TFIIK, a subcomplex of TFIIH, for phosphorylation [102].

1.6.3 Functions of the Mediator complex in transcription

1.6.3.1 Role of Mediator in transcription initiation

The Mediator complex plays an important role in both basal and activated transcription. Transcriptional activity in either *in vitro* or *in vivo* systems is designated using the terms ‘basal’ and ‘activated’. In an *in vitro* transcription system, RNAPII requires a set of purified GTFs to transcribe a naked DNA template. The level of transcription achieved with this *in vitro* system is termed as basal transcription. Addition of transcriptional activators which bind to sequences close to the promoter region leads to activated transcription *in vitro*. Activated transcription requires the addition of the Mediator, which helps specific activators to interact with the general transcriptional machinery. Transcription is much more complicated in eukaryotes than prokaryotes due to packaging of DNA into nucleosomes. Basal levels of transcription that occur with a naked DNA template *in vitro* are rarely achieved *in vivo*; basal transcription *in vivo* still relies on specific transcription factors that are present in the promoter regions constitutively [103].

Support for the requirement of Mediator in basal transcription came from studies in yeast. Temperature-sensitive Srb4 (Med17) and Srb6 (Med22) mutants showed a marked decline in total mRNA levels compared to wild-type cells. Since most of the *srb* gene products were found to be in a complex with RNAPII, it was concluded that the *srb*-RNAPII holoenzyme is required for the transcription of the majority of mRNA synthesis *in vivo* [104]. A genome-wide expression analysis in yeast also showed that the expression of all genes was as much dependent on Srb4 (Med17) as it was on the Rpb1 subunit of RNAPII. Therefore, the Srb4-containing RNAPII holoenzyme was considered to be a general requirement for transcription. In the same study, the effect of *srb10* (CDK8), a cyclin dependent kinase was also studied. mRNA levels were found to be higher in *srb10* point mutants compared to wild-type cells, implying that the CDK8 component of the Mediator complex has a negative role in transcription [105].

The human Mediator complex was also shown to be required for basal transcription. Depletion of the Mediator from crude cellular extracts using a monoclonal antibody showed that RNAPII transcription was eliminated. This was observed for all genes, on TATA-containing and TATA-less promoters and both in the presence and absence of activators [106]. In another study, whole-cell extracts prepared from wild type and Srb4 (Med17) temperature-sensitive mutants added to a reconstituted transcription system showed that there was much less transcription in the mutant compared to wild type at restrictive temperatures. All RNAPII promoters were found to be impaired in the mutant yeast strain. Mediator was found to act in concert with other general transcription factors, prior to transcription initiation. These observations highlighted the importance of Mediator in basal transcription *in vivo* and argued that Mediator should be viewed as a general transcription factor indispensable for transcription [107].

Another study showed that Mediator could be recruited to promoters under *in vivo* conditions, based on its interaction with RNAPII and GTFs and independently of the activator. TFIIB artificially tethered to a methionine biosynthetic gene promoter (Met17) was used to study transcription initiation mechanisms in the absence of activator. Mediator was found to be recruited to the Met17 promoter based on its interaction with RNAPII and GTFs. Mutations in the tail module subunits Med2, Med3 and Med14 did not

affect transcriptional activation of the Met17 gene [108].

However, other studies have argued that Mediator is specifically recruited by certain activators and is not a common component of the RNAPII PIC *in vivo*. When cells were grown under optimum growth conditions, Mediator was seen to be poorly associated with highly active promoters of genes encoding glycolytic enzymes or ribosomal proteins. However, activators associated with genes that are activated in response to environmental stress or non-optimal growth conditions were shown to recruit Mediator. Therefore, several, but not all, activators could recruit Mediator and the level of Mediator recruitment did not correlate strictly with RNAPII occupancy and transcriptional output [109]. Promoters, which don't rely on Mediator, might be dependent on TFIID and some promoters may require the presence of both complexes [110].

1.6.3.2 Mediator as a complex with RNAPII – the holoenzyme hypothesis

Mediator was initially purified as a complex with RNAPII and TFIID. This gave rise to the holoenzyme hypothesis according to which this pre-formed complex consisting of Mediator, RNAPII and other transcription factors would be recruited to promoters via Mediator-activator interaction [82]. However, *in vivo* studies have ruled out the role of such pre-formed complexes in transcription. Studies on *Drosophila* heat shock factor (HSF) showed that upon heat shock, the Mediator is recruited to HSF binding sites independently of RNAPII. Indirect immunofluorescence of *Drosophila* polytene chromosomes also revealed different distributions of Mediator and RNAPII [111]. Studies of the HO endonuclease gene in yeast show that the Mediator is recruited without RNAPII by the Swi5 zinc finger protein. Transcriptional activation of the HO gene is complex and involves the sequential recruitment of the Swi5 DNA-binding protein, Swi/Snf remodeling complex, SAGA histone acetyltransferase, and the SBF DNA-binding factor. Swi5 protein was found to interact directly with Mediator and recruit it in an orderly fashion, at the same time as Swi/Snf [112].

The recruitment of proteins to GAL genes in yeast upon induction indicated a staged recruitment process by the activator Gal4. First the SAGA complex was recruited, then Mediator and finally RNAPII along with TBP and TFII-E, F, H. Interestingly, of all the

recruited components, only RNAPII moved into the gene downstream from the promoter. Disruption of the SAGA complex showed that the Mediator was still recruited, pointing towards an independent recruitment of the two regulatory complexes to GAL genes in yeast [113].

These *in vivo* studies were further validated by the purification of yeast Mediator using a combination of ion-exchange and affinity chromatography. The complex contained 21 consensus subunits without the dissociable cdk8/Cyclin C module. Purification of yeast Mediator showed that it existed predominantly in free-form and not as a holoenzyme complex with RNAPII [114]. Another recent study by Petrenko et al., [115], has demonstrated that Mediator stimulates PIC formation and is essential for RNAPII transcription but it is not a stable component of the PIC *in vivo*. Individual yeast Mediator subunits were depleted from the nucleus using the anchor-away technique. It was seen that depletion of individual subunits only led to a modest decrease in global transcription, in contrast to previous studies. The effect of Mediator depletion was more prominent at SAGA-dependent genes than TFIID-dependent genes. The study suggests that RNAPII transcription proceeds even in the absence of Mediator at the PIC indicating that Mediator is not an essential component of the PIC *in vivo*. Although Mediator has an activity that stimulates transcription, it is not absolutely essential for association of RNAPII with promoters and therefore cannot be defined as a general transcription factor.

1.6.3.3 Genome-wide occupancy of the Mediator complex

Mediator acts as a bridge between certain transcription factors bound to distal regulatory elements and the general transcriptional machinery assembled at the core promoter. Mediator does not bind DNA directly and interacts with regions of the genome indirectly via activator and basal transcriptional proteins. These interactions can be captured by chromatin immunoprecipitation experiments.

Studies with GAL genes regulated by the Gal4 activator showed that Mediator associates with upstream activation sequences (UAS) and not with the core promoter. This association with the UAS was found to be independent of the RNAPII machinery

assembled at the core promoter. ChIP-PCR analysis using Srb6 (head), Med7 and nut1/Med5 (middle) and Rgr1/Med14 (junction between middle and tail module) subunits showed that all the four subunits were preferentially associated with the UAS, showing that the entire Mediator complex is involved in activation [116].

Genome-wide occupancy of Mediator was studied by ChIP-microarray analysis of subunits representing each of the four modules. The entire core Mediator complex was found to occupy regions upstream of both active and inactive genes and also gene coding regions (~1200 binding targets). This highlighted the fact that the role of Mediator is not limited to activated transcription, but it may also be pre-recruited to inactive genes for prompt activation of genes in response to external cues [117]. However, Mediator binding to numerous genes at the promoter and coding regions was challenged by another study by Fan et al., [109], who found that Mediator is bound to a small number of genomic regions in synthetic complete medium. Observations from their study showed that the targets identified by Andrau et al.,[117], could be false-positives arising due to artefacts from immunoprecipitation. Mediator presence in coding regions was attributed to non-specific effects caused by accessibility due to nucleosome depletion [118]. To further investigate the involvement of Mediator in basal transcription *in vivo*, Mediator binding to constitutively expressed genes in yeast was examined using Med17/srb4 temperature-sensitive mutants. Mediator was associated with constitutively expressed genes and shown to be important for recruitment of RNAPII to these genes. This study also argued that different growth conditions for yeast had little effect on Mediator association with constitutively expressed genes [119].

It has been difficult to reach a consensus in the field as to whether Mediator has a universal role in transcription or if it has a defined role in activation of specific genes. Conflicting observations of Mediator binding could have been due to stable interactions with activators and transient interactions with the general transcription machinery *in vivo*. This was experimentally demonstrated by another study of Mediator occupancy in the yeast genome. It was shown that Mediator stably associates with UASs but only transiently interacts with core promoters. The observation that it could affect PIC assembly and occupied promoters in a manner similar to RNAPII was however consistent

with its role as a general coactivator complex. Mediator was found to accumulate at core promoters together with RNAPII, by depletion of kin28 kinase activity or in cells in which the RNAPII CTD was mutated at Ser5 residue. Mediator occupancy at the promoter is transient due to kin28-mediated phosphorylation of RNAPII CTD at Ser5. The Ser5 phosphorylation leads to detachment of Mediator from the core promoter after which the polymerase escapes the promoter and begins transcription [120] [121].

Ambiguous results have been obtained for genome-wide binding pattern of Mediator using chromatin-immunoprecipitation. A ChEC-Seq (chromatin endogenous cleavage and high-throughput sequencing) approach showed that Mediator mostly occupied the UASs rather than core promoter or gene coding regions. At most genes, Mediator binding to UAS correlated weakly with expression levels, suggesting that Mediator occupancy is largely uncoupled from gene expression. Mediator and TFIID were also found to have a mutual dependence in recruitment to several genes [122].

Only 15% of yeast promoters contain the consensus TATA element. Such promoters are found in stress-response genes and are regulated by the SAGA complex. The remaining TATA-less promoters are enriched for housekeeping genes and are regulated by the TFIID complex. Genome-wide ChIP-Seq analysis of Mediator and RNAPII across the yeast genome using Med17 temperature-sensitive yeast, med3 Δ med15 Δ yeast and wild-type yeast, showed that SAGA-dependent genes have a greater dependence on Mediator tail module subunits for transcription, compared to TFIID-dependent genes. It was also demonstrated that the tail module could still be recruited to genes in the absence of intact mediator [123].

While all modules of the complex are recruited to UAS, association of the Mediator with the PIC is accompanied by release of CKM, the CDK8 kinase module. Several studies have led to the view of CKM as a negative regulator of transcription. Human Mediator complexes containing CKM were found to repress *in vitro* transcription. Genome-wide profiling of expression showed an upregulation of gene expression in CKM mutants. Biochemical and structural studies have shown that CKM and RNAPII bind in a mutually exclusive manner with the Mediator [124].

The CKM consists of the subunits Med12, Med13, CDK8 and cyclinC. Studies have shown that the CKM is a dissociable module and associates with the core Mediator complex only under certain conditions. Two types of Mediator complex are known to occur in cells, one with the CKM and one without it [125]. *In vitro* reconstitution assays showed that Mediator lacking the CKM module stimulates basal transcription whereas Mediator containing the CKM represses transcription. Interaction of CKM with Mediator inhibits CTD-dependent binding of RNAPII to a binding site in Mediator middle module. Thus, CKM module has a repressive role in holoenzyme formation [126]. A recent study has shown that the kinase module associates with the Mediator at enhancer regions and dissociates from Mediator to allow binding to RNAPII and other PIC components [127]. Jeronimo et al., analysed wild-type and mutant yeast in which RNAPII promoter escape was blocked to study the composition of transient Mediator forms [128]. They found that the CKM is essential for regulating Mediator association with the UAS rather than the promoter. In addition, tail-less Mediator was capable of interacting with the core promoter. Therefore, the head and middle modules carry out basic functions of Mediator at the core promoter, while the CKM and tail modules play regulatory roles.

1.6.3.4 Protein-protein interactions of the Mediator complex

Although Mediator was discovered as a protein complex required for activation of transcription, several studies over the years have shown that it has functions beyond transcription initiation. It is also linked to several other processes such as transcription elongation and termination, mRNA processing and chromatin remodeling.

Apart from initiation, release of RNAPII from promoter proximal sites to the gene body is an important rate-limiting step in the mechanism of gene regulation. Genome-wide analysis in *Drosophila* and human cells have shown that the polymerase pauses at sites 30-50 nucleotides downstream from the transcription start site after escape from the promoter, for several genes [129]. The negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF) function in promoter-proximal pausing of RNAPII. This negative regulation is overcome by the positive transcription elongation factor b (P-TEFb), which functions as part of a larger multi-subunit complex called the

Super Elongation complex (SEC) comprising ELL and EAF family of proteins.

Multiple studies have pointed towards the post-initiation functions of the Mediator. While paused RNAPII was found to be constitutively present in *Drosophila* heat shock genes, Mediator was recruited only after heat shock and was concomitant with the release of paused RNAPII. Therefore, it was proposed that Mediator might act co-operatively with positive elongation factors such as P-TEFb to release paused RNAPII. Furthermore, genome-wide ChIP studies have also shown Mediator to occupy not just the promoter regions, but also the gene coding regions in yeast and human cells.

Mediator has also been shown to overcome the negative inhibition by DSIF in a reconstituted system. The metazoan-specific subunit Med26 was reported to interact with TFIID and elongation complexes such as ELL/EAF and p-TEFb thereby facilitating the release of paused RNAPII from promoter proximal sites. The activation domain of the tumor suppressor gene, p53, has been shown to activate paused RNAPII in an *in vitro* setting. Interestingly, p53 activation domain has been shown to cause structural changes in the conformation of the Mediator highlighting an activator-specific role in transcription elongation [130]. Also, the free CDK8 module and CDK8-containing Mediator complex has been shown to interact with P-TEFb leading to elongation by RNAPII [131]. Studies have also shown that Mediator interacts with TFIIS, an elongation factor that has been implicated in both transcription initiation and elongation. In the absence of Med31, the TFIIS RNAPII binding domain was found to be essential for the recruitment of RNAPII to several genes in yeast [132]. PIC assembly occurs within a nucleosome-free region located near the +1 nucleosome downstream of the transcription start site. Mediator co-operatively functions along with elongation factor SII, which alters the catalytic properties of promoter-proximal RNAPII and enhances transcription of nucleosomal DNA [133].

Mediator is a key element of the PIC and is known to recruit factors that can modify or remodel chromatin. Interactions of the Mediator with the nucleosome have been reported in several studies. The yeast Med5 subunit (Nut1) was shown to possess a histone acetyltransferase activity [134]. Pure wild-type Mediator has also been shown to form stable complexes with the tails of histones H3 and H4. Mediator and nucleosome occupancy has been shown to have a broad correlation *in vivo*, except for nucleosomes

acetylated at histone H4 lysine 16. Therefore, structures of chromatin and acetylation states of histones were found to be important for Mediator occupancy [135].

Mediator also works in co-operation with other co-activator complexes. Mediator interacts with both the chromatin-remodeling complexes swi/snf and SAGA in yeast. In human cells, core Mediator was found to interact with the HAT complex STAGA (SPT3-TAF9-GCN5-acetylase). Mediator also acts synergistically with p300 complex in activation by HNF4. Under *in vitro* conditions, Mediator is known to interact with both p300 complex and TFIID in Gal4-VP16 mediated activation. Murine Mediator also interacts with Chd1, a chromatin remodeling protein, which binds to nucleosomes with a H3K4 trimethylation marking the beginning of active genes [131].

Evidence for how Mediator is able to interact with multiple proteins at distant sites came from studies, which demonstrated the role of Mediator in gene looping. Gene looping refers to the juxtaposition of 5' and 3' ends of genes that are transcribed by RNAPII and the connections formed between enhancer and promoter regions. Another key Mediator interaction in this regard is with cohesin, which is capable of forming rings that can connect two DNA segments. Together, they have been shown to form a loop between enhancers and core promoter regions in murine embryonic stem cells, which is particularly important for the expression of cell-type specific genes. Mediator-cohesin complexes are linked with the cohesin-loading factor, Nipbl responsible for loading of cohesin at promoters [136]. The importance of such gene loops has been strengthened by characterization of super-enhancers, which depend on Mediator and master transcription factors and determine expression of lineage-specific genes in embryonic stem cells [137]. Mediator has also been shown to interact with ncRNAs to facilitate chromatin looping and activation of ncRNA targets in a *cis*-mediated manner [138].

Processing of nascent mRNA transcripts includes capping, polyadenylation, removal of intron sequences and transcript termination. Splicing processes occur co-transcriptionally as splicing complexes are recruited based on recognition of splicing signals. The contribution of Mediator to the splicing process was discovered by a study, which found that the tail subunit Med23 directly interacts with the RNA binding splicing regulator, hnRNP L. Med23 also partially co-localized with hnRNP L and components of

the splicing machinery [139]. Another subunit, Med18, has also been implicated in proper termination of transcription as recruitment of polyadenylation factors involved in RNA cleavage was affected in Med18 null yeast [140]. Based on chromosome conformation capture analysis, it was found that gene looping was critical for termination of transcription by Med18 [141]. In addition to its classical role in the establishment of PIC, Mediator also has diverse roles in multiple stages of transcription which include elongation, termination, mRNA processing, gene looping and epigenetic modifications.

1.7 Evolutionary conservation of the Mediator complex

Initially discovered in yeast, the Mediator complex has since then been purified from other mammalian systems. Structural analysis showed that Mediator complexes purified from yeast and mammals look strikingly similar. However, during preliminary studies, only 8 out of 25 proteins of the human TRAP/SMCC were found to be orthologs of *S. cerevisiae* Mediator subunits. After the complete genome sequences were made available for several organisms, an extensive search was carried out to identify metazoan orthologs of yeast Mediator subunits. Eukaryotic species such as yeast (*S. cerevisiae*, *Candida albicans*, and *Schizosaccharomyces pombe*), filamentous fungi (*Neurospora crassa* and *Aspergillus fumigatus*), the Plant *Arabidopsis thaliana*, the nematode *C. elegans*, and the fly *Drosophila melanogaster*, as well as mice and humans were considered for the analysis. This extensive search among multiple species led to the identification of 12 additional Mediator subunits. The conservation of 20 out of 24 Mediator subunits across diverse organisms such as yeast, insects, worms, plants and mammals showed that a universal Mediator complex was formed early in evolution of eukaryotes [142].

There were some doubts about Med31 being a *bona fide* component of the Mediator complex, as it was not identified as a subunit in the *S. cerevisiae* Mediator complex preparations. However, specific analysis of Med31 based on affinity purification and antibody against the subunit showed that it is indeed a stable component of both *S. cerevisiae* and *S. pombe* Mediator complexes. Med31 orthologs were detected across different species using BLAST and ClustalX and a highly conserved 69 amino acid

sequence motif, rich in aromatic residues and leucine was identified. This Med31 motif was required for stable association with rest of the Mediator complex; however, the motif alone is not enough for association with the entire complex [143].

The ancient origins of the Mediator complex have been revealed by a seminal study carried out by Henri Marc Bourbon [144]. He studied genome sequences from nearly 70 eukaryotes, including fungi, algae, amoeba, diatoms, parasitic protists, oomycetes and animals using tools such as PSI-Blast, BlastP and tBlastN. Primary sequences or signature sequence motifs (SSMs) of known Mediator subunits were identified and used as queries for the search. SSMs used in this study were conserved sequence motifs found within subunits containing at least 7 amino acid residues and satisfied the following conditions: (i) SSMs were present in a minimum of 34 out of 40 aligned sequences for fungi and animal subunits, (ii) SSMs were present in 16 out of 20 aligned sequences of subunits, Med2, Med3 and Med5, (iii) SSMs were present in 17 out of 20 aligned sequences of animal-specific subunits Med23-30.

Sequences of known Mediator subunits were used as queries against the genome and proteome databases of 20 metazoans and 20 fungal species in analysis carried out with PSI-Blast, BlastP and TBlastN. Protein sequences obtained with this approach were labeled as putative Mediator subunits. A total of 22 subunits were conserved in animals as well. Complete primary sequences or SSMs of *bona fide* or putative fungal or metazoan Mediator subunits were then used as queries for search strategies using PSI-Blast, BlastP and TBlastN. The results showed the presence of a core Mediator complex composed of 17 subunits, conserved throughout evolution.

Interestingly, Mediator complex subunits were detected even in organisms such as *T. vaginalis* and *E. histolytica*, which lack RNAPII heptad CTD repeats. Also, CDK module components such as Cdk8, CycC, Med12 and Med13 have been identified in these organisms. These observations indicate that Rpb1 CTD heptad repeats were not initially required for assembly of RNAPII -Mediator holoenzyme. Cdk8 also does not seem to be incorporated into the Mediator complex as a designated CTD kinase.

Overall, the study by Bourbon, demonstrated the presence of a core Mediator

complex composed of 17 subunits, conserved throughout evolution. Mediator as a four-module complex has ancient origins and may have been responsible for species-specific differences [144].

1.8 The Mediator complex in *P. falciparum*: What do we know?

The overall structure of the Mediator complex is conserved across eukaryotes. Twenty out of 24 yeast Mediator subunits were conserved through a large spectrum of eukaryotic species including plants, insects, worms and mammals [142]. This is further validated by the structural similarities between yeast and mammalian mediator complexes as visualized by electron micrography studies [142]. Based on sequence and structural similarity, few putative Mediator subunits have been identified in *P. falciparum*. *In silico* predictions using 109 Pfam HMM profiles led to the identification of 202 transcription-associated proteins (TAPs) in *P. falciparum* [52]. This study identified two Mediator subunits, Med7 and Med31 as polymerase accessories. The mediator subunit, Med6, has been annotated by PlasmoDB. In the study by Bourbon mentioned above, 14 Mediator subunits were identified in *P. falciparum* as summarized in Table 1.1. A comparison of yeast, human and *P. falciparum* Mediator subunit composition has been shown in Fig 1.4. All the head subunits present in yeast are conserved in the parasite. Middle subunits except Med1 and Med19 are conserved in *P. falciparum*, compared to yeast. Tail subunits are known to contact activators for specific gene regulatory pathways. Intriguingly, Med14 is the only tail subunit that has been identified in the parasite. The dissociable CKM module has not been identified in *P. falciparum*.

	Subunit	PlasmoDB Gene ID	Function
Head	Med6	PF3D7_1469700	putative MED6
	Med8	PF3D7_0411100	unknown function
	Med11	PF3D7_0505900	unknown function
	Med17	PF3D7_0520200	unknown function
	Med18	PF3D7_1213200	unknown function
	Med20	PF3D7_1463000	unknown function
Middle	Med22	PF3D7_1469800	unknown function
	Med4	PF3D7_1465200	unknown function
	Med7	PF3D7_0822100	putative MED7
	Med9	PF3D7_1446700	unknown function
	Med10	PF3D7_0707600	unknown function
Tail	Med21	PF3D7_1126400	large ribosomal subunit processing factor, putative
	Med31	PF3D7_1475000	putative MED31
	Med14	PF3D7_0709300	Cg2 protein

Table 1.1 - Predicted Mediator subunits in *P. falciparum*

Subunits belonging to the head, middle and tail modules have been identified in the parasite. The corresponding PlasmoDB gene ID and function has been indicated in the table [52, 144].

Yeast				Human				<i>P. falciparum</i>		
Head	Middle	Tail	CKM	Head	Middle	Tail	CKM	Head	Middle	Tail
Med6	Med1	Med2	Med12	Med6	Med1	Med14	Med12/12L	Med6	Med4	Med14
Med8	Med4	Med3	Med13	Med8	Med4	Med15	Med13/13L	Med8	Med7	
Med11	Med7	Med5	Cdk8	Med11	Med7	Med16	Cdk8/Cdk19	Med11	Med9	
Med17	Med9	Med14	CycC	Med17	Med9	Med23	CycC	Med17	Med10	
Med18	Med10	Med15		Med18	Med10	Med24		Med18	Med21	
Med20	Med19	Med16		Med20	Med19	Med25		Med20	Med31	
Med22	Med21			Med22	Med21			Med22		
	Med31			Med27	Med31					
				Med28	Med26					
				Med29						
				Med30						

Figure 1.4 - Conservation of Mediator subunits in *P. falciparum* compared to yeast and human Mediator

Subunits belonging to the head, middle, tail and CKM modules have been indicated in the figure. (Subunit composition of yeast and human Mediator has been adapted from [99]; Predicted Mediator subunits in *P. falciparum* identified by Bourbon, H.M [144], have been shown in the figure for comparison.)

Head module subunits such as Med17, Med18 and Med22, which are required for interaction with RNAPII seem to be conserved in the *Plasmodium* Mediator complex [145], [93]. Med1 has been shown to interact with the GATA family of transcription factors [146]. Interestingly, both the Med1 subunit and GATA zinc finger motifs have not been identified in the *Plasmodium* genome. The RNAPII in *P. falciparum* does not contain the canonical CTD repeat structure and contains fewer tandem heptad repeats, indicating that these repeats originated recently within the parasite or in an immediate ancestor [147]. Interestingly, Med31 is highly conserved in the parasite. Therefore, Mediator is an ancient factor that could have been involved in relaying signals between activators or repressors to RNAPII in parasites that diverged before the establishment of the canonical heptad CTD structure in eukaryotes [143].

The tail module of Mediator is involved in making contacts with gene-specific activators bound to distal enhancer regions. The only tail module subunit that has been identified in the *Plasmodium* genome is Med14, which is known to contact all the Mediator modules. Med14 in *P. falciparum* contains a very large C-terminal extension. It is one of the largest subunits and is composed of 2729 amino acid residues. This large size supports the fact that it acts as a scaffold for the entire complex and connects the head, middle and tail modules. The Med2-Med3-Med5 tail triad has not been detected in the parasite. This tail triad has been implicated in species-specific regulatory signals [144]. It is unknown whether additional tail subunits exist or if Med14 is able to contact activators on its own. The lack of most of the tail subunits raises important questions about transcriptional regulatory mechanisms in *Plasmodium*. It leads us to wonder whether the parasite lacks gene-specific activators similar to the ones identified in other eukaryotes or if it has highly diverged tail subunits that contact activator proteins required for specialized functions.

None of the subunits of the CDK8 kinase module have been identified in *Plasmodium*. Studies in *yeast*, *Drosophila* and human cells have highlighted the role of CDK8 module in activation and repression of transcription. Phosphorylation of CTD by Srb10 disrupts the formation of pre-initiation complex. Phosphorylation of RNAPII CTD at serine 2 and serine 5 was found to be inhibited when CDK8 of human Mediator

complex was knocked down using RNAi, indicating the involvement of the CKM in transcription elongation as well [146, 148]. The subunits Cdk8, CycC, Med12 and Med13, belonging to the CKM, have been identified in organisms such as *T. vaginalis* and *E. histolytica* which lack RNAPII heptad CTD repeats. These observations point to the presence of a separable Cdk8 kinase module that appeared early in evolution but was subsequently lost in some phyla or species, similar to tail subunits [144]. A study by Jeronimo et al., found that the CKM is essential for regulating Mediator association with the UAS in yeast rather than with the promoter. They also observed that tail-less Mediator was capable of interacting with the core promoter. Therefore, the head and middle modules carry out basic functions of Mediator at the core promoter, while the CKM and tail modules play regulatory roles. It appears as if *P. falciparum* has retained Mediator subunits present in the head and middle modules that are essential for the basic functions of the Mediator complex while the regulatory CKM and tail module subunits may have been lost or diverged extensively.

1.9 Project rationale & research objectives

The general transcriptional machinery is largely conserved in *P. falciparum*, indicating that it might regulate gene expression in a conventional manner. However, very few gene-specific transcription factor families and DNA regulatory motifs have been identified in its genome. Furthermore, the open euchromatic conformation of the genome and the delay observed between mRNA transcript and protein levels suggest that post-transcriptional mechanisms may also play a significant role in parasite gene expression. Gene expression in the parasite therefore, may be a combination of different gene regulatory mechanisms, both conventional and unique to the parasite. Understanding of gene regulation mechanisms and drug resistance at the molecular level may help in understanding the pathogenesis of malaria and in identifying effective drug targets in the future. Hence it is of utmost importance to study components that are important for gene regulation.

One such component is the multi-subunit Mediator complex, which provides a mechanical and dynamic interface for the interaction of multiple regulatory factors that

affect transcription. Functional analysis of this complex may reveal novel insights about gene regulation in *Plasmodium* as it interacts with multiple proteins involved in transcription, chromatin modifications and RNA processing. Detailed functional or structural analysis of the *P. falciparum* Mediator complex has not been carried out to date. Although a single Mediator subunit can provide an interface for interactions with multiple transcription factors, it remains to be seen whether other Mediator subunits exist in *P. falciparum*. Characterization of the Mediator complex in *P. falciparum* is essential to understand regulation of gene expression as it is known to influence chromatin architecture, transcription, elongation and RNA processing in other organisms. Detailed analysis of the *P. falciparum* Mediator complex may help in answering questions related to the basic biology of the parasite.

This thesis aims to characterize the Mediator complex in *P. falciparum*. The main objectives of the thesis are:

1. To analyze the composition of the Mediator complex in *P. falciparum* through affinity purification and mass spectrometric analysis of Mediator subunits.
2. To assess the genome-wide occupancy pattern of the Mediator complex in *P. falciparum* using chromatin immunoprecipitation followed by high-throughput sequencing.
3. To develop a conditional knockdown system for the functional analysis of the highly conserved Mediator subunit, Med31.

2 MATERIALS & METHODS

2.1 *P. falciparum* culture and maintenance

P. falciparum T996 strain parasites were cultured at 2% hematocrit in RPMI 1640 medium containing 0.25% Albumax, 0.2% sodium bicarbonate and 0.01 mg/ml gentamicin. Human blood was collected from healthy volunteers by trained healthcare professionals and was approved by the NTU institutional review board. The blood was washed thrice with complete medium before addition to culture. Parasites were cultured in a sterile cell culture flask, in a gaseous environment consisting of 5% CO₂, 3% O₂ and 92% N₂ and incubated at 37°C.

2.2 Genomic DNA extraction from *P. falciparum*

Genomic DNA was extracted from *P. falciparum* parasites using DNeasy® Blood & Tissue Kit (Qiagen). Parasite cultures (>5% parasitemia) were lysed with 0.05% saponin in phosphate buffered saline (PBS), incubated at room temperature for 5 min and centrifuged at 6000 g for 5 min. DNA extraction was then carried out according to manufacturer's instructions. Genomic DNA was eluted in 100-200 µL of buffer AE. Roughly 40 µg of genomic DNA was obtained from 500 µL of infected RBCs.

2.3 Generation of plasmid constructs

2.3.1 Amplification of insert sequences

For recombinant expression in *E. coli*, the gene encoding Mediator complex subunit 31 (PF3D7_1475000), codon optimized for expression in yeast, was amplified from pHGF col22 plasmid vector using the primers Med31-pNIC28-Bsa4-FP and Med31-pNIC28-Bsa4-RP (Table 2.1). In order to achieve conditional knockdown of Med31 protein expression, the degron derived from the *E. coli* dhfr gene (ecDDD) was fused to the carboxy terminus of endogenous Med31. A 5' flanking region, 429 bp upstream of the Med31 coding region was used for single crossover homologous recombination. The genomic locus including upstream 5' flanking region and Med31 coding sequence (excluding the stop codon) was amplified from *Pf*T996 strain genomic DNA using the primers Med31-HA-DDD-HR-Fwd and Med31-HA-DDD-HR-Rev. For episomal expression, the Med31 gene (PF3D7_1475000, excluding the stop codon) was amplified

from *P. falciparum* T996 genomic DNA using the primers Med31-pBcamR-3HA –FP and Med31-pBcamR-3HA –RP (Table 2.1). Polymerase chain reaction (PCR) was carried out using the Bio-Rad thermocycler. A 25 µL reaction mixture comprising *Pfu* Polymerase buffer (1X), 0.2 mM each of dATP, dTTP, dGTP and dCTP, 0.5 µM each of forward and reverse primer, 2 mM of MgCl₂, 50-75 ng of template DNA and *Pfu* Polymerase (1.25 units/µL) was prepared in 0.2 ml tubes. The PCR cycling conditions were as follows: Initial denaturation at 95°C for 3 min followed by 30 cycles of 95°C for 1 min, 55°C for 1 min, 60°C for 1 min, followed by a final extension at 60°C for 10 min and pausing at 4°C. For the Med31^{HA-DDD} construct, the PCR cycling conditions were as follows: Initial denaturation at 95°C for 3 min followed by 35 cycles of 95°C for 30 s, 48°C for 30 s, 60°C for 2 min, followed by a final extension at 60°C for 10 min and pausing at 4°C. The PCR products were resolved using 1% agarose gels. The PCR products were purified and eluted in 15-30 µL using PCR purification kit (Qiagen) according to manufacturer's instructions. Primers sets for plasmid constructs used for Med31 recombinant protein expression, conditional knockdown of Med31 and episomal expression of HA-tagged Med31 have been shown in Table 2.1.

Plasmid construct	Primer name	Primer Sequence (5' → 3')
his-Med31 recombinant protein expression	Med31-pNIC28-Bsa4-FP	TACTTCCAATCCATGGGTATCT CTCAAAAGAAG
	Med31-pNIC28-Bsa4-RP	TATCCACCTTTACTGTTAGATT TGGTAAGAGAAG
Conditional knockdown of Med31	Med31-HA-DDD-HR-Fwd	GTGACACTATAGAACTCGAGA GGGAATAATGGAGGAAAGTAT ACC
	Med31-HA-DDD-HR-Rev	ACATCGTATGGGTACCTAGGT ATTTGGTAACTAAAATATAAC
Med31-HA episomal expression	Med31-pBcamR-3HA -FP	AATAGAAATATATCAGGATCAT GGGAATAAGCCAAAAAAG
	Med31-pBcamR-3HA -RP	GTACCTAAGCACCGCTAGC TAT TTGGTAACTAAAATATAAC

Table 2.1 - Primers used for amplifying insert sequences for cloning

2.3.2 Restriction digestion of plasmid vector

The plasmids used for cloning are pNIC28-Bsa4 for recombinant Med31 expression, pHSP101-HADB for conditional knockdown of Med31 protein expression and pBcamR_3XHA_X for episomal expression of Med31. Five micrograms of plasmid DNA was digested with specific restriction enzymes indicated in Table 2.2. The reaction mixture was supplemented with BSA when using enzymes produced by NEB. The plasmids were digested using 0.25 – 0.5 units of enzyme in a 30 - 50 μ L reaction volume and incubated at 37°C for 3 h. The restriction digestion products were resolved on 1% agarose gels and the band corresponding to the digested plasmid was purified from the agarose gel using gel extraction kit (Qiagen) according to manufacturer's instructions. The purified plasmid DNA was eluted in 30 μ L of elution buffer and the DNA concentration was measured using NanodropTM 2000 spectrophotometer (ThermoFisher Scientific).

Experiment	Plasmid	Restriction enzymes
<i>his</i>-Med31 recombinant protein expression	pNIC28-Bsa4	BsaI
Conditional knockdown of Med31	pHSP101-HADB	XhoI and AvrII
Med31-HA episomal expression	pBcamR_3XHA_X	BamHI & NheI

Table 2.2 - Restriction enzymes used for cloning into plasmid constructs

2.3.3 Ligation-independent cloning (LIC)

The linearized vector (10 ng/ μ L) and insert were mixed together in molar ratios of 1:2, 1:5 or 1:7. A 10 μ L reaction mixture comprising the linearized vector, insert, BSA (1X), Buffer 2 (NEB, 1X) and sterile water was prepared. A volume of 0.5 μ L of T4 DNA polymerase (3000 units/ml, NEB) was added to the mixture and incubated at room temperature for 2.5 min. The reaction mixture was placed immediately on ice to stop the

reaction and incubated for 10 min on ice. Competent *E. coli* TOP10 cells were then transformed with the reaction mixture.

2.3.4 Transformation of LIC product

Two microlitres of the LIC reaction mixture was mixed with 50 μ L of competent *E. coli* TOP10 cells and incubated on ice for 30 min. The mixture was subjected to heat-shock by incubation at 42°C for 1 min and placed on ice for 5 min. A volume of 1 ml of Luria-Bertani (LB) broth was added to the reaction mixture and incubated at 37°C for 1 h. The cells were centrifuged at 13,000 rpm for 1 min and the supernatant was discarded leaving 100 μ L volume in the tube. The pellet was resuspended and spread onto LB agar plates containing the desired antibiotic (Ampicillin: 100 μ g/ml, Kanamycin: 50 μ g/ml). The plates were incubated at 37°C overnight.

2.3.5 Screening of recombinant clones

Colonies were picked from the LB plate and inoculated in 3 ml of LB broth containing the desired antibiotic (ampicillin/kanamycin) and incubated at 37°C and 200 rpm for 8 h. Plasmid DNA was isolated using E.Z.N.A.[®] Plasmid Mini Kit I (Omega) and eluted in 30 μ L of sterile water. The plasmid DNA was analysed by restriction digestion and PCR using specific restriction enzymes and primer pairs mentioned in Tables 2.1 and 2.2. The recombinant constructs were further confirmed by sequencing using the primers mentioned in Table 2.3. Glycerol stocks of the recombinant plasmid constructs were prepared and stored at -80°C.

Recombinant plasmid construct	Primer name	Primer Sequence (5'-3')
his-Med31	pNIC28-Bsa4-Seq-FP	TGTGAGCGGATAACAATTCC
	pNIC28-Bsa4-Seq-RP	AGCAGCCAACTCAGCTTCC
Med31 episomal expression	Med31-pBcamR-3HA-Seq-FP	GTATTATTCCAATGTGCATGA TAAAAG
	Med31-pBcamR-3HA-Seq-RP	CAGACAGTAAAAAAAATCGCT ATCCCA
Conditional knockdown of Med31 expression	Med31-pHADB-Seq-FP	CTTA ACTATGCGGCATCAGA GC
	Med31-pHADB-Seq-MP	GCCGCAATCAGACTGAT
	Med31-pHADB-Seq-RP	ATGTATATTGGGGTGATGA

Table 2.3 - Sequencing primers used for confirming recombinant constructs

2.3.6 Preparation of recombinant plasmid constructs

Positive colonies were picked and inoculated in 3 ml LB broth containing the desired antibiotic (ampicillin 100 µg/ml or kanamycin 50 µg/ml) and incubated at 37°C, 200 rpm for 8 h. One ml of this starter culture was used to inoculate 500 ml of LB broth containing the desired antibiotic and incubated at 37°C, 200 rpm overnight. Plasmid DNA was isolated using Maxiprep plasmid isolation kit (Qiagen) according to manufacturer's instructions. DNA was eluted in 500 µL of sterile water and its concentration was quantified by UV spectrometry using NanodropTM 2000 spectrophotometer (ThermoFisher Scientific). Plasmid DNA was stored at -20°C.

2.4 Recombinant expression of his-tagged Med31 in *E. coli*

2.4.1 Induction of recombinant protein expression

Competent BL21 cells, BL21-CodonPlus (DE3) -RIL and Rosetta-gamiTM 2 (DE3) cells were transformed with the recombinant plasmid. Colonies were grown in LB medium supplemented with kanamycin (50 µg/ml) at 37°C until OD 600 was 0.6-0.8.

Protein expression was induced with 1 mM IPTG at different temperatures (37°C and 18°C) and time-points (2 h, 4 h and 19 h).

2.4.2 Solubility assay of recombinant protein

The solubility of the recombinant protein was tested using five different buffers: 50 mM Tris/HCl, 200 mM NaCl, pH 8.5; 50 mM Tris/HCl, 200 mM NaCl, pH 7.5; 50 mM HEPES, 150 mM NaCl, pH 7.0; 50 mM MES, 150 mM NaCl, pH 7.0; 25 mM phosphate buffer, 200 mM NaCl, pH 6.8. BL21 cells expressing recombinant his-tagged Med31 were induced with 1 mM IPTG and cultured at 37°C and 18°C for 4 h. The cells were collected by centrifugation and resuspended in an appropriate volume of each buffer and vortexed. The cells were then subjected to sonication thrice using 18% power for a maximum of 1 min. The cells were centrifuged and the supernatant was transferred to another tube. The pellet was resuspended in an equal volume of buffer. The supernatant (soluble fraction) and the pellet (insoluble fraction) were analyzed by 17% SDS-polyacrylamide gel stained with Coomassie blue.

2.4.3 Preparation of *E. coli* lysate under denaturing conditions

The cell pellet was resuspended in 5 ml of buffer (100 mM NaH₂PO₄, 10 mM Tris-Cl, pH 8.0) per gram of wet weight of cells. Two buffers were prepared either containing urea (4 M, 6 M, 8 M) or guanidinium hydrochloride (6 M). The cell suspension was mixed well and placed on a gyrosaker for 50 min at room temperature. The lysate was then centrifuged at 10,000 g for 30 min at room temperature. The supernatant and pellet fractions were analysed by 17% SDS-polyacrylamide gel stained with Coomassie blue.

2.4.4 Ni-NTA purification under denaturing conditions

The cell pellet was resuspended in buffer (50 mM Tris/HCl, 200 mM NaCl, pH 7.5) containing 8 M urea and protease inhibitor cocktail (Roche). Cells were sonicated using KE76 cone tip and 30% power (1 min pulse and 2 min pause, thrice). Cells were then centrifuged and the resulting supernatant was filtered using a 0.45 µm filter. The clear filtrate was mixed with 2 column volumes of Ni-NTA resin, washed with the same buffer

and incubated for 2 h at 4°C. The protein was eluted from the Ni-NTA column using a gradient of imidazole concentrations (0 mM, 10 mM, 25 mM, 50 mM, 75 mM, 100 mM, 125 mM, 200 mM, 250 mM, 300 mM, 400 mM and 600 mM). The fractions were analysed by 17% SDS-polyacrylamide gel stained with Coomassie blue.

2.5 Generation of antibodies against *Pf*MED31

2.5.1 Epitope design

The Med31 amino acid sequence was analyzed using the Immune Epitope Database (IEDB) analysis resource software. Based on features such as surface accessibility, antigenicity and hydrophilicity, a 15 amino acid peptide was chosen near the N-terminus. A cysteine tagged version of the peptide, CGISQKKYIVNNIIDDNK-OH, was synthesized to enable coupling to an immunogenic carrier.

2.5.2 Conjugation of peptide antigen with an immunogenic carrier

The Med31 peptide antigen was coupled to keyhole limpet hemocyanin (KLH) protein using the Inject Maleimide activated mKLH kit (Thermo Scientific). Two milligrams of the synthetic peptide were dissolved in 300 µL of nuclease-free water. The hapten solution was then added to 2 mg of activated mKLH reconstituted in 200 µL water. The peptide and the activated mKLH were mixed well and allowed to react for 2 h at room temperature. The conjugate was purified by desalting, with one desalting column used for 0.5 ml of sample. The column was washed with 5 volumes of purification buffer. A total of 0.5 ml of the conjugated peptide was added to the column and eluted using the purification buffer into 15 fractions. The absorbance of the fractions was observed at 280 nm and the conjugated peptide was stored at -20°C until further use.

2.5.3 Polyclonal antibody production

Rabbit polyclonal antibody production was carried out at the Biological resource centre (BRC) at Agency for Science, Technology and Research (A*STAR), Singapore. KLH-conjugated peptide was emulsified with complete Freund's adjuvant and used to

immunize two rabbits. A total of 0.5 mg/ml of antigen was used per rabbit during each injection. Two booster doses were administered and the polyclonal sera was analysed by western blotting.

2.6 Synchronization of ring-stage parasites

Parasites were synchronized in the early ring stage using 5% sorbitol. Parasite cultures were centrifuged at 2000 rpm for 5 min at room temperature. The pellet was harvested in 4 ml of media and an equal volume (1:1) of 5% sorbitol was added. The mixture was incubated at room temperature for 10 min and then centrifuged at 2000 rpm for 5 min. The pellet was washed again with complete medium and then transferred to a culture flask. This procedure was repeated again after each infection cycle in order to obtain highly synchronous ring-stage parasites.

2.7 Transfection of *P. falciparum*

Parasites were transfected with 100 µg of plasmid DNA by electroporation. The electroporation conditions were 0.32 kV and 950 µF. Parasites were cultured in media containing 2.5 µg/ml of blasticidin and grown at 37°C in an atmosphere of 3% oxygen, 5% carbon dioxide and 92% nitrogen. Fresh media was added to the culture every day and 100 µL of packed blood was added to the culture each week, until drug-resistant parasites were observed.

2.8 Total RNA extraction from *P. falciparum*

Total RNA was extracted using Trizol reagent. Parasite infected erythrocytes were lysed in Trizol (10 volumes of trizol for 1 volume of infected RBC volume) and the mixture was mixed well. Chloroform was then added in a 5:1 (trizol:chloroform) ratio. The mixture was mixed well, incubated on ice for 5 min and centrifuged at 3900 g for 10 min at 40°C. The aqueous supernatant was transferred to a fresh tube. An equal volume of isopropanol was added to precipitate RNA and the mixture was stored at -20°C overnight. The solution was centrifuged at 9000 rpm for 1 h at 40°C. The pellet was washed with

70% ethanol and incubated on ice for 10 min. The solution was centrifuged again and the supernatant was aspirated. The pellet was air-dried and resuspended in 50 µl of DEPC treated water. The concentration and purity of RNA were determined by measuring the absorbance at 260/280 nm wavelength using NanodropTM 2000 spectrophotometer (ThermoFisher Scientific).

2.9 Reverse-transcription PCR

Total RNA was treated with DNase using RNase-free DNase (Fermentas) according to manufacturer's instructions. The RNA was then purified by phenol-chloroform extraction. The DNase-treated reaction mixture was topped up with DEPC-treated water to make up the volume to 400 µL. An equal volume of phenol:chloroform (25:24:1) was added and vortexed for 15 s. The mixture was centrifuged at 14,000 rpm for 5 min at 4°C and the supernatant was transferred to a new tube. One-tenth of reaction volume of sodium acetate (pH 5.2) and 2.5 volumes of pre-cooled 100% ethanol was added and stored at -80°C for 30 min. The mixture was centrifuged at 12,000 g for 15 min at 4°C. The pellet was washed with 1 ml of 75% ethanol prepared in DEPC-treated water and centrifuged at 12,000 g for 5-10 min. The RNA pellet was air-dried and resuspended in 25 µL of sterile DEPC-treated water.

A total of 0.5 µg each of total RNA and DNase treated RNA were taken as a template for cDNA synthesis. CDNA synthesis was performed using the SuperScript® III First-Strand Synthesis System for RT-PCR (Invitrogen). Oligo-dT primers were used and the reaction was performed according to the manufacturer's protocol. A control reaction without reverse transcriptase was also included. cDNA was subjected to PCR using gene-specific primers.

2.10 Genetic analysis of parasite transgenic lines

For episomal expression of epitope-tagged Med31, genomic DNA was amplified by PCR using primers for amplifying the endogenous Med31 gene (Med31- BamHI-Fwd and Med31- BamHI-Rev) and episomal HA-tagged Med31 gene (Med31-BamHI-Fwd and

Med31-HA epi RP) (Table 2.4) The PCR conditions were as follows: Initial denaturation at 95°C for 3 min followed by 95°C for 1 min, 55°C for 1 min, 60°C for 1 min repeated for 31 cycles, followed by a final extension at 60°C for 10 min. For conditional knockdown of Med31 protein expression, transfectants were analysed to check for integration of plasmid DNA into the genomic locus by PCR. Primers were designed to amplify the endogenous locus, episomal locus and the integrated locus (Table 2.5). The PCR conditions were as follows: initial denaturation at 95°C for 3 min followed by 95°C for 30 s, 48°C for 30 s, 60°C for 2 min repeated for 30 cycles, followed by a final extension at 60°C for 10 min. For episomal expression of HA-tagged putative Mediator subunits, PF3D7_0526800-HA and P3D7_1363600-HA, genomic DNA was amplified by PCR using primers for amplifying the endogenous genes (PF3D7_0526800: pMedA-Fwd & pMedA-Rev; P3D7_1363600: pMedB-Fwd & pMedB-Rev) and episomal HA-tagged versions of the genes (PF3D7_0526800-HA: pMedA-Fwd & Med31-HA epi RP; P3D7_1363600-HA: pMedB-Fwd & Med31-HA epi RP). The primer sequences are mentioned in Table 2.6. The PCR conditions were as follows: initial denaturation at 95°C for 3 min followed by 95°C for 1 min, 48°C for 1 min, 60°C for 1.5 min repeated for 24 cycles, followed by a final extension at 60°C for 10 min.

Reverse-transcription PCR was carried out to check transcript expression. For episomal expression of epitope-tagged Med31, cDNA was subjected to PCR using primers for endogenous Med31 (Med31-HA epi Transcript Fwd and Med31 endo transcript rev) and HA-tagged Med31 transcripts (Med31-HA epi Transcript Fwd and Med31-HA epi Transcript Rev) shown in Table 2.4. The PCR conditions were as follows: initial denaturation at 95°C for 3 min followed by 95°C for 1 min, 55°C for 1 min, 60°C for 1 min repeated for 31 cycles, followed by a final extension at 60°C for 10 min.

For conditional knockdown of Med31 protein expression, cDNA was subjected to PCR using the primers, Med31-HA epi transcript fwd and Med31-pHADB-Seq-MP, shown in Table 2.5. The PCR conditions were as follows: initial denaturation at 95°C for 3 min followed by 95°C for 1 min, 48°C for 1 min, 60°C for 1 min repeated for 29 cycles, followed by a final extension at 60°C for 10 min.

	Primer name	Primer Sequence (5'-3')
Episomal DNA analysis	Med31- BamHI-Fwd	AGTCGGATCCATGGGAATAAGCCA AAAAAAG
	Med31Nhe1-3HA-RP	AGTCGCTAGCTATTTGGTAACTAAA ATATAACC
	Med31-HA epi RP	CATAAAGTTGTTAGAGCTCGGCAT AATCTGG
Transcript expression	Med31-HA epi Transcript Fwd	ATGGGAATAAGCCAAAAAAGTAC
	Med31 endo transcript rev	CTATATTTGGTAACTAAAATATAAC CAATG
	Med31-HA epi Transcript Rev	TTAGAGCTCGGCATAATCTGGAAC ATCG

Table 2.4 - List of primers used for genetic analysis of transfectants for episomal expression of epitope-tagged Med31

	Primer name	Primer Sequence (5'→ 3')
Endogenous locus	Intg-5'flank2-Fwd	ATGGGGAAATATAAGCAAGGC
	Med31 transcript-3'UTR-Rev	CACATTAGGAAAATATATATATAT ATATATATGGTCTA
Integragt locus	Intg-5'flank2-Fwd	ATGGGGAAATATAAGCAAGGC
	Intg-DDD2-Rev	AATTCGCTGAATACCGATTG
Episomal locus	Epi-5' up-HADB-Fwd	GGACATATTGTCGTTAGAACG
	Epi-DDD-Rev	GCCGCAATCAGACTGAT
Transcript expression	Med31-HA epi transcript fwd	ATGGGAATAAGCCAAAAAAGTA C
	Med31-pHADB-Seq-MP	GCCGCAATCAGACTGAT

Table 2.5 - List of primers used for genetic analysis of transfectants for conditional knockdown of Med31 expression

Primer name	Primer Sequence (5'-3')
pMedA-Fwd	ATGAGTAAAGAAAAAGAAGATAAACATAAG
pMedA-Rev	GACGTATCGTCTTGATGATCTGC
pMedB-Fwd	ATGGAGGAATCCATTTTAATTAAAAATTTTC
pMedB-Rev	CTTCTTCATATTTTTTCATATGTAG
Med31-HA epi RP	CATAAAGTTGTTAGAGCTCGGCATAATCTGG

Table 2.6 – List of primers used for genetic analysis of transfectants for episomal expression of putative HA-tagged Mediator subunits

2.11 Cloning by limiting dilution

For conditional knockdown of Med31 protein expression, single parasite clones were obtained by the limiting dilution technique. Highly synchronous parasite cultures were used. The culture was placed at 37°C on a shaker, 12 h prior to the invasion of RBCs in the new cycle, such that only one merozoite can invade one RBC. Parasite smears were prepared and examined by light microscopy to ensure that most of the infected RBCs contained only one ring-stage parasite.

The parasite culture was diluted such that it contains on an average, only 0.3 parasites per 0.2 ml volume. The parasitemia of the culture was calculated. The parasite culture was harvested, the volume of RBCs was measured carefully and re-suspended at 2% hematocrit for further use. The expected number of parasitized red blood cells per ml was calculated based on the parasitemia and hematocrit of the culture. Parasites were first diluted with fresh RBCs (2% hematocrit) such that 1000 parasites were present per 0.01 ml volume (diluent A). Diluent A was then diluted with fresh RBCs such that 20 parasites were present per 0.02 ml volume (diluent B). Diluent B was then diluted with fresh RBCs such that 0.3 parasites were present per 0.2 ml volume (diluent C). A volume of 200 µL of diluent C was transferred to 80 wells in two 96-well plates (40 wells each). Diluent A and diluent B solutions (200 µL each) were also transferred to the 96-well plate as positive

controls. All the outer wells of the 96-well plate were filled with 100 μ L of incomplete RPMI medium to maintain humidity in the plate.

The plates were gassed and placed in an incubator at 37°C. Medium was changed every 48 h. Positive controls of diluent A and diluent B were checked after 4 and 6 days respectively to confirm the presence of parasites. On day 6, fresh RBCs were added. Giemsa-stained smears were prepared for all the wells, 12- 14 days after limiting dilution, to check for wells containing parasites. A total of 15 out of 80 wells were found to be positive. The parasite-containing wells were transferred to a cell culture flask and maintained at 1% hematocrit. All parasite clones were cryopreserved in liquid nitrogen for future use. Clone B2 was randomly selected for further experiments.

2.12 Western blot analysis

2.12.1 Preparation of *P. falciparum* total protein lysate

Parasite cultures with a parasitemia greater than 5% were centrifuged at 1800 rpm for 5 min. The supernatant was removed and the parasites were washed with PBS. The infected RBCs were mixed with 0.1% saponin (Sigma) and incubated at room temperature for 5 min. The parasites were pelleted by centrifugation at 6000 rpm for 5 min and washed 3 times with PBS until the supernatant lost its reddish color. A total of 50-200 μ L of parasite lysis buffer (4% SDS, 0.5% Triton X-100) and protease inhibitor cocktail (Roche) was added depending on the size of the parasite pellet. The mixture was incubated on a rotating shaker for 10 min and then centrifuged at 17,500 g for 15 min at 4°C. The supernatant was transferred to a fresh tube and stored at -20°C.

2.12.2 Estimation of protein concentration by Bradford assay

Protein concentration was estimated using the Bio-rad quickstart Bradford reagent in a microplate assay format. BSA was used as the protein standard and different concentrations of the standard were prepared in sterile water: 1.25, 2.5, 5, 7.5 and 10 μ g/ml. Blank samples (0 μ g/ml) were also prepared using sterile water. The 1X dye reagent was removed from 4°C storage and warmed to room temperature. The protein samples of unknown concentration were diluted in a ratio of 1:1000 in sterile water.

Protein samples were assayed in duplicates. 100 μ L of standards and unknown samples were pipetted into the microplate wells. A total of 100 μ L of the Quickstart Bradford reagent was added to all the wells, mixed well and incubated at room temperature for 5 min. The absorbance of the standards and unknown samples were measured at 595 nm using a Tecan microplate reader. A standard curve was plotted with the absorbance values in the y-axis and the known protein concentration in x-axis. The unknown sample concentration was determined using the standard curve. Since the unknown protein samples were diluted 1000 times, the final concentration of unknown protein samples was multiplied by the dilution factor.

2.12.3 SDS-PAGE and western blot

Protein samples were resuspended in Laemmli buffer, boiled at 95°C for 5 min and resolved on SDS-polyacrylamide gel. The reagent composition for preparing SDS-polyacrylamide resolving and stacking gels has been provided in Table 2.6. Proteins were transferred to a polyvinyl difluoride (PVDF) membrane, using tris-glycine buffer containing 20% methanol as the transfer buffer. After transfer, the membrane was blocked using 5% non-fat milk for 1 h at room temperature. Incubation with primary antibody was done using 5% non-fat milk in PBS. Primary antibody dilutions were as follows: mouse monoclonal anti-*his* antibody (Qiagen): 1: 4000; rat monoclonal anti-HA antibody (Roche, clone 3F10): 1:1000; rabbit anti-Med31 polyclonal sera: 1: 200; mouse monoclonal anti-Med7 hybridoma supernatant (GenScript): 1:200; mouse monoclonal Med7 antibody (GenScript): 1: 2000.

Membrane was incubated with primary antibody either for 2 h at room temperature on a shaker or overnight at 4°C. Membrane was washed after primary and secondary antibody incubation with PBST buffer containing 0.1% tween 20. Secondary antibody dilutions used were: 1: 10,000 anti-rabbit HRP, 1: 10,000 anti-mouse HRP (Cell Signaling technology), 1: 10,000 anti-rat HRP (Abcam). Membrane was incubated with secondary antibody for 1 h at room temperature. Chemiluminescence from HRP conjugated secondary antibodies was detected using WesternBright ECL HRP substrate (Advansta Inc.). The membrane was exposed to Kodak XE films from 10 s to 10 min.

Components	12% resolving gel (ml)	15% resolving gel (ml)	5% stacking gel (ml)
30% acrylamide mix	2	2.5	0.67
1.5M Tris pH 8.8	1.3	1.3	-
1M Tris pH 6.8	-	-	0.5
Distilled water	1.6	1.1	2.7
10% SDS	0.05	0.05	0.04
10% ammonium persulphate	0.05	0.05	0.04
TEMED	0.002	0.002	0.004

Table 2.7 - Reagent composition for preparing SDS-polyacrylamide resolving and stacking gels

2.13 Preparation of cytoplasmic and nuclear protein fractions

Parasite culture (> 5% parasitemia, 2% hematocrit, 1.5 ml packed blood) was harvested in a 50 ml conical tube by centrifugation at 1800 rpm for 5 min. The parasite cells were washed with PBS once. RBCs were lysed using 0.1% saponin. The parasite pellet was resuspended in 5 pellet volumes of ice-cold lysis buffer (20 mM HEPES, pH 7.8, 10mM KCl, 1mM EDTA, 1mM PMSF, Protease inhibitor cocktail, 1mM DTT, 0.65% NP-40), divided into two microcentrifuge tubes and incubated on ice for 10 min. The mixture was centrifuged at 5200 rpm for 5 min at 4°C. The resulting supernatant containing cytoplasmic proteins was transferred to a separate tube and stored at -80°C. The nuclear protein pellet was washed twice with lysis buffer and resuspended in 1 pellet volume of extraction buffer (20 mM HEPES pH 7.8, 800 mM KCl, 1mM EGTA, 1mM EDTA, 1mM DTT, 1mM PMSF, 1X Protease inhibitor (Roche)). The mixture was vortexed at 4°C for 20 min and centrifuged at 13,000 rpm for 20 min. The resulting supernatant was pooled together into a single tube. For immunoprecipitation, the nuclear fraction was diluted 5 times with dilution buffer (20 mM HEPES pH 7.8, 1mM EDTA, 2mM DTT, 1mM PMSF, 0.2% Triton X-100, 30% glycerol) such that the salt

concentration was around 160 mM. Glycerol was added such that the final concentration was 15%. For immunoprecipitation from the cytoplasmic fraction, KCl and glycerol were added such that the final concentration was 150 mM KCl and 15% glycerol. An aliquot of 60 μ L was taken as input for analysis of immunoprecipitation efficiency. Immunoprecipitation was then carried out using Anti-HA magnetic beads (Pierce, #88836).

2.14 Immunofluorescence analysis

Parasite cultures (> 5% parasitemia) were washed with PBS. Parasite smears were prepared on microscopic slides, air-dried and fixed in methanol for 10 s. The slide was then immersed in PBS. A total of 1 ml of blocking solution (3% BSA, 0.01% saponin in PBS) was applied over the slide and incubated at room temperature for 1 h. The slide was washed with PBS thrice, with incubation for 5 min on a rotating shaker for each wash. Primary antibody diluted in PBS was applied over the slide and incubated overnight at 4°C. Primary antibodies used were rat monoclonal anti-HA (Roche, clone 3F10): 1:100 and rabbit polyclonal anti-aldolase (Abcam): 1:250. Following incubation with the primary antibody, the slide was washed with PBS thrice, with incubation for 5 min on a rotating shaker for each wash. Secondary antibodies used were anti-rat Alexa Fluor 488 (1: 1000, Invitrogen), anti-rabbit Alexa Fluor 555 (1: 1000, Invitrogen). DAPI was used at a dilution of 1: 1000. Secondary antibody containing DAPI was applied over the slide and incubated at room temperature for 1 h in the dark. The slide was then washed with PBS thrice, with incubation for 5 min on a rotating shaker for each wash. The washing step was carried out in the dark. After the final wash, the slides were dried and mounted in VECTASHIELD® Mounting Media using a coverslip. The slides were analysed by a fluorescence microscope and analysed using the Axiovision software.

2.15 Co-immunoprecipitation

2.15.1 Preparation of protein lysate

Parasite cultures (> 10% parasitemia, 1.5 ml packed blood) were harvested in 50 ml conical tubes by centrifugation at 1800 rpm for 5 min. The cells were washed once

with PBS. Infected RBCs were lysed with 0.1% saponin and washed multiple times with PBS until the supernatant lost its reddish color. The parasite pellet was resuspended in 5 pellet volumes of lysis buffer (50 mM Tris-Cl pH 7.5, 150 mM NaCl, 1% NP-40, 1mM EDTA, 1mM PMSF, 1X protease inhibitor cocktail (Roche)). The mixture was incubated at 4°C on a rotating shaker for 1 h. The solution was centrifuged at 14,500 g for 10 min at 4°C. The resulting supernatant was transferred to a new tube and 10% of the solution was taken as input for immunoprecipitation. The remaining solution was used for immunoprecipitation using the HA antibody.

2.15.2 Immunoprecipitation

Anti-HA magnetic beads (Pierce, Cat. No. 88836) were used for immunoprecipitation and were pre-washed according to the following protocol: A total of 25 µL of Pierce anti-HA magnetic beads were placed in a 1.5 ml microcentrifuge tube. A total of 175 µL of 0.05% TBS-T (25 mM Tris pH 7.5, 0.15M NaCl pH 7.5, 0.05% Tween-20) was added to the beads and mixed well. The tube was placed in a magnetic stand to collect the beads against the side of the tube and the supernatant was discarded. One ml of TBS-T was added to the beads, mixed well and magnetized to collect the beads. The supernatant was removed. The sample containing the HA-tagged protein was added to the pre-washed magnetic beads and incubated overnight on a rotating shaker at 4°C. The beads were magnetized using the magnetic stand and the supernatant containing the flowthrough was transferred to another tube for further analysis. A total of 300 µL of TBS-T was added to the beads and mixed well. The beads were magnetized and the supernatant was discarded. This wash step was repeated twice. A total of 300 µL of ultrapure water was added to the beads and mixed well. The beads were collected on a magnetic stand and the supernatant was discarded.

2.15.3 Elution of HA-tagged protein

A total of 50 µL of SDS-PAGE Laemmli buffer was added to the tube, mixed well and incubated at 97°C for 5 min. The beads were magnetized and the supernatant containing

the immunoprecipitated proteins was collected for further analysis by western blotting. The eluate sample was stored at -20°C.

2.16 Mass spectrometric analysis

Eluate samples obtained after immunoprecipitation were separated using a 12% SDS-polyacrylamide gel and stained with coomassie blue. The lanes corresponding to the samples of interest were cut using a scalpel and subjected to in-gel digestion protocol using trypsin.

The peptides were analyzed using a Dionex Ultimate 3000 RSLCnano system coupled to a Q Exactive instrument (Thermo Fisher Scientific, MA, USA). Separation was achieved with a Dionex EASY-Spray 75 $\mu\text{m} \times 10\text{ cm}$ column packed with PepMap C18 3 μm , 100 Å (Thermo Fisher Scientific) using solvent A (0.1% formic acid in 5% ACN) and solvent B (0.1% formic acid in 90% ACN) at flow rate of 300 nL/min with a 60 min gradient. A Q Exactive apparatus with an EASY nanospray source (Thermo Fisher Scientific) was used to analyse peptides at an electrospray potential of 1.5 kV. A full MS scan (350–1,600 m/z range) was obtained at a resolution of 70,000 at m/z 200 and a maximum ion accumulation time of 100 ms. Dynamic exclusion was set as 15 s and the resolution of the higher energy collisional dissociation (HCD) spectra was set to 17,500 at m/z 200. Single and unassigned charged ions were excluded from MS/MS. The raw LC-MS/MS data files were loaded into Spectrum Files (default parameters set in Spectrum Selector). The LC-MS/MS spectra of the raw data was extracted and converted into dta files using `extract_msn` and then combined into a mascot generic file using an in-house program as previously described.

The SEQUEST algorithm was then used to identify proteins using the following parameters; missed cleavage of two; dynamic modifications were oxidation (M), deamidation (N) and phosphorylation (S, T, Y). The static modifications were TMT-6plex (any N-terminus and K) and Carbamidomethyl (C). The false discovery rate for protein identification was <1%. The obtained peptide/protein list for each sample was exported to Microsoft Excel or processed using an in-house script for further analysis [149].

2.17 Chromatin immunoprecipitation-Sequencing (ChIP-Seq)

2.17.1 Saponin lysis and formaldehyde cross-linking

Synchronised parasite cultures were centrifuged at 2000 rpm for 5 min. Infected RBCs were lysed with 0.1% saponin and washed thrice with PBS warmed to 37°C. The supernatant was removed and the pellet volume was made up to 3.65 ml with PBS. A total of 50 µL of 37% formaldehyde was added and incubated for 10 min. A total of 300 µL of 1.67M glycine/PBS was added and incubated for 1 min. Ice-cold PBS was added to make up the volume to 40 ml. The mixture was centrifuged at 4500 rpm for 5 min and the supernatant was removed. The pellet was resuspended in 4 ml of ice-cold PBS and split into 1 ml aliquots. The mixture was centrifuged at 4500 rpm for 4 min at room temperature and the supernatant was removed. The pellet was snap-frozen in liquid nitrogen and stored at -80°C.

2.17.2 Chromatin immunoprecipitation

The parasite pellets were mixed with 1.5 ml of buffer A (1 mM EDTA pH 8.0, 0.25% NP-40, 2 mM phenylmethylsulfonyl fluoride and 1X protease inhibitor) and incubated on ice for 60 min. The parasites were lysed by 200 strokes using a dounce homogenizer. Nuclei were pelleted down by centrifugation at 5000 rpm for 5 min at 4°C. The supernatant was discarded and the pellet was resuspended in 160 µL of lysis buffer (1% SDS, 50 mM Tris pH 8.1, 10 mM EDTA, 2 mM phenylmethylsulfonyl fluoride and 1X protease inhibitor). The mixture was incubated on ice for 15 min and sonicated for 24 cycles at 25% amplitude, with each cycle consisting of a 10 s pulse and 50 s interval before the next pulse, using a Sonics Vibra Cell to obtain DNA fragments in the range of 250-500 bp. After sonication, the mixture was centrifuged at 13,000 rpm for 10 min at 4°C. The supernatant was collected in a separate tube and diluted 10 times with ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-Cl pH 8.1, 167 mM NaCl, 2 mM phenylmethylsulfonyl fluoride and 1X protease inhibitor). An aliquot of 100 µL was taken as input DNA and stored at 4°C. The remainder of the solution was used for chromatin immunoprecipitation.

A volume of 80 μL of protein A agarose/salmon DNA beads were added to the sheared chromatin solution and incubated on a rotating shaker for 2 h at 4°C. The mixture was then centrifuged at 2500 rpm for 1 min at 4°C and the supernatant was transferred to another tube. A total of 3 μg of anti-HA antibody (Rat MAb, Roche clone 3F10) was added and incubated overnight on a rotating shaker at 4°C. A total of 80 μL of protein A agarose/salmon DNA beads were preblocked in 3% BSA ChIP dilution buffer for 30 min at room temperature and rinsed with ChIP dilution buffer. The pre-blocked protein A agarose beads were added to the immunoprecipitated chromatin solution and incubated on a rotating shaker at 4°C for 2 h. The mixture was centrifuged at 2500 rpm for 1 min at 4°C. The unbound material was transferred to another tube for further analysis. The beads were washed using the buffers mentioned below. All the washing steps were carried out at 4°C using ice-cold wash buffers. A total of 1 ml of wash buffer was added for each wash, incubated on a rotating shaker for 5 min and centrifuged at 2500 rpm for 1 min to collect the beads. The supernatant was discarded. The beads were washed twice with low-salt wash buffer (0.1% SDS, 2 mM EDTA, 20 mM Tris-Cl, pH 8.1, 150 mM NaCl), 3 times with high-salt wash buffer (0.1% SDS, 2 mM EDTA, 20 mM Tris-Cl, pH 8.1, 500 mM NaCl), once with LiCl wash buffer (0.25 M LiCl, 1% NP-40, 1% deoxycholic acid, 1 mM EDTA, 10 mM Tris pH 8.1) and two times with TE buffer (10 mM Tris-Cl pH 8.1, 1 mM EDTA).

A total of 250 μL of elution buffer (1% SDS, 50 mM Tris-Cl pH 8.0, 10 mM EDTA pH 8.0) was added to the protein A agarose/antibody complex, vortexed briefly and incubated at room temperature on a rotating shaker for 20 min. The beads were collected by centrifugation at 2500 rpm for 5 min and the supernatant fraction containing the eluate was transferred to another tube. The elution was repeated and the eluates were combined. A total of 20 μL of 5M NaCl was added to the eluate (Volume: 500 μL) and 4 μL of 5M NaCl was added to 100 μL of input DNA. The protein-DNA cross-links were reversed by heating at 65°C overnight. A total of 5 μL of proteinase K (Fermentas, 20 mg/ml) was added and incubated at 37°C for 2 h. The immunoprecipitated DNA was purified using the mini-elute PCR clean up kit (Qiagen) and eluted in 15 μL of elution buffer. The DNA concentration was quantified using Qubit fluorometer (ThermoFisher Scientific).

2.17.3 ChIP-Seq library preparation

For ChIP libraries, 5 ng of ChIP or Input material was used to prepare indexed Illumina sequencing libraries using the NEB Next Ultra DNA library prep kit (New England Biolabs). Adapter ligated ChIP DNA libraries were amplified for 15 cycles. The amplification of the ChIP DNA libraries was performed using Kapa DNA polymerase enzyme. The MiSeq V2 Run (150 PE) generated 150 bps paired end reads. The libraries were then pooled at equimolar concentrations and sequenced in one lane on an Illumina HiSeq2500 sequencer in rapid mode at a read-length of 100bp paired-end.

2.17.4 ChIP-Seq data analysis

2.17.4.1 Quality check and data pre-processing

Paired-end reads were checked for base call quality, GC content, duplication, and adapter content using the program, *FastQC*. Index adapter used for the individual libraries in the multiplexed experiment was identified as the most dominant adapter content (read-through adapter contamination when the fragment size was less than the read length). Therefore, the program, ‘*trim_galore*’ was used to pre-process the fastq files based on the following conditions:

- 1.) Reads were trimmed when the base quality started declining below a threshold of 20;
- 2.) Read through adapter from 3’ ends were removed with a similarity constraint of at least 10 bp between the read and adapter sequence;
- 3.) Reads with length below 76 bp after applying steps 1) and 2) were removed;
- 4.) Only pairs where both the reads passed all the above steps were retained.

The counts of raw and trimmed reads obtained from Mi-Seq paired-end sequencing have been indicated in Table 2.8.

Sample	Raw read counts	Trimmed read counts
Med31	6190142	5971200
Med31 Input	9853700	9603512
T996	8152496	7613636
T996 Input	10078136	9854608

Table 2.8 - Counts of raw and trimmed reads obtained from Mi-Seq paired-end sequencing

2.17.4.2 Mapping and peak calling

Pre-processed reads were mapped to *P. falciparum* 3D7 sequences (PlasmoDB version 9.0) using the software, *Bowtie2*. Default options were used except for the mismatch criteria and maximum insert size between the paired ends (-N 1, -X 1000). In Table 4.3 (chapter 4), the column labeled as ‘mapped’ represents the number of reads mapped to the genome. The program, *samtools*, was used for filtering alignments and the column named ‘uniquely mapped’ is the final set of reads remaining after filtering for,

- i) Alignment quality (MAPQ) > 10 for reliable unique mapping
- ii) Reads that aligned as a pair on the same chromosome and within in a distance of 1 kb from each other (concordant alignment)
- iii) Reads where only one end aligned uniquely, while the other failed or multi-mapped (single-end or orphan reads)

The software, *MACS2* (v2.1.1) was used to call narrow and broad peaks for the ChIP data with corresponding genomic DNA as input. The last 2 columns for the ChIP sample in Table 4.3 indicate the number of peaks identified. The general report on mapping of input samples has also been provided in Table 4.3. The columns labelled as ‘% genome covered’ indicates the proportion of genome bases covered (number of genome bases with at least 1 read covering) while the column, ‘Sequencing depth’ indicates the average number of reads covering each base in the genome.

2.17.4.3 Features

A total of 5400 (nuclear genes excluding antigenic and pseudogenes) genes have been used from PlasmoDB v9.0 for generating profile plots and heat maps. Genomic regions have been binned to facilitate computational analyses. For the purpose of data analysis, 5' UTR has been arbitrarily defined as the region 1 kb upstream of the translation start codon (ATG) while 3' UTR has been arbitrarily defined as the region 1 kb downstream of translation stop codon (STOP). The UTR regions are binned into 10 bins of equal size (100bp in a bin). The gene body region is binned into 20 bins with each bin size proportional to the actual gene length. In profile plots or heat map, when all the three gene elements (i.e. 5' UTR, gene body and 3' UTR) are shown, the regions have been binned.

2.17.4.4 Extending reads to fragments

Paired-end reads were extended to fragments using the left and right mate alignment information. The average fragment size derived from the paired-ends was used to extend the single-end mapped reads in a 5' to 3' direction. Enrichment in each feature (binned or whole) was calculated as a ratio of ChIP by input, after adjusting for sequencing depth difference between the ChIP and input (as described in the 'normalization' section below).

2.17.4.5 Normalization

Normalization was done using a linear scaling, where the larger sample is down-scaled to the size of smaller sample, and the final values are expressed in million reads (Fragments Per Million). The values are then log₂ transformed.

Enrichment score =

$$\frac{(\text{Frag counts in a feature from ChIP} \times 10^6) / \text{Total uniquely mapped read count in the ChIP}}{(\text{Frag counts in a feature from Input} \times 10^6) / \text{Total uniquely mapped read count in the Input}}$$

This is same as using a scaling factor ‘s’, and dividing the fragment counts from larger sample by ‘s’. If we assume that input sample is larger than ChIP,

$$s = \frac{\text{Total uniquely mapped frag counts in Input (larger sample)}}{\text{Total uniquely mapped frag counts in ChIP (smaller sample)}}$$

Down-scaling the counts from input sample as,

$$\text{Normalized read counts in Input} = \frac{\text{Frag counts in a feature from Input}}{s}$$

While the fragment counts from the smaller sample (ChIP) is same as raw counts.

A pseudo count of ‘1’ fragment is added to all counts to avoid division by zero error.

2.18 Real-time PCR

Immunoprecipitated and input DNA were subjected to qPCR using the Applied Biosystems Step One Plus Real-time PCR System. ChIP DNA was diluted to 4 ng/ μL and 16 ng of DNA sample was mixed with 0.5 μM of forward and reverse primer (Table 2.8) and 10 μL of 2X SYBR Green PCR Master Mix (Roche), in a total volume of 20 μL . The thermal cycling parameters used were as follows: 1 cycle of pre-incubation at 95°C for 3 min, 45 cycles of amplification at 95°C for 5 s, 55°C for 30 s and 60°C for 30 s. This was followed by 1 cycle of 95°C for 5 s and 65°C for 1 min. All reactions were carried out in triplicates. ChIP enrichment was calculated using $\Delta\Delta\text{Ct}$ method wherein,

$$\Delta\Delta\text{Ct} = (\text{Ct of Med31 ChIP} - \text{Ct of Med31 input}) - (\text{Ct of T996 ChIP} - \text{Ct of T996 input}),$$

Ct refers to the threshold cycle. The value, $2^{-\Delta\Delta\text{Ct}}$ denotes the ratio of target gene compared to reference gene.

Gene ID	Primer name	Primer Sequence (5'-3')
PF3D7_1368200	ATG-1368200-Fwd	ATAACCCACCTGAATGGCGA G
	ATG-1368200-Rev	GTAATTCTCCTCCACTTAAAT CTTCC
PF3D7_0819400	ATG-0819400-Fwd	CAAGATCTTAGCATTAATGTA GATGGTCAC
	ATG-0819400-Rev	TCAATGAATACAAGTCAAGAT GTATAGGAAC
PF3D7_0823400	ATG-0823400-Fwd	GATCCAATTAATAATACACTC C
	ATG-0823400-Rev	AGGGTGTATTCATTACATATA ATCACATC
PF3D7_0529900	ATG-0529900-Fwd	TCCAGATAAGAGAAAAAAAT ATATATGCC
	ATG-0529900-Rev	TGAATATATAAGGAGCACA
PF3D7_0902000	ATG-0902000-Fwd	GGTTTTACAATCATAACTGGT TATTGC
	ATG-0902000-Rev	CCATAACTCATATTTATCTGC ACTTTATAACTTC
PF3D7_1474800	STOP-1474800-Fwd	GTACAGAATGAAGCAAACGA ATAAAATAAATCACTCAATTG
	STOP -1474800-Rev	GACATCTTTTTAATATAAAGT AGTTTTTCATG
PF3D7_0306200	STOP -0306200-Fwd	GAAAAGTGAATTTTTATCCGT ACTACG
	STOP -0306200-Rev	CCTTGTAAGTGATGTGTATTTA TATGTATG
PF3D7_1108600	STOP -1108600-Fwd	GTTATCTTACCAACAGCTCGT AGTAGAG
	STOP -1108600-Rev	TTTATTTGTGGGGTGTACGAA AT
PF3D7_0323900	STOP -0323900-Fwd	CATGGAGAATTTAAATGTAGA TAGCGAGGG
	STOP -0323900-Rev	TGTTCCCTTTAATTATATCGCG AAGATCTGC
PF3D7_0606100	STOP -0606100-Fwd	GCGTGAAATGGCAATTAGAG
	STOP -0606100-Rev	CGTTCGGTTTATATATTCTGT GATTAC
PF3D7_1478600	STOP -1478600-Fwd	GATTATAAGGATGAATATGAT GGAGAGTTTAATAAAGAG
	STOP -1478600-Rev	CTCTTCATCATCTATACCTTTT AATACAGTCTTTGC

Table 2.9 - List of primers used for quantitative real-time PCR for validating ChIP-Seq results

3 GENERATION OF TOOLS TO STUDY MEDIATOR SUBUNITS

3.1 Introduction

Functional analysis of the Mediator complex requires tools such as antibodies against the Mediator subunits themselves or against epitopes to which they have been fused.

The subunit Med31 was chosen for this study, as it is the most conserved Mediator subunit across multiple species [143] and is readily detectable by simple bioinformatic searches as encoded by the *Plasmodium* spp. genome. EM analysis of yeast and human Mediator complexes have shown that Med31 is present in the middle module and is located close to the CTD-binding patch located on the head module [99]. The crystal structure of a sub-module consisting of Med31 and the N-terminal region of Med7 has also been determined [94], and has illuminated the physical basis for the requirement of this sub-module for activated transcription in yeast.

Therefore, attempts were made to raise antibodies against the subunits, Med31 and Med7 and the following transgenic parasite lines were also generated for the functional analysis of Med31:

- Episomal expression of epitope-tagged Med31
- Conditional knockdown of Med31 protein expression

The generation of tools to study Mediator subunits Med31 and Med7 is discussed in detail in the following sections.

3.2 Generation of antibodies against Med31 and Med7

3.2.1 Recombinant expression of *PfMed31* protein in *E. coli*

Since large amounts of antigen are required in pure form for antibody production, *E. coli* was chosen as a suitable host for the heterologous expression of *PfMed31*. The full-length *PfMed31* protein (132 amino acids), codon-optimized for expression in the yeast *Saccharomyces cerevisiae*, was cloned into pNIC28-BsaI vector (Fig 3.1) to express the cloned gene as a fusion protein with an N-terminal hexa-histidine (*his*) tag. The pNIC28-BsaI plasmid was obtained from Dr. Opher Gileadi (Structural Genomics Consortium).

The recombinant plasmid was analyzed by PCR, restriction digestion and confirmed by sequencing.

Figure 3.1 - pNIC28-Bsa4 vector used for recombinant protein expression

Full-length *PfMed31* was cloned between *BsaI* restriction enzyme sites using ligation-independent cloning. The pET expression vector contains a 6X-histidine tag as a 22-AA N-terminal fusion peptide and also contains the *SacB* gene which allows negative selection on 5% sucrose.

3.2.1.1 Induction of protein expression using BL21 cells

BL21 cells expressing the recombinant plasmid induced with 1 mM IPTG were cultured at two different temperatures (37°C and 18°C) and harvested at different time-points (2, 4 and 19 h) to check for levels of protein expression. Although the expected size of the *his*-Med31 protein was 19 kDa, it was found to move closer to the 15 kDa molecular weight marker on the polyacrylamide gel (Fig 3.2). The identity of the overexpressed protein was confirmed to be Med31 based on mass spectrometric analysis.

Highest level of recombinant protein expression was obtained when BL21 cells were

cultured at 37°C and harvested 4 h after induction (Fig 3.2 B). A comparable level of protein expression was also observed with cells cultured at 18°C and harvested after 19 h (Fig 3.2 A). However, owing to lower amount of time (4 h) required for expression at 37°C, these conditions were chosen for further applications.

Figure 3.2 - *his*-Med31 protein expression in BL21 cells cultured at 18°C (A) and 37°C (B)

BL21 cells were cultured either with 1 mM IPTG (+) or without IPTG (-) for 2, 4 and 19 h at 18°C and 37°C. Crude cell lysates were separated using a 17% SDS-polyacrylamide gel and stained by coomassie blue. Molecular weight markers (kDa) are indicated in the figure. Band corresponding to the induced protein is indicated with an arrow.

3.2.1.2 Assay for solubility of *his*-Med31 protein in BL21 cells

Solubility of the *his*-Med31 protein was tested using five different buffers (described in the section 2.4.2). BL21 cells cultured at 37°C and 18°C and harvested 4 h after induction with IPTG were considered for the solubility assay. The recombinant *his*-Med31 protein was found to be insoluble in all the buffers tested (Fig 3.3).

Figure 3.3 - Solubility of *his*-Med31 protein in BL21 cells at 18°C (A) and 37°C (B)

Cellular fractions were separated using a 17% SDS-polyacrylamide gel and stained by coomassie blue. Solubility test showed that *his*-Med31 protein was mostly present in inclusion bodies in the insoluble fraction under all conditions tested. **I**: Insoluble fraction,

S: Soluble fraction. Molecular weight markers (kDa) are indicated in the figure. Band corresponding to the induced protein is indicated with an arrow.

3.2.1.3 Induction of protein expression using alternate BL21-derived cells

Due to the disparity in the codon usage of *P. falciparum* and *E. coli*, translation of recombinant *Plasmodium* proteins may be stalled due to the unavailability of certain codons that are rarely used in *E. coli*. For this purpose, BL21-CodonPlus (DE3) –RIL and Rosetta-gami™ 2 (DE3) strains that contain extra copies of rare tRNA genes, were used. BL21-CodonPlus (DE3) –RIL strain contains additional copies of tRNA genes that recognize the arginine codons (AGA, AGG), the isoleucine codon (AUA), and the leucine codon (CUA). Rosetta-gami™ 2 (DE3) cells harbor the chloramphenicol-resistant plasmid, pRARE2, which supplies tRNAs for seven rare codons (AUA, AGG, AGA, CUA, CCC, GGA, and CGG).

The alternate BL21-derived cells were used for overexpression of *his*-Med31 with the aim of increasing the yield of the expressed protein. Induction assay showed that higher level of protein expression was observed in BL21-CodonPlus-RIL cells induced at 37°C for 4 h. In the case of Rosetta-gami™ 2 (DE3) cells, higher level of protein expression was found when cells were induced at 18°C for 4 h (Fig 3.4).

Figure 3.4 – *his*-Med31 protein expression in BL21-CodonPlus-RIL strain (RIL) and Rosetta-gami™ 2 (DE3) cells cultured at 37°C and 18°C

The alternate BL21-derived cells were cultured either with 1 mM IPTG (+) or without IPTG (-) for 4 h at 37°C and 18°C. Crude cell lysates were separated using a 17% SDS-polyacrylamide gel and stained by coomassie blue. **M**: Protein marker. Molecular weight markers (kDa) are indicated in the figure. Band corresponding to the induced protein is indicated with an arrow.

3.2.1.4 Assay for solubility of *his*-Med31 protein in alternate BL21-derived cells

The solubility of the *his*-Med31 protein expressed in BL21-CodonPlus –RIL and Rosetta-gami™ 2 cells was tested using 5 different buffers described in section 2.4.2. The protein was found to be insoluble in all the buffers tested using both the cells (Fig 3.5 & 3.6).

Figure 3.5 - Solubility of his-Med31 protein in BL21-CodonPlus-RIL cells at 18°C (A) and 37°C (B)

Cellular fractions were separated using a 17% SDS-polyacrylamide gel and stained by coomassie blue. Solubility test showed that *his*-Med31 protein was mostly present in inclusion bodies in the insoluble fraction under all conditions tested. **I**: Insoluble fraction, **S**: Soluble fraction. Band corresponding to the induced protein is indicated with an arrow.

A **B**

Figure 3.6 - Solubility of *his*-Med31 protein in Rosetta-gami™ 2 cells at 18°C (A) and 37°C (B)

Cellular fractions were separated using a 17% SDS-polyacrylamide gel and stained by coomassie blue. Solubility test showed that *his*-Med31 protein was mostly present in inclusion bodies in the insoluble fraction under all conditions tested. **I**: Insoluble fraction, **S**: Soluble fraction. Band corresponding to the induced protein is indicated with an arrow.

3.2.1.5 Purification of recombinant *his*-Med31 under denaturing conditions

Protein lysate was prepared using urea (4M, 6M & 8M) and guanidinium chloride (6M) to solubilize the protein from inclusion bodies. Whole cell lysate was prepared from BL21 cells expressing *his*-Med31 induced with 1 mM IPTG and cultured at 37°C for 4 h. The protein was found to be present in the pellet fraction using both denaturants (Fig 3.7)

Figure 3.7 - Protein solubilisation from inclusion bodies using urea (A) and guanidinium chloride (B)

Soluble and insoluble fractions from BL21 cell lysate prepared under denaturing conditions were separated using a 17% SDS-polyacrylamide gel and stained by coomassie blue. **M**: Protein marker. Molecular weight markers (kDa) are indicated in the figure. The protein was present in the insoluble fraction using both denaturants. Band corresponding to *his*-Med31 protein is indicated with an arrow.

3.2.1.6 Ni-NTA Purification of *his*-tagged MED31 protein

Solubilisation of *his*-Med31 using denaturants showed that the recombinant protein was always present in the insoluble fraction. In the case of both denaturants, the cells were resuspended in buffer containing urea or GuHCl and incubated at room temperature for 1 h. Sonication was not performed to break open the cells. Therefore, the crude cell lysate was prepared with buffer (50 mM Tris/Hcl, 200 mM NaCl, pH 7.5) containing urea (8 M)

and subjected to sonication. The supernatant fraction was then purified using a batch Ni-NTA purification under denaturing conditions. The protein was eluted using an imidazole gradient. Bands corresponding to 16 kDa were observed in all the fractions except those eluted with 400 and 600 mM imidazole (Fig 3.8).

A

Figure 3.8 - Ni-NTA purification of *his*-Med31 under denaturing conditions (A & B)

Protein was eluted using an imidazole gradient (0–600 mM) after Ni-NTA purification. Protein samples were separated using a 17% SDS-polyacrylamide gel and stained by coomassie blue. **M**: Protein marker; **Wash fraction**: fraction collected after washing the column with buffer containing urea; **Unbound**: Flow-through fraction collected after loading of supernatant onto the Ni-NTA resin column. **Supernatant**: Cell lysate supernatant fraction, which was subjected to Ni-NTA purification. Molecular weight markers (kDa) are indicated in the figure. Band corresponding to *his*-Med31 protein is indicated with an arrow.

The protein fractions eluted with 10, 25 and 50 mM imidazole were concentrated using centrifugal columns. A faint band corresponding to ~16 kDa was also observed with fractions eluted with 250 and 300 mM imidazole which seemed relatively pure. These two fractions were also concentrated (Fig 3.9). However, western blot analysis indicated that the *his*-Med31 protein was not present in either of these concentrated fractions (results not shown).

Figure 3.9 - Concentration of imidazole eluted protein fractions

Protein fractions eluted with: 10, 25 and 50 mM imidazole ($C_{10}+C_{25}+C_{50}$); 250 and 300 mM imidazole ($C_{250}+C_{300}$) were separated using a 17% SDS-polyacrylamide gel and stained by coomassie blue. **Wash fraction**: fraction collected after washing the column with buffer containing urea; **Unbound**: flow-through fraction collected after loading of supernatant onto the Ni-NTA resin column; **Supernatant & Pellet**: Protein supernatant and pellet fraction obtained after cell sonication. Molecular weight markers (kDa) are indicated in the figure. Band corresponding to *his*-Med31 protein is indicated with an arrow.

3.2.2 Polyclonal antibody against Med31

The Med31 amino acid sequence was analyzed using the IEDB analysis resource software and based on features such as surface accessibility, antigenicity and hydrophilicity, a 15 amino acid peptide closer to the N-terminus was chosen for peptide

synthesis. A cysteine tagged peptide was synthesized to enable coupling to an immunogenic carrier, CGISQKKYIVNNIIDDNK-OH. The Med31 peptide antigen was coupled to keyhole limpet hemocyanin (KLH) protein in order to increase its immunogenicity and used to immunize two rabbits. Rabbit polyclonal antibody production was carried out at the Biological resource centre (BRC) at Agency for Science, Technology and Research (A*STAR), Singapore. Two booster doses were administered and the polyclonal sera was analysed by western blotting.

Rabbit polyclonal sera against Med31 were tested against parasite protein lysate by western blotting analysis. BL21 cells overexpressing *his*-Med31 and untransformed BL21 cells were used as positive and negative control respectively.

Figure 3.10 - Western blot analysis of polyclonal serum against Med31, obtained from two rabbits (A & B)

Protein lysates were prepared ring, trophozoite and schizont stages of *P. falciparum* (*PfT996*). Total cell lysate of non-transformed BL21 cells and total cell lysate of BL21 cells overexpressing *his*-tagged *Pf* Med31 were taken as a negative and positive control respectively (*E. coli*). Molecular weight markers (kDa) are indicated in the figure.

Protein bands in the range 15-20 kDa were observed for ring, trophozoite and schizont stages of the parasite and for the *his*-Med31 overexpressing BL21 cell lysate (Fig 3.10). However, a band of the expected size (15-20 kDa) was also observed for the control non-transformed BL21 cell lysate, pointing towards non-specific binding of the polyclonal serum. The polyclonal sera can be further purified using antigen-specific immunoaffinity purification and tested by western blot for specificity. However, in the course of this

project, this polyclonal serum was not used further for any experiments due to its non-specific binding.

3.2.3 Monoclonal antibody against Med7

A monoclonal antibody against the *P. falciparum* Mediator subunit, Med7, was produced by the company, GenScript, under contract with our laboratory. Med7 recombinant protein synthesized by GenScript was used to immunize BALB/c mice. SP2/0 myeloma cells were used for creation of hybridomas, which were screened by indirect ELISA. The ELISA positive supernatants provided by GenScript were analysed by western blot using wild-type *P. falciparum* total protein lysate and *P. falciparum* strain expressing episomal HA-tagged Med31 protein.

Figure 3.11 - Western blot analysis of Med7 monoclonal cell line

Total protein lysate prepared from wild-type T996 parasite strain (**T996**) and strain expressing Med31-HA episomally (**Med31-T996**) were separated using 15% SDS-polyacrylamide gel and analyzed by western blot using Med7 monoclonal cell line

supernatants. Molecular weight markers (kDa) are indicated in the figure. Band corresponding to Med7 protein is indicated with an arrow.

Ten hybridoma supernatants were tested by western blot (data not shown) and one hybridoma supernatant (Clone 1H2H2) was chosen for antibody purification. The purified antibody was provided by GenScript, which was used for further experiments.

3.3 Conditional Knockdown of Med31 protein

3.3.1 Endogenous epitope tagging approach

In order to conditionally knockdown MED31 protein expression, a destabilization domain derived from the *E. coli* dhfr gene (ecDDD) was fused to the carboxy terminus of endogenous Med31. When the ecDDD is fused to any protein of interest, the fusion protein becomes unstable and is targeted for degradation by the proteasome. However, a ligand called trimethoprim (TMP) can stabilize the fusion protein in a dose-dependent manner [150]. Therefore, the addition and removal of TMP can be used to conditionally alter protein levels to assess protein function.

The destabilization domain was fused to the endogenous Med31 gene using single crossover homologous recombination. The Med31 coding region and a 5' flanking region comprising 429 bases upstream of the Med31 coding region was used as the homology arm for mediating single crossover. The genomic locus including the 5' flanking region and Med31 coding sequence (excluding the stop codon) was amplified from *Pf*T996 strain genomic DNA using the primers, Med31-HA-DDD-HR-Fwd and Med31-HA-DDD-HR-Rev (Table 2.1). Med31 was cloned between the restriction sites XhoI and AvrII in pHSP101-HADB plasmid, upstream of the 3X-Hemagglutinin (HA) tag and the destabilization domain (ecDDD) to yield the Med31-HADB plasmid (Fig 3.12). pHSP101-HADB plasmid was a kind gift from Daniel E. Goldberg's lab [150]. Positive clones were confirmed by both restriction digestion and PCR analysis. The recombinant plasmid was further confirmed by sequencing.

The drug, TMP, acts by inhibiting the enzyme dihydrofolate reductase. Therefore, the ecDDD system requires the use of parasite strains resistant to this antifolate drug. Therefore, the recombinant plasmid Med31-pHADB was co-transfected with another

plasmid encoding the hDHFR cassette conferring resistance to TMP. The parasites were selected using two drugs, blasticidin (2.5 $\mu\text{g/ml}$) and WR99210 (2.5 nM). After the appearance of drug-resistant parasites around 14 days post transfection, WR99210 was replaced with TMP (10 μM) in the culture medium. To select for parasites with homologous recombination by single crossover, parasites were subjected to two rounds of drug cycling using blasticidin (2.5 $\mu\text{g/ml}$). After two rounds of cycling, genomic DNA was extracted from the parasites and analysed by PCR to check for the desired integration into the endogenous locus.

Figure 3.12 - Schematic diagram of Med31-pHADB plasmid

The *P. falciparum* Med31 gene was cloned upstream of hemagglutinin tag (3X-HA) and the *E. coli* derived DHFR-based destabilization domain (ecDDD), using the restriction enzyme sites, XhoI and AvrII. The plasmid contains the ampicillin (amp^R) and blasticidin (BSD) resistance gene cassette for selection in *E. coli* and *P. falciparum* respectively.

3.3.2 Genomic loci analysis

Genomic DNA extracted from the parasites was analysed by PCR to check for integration into the endogenous locus. The strategy used for incorporating the HA tag and the ecDDD to the 3' end of endogenous Med31 locus by single-crossover homologous recombination has been shown in Fig 3.13. It is to be noted that the full-length Med31 coding region and a 5' flanking region comprising 429 bases upstream of the coding region were used for single-crossover homologous recombination. This is because the size of Med31 coding region is too small (399 bp) for considering only the 3' end of the gene for homologous recombination. Therefore, to increase the length of the homology region, the 5' upstream flanking region was used in addition to the Med31 coding sequence. However, this leads to a duplicate copy of the Med31 gene, which could potentially be expressed if its promoter region lies within the 429 bp upstream 5' flanking region.

Figure 3.13 - Endogenous tagging of Med31 by homologous recombination

A 429 bp region (5' upstream), upstream of the Med31 coding region along with the Med31 ORF was cloned upstream of the HA-DDD tag present in the plasmid. A single

crossover event resulted in insertion of the Med31-pHADDB expression construct and tagging of endogenous Med31 gene. The arrowheads labeled as P1, P2, P3, P4 and P5 denote the primers used for amplifying the wild-type, integrant and plasmid locus. The primer combination used for amplifying each locus is shown in the table. Primer P5 is designed such that it binds to the first 17 bases of the ecDDD sequence, whereas primer P3 is designed to cover 20 bases towards the end of the ecDDD sequence. Although both primers could have been used interchangeably, primer P3 was used to amplify majority of the Med31-HA-DDD coding sequence for confirming the desired integration into the endogenous genomic locus.

Figure 3.14 - Diagnostic PCR shows integration into the endogenous locus

Genomic DNA obtained from wild-type and drug resistant parasites after two rounds of

cycling with blasticidin (2.5 $\mu\text{g/ml}$) were analysed by PCR using primer combination P1+P2 for wild-type locus, P4+P5 for episomal DNA and P1+P3 for integrant locus.

A band corresponding to the integrated loci was observed for the parasites (Fig 3.14). However, the parasite culture still contains a mixed population of parasites having the plasmid integrated into the genomic locus and parasites without integration but carrying the plasmid episomally. Single parasite clones having the plasmid integrated genomic locus can be obtained by a technique called limiting dilution, which will be described, in the section below.

Dilution plating of the parasite culture was carried out in a 96-well plate. A total of 80 wells were prepared, out of which 15 wells contained parasites. Genomic DNA was extracted from these 15 clones and analysed by PCR.

Figure 3.15 - Genomic DNA analysis of clones obtained from limiting dilution

Genomic DNA extracted from the clones was amplified using primers for integrant loci (A) and endogenous loci (B & C). Parental T996 strain was used as a control (**wild-type**). While the band corresponding to the integrant loci was observed (A), the endogenous locus was not detected in any of the 15 clones (B & C), showing that all the clones contained the desired integration.

All 15 clones showed the presence of the integrated plasmid at the genomic locus (Fig 3.15). PCR was also done to check for parasites that did not contain the integration but carried the plasmid episomally. All 15 clones showed an absence of the endogenous loci, indicating that they are pure integrant clones.

Clone B2 was chosen at random and used for further experiments. Total RNA was extracted from these parasites, cDNA was synthesized and analysed by RT-PCR using primers designed to amplify the integrated locus (Fig 3.16). RT-PCR analysis showed the expression of the Med31^{HA-DDD} transcript.

Figure 3.16 - Transcript analysis by reverse transcription-PCR

The figure shows the PCR analysis of cDNA preparations with (+DNase) and without (-DNase) RNase-free DNase. RT-PCR was carried out in the presence (+RT) and absence (-RT) of reverse transcriptase enzyme. The Med31^{HA-DDD} transcript was detected and no bands were observed in the -RT control.

Both RNA treated with and without DNase was used for the cDNA synthesis reaction. RT-PCR was carried out in the presence and absence of reverse transcriptase

enzyme. As expected, no bands were observed in the –RT control (Fig 3.16), indicating the absence of any appreciable genomic DNA contamination.

3.4 Episomal expression of epitope-tagged Med31

In order to study Mediator protein interactions by means of epitope tagging, *P. falciparum* Med31 full-length gene (~399 bp), without the stop codon, was cloned upstream of the hemagglutinin (3X-HA) tag already present in the pBcamR_3XHA-X vector (Fig 3.17). The recombinant plasmid was checked by restriction digestion analysis and PCR. The clones were further confirmed by sequencing.

Figure 3.17 - Plasmid vector map of pBcamHA

Med31 coding region was cloned upstream of the triple hemagglutinin tag (3x-HA) into the multiple cloning site (MCS) between the restriction enzyme sites of BamHI and NheI.

The calmodulin promoter (cam) drives transgene expression. The plasmid contains ampicillin (amp^R) and blasticidin (BSD) resistance cassettes for selection in *E. coli* and *P. falciparum* respectively.

P. falciparum T996 wild-type parasites were transfected with the recombinant plasmid, such that the plasmid could be episomally maintained. The parasites were grown under drug pressure (blasticidin, 2.5 µg/ml) and drug-resistant parasites were observed 21 days post-transfection. PCR was performed to confirm the presence of episomally expressed Med31-HA gene.

3.4.1 Episomal DNA and transcript analysis

DNA was extracted from drug-resistant parasites and subjected to PCR using primers designed to amplify the endogenous Med31 gene as well as the episomally expressed HA-tagged Med31 gene (Fig 3.18).

Figure 3.18 - Schematic representation of the HA-tagged Med31 gene present on the pBcamHA plasmid

The arrows (blue) indicate the primers designed to amplify endogenous Med31 and episomally expressed Med31-HA.

PCR analysis confirmed the presence of HA-tagged Med31 gene (484 bp) in the drug-resistant parasites. As expected, wild-type T996 parasites showed the presence of endogenous Med31 (399 bp) but not the HA-tagged Med31 gene (Fig 3.19).

Figure 3.19 - PCR analysis of DNA extracted from wild-type (T996) and drug resistant parasites (Med31-T996)

PCR was carried out with primers designed to amplify endogenous Med31 (Med31) and HA-tagged Med31 gene (Med31-HA).

Transcript expression was analysed by reverse transcription-PCR (RT-PCR) using primers designed to amplify endogenous Med31 and episomally expressed HA-tagged Med31. Transcripts corresponding to endogenous Med31 (~399 bp) and 3X-HA tagged Med31 (~483 bp) were found to be expressed. RNA either treated or not treated with a DNase was used for cDNA synthesis. RT-PCR was carried out in the presence and absence of reverse transcriptase enzyme. As expected, no bands were observed in the -RT control (Fig 3.20), indicating the absence of any appreciable genomic DNA contamination.

This parasite strain expressing Med31-HA episomally was used for co-immunoprecipitation and mass spectrometric analysis to identify protein interaction partners of Med31. It was also used for ChIP-Seq analysis using the HA antibody to assess Mediator occupancy across the genome.

Figure 3.20 - RT-PCR analysis for detection of endogenous Med31 & episomally expressed HA-tagged Med31 transcript

The figure shows the PCR analysis of cDNA preparations with (+DNase) and without (-DNase) RNase-free DNase. RT-PCR was carried out in the presence (+RT) and absence (-RT) of reverse transcriptase enzyme. The endogenous Med31 and Med31-HA transcripts were detected and no bands were observed in the -RT control.

3.5 Episomal expression of epitope-tagged putative Mediator subunits

Two putative Mediator subunits PF3D7_0526800 and P3D7_1363600 were identified in the co-immunoprecipitation and mass spectrometric analysis of episomal Med31-HA protein, which will be discussed in detail in chapter 4. To further characterize them, these two putative Mediator subunits were each cloned into pBcamHA vector upstream of the haemagglutinin (3X-HA) tag, for episomal expression (Fig 3.21). Cloning and generation of the constructs was carried out by the company, GenScript, under contract with our laboratory.

Figure 3.21 - Schematic representation of the PF3D7_0526800 and PF3D7_1363600 expression constructs

Full-length PF3D7_0526800 PF3D7_1363600 genes were cloned upstream of the triple hemagglutinin tag in the pBcamHA plasmid. The plasmid contains ampicillin (amp^R) and blastidin (BSD) markers for selection in *E. coli* and *P. falciparum* respectively. The arrows indicate the primers designed to amplify the endogenous gene and episomally expressed HA tagged gene.

The clones provided by GenScript were confirmed in the laboratory by restriction digestion and PCR analysis (Fig 3.22).

Figure 3.22 - Restriction digestion and PCR analysis of recombinant PF3D7_0526800-HA and P3D7_1363600-HA clones

pBcamHA empty vector was used as a control for the restriction digestion. Restriction digestion with BamHI and NheI: PF3D7_0526800-HA (7598 bp & 1116 bp) and P3D7_1363600-HA (7598 bp & 693 bp). The expected PCR product for both the genes PF3D7_0526800-HA (1116 bp) and P3D7_1363600-HA (693 bp) was observed.

P. falciparum T996 wild-type parasites were transfected with the recombinant plasmids, such that the plasmid could be episomally maintained. The parasites were grown under drug pressure (blasticidin, 1 µg/ml) and drug-resistant parasites were observed 21 days post-transfection. PCR was performed to confirm the presence of episomally expressed genes (Fig. 3.23).

Figure 3.23 - PCR analysis of DNA extracted from parental T996 strain (wild-type) and drug resistant parasites (transfectants)

PCR was done with primers designed to amplify the wild-type PF3D7_0526800 (1104 bp) and P3D7_1363600 (681 bp) genes and HA tagged PF3D7_0526800 (1116 bp) and P3D7_1363600 (693 bp) genes.

PCR analysis confirmed the presence of HA-tagged PF3D7_0526800 and P3D7_1363600 gene in the drug-resistant parasite lines. As expected, wild-type T996 parasites showed the presence of endogenous PF3D7_0526800 and P3D7_1363600 gene but not the HA-tagged version of the genes.

4 FUNCTIONAL VALIDATION OF THE MEDIATOR COMPLEX IN *P. FALCIPARUM*

4.1 Conditional knockdown of endogenous Med31 protein

In order to understand the function of Med31 protein in *P. falciparum*, a conditional knockdown system was employed, wherein an *E. coli* dihydrofolate reductase (DHFR)-based destabilisation domain (ecDDD) was fused to endogenous Med31 by single-crossover homologous recombination. The ecDDD can be stabilised by addition of the drug, TMP and thus provides a means to regulate Med31 accumulation to better understand its function. The generation of the endogenous epitope-tagged cell line, Med31^{HA-DDD}, has been discussed in detail in chapter 3 (section 3.3). The Med31^{HA-DDD} cell line was further analysed for growth, protein expression and protein stability in the presence and absence of TMP.

4.1.1 Expression of Med31^{HA-DDD} protein

In order to determine expression of Med31^{HA-DDD} protein, total protein lysate was prepared from asynchronous cultures with a parasitemia greater than 5% and analysed by western blot using an anti-HA antibody. Total protein lysate from wild-type untransfected T996 strain was used as a negative control. A band corresponding to the expected molecular weight of Med31^{HA-DDD} protein (~38 kDa) was observed in the Med31^{HA-DDD} cell line (Fig 4.1 A). No band was observed with the wild-type T996 strain, which is expected as it does not express any HA-tagged protein. A lower molecular weight band migrating between 16 and 30 kDa was also observed with the Med31^{HA-DDD} strain.

To examine the expression of endogenous Med31 across the intraerythrocytic developmental cycle (IDC), protein samples were harvested from synchronous parasite cultures at ring, trophozoite and schizont stages. Western blot analysis showed that the endogenous Med31^{HA-DDD} protein is expressed at all three stages of the parasite IDC (Fig 4.1 B).

Figure 4.1 - Western blot showing endogenous Med31^{HA-DDD} protein expression (A) & Med31^{HA-DDD} protein expression across the IDC (B)

(A): Total protein lysates were prepared from the Med31 conditional knockdown strain (**Med31^{HA-DDD}**) cultured in the presence of TMP. Wild-type untransfected strain (**T996**) was used as a negative control. Epitope-tagged Med31 protein was detected only in the Med31^{HA-DDD} cell line and not in the wild-type control. **(B):** Protein samples harvested at ring, trophozoite and schizont stages of the parasite were analysed by western blot. Med31^{HA-DDD} protein was expressed at all the three stages.

The protein samples were separated using a 12% SDS-polyacrylamide gel and analysed by western blot using anti-HA antibody. β -actin was used as the loading control. Molecular weight markers (kDa) have been indicated in the figure. Band corresponding to Med31^{HA-DDD} protein is indicated with an arrow.

4.1.2 Analysis of protein expression in the presence and absence of stabilizing ligand

To conditionally knockdown Med31 protein expression, the stabilizing ligand, TMP, was removed from the culture medium. Parasites were cultured for 24 h in the presence and absence of 10 μ M TMP and analysed by western blot using an anti-HA antibody. The results showed that the Med31^{HA-DDD} protein is not degraded in the absence of the stabilizing ligand. A band corresponding to the molecular weight of Med31^{HA-DDD} (38 kDa) was observed in parasites cultured in the presence and absence of TMP (Fig 4.2 A). However, a lower molecular weight protein migrating between 16 and 30 kDa was observed only in parasites cultured with TMP.

Protein stability in the presence or absence of TMP was further assessed by immunoprecipitation of Med31^{HA-DDD} protein from asynchronous parasites grown in the presence or absence of 10 μ M TMP for 24 h. Magnetic beads coupled with HA antibody were used for immunoprecipitation and the proteins were eluted using laemmli buffer. The results showed that the full-length Med31^{HA-DDD} protein was immunoprecipitated from both cultures, grown with and without TMP (Fig 4.2 B). Again, a lower molecular weight protein migrating between 16 and 30 kDa was observed only in the parasite culture grown with TMP. Another protein with a molecular weight lower than 16 kDa was also observed in both eluate samples.

Since the endogenous destabilisation domain-tagged Med31 protein is observed even in cultures grown without TMP, it appears that the Med31 protein is not being targeted to the proteasome for degradation.

Figure 4.2 - Med31^{HA-DDD} protein stability in the presence or absence of TMP (A) & Immunoprecipitation of endogenous Med31^{HA-DDD} protein from asynchronous cultures grown in the presence or absence of TMP (B).

A. Protein samples harvested from Med31^{HA-DDD} cell line cultured in the presence (+TMP) or absence (-TMP) of TMP for 24 h were analysed by western blot. **B.** Immunoprecipitation of endogenous Med31^{HA-DDD} protein from asynchronous cultures grown in the presence (+TMP) or absence (-TMP) of 10 μ M TMP for 24 h.

Proteins were separated on a 12% polyacrylamide gel and analysed by western blot using an anti-HA antibody. Molecular weight markers (kDa) are indicated in the figure. Band corresponding to Med31^{HA-DDD} protein is indicated with an arrow.

4.1.3 Growth analysis of parasites in the presence and absence of stabilizing ligand

Protein expression and immunoprecipitation experiments show that the Med31^{HA-DDD} protein is not degraded upon removal of TMP. It could be that the destabilisation domain still interferes with protein function by affecting its interactions with other proteins [151] and has an indirect effect on parasite growth and viability. Therefore,

parasite growth was monitored in the presence or absence of TMP. Med31^{HA-DDD} parasites were grown in the presence and absence of 10 μ M TMP for 4 days and the parasitemia was calculated by microscopic examination of geimsa stained smears. Asynchronous parasite cultures were used, with a starting parasitemia of 1% on day 0. Medium was changed every day and fresh erythrocytes were added from day 2 onwards. Parasites were subcultured 1:2 each day so that the cultures were not adversely affected by high parasitemia. Growth curve of Med31^{HA-DDD} parasites with and without TMP (Fig 4.3) does not show any significant decrease in parasitemia in the absence of the stabilizing ligand.

Figure 4.3 - Growth analysis of asynchronous Med31^{HA-DDD} parasites in the presence and absence of TMP

Asynchronous Med31^{HA-DDD} parasites were cultured with or without 10 μ M TMP and their growth was monitored by measuring parasitemia over 4 days. Error bars represent standard deviation of three independent biological replicates.

4.2 Analysis of the Mediator complex subunit Med7

The purified Med7 monoclonal antibody produced by GenScript, was tested by

western blot analysis. Total protein lysate prepared from asynchronous wild-type T996 strain was tested. The recombinant *Pf*Med7 protein (produced by GenScript) was used as a positive control, while uninfected RBC protein lysate was used as a negative control. A band corresponding to endogenous Med7 (~28 kDa) was observed for the T996 total protein lysate and the Med7 recombinant protein, but not in the uninfected RBC lysate (Fig 4.4 A). Although endogenous Med7 is predicted to have a molecular weight of 28 kDa, the band corresponding to Med7 in the T996 total protein lysate was found to migrate between 30 and 37 kDa on a 12% polyacrylamide gel. However, for the recombinant Med7 protein, a band migrating close to 30 kDa was observed.

Figure 4.4 - Evaluation of Med7 monoclonal antibody by western blot analysis (A) & Expression of endogenous Med7 across the IDC (B)

(A): Total protein lysate prepared from *P. falciparum* T996 strain (**T996**) was used. Recombinant Med7 protein produced by GenScript (**Med7-his**) and uninfected RBC

protein lysate were taken as positive and negative control respectively. **(B):** Protein lysates prepared from the ring, trophozoite and schizont stages of the parasite were tested by western blot using the Med7 monoclonal antibody. The endogenous protein was found to be expressed in all stages.

To test the stage-specific expression of the endogenous Med7 protein, protein lysates were prepared from the ring, trophozoite and schizont stages and tested by western blot using the Med7 monoclonal antibody. Western blot analysis showed that the endogenous protein was expressed in all stages (Fig 4.4 B). The Med7 antibody was further used to immunoprecipitate endogenous Med7 protein from *P. falciparum* T996 strain. However, several non-specific bands were observed in the eluted fraction (data not shown). The antibody also produced non-specific signals in immunofluorescence assays. Hence, the Med7 monoclonal antibody was not used for any further experiments.

4.3 Analysis of the Mediator complex subunit Med31

4.3.1 Molecular characterization of episomally expressed Med31-HA protein

4.3.1.1 Protein expression of Med31

P. falciparum T996 strain expressing Med31-HA protein episomally was established during the course of this study. The parasites were cultured in medium containing blasticidin at a concentration of 2.5 µg/ml and protein expression was determined by western blot. Total protein lysate prepared from wild type T996 was used as a negative control. Med31-HA protein (~19 kDa) was observed only in the Med31-HA episomal strain and not in the wild-type T996 strain (Fig 4.5 A). The copy number of expression constructs has been found to increase with increasing blasticidin concentration [152] leading to an increase in transgene expression. Therefore, Med31-HA parasites were cultured with varying concentrations of blasticidin (2.5, 5 and 7.5 µg/ml) and the protein expression was analysed by western blot (Fig 4.5 B). The levels of episomal Med31-HA protein were found to increase with increasing blasticidin drug concentration.

To determine the expression of episomal Med31-HA protein across the IDC, protein samples were harvested at the ring, trophozoite and schizont stages from Med31-

HA parasites and analysed by western blot.

Western blot analysis showed that the episomal Med31-HA protein is expressed in ring, trophozoite and schizont stages (Fig 4.6). However, the protein expression is highest in trophozoite stage. The episomal Med31-HA gene is under the control of the non-endogenous calmodulin promoter. The calmodulin promoter in *P. falciparum* is known to display maximum gene expression between 14 - 40 h [153]. As expected, the expression of episomal Med31-HA protein is lowest in the ring stage and highest in the trophozoite stage. In contrast, the endogenous Med31 protein is expressed uniformly throughout all parasite life stages, as seen with the Med31^{HA-DDD} cell line (Fig 4.1 B).

Figure 4.5 - Episomal Med31-HA protein expression (A) & Expression of episomal Med31-HA protein with increasing blasticidin concentration (B)

(A) Western blot analysis was carried out on wild-type T996 strain (T996) and parasite

strain expressing Med31-HA episomally (**T996-Med31-HA**). Image is representative of two independent experiments. **(B)** Western blot analysis was carried out on T996-Med31-HA cell line cultured with 2.5, 5 and 7.5 $\mu\text{g/ml}$ blasticidin. Band corresponding to Med31-HA protein is indicated with an arrow. Image is representative of two independent experiments.

Total protein lysate was prepared and the protein samples were separated using a 12% SDS-polyacrylamide gel. Western blot was performed with anti-HA antibody and β -actin was used as the loading control. Molecular weight markers (kDa) are indicated in the figure.

Figure 4.6 - Episomal Med31-HA protein expression across the IDC

Western blot analysis was carried out on T996-Med31-HA cell lines across ring, trophozoite and schizont stages. Protein samples were separated using a 15% SDS-polyacrylamide gel and western blot was carried out using the anti-HA antibody. β -actin was used as the loading control. Band corresponding to Med31-HA protein is indicated

with an arrow. The experiment was carried out once.

4.3.1.2 Intracellular localization of Med31

As a presumed component of the *P. falciparum* Mediator complex, Med31 would be expected to show significant nuclear localization. The subcellular localization of the episomal Med31-HA protein was determined by immunofluorescence analysis using an anti-HA antibody. Parasite smears were prepared from asynchronous cultures grown with 2.5 µg/ml blasticidin. DAPI was used to stain the nucleus. Fluorescence signal observed from anti-HA antibody was mainly localized to the nucleus with a distinct punctate pattern (Fig 4.7 A). The fluorescence signal intensity was low for parasites cultured at 2.5 µg/ml blasticidin. Therefore, they were cultured at a higher blasticidin concentration of 7.5 µg/ml to select for increased episome copy number. Fluorescence intensity was much higher and it could be clearly observed that Med31-HA protein mostly had a nuclear localization (Fig 4.7 B). Because some signal was also observed outside the nucleus, a co-localisation IFA was performed with aldolase as a cytoplasmic marker. This showed that Med31 protein was found to localize to the cytoplasm as well. However, there is an overlap between nuclear DAPI and cytoplasmic aldolase signals, as observed in Fig 4.8. Therefore, the results of the co-localisation analysis are inconclusive. This experiment was carried out only once and needs to be repeated with optimized conditions to confirm the cytoplasmic localization of Med31. Nonetheless, the significant nuclear accumulation of Med31 observed here (Fig 4.7) is consistent with its presumed role as a transcriptional regulator.

Figure 4.7 - Intracellular localization of Med31-HA protein in Med31-HA cell lines cultured with BSD 2.5 µg/ml (A) and 7.5 µg/ml (B).

Immunofluorescence assay (IFA) was carried out to determine the localization of episomal Med31-HA protein in Med31-HA cell lines cultured with blasticidin concentration of 2.5 $\mu\text{g/ml}$ (**A**) and 7.5 $\mu\text{g/ml}$ (**B**). Anti-HA antibody (**green**) was used to detect the HA tagged protein and DAPI was used to stain the nucleus (**blue**). Fluorescence signal was detected mainly in the nucleus. Scale bar = 2 μm . Each image is representative of two independent experiments.

Figure 4.8 - Co-localisation of Med31-HA with a cytoplasmic marker

Immunofluorescence assay (IFA) was carried out to determine the co-localization of episomal Med31-HA protein with a cytoplasmic marker, aldolase. Anti-HA antibody (**green**) was used to detect the HA tagged protein while anti-aldolase antibody was used to detect aldolase (**red**). DAPI was used to stain the nucleus (**blue**). Med31-HA protein is localized predominantly to the nucleus and is also present in the cytoplasm. Scale bar = 2 μm . The experiment was carried out once.

The episomal Med31 protein was found to be present in both nucleus and cytoplasm based on immunofluorescence assays. A western blot analysis was done to further confirm this observation. Nuclear and cytoplasmic protein extracts were prepared from Med31-HA cell line and subjected to immunoprecipitation using anti-HA antibody coupled to magnetic beads. Western blot analysis was done using the HA antibody. A heavy-chain specific secondary antibody was used for western blot, such that the signal from antibody light chain (~25 kDa) does not interfere with the signal from Med31-HA protein (~19 kDa).

Med31-HA protein (~19 kDa) was observed in both the nuclear and cytoplasmic pull down samples (Fig 4.9), consistent with the observation by immunofluorescence. Cellular fractionation can be further verified by western blot analysis of known nuclear and cytoplasmic markers in each of the sub-cellular fractions. This would give us greater confidence in interpreting these localization results. Since episomal expression of Med31 is essentially an overexpression system, we cannot rule out the possibility that cytoplasmic levels of Med31 could be an artefact of overexpression of the fusion protein. A higher molecular weight band corresponding to 66 kDa was also observed in the cytoplasmic eluate sample.

Figure 4.9 - Episomal Med31-HA protein is present in both nucleus and cytoplasm

Nuclear and cytoplasmic fractions were prepared from asynchronous cultures grown with blasticidin (2.5 $\mu\text{g/ml}$). Med31-HA protein was immunoprecipitated from nuclear and cytoplasmic fractions using anti-HA antibody. The input, eluate and flow-through samples from IP were separated using a 12% SDS-polyacrylamide gel and western blot was carried out using the anti-HA antibody. Med31 protein was found to be present in both nucleus and cytoplasm. Purity of the cellular fractions needs to be further confirmed by western

blot analysis of known nuclear and cytoplasmic markers. Molecular weight markers (kDa) are indicated in the figure. Band corresponding to Med31-HA protein is indicated with an arrow. The experiment was carried out once.

4.3.1.3 Immunoprecipitation of Med31 from *P. falciparum* total protein lysate

After characterizing appropriate protein expression and nuclear localization, the next step was to immunoprecipitate the episomal Med31-HA protein in order to identify its protein interaction partners by mass spectrometric analysis. Asynchronous cultures grown in the presence of blasticidin (2.5 µg/ml) and having a parasitemia greater than 10% were harvested for immunoprecipitation with an antibody against the hemagglutinin tag. A non-specific antibody against rabbit IgG was used as a negative control. A fraction of the parasite cell lysis supernatant was taken as input and the remaining amount was split equally into two portions and subjected to affinity purification with the HA antibody and the non-specific IgG antibody respectively.

Western blot analysis using anti-HA antibody showed that the episomal Med31-HA was successfully pulled down from parasite lysate (Fig 4.10). A band corresponding to Med31 (19 kDa) was observed only in the IP with HA antibody and not with the non-specific IgG antibody. Since a heavy-chain specific secondary antibody was used for the western blot analysis, only the band corresponding to antibody heavy chain (~55 kDa) was observed in both eluate fractions using HA and IgG antibodies.

Figure 4.10 - Immunoprecipitation of Episomal Med31-HA protein

Asynchronous cultures grown with blasticidin (2.5 $\mu\text{g/ml}$) were subjected to immunoprecipitation using anti-HA and anti-IgG antibodies. Input, eluate and flow-through fractions from the IP have been shown in the figure. Protein samples were separated using a 12% SDS-polyacrylamide gel and western blot was carried out using the anti-HA antibody. Molecular weight markers (kDa) have been indicated in the figure. A heavy chain specific secondary antibody was used for western blot. Band corresponding to Med31-HA protein is indicated with an arrow. Image is representative of two independent experiments.

4.3.2 Identification of protein interaction partners of epitope tagged Med31

4.3.2.1 Co-immunoprecipitation of Med31 and mass spectrometric analysis

Protein interaction networks and protein complex composition can be studied by affinity purification and mass spectrometric analysis. To identify the potential binding partners of Med31, a pull down experiment was performed on the Med31-HA parasites using HA antibody. To rule out non-specific interactions, wild-type T996 strain was used as a negative control. IP was scaled up and 6 ml of infected blood ($>10\%$ parasitemia, 2% hematocrit) was harvested for the experiment. Western blot analysis of the IP samples

showed that the Med31-HA protein was successfully immunoprecipitated from Med31-HA parasites (Fig 4.11, Med31-HA Eluate I). The eluate obtained from T996 strain did not show any bands indicating the specificity of the HA antibody. A two-step elution was done using Laemmli buffer. However, most of the proteins were extracted in the first elution step (Fig 4.11).

Figure 4.11 - Co-immunoprecipitation of Episomal Med31-HA protein

Asynchronous Med31-HA and wild-type T996 cultures (> 10% parasitemia, 2% hematocrit) grown with blasticidin (2.5 µg/ml) were subjected to immunoprecipitation using anti-HA antibody. Input, eluate and flow-through fractions from the IP are shown in the figure. Protein samples were separated using a 12% SDS-polyacrylamide gel and western blot was carried out using the anti-HA antibody. Molecular weight markers (kDa) have been indicated in the figure. Band corresponding to Med31-HA protein is indicated with an arrow. Image is representative of two independent experiments.

In addition to the Med31-HA protein, a higher molecular weight protein migrating to 66 kDa was also observed in the Med31-HA eluate sample. A protein of similar size was observed previously in the immunoprecipitated cytoplasmic fraction (Fig 4.9) indicating that this protein may be present in the cytoplasmic fraction.

The IP eluate samples (T996 and Med31-HA) were further analysed by LC-MS/MS (liquid chromatography coupled with tandem mass spectrometry) based sequencing at the NTU mass spectrometry facility. The peptide spectra were searched against the *P. falciparum* database to identify proteins associated with Med31.

The amount of protein in a complex protein mixture can be determined by calculating its emPAI score. emPAI refers to exponentially modified protein abundance index, which is used for quantifying the amount of protein in a complex protein mixture. The emPAI is defined as follows:

$$\text{emPAI} = 10^{\text{PAI}} - 1$$

PAI refers to the protein abundance score and is calculated as follows:

$$\text{PAI} = \left(\frac{\text{Number of observed peptides per protein}}{\text{Number of observable peptides per protein}} \right)$$

emPAI is the exponential form of PAI minus 1 and provides an absolute quantification of protein abundance in a mixture [154].

emPAI scores were calculated for Med31 and the T996 negative control samples and the ratio of emPAI scores of Med31 compared to T996 was deduced. Proteins with high emPAI score ratios were considered to be strong potential interacting partners. A total of 846 proteins were identified in the pull down experiment, out of which 13 proteins were found to have high emPAI score ratios greater than 500. The high scoring proteins have been shown in Table 4.1. The individual emPAI scores and the emPAI score ratios have also been indicated for each protein.

The mass spectrometric results showed that the annotated *P. falciparum* Mediator subunits, Med6, Med7 and Med31 were present in the pull down results as high-scoring proteins (Table 4.1). Importantly, 11 out of 14 Mediator subunits that had been identified computationally by Bourbon [144] were found to have co-immunoprecipitated with Med31 and were all among the highest scoring “hits” in this experiment. This study therefore provides a physical validation of the Mediator complex in *P. falciparum*, which has only been computationally predicted. The computationally predicted subunits, Med9,

Med11 and Med20 were not identified in the pull down experiment. Interestingly, two uncharacterized proteins encoded by PF3D7_0526800 and PF3D7_1363600 were among the highest scoring proteins detected by mass spectrometry.

S. No	Subunit	Module	Gene ID	Detected by MS	emPAI (Med31)	emPAI (T996)	emPAI ratio (Med31/T996)
1	Med6	head	PF3D7_1469700	Yes	2.13	0.001	2130
2	Med8	head	PF3D7_0411100	Yes	1.48	0.001	1480
3	Med11	head	PF3D7_0505900	No	NA	NA	NA
4	Med17	head	PF3D7_0520200	Yes	1.86	0.001	1860
5	Med18	head	PF3D7_1213200	Yes	2.78	0.001	2780
6	Med20	head	PF3D7_1463000	No	NA	NA	NA
7	Med22	head	PF3D7_1469800	Yes	1.09	0.001	1090
8	Med4	middle	PF3D7_1465200	Yes	10.88	0.001	10880
9	Med7	middle	PF3D7_0822100	Yes	7.35	0.001	7350
10	Med9	middle	PF3D7_1446700	No	NA	NA	NA
11	Med10	middle	PF3D7_0707600	Yes	1.96	0.001	1960
12	Med21	middle	PF3D7_1126400	Yes	10.31	0.001	10310
13	Med31	middle	PF3D7_1475000	Yes	1.09	0.001	1090
14	Med14	tail	PF3D7_0709300	Yes	1.34	0.001	1340
15	-	-	PF3D7_0526800	Yes	2.18	0.001	2180
16	-	-	PF3D7_1363600	Yes	2.15	0.001	2150

Table 4.1 - Mediator subunits of *P. falciparum* identified by Co-IP and LC-MS/MS analysis of Med31

The table shows the list of predicted Mediator subunits in *P. falciparum* identified by our mass spectrometry results. Eleven out of 14 Mediator subunits were confirmed to be present in *P. falciparum* and two additional proteins, PF3D7_0526800 and PF3D7_1363600 were also identified in our study. The individual emPAI scores and the ratio of emPAI scores have been indicated in the table. The PlasmoDB gene IDs and module assignment for the known Mediator subunits have also been included in the table.

4.3.2.2 Gene ontology analysis of interacting partner proteins

Proteins identified in the mass spectrometric analysis were classified based on molecular function using the Gene Ontology (GO) analysis tool available on PlasmoDB.

A total of 142 proteins with an emPAI score greater than 10 were considered for the analysis. Based on the GO results, the Mediator complex seems to interact with proteins involved in key cellular processes such as gene expression, metabolism and protein turnover (Table 4.2).

Apart from subunits of the Mediator complex itself, several other proteins with transcriptional functions were identified. These included subunits of RNAPII and some general transcription factors (GTFs). RPB2, RPB3 and RPB9 subunits of RNAPII and TFIID subunit 2 were some of the prominent transcription-related proteins identified. This is consistent with studies that have shown the role of the Mediator head module in recruitment of TFIID to the PIC [155]. Although Mediator is known to interact with the C-terminal domain of RNAPII [124], the CTD was not identified in the pull down results. Chromatin related proteins such as chromatin assembly protein (ASF1), chromodomain protein (PF3D7_1220900) and a bromodomain protein (PF3D7_1475600) were also found to interact with the Mediator complex. Mediator also seems to play a role in protein turnover as it interacts with the 26S proteasome complex regulatory subunits such as RPN9, RPN10, subunit 7 and ubiquitin-related proteins such as E3 ubiquitin-protein ligase and ubiquitin conjugation factor E4 B. Interaction of Mediator with proteasome components could also reflect its own degradation.

Mediator is known to act as a bridge between transcription factors bound to enhancer sequences and the general transcription machinery assembled at the core promoter. Consistent with its function, Med31 was also found to interact with a few transcription factors. These include a putative AP2 domain transcription factor (PF3D7_1449500) and a putative zinc finger protein (PF3D7_1464200). Apart from transcription initiation, several studies have shown Mediator to interact with post-transcription events such as RNA processing and splicing [139]. In line with this, splicing factor 3B subunit 2 & 3 (PF3D7_1461600 and PF3D7_1234800) and a serine/arginine-rich splicing factor 4 (PF3D7_1022400) were also identified as interacting partners. Mediator complex is also known to play an important role in regulation of metabolism related genes [156] and several enzymes involved in metabolic pathways were identified in this study as well. Purine nucleoside phosphorylase, an enzyme involved in purine and

pyrimidine metabolism had a high emPAI score of 260. Other metabolic pathway enzymes such as protein-L-isoaspartate(D-aspartate) O-methyltransferase, dephospho-CoA kinase, ribonucleoside-diphosphate reductase and sphingomyelin phosphodiesterase were also identified in the mass spectrometry results.

Gene ontology ID	Molecular function	No. of genes with this GO term	No. of genes in sample with this GO term	Bonferroni adjusted p-value
GO:0008150	Biological process	2640	102	7.18×10^{-4}
GO:0050789	Regulation of biological process	552	38	2.31×10^{-5}
GO:0065007	Biological regulation	574	38	5.84×10^{-5}
GO:0050794	Regulation of cellular process	525	36	5.47×10^{-5}
GO:0010468	Regulation of gene expression	141	32	1.11×10^{-16}
GO:0060255	Regulation of macromolecule metabolic process	153	32	8.93×10^{-16}
GO:0019222	Regulation of metabolic process	178	32	4.21×10^{-14}
GO:2000112	Regulation of cellular macromolecule biosynthetic process	134	31	2.52×10^{-16}
GO:0010556	Regulation of macromolecule biosynthetic process	134	31	2.52×10^{-16}
GO:0031326	Regulation of cellular biosynthetic process	135	31	3.03×10^{-16}
GO:0009889	Regulation of biosynthetic process	135	31	3.03×10^{-16}
GO:0031323	Regulation of cellular metabolic process	164	31	3.71×10^{-14}
GO:0080090	Regulation of primary metabolic process	166	31	5.00×10^{-14}
GO:0006355	Regulation of transcription, dna-templated	108	29	1.09×10^{-16}
GO:2001141	Regulation of RNA biosynthetic process	108	29	1.09×10^{-16}
GO:0051252	Regulation of RNA metabolic process	110	29	1.66×10^{-16}
GO:0019219	Regulation of nucleobase-containing compound metabolic process	134	29	1.60×10^{-14}
GO:0051171	Regulation of nitrogen compound metabolic process	134	29	1.60×10^{-14}
GO:0006807	Nitrogen compound metabolic process	715	29	1

Gene ontology ID	Molecular function	No. of genes with this GO term	No. of genes in sample with this GO term	Bonferroni adjusted p-value
GO:0044248	Cellular catabolic process	148	10	4.68×10^{-1}
GO:0006508	Proteolysis	160	10	0.74
GO:1901575	Organic substance catabolic process	180	10	1
GO:0044265	Cellular macromolecule catabolic process	103	8	0.44
GO:0009057	Macromolecule catabolic process	108	8	0.56
GO:0030163	Protein catabolic process	93	7	0.76
GO:0000375	RNA splicing, via transesterification reactions	56	5	0.94
GO:0009116	Nucleoside metabolic process	62	5	1
GO:1901657	Glycosyl compound metabolic process	63	5	1
GO:0042540	Hemoglobin catabolic process	7	2	0.81
GO:0020027	Hemoglobin metabolic process	7	2	0.81
GO:0016567	Protein ubiquitination	10	2	1

Table 4.2 - Gene ontology (GO) analysis of proteins identified by LC-MS/MS analysis

142 proteins with an emPAI score greater than 10 were analysed using the GO tool on PlasmoDB.

4.3.2.3 Bioinformatic analysis of the putative Mediator subunits

Two proteins, PF3D7_0526800 and PF3D7_1363600 were identified in the LC-MS/MS analysis, with high emPAI score ratios. It led us to question whether these could be potential Mediator subunits. To ascertain their identity based on homologs in other species, a bioinformatic analysis of these two genes was carried out. Ms. Aarthi Mohan, a bioinformatician in the laboratory, carried out the bioinformatics analysis. The full-length protein sequences of the two genes were used as query sequences and analysed by BLASTp and PSI-BLAST. The E-value was varied from 10 to 1000, for BLASTp

analysis. For PSI-BLAST, the E-value threshold was varied from 10 to 1000 while the PSI-BLAST inclusion threshold was varied from 0.005 to 0.1. The BLASTp and the PSI-BLAST search strategies did not return any results for the two putative Mediator genes tested. The search databases used were Pfam, TIGRfam, Gene3D, Superfamily and PIRSF. Only the top hits with low E-values were considered as a good match. The two putative Mediator subunits do not contain any conserved Mediator domains. The search was repeated using Prosite scan, which also yielded similar results.

Since sequence and profile-based searches did not yield any results for the putative Mediator subunits, an alternative approach was adopted. Mediator complex is known to undergo structural rearrangements upon interaction with activator proteins or with components of the basal transcription machinery. This structural flexibility is conferred by intrinsically disordered regions (IDRs) present in different subunits of the Mediator complex. Agnes Toth-Petroczy et al., [157] carried out an extensive analysis of IDRs in Mediator of *Saccharomyces cerevisiae* and *Homo sapiens*. The IDRs have important functional roles:

1. They provide flexible target sites for multiple partners
2. They act as flexible linkers joining globular domains of proteins
3. They enable formation and disassembly of complexes in response to regulatory cues.

Similar to the work done by Agnes Toth-Petroczy et al., IUPred was used to predict the disorder forming residues and also disordered regions of at least 30 residues (inclusive of 3 residue gaps) length inside the proteins. The residue wise probability score for forming disordered regions was obtained from the IUPred analysis. Scores from every amino acid was averaged for full length. For disordered segments, regions with a position-wise score of ≥ 0.5 of 30 amino acid length were filtered. Although, the average probability based on full-length protein was not indicative of disordered proteins, there was at least one 27-30 AA disordered regions within four proteins namely, Med14, Med22 and the two uncharacterized proteins, PF3D7_0526800 and PF3D7_1363600. While PF3D7_1363600 contained one such region, PF3D7_0526800 was found to contain two such segments. This was the only information that we could obtain through the

bioinformatics approaches that we used. In conclusion, the two putative Mediator subunits, PF3D7_0526800 and PF3D7_1363600 did not show any significant matches with known Mediator subunits from other species.

4.3.2.4 Episomal expression of putative Mediator subunits

In order to test if the uncharacterized protein PF3D7_0526800 identified in our study is a putative Mediator complex subunit, the gene was tagged with a triple hemagglutinin tag and expressed episomally in *P. falciparum* T996 strain. The protein was successfully immunoprecipitated from parasite protein lysates using anti-HA antibody (Fig 4.12).

Figure 4.12 - Immunoprecipitation of episomal PF3D7_0526800-HA parasite line

Western blot of immunoprecipitated PF3D7_0526800-HA protein from parasite protein extracts. Untransfected parental strain (T996) was used as a negative control. Protein samples were separated using a 12% SDS-polyacrylamide gel and western blot was carried out using the anti-HA antibody. Band corresponding to PF3D7_0526800-HA

protein is indicated with an arrow. Image is representative of two independent experiments.

The PF3D7_0526800-HA protein which has a predicted molecular weight of ~48 kDa was observed in the eluate fraction for only PF3D7_0526800-HA strain and not the wild-type strain. The protein can be further analysed by co-immunoprecipitation and mass spectrometry to test whether interaction with predicted Mediator subunits can be further substantiated.

4.3.3 Mediator occupancy across the *P. falciparum* genome

DNA regions occupied by the Mediator complex can be identified by ChIP-seq, a technique that involves sequencing of chromatin-immunoprecipitated DNA and mapping the sequencing reads to a reference genome. Therefore, episomal Med31-HA parasites were tightly synchronized and harvested at the schizont stage (36 - 40 h) for the ChIP-seq experiment using anti-HA antibody. The episomal Med31-HA gene is under the control of the non-endogenous calmodulin promoter, which is known to display maximum gene expression between 14 - 40 h. Therefore, the schizont stage was selected for the ChIP-seq analysis. To test for non-specific binding, anti-HA antibody was used in a control ChIP-Seq experiment with wild-type T996 strain that does not express any hemagglutinin-tagged protein. ChIP-Seq experiments carried out with the Med31-HA parasite line have been denoted as Med31 and those carried out with wild-type T996 expressing no HA tag have been denoted as T996.

High-throughput sequencing was done using a paired-end MiSeq run. Approximately, 6-10 million reads including forward and reverse reads were obtained and the read length was 151 bp. The ChIP-seq data was analysed by Ms. Aarthi Mohan, a bioinformatician in our laboratory.

4.3.3.1 Enrichment-based analysis

The paired-end reads obtained were pre-processed (section 2.17.4.1) and mapped to *P. falciparum* 3D7 sequences obtained from the database, PlasmoDB (version 9.0). A

total of 76% and 65% of the reads mapped out of which 68% and 58% of the reads were uniquely mapped for Med31 and T996 ChIP samples respectively (Table 4.3). Only uniquely mapped reads were considered for further analyses.

ChIP

Sample	Total reads	Mapped	Uniquely mapped	Sequencing depth	% of genome covered	Mapped (Human hg19)	Mean Fragment size	No. of narrow peaks called
Med31	5971200	4581269 (76.722753%)	4114536 (68.906350%)	25.6099	91.868%	138205 (2.314526%)	303	4134
T996	7613636	5000386 (65.676715%)	4430757 (58.195020%)	27.176	92.1609%	115622 (1.518617%)	264	5066

Input

Sample	Total reads	Mapped	Uniquely mapped	Sequencing depth	% of genome covered	Mapped (Human hg19)	Mean Fragment size
Med31	9603512	8672988 (90.310586%)	7645655 (79.613114%)	46.6755	92.8207%	114798 (1.195375%)	251
T996	9854608	8912250 (90.437387%)	7863172 (79.791829%)	48.1056	93.0116%	76078 (0.772004%)	250

Table 4.3 - Mapping statistics using pre-processed reads for ChIP and input DNA

Total reads: Total number of sequencing reads retained after trimming; Mapped: Number of reads mapped to genome; Uniquely mapped: Number of reads (alignment) with a mapping quality of 10 and mapped as a pair; Sequencing depth: Average number of time each base in the genome was covered; % of genome covered: Number of bases in the genome with at least one read coverage; Mapped (human hg19): Amount of reads mapped to human genome; Mean fragment size: Estimated from the alignment distance between the paired-end reads; No. of narrow peaks called: Number of significant binding locations determined by MACS2 using corresponding input DNA as control.

Profile plots show the overall binding pattern of the protein in the gene regions. Gene regions have been binned to facilitate computational analyses. For the purpose of data analysis, 5' UTR has been arbitrarily defined as the region 1 kb upstream of the

translation start codon (ATG) while 3' UTR has been arbitrarily defined as the region 1 kb downstream of translation stop codon (STOP). Since gene lengths vary across the genome, coding regions were binned into 20 bins, and each of the 5' and 3' UTR regions were binned into 10 bins, scaling all genes uniformly to 40 bins. An overall profile plot for 5400 genes was thus generated. The overall profile plot shows that Mediator occupies the coding region of most genes during the schizont stage (Fig 4.13). Although, the occupancy profile for Med31 looks similar to the T996 control, the log₂ transformed values of ChIP vs input were higher for the Med31 sample compared to T996. Spearman's correlations were determined between all sequenced libraries (Med31 & T996 ChIP & input samples). The Med31 and T996 ChIP samples were found to be strongly correlated (Spearman's correlation co-efficient: 0.9).

Figure 4.13 - Mediator occupancy across the transcription unit for Med31-HA parasites (red) and wild-type T996 parasites (green) during schizont stage

The occupancy profile for each cell line has been plotted as a line plot. y-axis: Log₂ (ratio of ChIP / ratio of Input); x-axis: 1 kb upstream of ATG codon (10 bins), gene body (20 bins), and 1 kb downstream of STOP codon (10 bins). The UTR regions are binned into 10 bins of equal size (100 bp in a bin). The gene body region is binned into 20 bins, for a total of 40 bins, with each bin size proportional to the actual gene length. Enrichment scores inside each bin were obtained and the averaged across all genes and plotted as lines.

A total of 5400 genes have been considered for generating profile plots and heat maps. Antigenic variant genes, pseudo genes, mitochondrial, and apicoplast genes have been excluded from the analysis. It should be noted that the 5' and 3' UTR regions may overlap with the 5' or 3' UTRS or coding regions of adjacent genes. It is hard to quantify the extent of this overlap between genes. Roughly, 2400 genes will remain if we choose to consider a 500 bp UTR region free of neighbouring genes. Approximately 400 genes will remain if we consider a 1 kb UTR region free of neighbouring genes. Therefore, it should be noted that approximately 5000 genes may contain overlap with the UTRs/gene-body of neighbouring genes.

Invasion-specific genes are known to be strongly expressed in the schizont stage from which our ChIP samples were derived. We therefore looked at the distribution of Mediator across the gene body and flanking regions of 19 invasion-specific genes. Mediator seems to be enriched in the 5' UTR and 3' UTRs of invasion-specific genes (Fig. 4.14 A). This enrichment is statistically significant with a p-value of 0.0007 and 9.13×10^{-5} in the 5' UTR and 3' UTR respectively (Mann-Whitney one-sided test). We also assessed Mediator occupancy for multi-gene families such as *var*, *rifin* and *stevor*. For *var* (Fig 4.14 D) and *rifin* (Fig 4.14 F) genes, enrichment of mediator is significantly different in the 3' UTR compared to T996. This difference is statistically significant with a p-value of 0.003 and 0.001 for *var* and *rifin* genes respectively (Mann-Whitney two-sided test). For *stevor* genes, there was no significant difference in Mediator enrichment in any of the regions compared to T996 (Fig 4.14 E). Mediator seems to be enriched in the 5' UTR and 3' UTRs of sporozoite- (Fig 4.14 B) and gametocyte-specific genes (Fig 4.14 C). This enrichment is statistically significant with a p-value of 9.13×10^{-5} (5' UTR) and 0.001 (3' UTR) for both sporozoite- and gametocyte-specific genes (Mann-Whitney one-sided test).

A

B

C

Figure 4.14 - Profile plot showing mediator occupancy for Med31-HA parasites (red) and wild-type T996 parasites (green) during schizont stage for special groups of genes

Special groups of genes included 19 invasion-specific genes (**A**), multi-gene families such as *var* (**D**), *rifin* (**F**) and *stevor* (**E**) and genes related to other stages such as sporozoite- (**B**) and gametocyte-specific (**C**) genes. The occupancy profile for each cell line has been plotted as a line plot. y-axis: Log_2 (ratio of ChIP / ratio of input) and x-axis: 1 kb upstream of ATG codon (10 bins), gene body (20 bins), and 1 kb downstream of STOP codon (10 bins). The UTR regions are binned into 10 bins of equal size (100 bp in a bin). The gene body region is binned into 20 bins, for a total of 40 bins, with each bin size proportional to the actual gene length. Enrichment scores inside each bin were obtained and the averaged across all genes and plotted as lines.

To gain a genome-wide perspective of Mediator association, a heat map of fragment distribution from 5' UTR to 3' UTR was generated for a total of 4,159 genes (Fig 4.15). The heat map shows that Mediator is enriched over the coding region for most genes, in comparison with T996 wild-type control. This enrichment is even more prominent in the line plot showing Mediator occupancy (Fig. 4.13).

Figure 4.15 - Overall fragment distribution in gene regions for Med31 cell line and the T996 wild-type control

Gene order is the same between the heat maps. The y-axis represents genes and the x-axis represents distance to gene [1 kb upstream of TSS (1 bp resolution/1000 bins), gene-body (1000 bins), and 1 kb downstream TES (1 bp resolution/1000 bins)]. Plotted values are: $\log_2$ enrichment ratio = $\log_2$ (ChIP/input). Only genes with at least 1 bin out of 3000 bins showing $\log_2$ enrichment of 0.5 and above have been shown in the heatmap. A total of 4,159 genes were obtained after applying this filter. This plot has 3000 bins (1000 in UTRs and 1000 in gene-body).

To better demarcate the regions occupied by the Mediator in the genome, genes

were clustered based on the location of Mediator binding. Fragment distribution around ATG or STOP codon as a reference point was considered and clustering was done using k means, where $k=4$. Clusters were generated based on Med31 ChIP only. For fragment distribution around the ATG codon, a total of 4 clusters were obtained and the number of genes in each cluster are as follows: cluster 1 = 968; cluster 2 = 891; cluster 3 = 672 and cluster 4 = 344. Fragment distribution around the ATG codon shows that cluster 1 genes display a pattern different from the T996 control for all four clusters (Fig 4.16 A & B). In particular, genes in cluster 1 show Mediator occupancy upstream of the ATG codon. Expression of genes belonging to the 4 clusters during the ring and schizont stages was analysed using RNA-Seq data (Fig 4.16 C & D). The distribution of gene expression between all four clusters was significantly different (Kruskal-Wallis test - Ring stage: $p < 2.09 \times 10^{-93}$; Schizont stage: $p < 3.89 \times 10^{-68}$). A post-hoc test was carried out to check if expression of genes from cluster 1 was significantly lesser than other three clusters during ring and schizont stages. Mann-Whitney test was used and Bonferroni adjusted p-values of $\text{Padj} < 0.01$ were considered statistically significant. p-values are shown in Table 4.4. The RNA-Seq expression data for the genes indicates that cluster 1 containing 968 genes have a lower level of expression in the schizont and ring stages compared to genes from other clusters (Fig 4.16 C & D).

For fragment distribution around the STOP codon, a total of 4 clusters were obtained and the number of genes in each cluster are as follows: cluster 1 = 1,237; cluster 2 = 704; cluster 3 = 694 and cluster 4 = 558. Fragment distribution around the STOP codon shows that cluster 1 genes display a pattern different from the T996 control for all four clusters (Fig. 4.17 A & B). Expression of genes belonging to the 4 clusters during the ring and schizont stages was analysed using RNA-Seq data (Fig 4.17 C & D). The distribution of gene expression between all four clusters was significantly different (Kruskal-Wallis test - Ring stage: $p < 7.94 \times 10^{-12}$; Schizont stage: $p < 6.67 \times 10^{-12}$). A post-hoc test was carried out to check if expression of genes from cluster 1 was significantly different from the other three clusters during ring and schizont stages. Mann-Whitney test was used and Bonferroni adjusted p-values of $\text{Padj} < 0.01$ were considered statistically significant. p-values are shown in Table 4.5. The RNA-Seq expression data

for the genes indicates that the expression for cluster 1 containing 1,237 genes in the schizont and ring stages is significantly different compared to cluster 2 (Fig. 4.17 C & D).

In summary, a total of 968 and 1,237 genes were identified to belong to cluster 1 centred around the ATG and STOP codon respectively. Cluster 1 genes represent 33% and 38% of all the genes considered for the analysis for fragment distribution around the ATG and STOP codon respectively. It is interesting to note that enrichment of Mediator is not associated with increased expression for cluster 1 genes associated with either the ATG or STOP codon. This validates our ChIP-Seq approach as Mediator occupancy is not an artefact caused by highly transcribed genes. The observation that Mediator enrichment is not directly linked to higher expression is interesting and needs to be further explored. A comparison of Mediator occupancy with that of RNAPII may further shed some light on the regulatory mechanisms of the Mediator complex.

Figure 4.16 - Pattern of fragment distribution in regions 1 kb upstream and downstream of ATG

K means clustering was done with $k=4$ to generate 4 clusters. Clusters are generated based on Med31 ChIP only and T996 has been shown in parallel (with same gene order) for

comparison. The average profile of each cluster is shown as line plots (**A**) above the heat map (**B**). Gene expression of these clusters in the schizont (**C**) and ring (**D**) stage is shown as box plot. **Heat map**: y-axis = genes; x-axis = 1 kb upstream and downstream of ATG codon (1 bp resolution/2000 bins), Plotted values are $\log_2$ enrichment ratio = $\log_2$ (ChIP/input). Genes with at least 1 bin out of 2000 bins showing $\log_2$ enrichment of 0.5 and above have been shown in the heat map. **Box plot**: For each cluster, expression data from RNA-Seq (triplicates) is shown as boxplot. $\log_2$ normalized RNA-Seq read counts were generated using DESeq2. Stage from which expression data was derived has been indicated next to the box plot.

		Cluster 2	Cluster 3	Cluster 4
Cluster 1	Ring	1.20×10^{-24}	5.91×10^{-35}	2.35×10^{-78}
	Schizont	1.06×10^{-19}	6.58×10^{-23}	1.82×10^{-59}

Table 4.4 - p-values from Mann-Whitney test for clusters with fragment distribution in regions 1 kb upstream and downstream of ATG

B

Figure 4.17 - Pattern of fragment distribution in regions 1 kb upstream and downstream of STOP

K means clustering was done with $k=4$ to generate 4 clusters. Clusters have been generated based on Med31 ChIP only and T996 is shown in parallel (with same gene

order) for comparison. The average profile of each cluster is shown as line plots (**A**) above the heat map (**B**). Gene expression of these clusters in the schizont (**C**) and ring (**D**) stage is shown as a box plot. **Heat map:** y-axis = genes; x-axis = 1 kb upstream and downstream STOP codon (1 bp resolution/2000 bins), Plotted values are $\log_2$ enrichment ratio = $\log_2(\text{ChIP}/\text{input})$. Genes with at least 1 bin out of 2000 bins showing $\log_2$ enrichment of 0.5 and above have been shown in the heat map. **Box plot:** For each cluster, expression data from RNA-Seq (triplicates) is shown as a box plot. $\log_2$ normalized RNA-Seq read counts were generated using DESeq2. The stage from which expression data was derived is indicated next to the box plot.

		Cluster 2	Cluster 3	Cluster 4
Cluster 1	Ring	2.12×10^{-10}	0.49	0.09
	Schizont	1.54×10^{-10}	1	0.24

Table 4.5 - p-values from Mann-Whitney test for clusters with fragment distribution in regions 1 kb upstream and downstream of STOP

4.3.3.2 Peak-based analysis

DNA obtained from chromatin immunoprecipitation is mapped onto the reference genome to identify high-coverage regions of the genome. These high-coverage genomic regions are termed as ‘peaks’ and the algorithms used to identify them are known as ‘peak-callers’. The previous section focused on enrichment-based analysis wherein enrichment refers to the amount of reads present in a location. In peak-based analysis, any location with higher enrichment in the ChIP sample over input is considered for statistical testing to assess whether it is a peak. Statistical testing involves calculation of p-values that are obtained from Poisson distribution. If the location has a higher p value, meaning

that it is random, it is not considered as a peak. Only if a location has a p-value of ≤ 0.05 , it is considered as a peak. Simply put, the heat maps generated previously represent a super set of genes and the peak calling data represents a subset of significant locations from the super set based on assessment of p-values. The peak-based analysis was done using a peak caller program named 'MACS2'. The peaks were annotated to nearby/overlapping genes with a minimum overlap of at least 1 bp within any regions (gene body, 5' UTR and 3' UTR). The number of peaks called from each ChIP and the number of genes where the peak is contained either within the gene-body or in the flanking regions has been indicated in Table 4.4. Gene to peak association has been done in a loose context, i.e. when a peak is found to overlap 2 genes, then both genes are considered to have that peak. In both Med31 and T996 ChIP, the peaks were predominantly found to be present within the gene body.

Sample	Number of peaks	Number of genes		
		Gene-body	5' UTR	3' UTR
Med31	4,134	2,742	40	240
T996	5,066	3,162	62	347

Table 4.6 - Peak annotation to nearest gene regions for narrow peaks

MACS2 was used for peak calling. Peaks were called separately for the two ChIP samples and the peak-to-gene annotation was performed. The number of peaks overlapping the different gene regions considered have been indicated.

The number of overlapping and specific sets of genes present in Med31 and T996 ChIP samples were visualised using a Venn diagram (Fig. 4.18). Four hundred and seven genes were found to be specific to Med31 ChIP, while 820 genes were found to be specific to T996 ChIP. A total of 2,362 genes were found to be common between both ChIP samples. In order to determine the specific binding pattern of the 407 genes specific to Med31, heat maps were generated showing T996 in comparison (Fig 4.19). Most of the

Med31-specific genes show occupancy in the coding region and 3' UTR regions.

Figure 4.18 - Venn diagram of gene names overlapping between Med31 and T996 ChIP

Figure 4.19 - Med31 and T996 ChIP fragment distribution on 407 genes specific to Med31

Heat map: y-axis = genes (407 genes specific to Med31); x-axis = Distance to gene: 1 kb upstream TSS (1 bp resolution/1000 bins), gene-body (1000 bins), and 1 kb downstream TES (1 bp resolution/1000 bins). Plotted values are $\log_2$ enrichment ratio = $\log_2$ (ChIP/input).

A total of 2,362 genes were found to be common to both Med31 and T996 samples. These genes common to both the samples were further classified into two groups based on which ChIP samples had a higher enrichment ratio ($\log_2$ (ChIP/input)) for the common genes. The classification resulted in 768 and 1,594 genes classified as Med31- and T996-specific, respectively (Fig. 4.20).

Figure 4.20 - Genes common between Med31 and T996

A total of 2,362 genes common to both Med31 and T996 ChIP samples were considered. By comparing the enrichment for MED31 and T996, these genes were further classified into either one of the ChIP with the highest enrichment. Based on this, 768 and 1,594 genes can be classified as Med31- and T996-specific, respectively. **Heat map:** y-axis =

genes; x -axis = Distance to gene: 1 kb upstream TSS (1 bp resolution/1000 bins), gene body (1000 bins), and 1 kb downstream TES (1 bp resolution/1000 bins). Plotted values are $\log_2$ enrichment ratio = $\log_2$ (ChIP/input).

All the genes which contain peaks in Med31 ChIP were analysed to better understand peak location in the transcription unit. These include 407 genes specific to Med31 ChIP and also the 2362 genes common to both Med31 and T996 ChIP. The peak location was also compared with gene expression to correlate Mediator occupancy and gene expression. The UTR regions were binned into 10 bins of equal size (100 bp in a bin) and the gene body region was binned into 20 bins, for a total of 40 bins. Peak frequency in the 40 bins over three gene regions (5' UTR to 3' UTR) were plotted as bar plot (Fig 4.21). Most of the peaks were found to be localized to the coding region of genes. The genes were classified based on location of peaks within the gene regions (5' UTR, gene body and 3' UTR) and their expression across ring, trophozoite and schizont stages was determined (boxplot, Fig 4.21). Frequencies were normalized to total number of peaks in each ChIP, and normalized and $\log_2$ -transformed read counts from T996 RNA-Seq were taken as the expression values.

In the analysis of ChIP-Seq data, 5' UTR has been arbitrarily defined as the region 1 kb upstream of the translation start codon (ATG) while 3' UTR has been arbitrarily defined as the region 1 kb downstream of translation stop codon (STOP). Genes that are less than 1 kb from a flanking region have not been excluded from the analysis. Therefore, there is a possibility that the peaks or enrichment that is observed in the 5' or 3' UTR could be due to the presence of the coding region of those flanking genes. For genes which show Mediator occupancy in the coding region (Fig. 4.21), expression across the three stages was found to be significantly different (Kruskal-Wallis test, p -value: 0.009). For genes with Mediator occupancy in the 5' UTR and 3' UTR, expression across the three stages was not significantly different. Overall, Mediator occupancy is not correlated with high gene expression.

Figure 4.21 - Peak frequency in bins from Med31 ChIP (Bar plot); Expression pattern of genes with peaks in the three genomic locations (Box plot)

Bar plot: Peak frequency in each bin from 1 kb upstream of ATG (5' UTR), gene body and 1 kb downstream of STOP (3' UTR) was calculated. Number of peaks in each bin was normalized to total number of peaks. y -axis = Frequency in % and x -axis = binned gene regions. **Box plot:** Genes are classified based on where the peaks are located within these three regions, and their expression during the three stages is shown as a box plot.

4.3.3.3 Expression pattern of genes that display exclusive binding by Mediator

A total of 407 genes specific to Med31 ChIP were clustered based on their

expression across different stages of the parasite. Gene expression data was obtained from RNA-Seq done in triplicates at ring, trophozoite, and schizont stages (The RNA-seq experiment was carried out by Dr. Ragini Rai, a research fellow in the laboratory). Hierarchical clustering based on stage of gene expression resulted in three clusters (Fig 4.22) and gene ontology was performed for the clusters in the heat map by using three pathway databases, namely Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Malaria Parasite Metabolic Pathway (MPM) [69]. Genes bound specifically by the Mediator were enriched in metabolic pathways such as selenocysteine biosynthesis and tRNA modifications. Both cluster 1 and cluster 2 genes were found to be enriched for inorganic cation transporter activity (Fig 4.22).

B

Pathway name	No. of genes in cluster	No. of genes in genome with the annotation	No. of genes in cluster with the annotation	Bonferroni adjusted p-value from hypergeometric test
Cluster 1				
GO:0015082 di-, tri-valent inorganic cation transporter activity	168	4	2	0.01
Cluster 2				
GO:0015077 monovalent inorganic cation transporter activity	165	21	3	0.02
Selenocysteine biosynthesis	165	11	2	0.04
tRNA modifications	165	26	3	0.04

GO
 MPM

Figure 4.22 - Expression of genes that exclusively bind Med31 in the Schizont stage

Genes with exclusive Med31 peaks were further clustered into three groups by hierarchical clustering of gene expression across three stages, producing clusters based on expression timing (A). The three clusters, cluster 1 (ring), cluster 2 (trophozoite) and cluster 3 (schizont) are shown in the heat map. The results of GO analysis for the clusters have also been indicated in the table below the heatmap (B).

4.3.3.4 Motif analysis in regions associated with Mediator

In other organisms, Mediator is recruited in part through interaction with sequence-specific DNA-binding transcriptional activators. Therefore, we were interested to find out if potential *cis*-acting elements were enriched in regions of DNA associated with Mediator. The potential nucleotide sequence motifs enriched in DNA associated with Mediator were determined using the FIRE program for the 407 genes specific to Med31 ChIP [158]. Motifs that were over/under represented in each cluster against the remaining genes in the genome were analysed (Fig 4.23). The DNA motif, [AT]ATG[GT]GT[CG]

was found to be enriched in the 5' UTR of genes in cluster 1, whereas an RNA motif, [CGU]UGUCAG[AG] was enriched in the 3' UTR of cluster 3 genes. The motif identified in cluster 3 genes has been termed as a RNA motif because it was only obtained by searching 3' UTR regions and was located only in one orientation on the transcribed strand of mRNA. Genes in cluster 2 did not contain any over-represented motifs.

Figure 4.23 - Gene ontology and DNA motif analysis for 3 clusters specific to Med31

Gene ontology and DNA motif discovery on the three clusters were done using FIRE tool [158]. FIRE tool reports the over/under-representation (motif-yellow/blue; GO-red/green) for each motif/GO term in the given cluster versus background (rest of the genes in the genome). Yellow boxes highlighted with red frames indicate significant over-representation with $p < 0.05$ after Bonferroni correction. Similarly, the blue frames indicate under-representation. Only cluster 1 and cluster 3 had motif enrichment in their 5' UTR and 3' UTR respectively. Additionally, cluster 1 also had an enriched ontology term.

The FIRE tool uses sequences 1 kb upstream of ATG and 1 kb downstream of

STOP codon and searches for both DNA and RNA motifs. It also allows us to identify gene clusters. Interestingly, gene ontology analysis using the FIRE tool also showed that cluster 1 genes were enriched for the copper ion transport pathway (GO:0006825) (Fig 4.23).

4.3.3.5 Validation of ChIP-Seq targets by real-time PCR

Mediator occupancy profiles obtained by ChIP-Seq analysis were further verified by using quantitative real-time PCR. As discussed in section 4.3.3.1.1, using fragment distribution around the ATG or STOP codon, genes were divided into 4 clusters. Since cluster 1 showed a distinct occupancy pattern compared to T996 (Fig 4.16 & 4.17), 11 genes were selected from cluster 1 that show Mediator occupancy either near the ATG or the STOP codon. For determining fragment distribution and clustering, a two kb region (1 kb upstream and 1 kb downstream of ATG codon) was divided into 20 bins of 100 bp each and the mean enrichment inside each bin was calculated. The bin location with the maximum enrichment was determined for each gene in cluster 1. For ChIP-qPCR, primers were designed to amplify the region flanking the bin with the maximum enrichment. The primers used for qPCR have been indicated in Table 2.8 (Chapter 2, section 2.18). A total of 5 genes were selected for ATG cluster 1, while 6 genes were selected for STOP cluster 1.

The genes for qPCR were selected such that they were present in the peak-calling dataset and were not annotated as unknown conserved *Plasmodium* proteins in PlasmoDB. Most of the selected genes (10 out of 11) either contained peaks only in the Med31 parasite line or had peaks in both T996 and Med31 ChIP samples but had a higher fold change in the Med31 sample. However, the gene, PF3D7_1478600 had peaks in both T996 and Med31 ChIP samples but had a higher fold change in the T996 sample.

The ChIP enrichment in the Med31 sample compared to the T996 control was determined by real-time PCR. The ChIP input samples were used as an internal reference control. The fold change was calculated using the $\Delta\Delta\text{CT}$ method. For Cluster 1 genes, having an enriched fragment distribution around the ATG codon, 4 out of 5 genes showed higher fold-change in Med31 ChIP compared to T996 wild-type control (Fig 4.24). For

Cluster 1 genes, having a fragment distribution around the STOP codon, 3 genes showed higher fold-change in Med31 ChIP compared to T996 wild-type control (Fig 4.24). As expected, the gene PF3D7_1478600, showed a fold change lower than Med31 ChIP. Overall, the ChIP-qPCR results showed a good agreement with the enrichment profiles obtained by ChIP-Seq.

Figure 4.24 - Validation of ChIP-Seq data by real-time PCR for Cluster 1 genes centered around the ATG and STOP codon

ChIP Enrichment of Med31 ChIP (Red) over wild-type T996 (blue) was determined by qPCR. qPCR experiments were carried out in triplicates. Student's t-test revealed non-significant p-values of $P > 0.05$. Fold change was calculated using the $\Delta\Delta C_t$ method. Error bars represent standard deviation of the mean.

5 DISCUSSION

Mediator is a large multi-subunit complex unique to eukaryotic organisms, which plays a central role in regulating RNAPII transcription. It interacts with both gene-specific activators bound to enhancer sites as well as with the general transcription machinery assembled at the promoter. Mediator has a broad influence on gene expression and is involved in a range of processes such as transcription initiation and elongation, epigenetic regulation and mRNA processing. Mediator, therefore, is capable of responding to developmental or environmental cues by fine-tuning gene expression programs.

Development of *P. falciparum* within multiple hosts is characterized by a highly co-ordinated pattern of gene expression. However, gene regulation in *Plasmodium* is still poorly understood and it is unclear whether transcriptional or post-transcriptional modes of control are predominant. Studying the Mediator complex in *P. falciparum* may therefore shed some light on the contribution of distinct gene regulatory mechanisms.

In this thesis, we have characterized the Mediator complex in *P. falciparum* using the highly conserved Mediator subunit, Med31. Protein-protein interactions and genome-wide occupancy of the Mediator were studied using a transgenic strain expressing HA-tagged-Med31 episomally. A conditional knockdown system based on the *E. coli* DHFR-based destabilization domain was also tested in an attempt to functionally analyze Med31. In addition to generating transgenic parasite strains, we also attempted to produce antibodies against Med31 using recombinant full-length protein or with commercially synthesized peptides as antigens. An antibody against Med7 produced by GenScript (under contract with our laboratory) was also tested in the current study. The results of these experimental approaches are discussed in detail in the following sections.

5.1 Production of antibodies against Mediator subunits

Antibodies directed against Mediator subunits can be used for affinity purification of a given Mediator subunit and can be used to identify protein interaction partners and to assess its genome-wide occupancy.

Commercially synthesized peptide designed against Med31 was used to raise rabbit polyclonal antibodies. However, western blot analysis of anti-Med31 polyclonal sera indicated non-specific binding (Fig. 3.10) and the antibody could not be used further for any functional assays. Additionally, a mouse monoclonal antibody against Med7 produced by GenScript was also tested against *P. falciparum* protein lysate in the laboratory. The ability of the Med7 mAb to recognize the Med7 protein was established by western blot analysis (Fig. 4.4). However, the antibody gave rise to non-specific signals in immunofluorescence and immunoprecipitation assays and therefore, was not used for any other experiments in the current study.

In order to use purified Med31 protein for polyclonal antibody production, Med31 was overexpressed in *E. coli* as a fusion protein with an N-terminal hexahistidine tag. While recombinant *his*-Med31 protein was expressed in high quantities, it was found to be present in the form of insoluble protein aggregates, not amenable to solubilisation with urea or guanidinium hydrochloride.

Heterologous production of *Plasmodium* proteins in *E. coli* has been fraught with difficulties including low levels of protein expression or production of insoluble protein aggregates. As a part of a structural genomics study, when 1000 *P. falciparum* open reading frames were studied for expression in *E. coli*, only 337 of the proteins were found to be expressed, while only 63 of the proteins were produced in soluble form [159]. Factors such as higher disorder, basic isoelectric point values and absence of homology to *E. coli* proteins were found to negatively impact protein expression. pI values of proteins were particularly strongly correlated to expression and solubility of *Plasmodium* proteins [160].

In this study, the *P. falciparum* Med31 protein was expressed at high levels in *E. coli*. The level of protein expression was comparable when BL21 cells were either cultured at 37°C for 4 h or 18°C for 19 h (Fig. 3.2). The solubility of the *his*-Med31 protein was tested using five different buffers with variations in pH (6.8, 7, 7.5 and 8.5), buffering components (Tris/HCl, HEPES, MES, phosphate buffer) and salt concentration

(150 and 200 mM NaCl). The protein was found to be insoluble in all the buffers tested at both temperatures 37°C and 18°C (Fig. 3.3).

One of the major limiting factors for the expression of *Plasmodium* proteins in *E. coli* is the disparity in codon usage. Due to the unavailability of certain rare codons, which are usually present in genes with low expression, the *E. coli* ribosome stalls during protein synthesis, resulting in translation termination thereby producing truncated products [161]. Therefore, the *his*-Med31 construct was expressed in BL21-CodonPlus (DE3) -RIL and Rosetta-gamiTM 2(DE3) cells, which contain additional copies of tRNA genes rarely found in *E. coli*. Higher level of protein expression was observed at 37°C in the case of RIL cells, whereas a lower growth temperature (18°C) gave better expression in the case of RG2 cells (Fig. 3.4). Though the protein was expressed in both the *E. coli* strains, it was still found to be insoluble in all the different buffers tested (Fig.3.5 & 3.6).

Often, the overexpressed recombinant protein forms insoluble aggregates due to multiple factors such as reduction of disulfide bonds, lack of post-translational modifications and lack of access to chaperones. The insoluble protein aggregates are sequestered into inclusion bodies in the cytoplasm of cells. Under these circumstances, proteins can be extracted from inclusion bodies using strong denaturants such as urea or guanidinium hydrochloride. In this study, *E. coli* lysates expressing *his*-Med31 were prepared under denaturing conditions using urea (2M, 4M and 6M) and guanidinium hydrochloride (6M). However, the *his*-Med31 protein was always found to be present in the insoluble fraction (Fig. 3.7), suggesting that the protein might be associated with cellular debris via multiple weak interactions with proteinaceous and non-proteinaceous components [162]. Since solubilisation of inclusion bodies in urea is pH-dependent, higher pH and urea concentration could be used to enhance the solubility of the inclusion bodies. [163]

When *E. coli* lysates were prepared under denaturing conditions using urea (8M) and purified using Ni-NTA resins, very small quantities of the recombinant *his*-Med31 protein were recovered by elution with imidazole (Fig. 3.8). This can be explained by the retention of the insoluble protein in pellet fractions even after solubilisation with urea

(Fig. 3.9, lane labeled as pellet). Since the starting amount in the supernatant fractions itself was low, it is understandable that very low levels of protein were recovered after Ni-NTA purification. Alternatively, *E. coli* lysates can be prepared under non-denaturing conditions and the pellet fraction can be subjected to Ni-NTA purification under denaturing conditions to recover relatively pure fractions of the recombinant protein.

In the X-ray crystallography studies of the Med7N-Med31 subcomplex, attempts to isolate recombinant Med31 were not successful and this was attributed to the high hydrophobicity of the Med31-Med7N interface [94]. Therefore, co-expression of Med31 with the N-terminal region of Med7 could result in a soluble protein. However, in this case, the polyclonal antibodies produced will recognize epitopes on both Med31 and Med7. As an alternate strategy, C-terminal or N-terminal deletion constructs can be used to achieve soluble protein expression. Solubility enhancing fusion partners such as MBP, GST or thioredoxin can also be added to the recombinant protein. Co-expression of molecular chaperones such as Pfhs70 and Pfhs40 along with the recombinant protein may also lead to proper protein folding and solubility [161].

In my study, since the recombinant Med31 protein was found to be highly insoluble, it could not be purified further for use in antibody production.

5.2 Conditional knockdown of Med31 protein expression

Reports over the years have shown that each Mediator subunit can regulate different sets of genes to perform distinct cellular functions. Knockout studies of individual Mediator subunits have provided insights into their function and role within the Mediator complex. In this study, we attempted to knock down protein expression of the Mediator subunit, Med31, using a conditional system. An *E. coli* DHFR based destabilization domain and a triple-hemagglutinin tag were fused to the C-terminus of endogenous Med31 gene using a homologous recombination approach. A ligand called TMP stabilizes the degron domain. In the absence of the ligand, the instability of the degron is coupled to the fused protein of interest and it is targeted for degradation by the proteasomal machinery [150].

Single parasite clones were obtained by the technique of limiting dilution. Integration into the endogenous genomic locus was confirmed by PCR analysis of genomic DNA and RNA extracted from transgenic parasites (Fig. 3.15 & 3.16). Expression of the transgenic Med31^{HA-DDD} protein was confirmed by western blot analysis (Fig 4.1 A.). The endogenous tagged Med31 protein was uniformly expressed throughout all red blood cell stages of the parasite (Fig. 4.1 B).

Transgenic parasites were then cultured in the absence of the drug, TMP, to test the conditional knockdown system. Removal of the drug did not lead to degradation of the Med31 fusion protein (Fig. 4.2 A). Protein stability in the presence or absence of TMP was further assessed by immunoprecipitation of endogenous Med31^{HA-DDD} protein from parasite protein lysates using an antibody against the hemagglutinin tag (Fig. 4.2 B). The full-length Med31 fusion protein was immunoprecipitated from cultures grown in the presence as well as absence of TMP. However, a lower molecular weight protein migrating between 16 and 30 kDa was observed only in the parasite culture grown with TMP. Another protein with a molecular weight lower than 16 kDa was also observed, in both eluate samples, with and without TMP. Since an anti-HA antibody was used for the western blot analysis of IP samples, the lower molecular weight bands appear to be degradation products of the full-length Med31^{HA-DDD} protein. However, it is unclear why the band migrating between 16 and 30 kDa is not prominently observed for parasites cultured without TMP.

We observed that the Med31^{HA-DDD} protein was not degraded upon removal of the drug. However, studies with the *E. coli* based destabilization domain have shown that even though the protein may be expressed, the domain may still interfere with protein interactions thereby indirectly affecting parasite growth and viability [151]. Therefore, parasite growth was monitored in the presence or absence of TMP. Asynchronous Med31^{HA-DDD} parasites were grown in the presence and absence of 10 μ M TMP for 4 days and the parasitemia was calculated by microscopic examination of geimsa stained smears. The results of the growth assay did not show any significant decrease in parasitemia in the absence of the stabilizing ligand (Fig. 4.3).

There could be multiple reasons for these observations. For the conditional knockdown approach, a sequence comprising the full-length Med31 coding region and a 5' flanking region comprising 429 bases upstream of the coding region was used for single-crossover homologous recombination. This is because the size of Med31 coding region is too small (399 bp) for considering only the 3' end of the gene for homologous recombination. In *Plasmodium*, cloned sequences, which are 600 – 1200 bp in length, have been found to function well for homologous recombination [164]. Therefore, to increase the length of the homology region, the Med31 coding sequence and the 5' upstream flanking region were used. However, this leads to a duplicate copy of the Med31 gene, which could potentially be expressed if its promoter region lies within the 429 bp upstream 5' flanking region.

RT-PCR analysis was therefore carried out to rule out the expression of the duplicate Med31 transcript (data not shown). RT-PCR primers were designed such that they flank the Med31 coding region and its corresponding 3' UTR region. Due to integration of the entire plasmid at the recombinant locus (Fig. 3.13), the Med31-HA-DDD locus is spaced far apart from the 3' UTR region and would not be amplified by the PCR primers designed to amplify the duplicate Med31 transcript. Although preliminary RT-PCR analysis showed that the duplicate Med31 transcript is not produced, this needs to be further confirmed by repeating the RT-PCR analysis or by performing qRT-PCR. Another alternative is to mutate the start codon of Med31 such that the production of full-length protein is affected. This strain can be used for conditional knockdown such that the duplicate Med31 transcript is not functional and does not eliminate the effects of Med31 fusion protein knockdown.

Even if the duplicate Med31 copy was not functional, the DDD may still have failed to regulate the protein despite successful integration of the DDD into the endogenous Med31 locus. This is not without precedent, as failure of regulation using conditional knockdown systems has been documented in some studies in *Plasmodium* [165, 166]. Nuclear proteins or proteins residing in unique cellular compartments may not be efficiently targeted for degradation. Failure to be degraded may also depend on specific functional features of the protein.

Another reason for failure to observe effects of Med31 knockdown could be because the loss of Med31 may not have a significant effect on parasite viability. The requirement of this subunit varies among different organisms. Although Med31 is not essential for viability in *Saccharomyces cerevisiae* [167], it is required for septation in *Saccharomyces pombe* and for embryogenesis in *Drosophila* and mammals [168]. Deletion of Med31 however, does not destabilize the Mediator complex entirely as the complex was found to remain intact even upon deletion of Med31 [94]. The Med7N-Med31 subcomplex present in the middle module contacts the neck domain of the head module forming a head-middle interface. Deletion of Med31 was found to disrupt the interaction with the head thereby causing increased mobility of the head module [99]. It is interesting to note that this head-middle interface also forms a CTD channel that brings the RNAPII in close proximity to TFIIF, a subcomplex of TFIIF, which phosphorylates CTD [102]. It could be that Med31 may not be essential for the function of core Mediator complex in *P. falciparum* and may not be absolutely required for the assembly of the Mediator complex.

Nevertheless, the endogenous tagged Med31 strain generated in this study can still be used for further experiments to assess protein interactions and genome-wide occupancy of Mediator complex.

5.3 Genome-wide occupancy of Mediator

5.3.1 Mediator binds predominantly to coding regions

Identification of *cis* regulatory DNA sequences such as enhancers is essential for understanding mechanisms behind activation and repression of gene expression. Analysis of genome binding sites of the Mediator complex may reveal novel *cis* regulatory motifs and may help in elucidation of specific features of promoter and enhancer architecture required for gene-specific recruitment of the Mediator complex. It may also give us insights about whether transcription is controlled at the level of initiation, elongation or termination. Mediator does not contain DNA binding domains and its interactions with the genome in ChIP experiments are captured based on interactions with activators or components of PIC.

In this study, genome-wide binding sites of *Plasmodium* Mediator were analysed by chromatin immunoprecipitation with high throughput sequencing (ChIP-Seq). The transgenic strain expressing Med31-HA protein episomally was used for ChIP-Seq using an antibody against the HA tag. The wild-type T996 strain was used in parallel, as a negative control, to filter out non-specific interactions. For initial mapping of Mediator genome occupancy, ChIP-Seq was carried out with parasites in the schizont stage (36 - 40 h). High-throughput sequencing was done using paired-end MiSeq sequencing. The paired-end reads obtained from high-throughput sequencing were pre-processed (chapter 2, section 2.17.4.1) and mapped to *P. falciparum* 3D7 sequences obtained from the database, PlasmoDB (version 9.0). Only uniquely mapped reads were considered for further analyses.

While we hypothesized that Mediator would most probably bind in the 5' regions upstream of genes, the overall profile plot showed that Mediator occupies the coding regions of most genes in the schizont stage (Fig. 4.13). Although the profile plot of Med31 looks similar to the T996 control, the log₂ transformed values of ChIP vs input were higher for the Med31 sample compared to T996.

Mediator occupancy in gene coding regions has been reported in yeast. Early studies by ChIP-microarray showed the association of Mediator with over 1200 binding sites located upstream of genes and within the coding regions [117]. Another study in *S. pombe* found Mediator to be associated with the coding region of nearly 60% of all genes [169]. Other studies have shown Mediator to mainly occupy UASs in yeast [116, 122]. However, the widespread binding of Mediator across the yeast genome has been challenged by other studies, which showed that Mediator does not occupy promoter regions of highly transcribed genes in yeast [109] and is recruited to promoters of a limited number of genes in an activator-dependent manner [118]. Association of Mediator with coding regions has been described as a ChIP artefact in these studies, which argue that binding of Mediator to highly transcribed gene coding regions may be a non-specific effect due to accessibility of DNA in these regions. Such genomic loci have been termed as 'hyper-ChIPable' in another study, calling into question recent genome-wide ChIP studies. Highly expressed genes were found to be particularly vulnerable to this ChIP

artefact. This could be due to nucleosome depletion in highly transcribed regions, which exposes DNA, making it susceptible to non-specific interactions with beads or antibodies used in the ChIP procedure. Furthermore, the antibodies used in ChIP could also interact non-specifically with certain nuclear proteins [170]. In the case of *P. falciparum*, studies have shown that coding regions have a higher nucleosome occupancy compared to intergenic regions [171] and nucleosome depleted regions are present near transcription start sites [172]. Also, our results show that genes occupied by Mediator are not necessarily highly transcribed (Fig. 4.22). Therefore, Mediator occupancy detected in the coding regions may not be an artefact due to nucleosome depletion associated with highly transcribed genes.

In my study, the use of wild-type T996 strain for ChIP-Seq using the anti-HA antibody provides a good control for normalization of the ChIP-Seq dataset. The T996 wild-type control helps in establishing the noise level caused due to hyper-ChIPable loci. Identification of peaks even in the negative control sample, which does not contain an epitope recognized by the antibody, could be due to some *in vitro* component of IP that interacts with open chromatin [170].

The regions occupied by the Mediator in the genome were defined by clustering genes based on the location of Mediator binding. Fragment distribution around the ATG or STOP codon was considered and four clusters were generated based on Med31 ChIP. Fragment distribution around the ATG codon showed that cluster 1 containing 968 genes displays a pattern different from T996, among all the four clusters (Fig. 4.16 A & B). Genes in cluster 1 show enriched Mediator occupancy upstream of the ATG codon. The RNA-Seq expression data for the genes (Fig. 4.16 C, Fig. 4.16 D & Table 4.4) indicates that cluster 1 genes have a lower level of expression in the schizont and ring stages compared to genes from other clusters. Mediator present at the PIC may have been captured at these loci. Mapping of TSSs in *Plasmodium* has shown that transcriptionally active genes contain TSSs that overlap with start of the coding region [173]. This may be the reason for Mediator occupancy detected around the ATG start codon. However, the exact location of TSSs within the 968 genes belonging to cluster 1 needs to be analyzed further.

Fragment distribution around the STOP codon showed that cluster 1 containing 1237 genes displays a pattern different from T996, among all the four clusters (Fig 4.17 A & B). Genes in cluster 1 show enriched Mediator occupancy downstream of the STOP codon. Expression of cluster 1 genes in the schizont and ring stages was significantly different compared to cluster 2 (Fig. 4.17 C, 4.17 D & Table 4.5). The presence of Mediator downstream of the coding region of genes may be indicative of its role in transcription termination for this subset of genes. In yeast, Mediator has been shown to be involved in terminating transcription at the 3' end for a subset of genes based on interactions of Med18 with the cleavage factor 1 (CF1) complex. Gene looping was identified as a mechanism behind Mediator's role in termination [141]. RNAPII has also been shown to pause downstream of the 3' ends of annotated genes which produce polyadenylated transcripts, thereby linking transcription termination with mRNA processing mechanisms [174]. Mediator may therefore be involved in transcription termination for this subset of genes.

Overall, there seems to be no correlation between enriched Mediator occupancy and mRNA levels. In the case of fragment distribution around the ATG codon, cluster 1 genes show enriched Mediator occupancy upstream of the ATG codon. However, cluster 1 genes have a lower level of expression in the schizont and ring stages compared to genes from other clusters (Fig. 4.16 C, Fig. 4.16 D & Table 4.4). This could indicate transcriptional repression by Mediator on these genes. Mediator has been shown to play a repressive role in transcription in mammalian cells by recruiting chromatin modifying enzymes. For example, Mediator represses the expression of neuron-related genes in non-neuronal cells by recruiting a H3K9 methyltransferase to the promoter of these genes. Specifically, the Med12 subunit was found to interact with a transcriptional repressor REST (RE1 silencing transcription factor) to recruit the methyltransferase and thus mediate repression [175]. Mediator has also been shown to recruit PRMT5, a histone arginine methyltransferase to genes that are regulated by the CCAAT enhancer-binding protein β (C/EBP β) in *HeLa* cell lines. Recruitment of PRMT5 was shown to cause transcriptional repression of C/EBP β target genes involved in immune responses and cellular differentiation [176]. Therefore, repressive mechanisms may be in place for

cluster 1 genes where Mediator binding upstream of ATG codon is accompanied by lower levels of transcript expression during the schizont stage. Alternatively, it may not be reasonable to expect that a positive role for Mediator would necessarily lead to exceptionally high levels of transcription. It may rather be that the products of genes associated with Mediator are not required at high levels and that Mediator serves to fine-tune expression within a narrow band of low-level expression. Further correlation with RNAPII occupancy at these genes needs to be carried out to see if the levels of RNAPII recruitment correlate more directly with Mediator association.

In the case of fragment distribution around the STOP codon, cluster 1 genes show enriched Mediator occupancy downstream of the STOP codon. Expression of cluster 1 genes in the schizont and ring stages was significantly different compared to cluster 2 (Fig. 4.17 C, Fig. 4.17 D & Table 4.5). Mediator may contribute to mechanisms downstream of transcription initiation at these genes. As mentioned previously, Mediator at these genes may have a specific role in transcription termination. Location of Mediator downstream of the STOP codon may also indicate its role in mRNA transcript processing as Mediator has been shown to physically interact with splicing factors such as hnRNPL [139]. Since enriched Mediator occupancy does not correlate with mRNA levels, post-transcriptional mechanisms affecting mRNA splicing, transport and stability may play a role in gene expression.

5.3.2 Mediator binds to genes involved in metabolic pathways

Further peak-based analysis of ChIP-Seq data showed that majority of the peaks was present within the gene body. Specifically, 2,742 peaks were localized to the gene body while 40 and 240 peaks localized to the 5' UTR and 3' UTR regions respectively, for the Med31 ChIP sample (Table 4.4). We found that 407 genes were found to be specific to Med31 ChIP, while 820 genes were specific to T996 ChIP. A total of 2,362 genes were found to be common between both the ChIPs. Only the 407 genes specific to Med31 ChIP were considered for further analysis.

Fragment distribution of the 407 genes specific to Med31 showed that most of the Med31 specific genes show enrichment over the coding region (Fig. 4.19). Med31-specific genes were clustered based on their expression across different stages of the parasite. Hierarchical clustering resulted in three clusters; cluster 1 – ring, cluster 2: trophozoite, cluster 3: schizont (Fig 4.22). Not all the genes bound by Med31 are highly transcribed, which increases our confidence in these results, as occupancy is not an artefact caused by highly transcribed genes. Based on gene ontology analysis, both cluster 1 and cluster 2 genes were enriched for inorganic cation transporter activity (Fig. 4.22). This is interesting because genes coding for channels and transporters for the movement of solutes, nutrients, and ions account for 2.5% of the parasite genome [177]. The parasite relies on ion transporters for maintaining the ion content of the host cell, cytosol and its organelles. Membrane transport proteins related to transporters, pumps and channels have also been implicated in mechanisms of antimalarial drug resistance. For example, the *P. falciparum* Na^+/H^+ exchanger, PfNHE, has been implicated in affecting parasite drug sensitivity to quinine [178].

In the current study, genes bound specifically by the Mediator were enriched in metabolic pathways such as selenocysteine biosynthesis and tRNA modifications (Fig. 4.22). Selenocysteine is an amino acid similar to cysteine, which contains the trace element Selenium in place of the sulfur group of cysteine. Selenocysteine tRNA and selenoproteins have been identified in *P. falciparum* [179, 180]. The four selenoproteins identified in the parasite are unique to it and have the potential to be used as drug targets.

The potential nucleotide sequence motifs associated with enriched Mediator occupancy were determined by using the FIRE program. The DNA motif, [AT]ATG[GT]GT[CG] was found to be enriched in the 5' UTR of genes in cluster 1, whereas an RNA motif, [CGU]UGUCAG[AG] was enriched in the 3' UTR of cluster 3 genes (Fig 4.23). Interestingly, gene ontology analysis using the FIRE tool also showed that cluster 1 genes, which were expressed in the ring stage, were enriched for the copper ion transport pathway (GO:0006825) (Fig 4.23). This observation is supported by several studies that have highlighted the role of Mediator in regulating genes involved in metabolic pathways. For instance, the Med7N-Med31 submodule has been found to

regulate a subset of genes involved in methionine metabolism, regulation of iron content and cellular transport [94]. In yeast, Med31 has also been shown to act in conjunction with TREX-2 (Transport & Export complex-2) to regulate genes involved in sulfate/methionine metabolism [181]. In the mass spectrometric analysis of Med31 in the current study, we have also found it to interact with several components of metabolic processes such as purine nucleoside phosphorylase, protein-L-isoaspartate(D-aspartate) O-methyltransferase, dephospho-CoA kinase, ribonucleoside-diphosphate reductase and sphingomyelin phosphodiesterase.

Although Mediator is enriched at genes involved in metabolic pathways, these genes are not expressed at higher levels. Nonetheless, these genes may have a functional dependence on Mediator due to regulation by a SAGA-like HAT complex rather than TFIID. In yeast, genes have been divided into two categories based on whether they are regulated by the SAGA complex or the TFIID complex. SAGA complex is involved in the transcription initiation of highly regulated genes whose core promoter contains the consensus TATA box motif. On the other hand, housekeeping or constitutive genes do not have conserved TATA box motifs and rely on the co-activator complex, TFIID for transcription initiation [182].

Previous studies have shown that Mediator is required for recruitment of TBP at SAGA-dependent genes. Mediator has been shown to directly interact with the SAGA complex in human cells for transcription activation mediated by *myc* protein [183]. In yeast, tail module of Mediator has been shown to be important for SAGA-regulated genes and those involved in responding to stress and metabolic pathways [184]. The link between Mediator and SAGA in transcription has been further explored by another study by Chen et al. [185]. They carried out *in vitro* transcription assays and showed that SAGA is not required for PIC formation but it is subsequently required for making chromatin accessible to transcriptional machinery. In *P. falciparum*, homologues of GCN5 and ADA2, which are components of SAGA, have been identified [53, 186]. PfGCN5 was shown to have HAT activity and could also interact with PfADA2 [187], indicating that they could form a putative HAT complex in the parasite. Therefore, a subset of genes

coding for components involved in metabolic pathways could be regulated by a distinct mechanism involving a SAGA-like complex, which acts in concert with Mediator.

Association of Mediator with genes involved in metabolic pathways could also be due to interaction with activators specific to these processes. In our study, Mediator has been found to bind to a fewer number of genes (407) compared to the total number of genes (~ 6372) in the genome. Mediator may be required only for highly regulated genes, which are required for the parasite to respond to environmental cues.

Overall, the current study provides a preliminary view of Mediator occupancy across the *P. falciparum* genome during the schizont stage of the IDC. However, for ChIP-seq datasets to reflect actual Mediator occupancy sites in *Plasmodium*, ChIP-Seq needs to be performed in multiple biological replicates. The noise level due to hyper-ChIPable genomic loci can also be determined by carrying out ChIP-Seq analysis of a heterologous protein such as GFP, in parallel. In this study, Med31 has been used for assessing Mediator occupancy across the genome. Alternatively, other Mediator subunits can be tagged and utilized for ChIP-Seq analyses since cross-linking efficiencies have been found to vary for different proteins. Furthermore, ChIP-Seq can also be performed at other time-points during the IDC such as ring and trophozoite stages to understand the dynamics of Mediator occupancy. Such approaches may yield a comprehensive understanding of Mediator occupancy across the *Plasmodium* genome.

5.4 Conserved core Mediator complex is present in *P. falciparum*

In this study, we have shown for the first time the physical presence of the Mediator complex in *P. falciparum* by affinity purification and mass spectrometric analysis of the subunit Med31. In a study in yeast, Med31 has been tagged with a triple hemagglutinin tag at the 3' end of the ORF to characterize subunit interactions within the Mediator complex [188]. Therefore, in this study, Med31 tagged with a triple-hemagglutinin tag at the C-terminus was expressed episomally in T996 strain of parasites. This transgenic strain was further characterized at the molecular level.

Immunofluorescence analysis showed that the episomal Med31-HA protein was predominantly localized in a punctate pattern to the nucleus, in line with its presumptive nuclear functions in transcription (Fig. 4.7 B). The Med31-HA protein was also detected in the cytoplasm. Cytoplasmic localization was verified using aldolase as a cytoplasmic marker (Fig 4.8). Cytoplasmic localization of Med31 could represent its levels before export to the nucleus. We cannot rule out the possibility that cytoplasmic localization of Med31 could be an artefact caused by episomal expression. Med31 may also have functions in the cytoplasm as well, which may be unknown. This is not surprising, as murine Med31 is known to bind to a cytoplasmic signaling protein called Elmo1. Interestingly, Elmo1 binds to Med31 in the same region involved in interaction with Med7 and Elmo1-Med31 binding is independent of the Mediator complex. Interaction between murine Med31 and Elmo1 was found to be required for LPS-activated gene expression in primary macrophages [189]. However, homologs of elmo proteins or elmo domain proteins have not been identified in *Plasmodium spp.* [190]. Nevertheless, Med31 might have subunit-specific interactions and roles in other cellular processes in the parasite cytoplasm.

Co-localisation of episomal Med31-HA protein with a cytoplasmic marker aldolase was tested by IFA (Fig 4.8). As observed in Fig. 4.8, there is an overlap between nuclear DAPI and cytoplasmic aldolase signals. The non-specific staining could be due to high concentration of anti-aldolase antibody used in the experiment. Non-specific signals could be reduced by using lower concentrations of the anti-aldolase antibody, by prolonging the washing steps or by increasing the concentration of BSA in the blocking solution. A negative control without the primary aldolase antibody can also be tested to see if there is non-specific binding by the secondary antibody. The optimal conditions for staining by anti-aldolase antibody could be established independently before performing a co-localisation analysis. Also, the co-localisation IFA was carried out only once and needs to be repeated with optimized conditions to confirm the pattern of localization. The IFA results from Fig. 4.7 show that Med31 has a predominantly nuclear localization, in line with its functions in transcription. However, the co-localisation IFA in Fig. 4.8 is inconclusive and needs to be repeated to confirm the cytoplasmic localization of Med31.

An extensive bioinformatics analysis of the genome sequence of around 70 eukaryotic species using known or putative Mediator sequences identified 14 putative Mediator subunits in *P. falciparum* [144]. To detect other members of the Mediator complex, the episomal Med31-HA protein was immunoprecipitated from parasite protein lysates using an anti-HA antibody and the associated proteins were analysed by LC-MS/MS analysis in this study. The wild-type T996 strain was used as a negative control to rule out non-specific interactions. emPAI scores were calculated for Med31 and the T996 negative control samples and the ratio of emPAI scores of Med31 compared to T996 was deduced. Proteins with high emPAI score ratios were considered to be strong potential interacting partners. A total of 846 proteins were identified in the pull down experiment and 13 proteins having emPAI score ratios greater than 500 were considered as top ‘hits’. Med31 as well as 10 of the 14 putative Mediator subunits postulated on the basis of the bioinformatics analysis of the *P. falciparum* genome [144] were present among the high-scoring proteins. The Mediator subunits identified by our mass spectrometric results have been indicated in Table 5.1.

Head	Med6, Med8, Med17, Med18 & Med22
Middle	Med4, Med7, Med10, Med21 & Med31
Tail	Med14

Table 5.1 - List of Mediator subunits identified by Co-IP and LC-MS/MS analysis

The subunits belonging to the head, middle and tail modules have been indicated.

Med31 is one of the most highly conserved Mediator subunits [143] and is a component of the Mediator middle module [99]. Yeast two-hybrid experiments have shown that Med31 specifically interacts with Med10 and Med21 [188]. These subunits have been identified in the current study as interacting partners of Med31, thus validating our results. Middle module subunits including Med1, Med4, Med7, Med9, Med10, Med21 and Med31 are involved in a grid of interactions within the human Mediator complex,

wherein Med31 directly interacts with Med7 and Med19 [99]. Med19, however, has not been identified bioinformatically in the *P. falciparum* genome. Med31 interacts with the N-terminal region of Med7 and an X-ray crystal structure is available for the Med7N/Med31 submodule. Med31 forms a compact right-handed four-helix bundle enclosed by two proline-rich stretches of Med7N [94]. Consistent with Med31/Med7N interaction, our mass spectrometry dataset also contains Med7 as one of the top interacting partners.

Med31 has been shown to be required for heat-shock induced activation of endogenous genes in *Drosophila melanogaster* [191]. Interestingly, a heat shock protein DNAJ homologue, Pfj4 (PF3D7_1211400), has been identified in the current study as an interaction partner of Med31, with a high emPAI score ratio of 400. Pfj4 is a type-II Hsp40 protein whose expression is increased in *P. falciparum* upon heat shock. Pfj4 interacts with another heat shock protein, PfHsp70-1 suggesting that it has an important role in parasite adaptation to high temperatures [192].

The head module of the parasite Mediator complex is similar to that of yeast and consists of seven subunits namely Med6, Med8, Med11, Med17, Med18, Med20 and Med22. X-ray crystal structure of the yeast head module is available which shows an extensive network of interaction between the head subunits [93]. Integrity of the head module complex was shown to reside in a helical bundle formed by five different head subunits, Med6, Med8, Med11, Med17 and Med22. Several important roles of Med11 have been documented. It plays an important role in the assembly of the head module [93]. Mediator is known to interact with TFIIF through the Med11 subunit. Mutations in the Med11 subunit have also been shown to affect the interaction of head module with TFIIF *in vivo*, resulting in decreased phosphorylation of RNAPII CTD at Ser5 residue [155]. The subunits Med17, Med18 and Med22 have been implicated in functional interactions with RNAPII [93]. TBP binds to Mediator via the N-terminal residues of Med8 subunit. Med8, Med18 and Med20 form a subcomplex that is important for initiation complex formation [193].

All the head module subunits except Med11 and Med20 have been identified in the current study. Both these subunits are known to interact with another head module subunit Med18, that is present in *P. falciparum*, as shown by computational analysis [144] supported by our mass spectrometry results. Immunoprecipitation using this subunit could perhaps shed some light on whether the Med11 and Med20 subunits are present, assuming their interactions with Med18 are preserved in the parasite.

The head and middle modules are connected together through the Med17 (head) - Med21 (middle) interaction [188]. Our study shows that these conserved subunits are indeed present in the parasite. A 3D model of the Mediator middle module shows that it is composed of a tetramer formed by the heterodimeric subcomplexes Med4/Med9 and Med7/Med21 [96]. Med9 has not been identified in the current study. Since it interacts with Med4, maybe affinity purification of Med4 can lead to its isolation from parasite protein lysates.

CKM module associates with the core Mediator only under certain conditions and is known to be involved in transcriptional repression. Computational studies of the *P. falciparum* genome have not detected genes encoding members of the Cdk8 kinase module (CKM), nor have any of the CKM components been identified in our MS results. CKM interaction with the core Mediator involves direct contact between the Med13 subunit of CKM and the Med19 subunit of the Mediator middle module [99]. Med19 has also not been identified in the *P. falciparum* genome.

The basic function of tail module subunits is to link Mediator to sequence-specific activators as most Mediator-activator interactions involve the tail subunits [128]. The only tail Mediator subunit to be predicted from the parasite genome, Med14, has been validated in our study. Structurally, Med14 spans the entire Mediator complex, bridging all the modules [99]. Apart from its architectural role, it is also required for Mediator to be transcriptionally active. Cevher et al., observed that although a complex of 13 subunits derived from the head and middle modules was stable, Med14 was required for basal and co-activator functions of the complex under *in vitro* assay conditions [194]. This study and another study by Plaschka et al., [101] have defined the minimum set of subunits

required to assemble a functional Mediator complex under *in vitro* conditions. This minimal complex, termed as the core Mediator, is composed of head and middle module subunits supported by the Med14 subunit. Med9 and Med19 have been identified as components of the core Mediator in yeast and not in humans, although these subunits are found in humans as well. Med30 is a component of human core Mediator and has not been identified in yeast. Therefore, the composition of the functional core Mediator complex seems to vary among species. The predicted Mediator complex in *P. falciparum* is similar to the yeast core Mediator except for the subunit Med19. It is unclear whether the parasite has only retained the minimal set of subunits, which are required for its basic functions or if it contains other subunits required for specialized functions.

In the current study, two uncharacterized *P. falciparum* proteins; PF3D7_0526800 and PF3D7_1363600 were also identified as interacting partners of Med31. These two proteins have high emPAI score ratios in the mass spectrometric analysis and seem to be strongly associated with the parasite Mediator complex.

To identify whether these protein sequences are related to known Mediator subunits, full-length protein sequences were analysed by BLASTp and PSI-BLAST. However, the search strategies did not yield any matches with known Mediator subunits from other species, nor were they detected in the careful bioinformatics approach of Bourbon [144]. Different subunits of the Mediator are known to contain intrinsically disordered regions (IDRs), which do not form any well-defined three-dimensional structures. These IDRs are thought to impart structural flexibility to the Mediator [157]. Disorder-forming regions were predicted for the two uncharacterized proteins. Although, the average probability based on full-length protein was not indicative of disordered proteins, 27-30 bp disordered regions were found within the proteins. While PF3D7_1363600 contained one such region, PF3D7_0526800 was found to contain two such segments. However, we could not glean any other information about these proteins through the bioinformatics approaches that we used.

Mediator interacts with several accessory proteins during the process of transcription. Therefore, it is difficult to determine whether its interacting protein partners

are *bona fide* Mediator subunits. One way to analyze this is by determining if the protein reproducibly associates with the complex. Therefore, parasite strains expressing HA-tagged versions of these two parasite proteins episomally were established to assess their protein interaction partners by Co-IP and mass spectrometric analysis. The PF3D7_0526800 transgenic line was chosen for initial experiments and the HA-tagged protein was successfully immunoprecipitated from parasite protein lysates. Further Co-IP and LC-MS/MS analysis of this protein may help in validating its interaction with Mediator subunits. An alternative biochemistry-based approach can be adopted to determine if these uncharacterized proteins might be putative Mediator subunits. Parasite protein lysates prepared from the transgenic strains could be subjected to immunoprecipitation using an anti-HA antibody and separated by chromatography on a gel filtration column. If these proteins are indeed a part of the Mediator complex, they might co-elute together, in a single high molecular weight peak [188].

It would be interesting to find out if these uncharacterized proteins are *Plasmodium*-specific Mediator subunits. It would not be surprising since some Mediator subunits have been identified only in certain species and have an important role in regulating species-specific gene expression. The metazoan-specific subunit, Med26, is one such example. Studies have shown the involvement of Med26 in the release of paused RNAPII into productive elongation. The same N-terminal region of human Med26 subunit that binds Super elongation complex (SEC) [195] also binds TFIID showing that Med26 first interacts with TFIID for PIC assembly and then exchanges TFIID for SEC for controlling elongation [196]. Similarly, several plant-specific Mediator subunits have also been identified. When Mediator complex was purified from *Arabidopsis thaliana*, several proteins co-purified with the complex. These plant-specific Mediator subunits were designated as Med34-37 [197]. However, there is uncertainty whether these proteins can be truly considered as plant-specific Mediator subunits, as all of them have orthologous proteins in humans and yeast [198].

Compared to RNAPII and TFIID, Mediator sequences have diversified extensively among eukaryotes throughout evolution [199]. This is seen in the limited sequence similarity between yeast and human Mediator subunits [124]. The number of Mediator

subunits has increased with the increasing complexity of eukaryotes. In some species, certain Mediator subunits have been lost during the course of evolution. For example, counterparts of Med16 and Med25 could not be detected in any of the worm species. These subunits were also missing from the malaria vector, *Anopheles gambiae* [144]. Conversely, some species have accumulated additional subunits. Since Mediator sequences have diverged extensively, it is not known if CKM module or other tail subunits are present in the parasite genome.

Based on computational studies, *Plasmodium* seems to have a minimum functional complex made of middle and head module subunits held together by Med14, which acts as a scaffold. A schematic diagram of the *P. falciparum* Mediator complex is shown in Fig 5.1. The current study has uncovered two additional proteins that could be putative Mediator subunits. These presumptive Mediator subunits could be components of the tail module that have probably evolved in the parasite to interact with AP2 family of transcription factors or for other parasite-specific interactions.

Figure 5.1 - Schematic diagram of the Mediator complex in *P. falciparum*

Subunits belonging to the head (blue), middle (orange) and tail (yellow) modules have been depicted. Subunits that have not been identified in the current study have been represented as dotted circles (Med9, Med11 & Med20). Med31 subunit used in this study has been outlined in red. The uncharacterized proteins, PF3D7_0526800 &

PF3D7_1363600, identified in the current study as putative *Plasmodium* Mediator subunits have been shown in gray. Figure adapted from [198].

Collectively, the reduced spectrum of conserved gene-specific transcription factors, *cis*-regulatory DNA motifs, and absence of the regulatory tail and CKM modules of Mediator in *Plasmodium* highlight the gaps in our understanding of transcriptional regulation in this organism. The putative Mediator subunits identified in this study could be potential tail subunits of the parasite Mediator complex and may have evolved for specialized interactions unique to parasite gene regulatory mechanisms. Since tail subunits are known to contact activator proteins in other organisms, further analysis of these presumptive Mediator subunits may lead to the identification of gene-specific activator proteins. It may also help in uncovering other unknown tail subunits of the *Plasmodium* Mediator complex. Furthermore, *Plasmodium*-specific Mediator subunits could have a significant role to play in the unique control of gene expression programs and could be explored as potential drug targets.

6 FUTURE DIRECTION

The findings of this study provide a preliminary characterization of the Mediator complex in *P. falciparum*, in terms of the proteins it interacts with and its association with genomic regions. However, further experiments need to be carried out to validate and extend the findings of this study.

Through Co-IP and mass spectrometric analysis of episomal HA-tagged Med31 protein, we have identified 11 out of 14 Mediator subunits predicted in the *P. falciparum* genome. However, the Co-IP and mass spectrometric analysis needs to be repeated to ensure that the interactions that we observe are reproducible. We observed two uncharacterized proteins, PF3D7_0526800 & PF3D7_1363600 among the top interacting partners of Med31. Based on their strong association with other Mediator subunits, we hypothesize that they could be putative Mediator subunits. However, further experiments are required to validate the interaction of these proteins with other Mediator subunits. Independent parasite strains expressing HA-tagged versions of either PF3D7_0526800 or PF3D7_1363600 episomally have already been established. Co-IP and mass spectrometric analysis of these proteins can help to determine if these proteins form a complex with other Mediator subunits. Another approach that can be tested is by immunoprecipitation of these proteins and separating them by chromatography on a gel filtration column. If these proteins are indeed a part of the Mediator complex, they might co-elute together, in a single high molecular weight peak. Thus, a reproducible association of these proteins with other Mediator complex subunits can help in ascertaining if they are *bona fide* Mediator subunits.

The only tail subunit that has been predicted in the *Plasmodium* genome and identified by our study is Med14. Analysis of protein interaction partners of this subunit could perhaps shed light on other tail subunits that may exist in the parasite. Although the subunits Med9, Med11 and Med20 have been predicted in the *Plasmodium* genome, they were not identified in our study. These subunits could be characterized individually or by analysis of their known interaction partners. For example, Med9 is known to form a heterodimer with Med4, while Med11 and Med20 are known to interact with Med18. Both

Med4 and Med18 have been identified in our study. Therefore, co-IP and mass spectrometric analysis of these subunits may help in confirming the presence of Med9, Med11 and Med20 subunits in the parasite.

Since the conditional knockdown system was not found to be functional in the current study, we could attempt to knock out Med31 by double crossover homologous recombination. This would help us in determining if Med31 is an essential Mediator subunit in *Plasmodium spp.* Transcripts that are regulated by Med31 can be further analysed by performing an RNA-Seq analysis of wild-type and Med31 knockout strains. We could also attempt to introduce mutations in Med31 or other Mediator subunits using the CRISPR-cas9 genome editing technique. Assuming that the mutant proteins are still stable, this would allow us to specifically test the functions of different domains of the proteins. Mutations in essential Mediator subunits may help in retaining partial function thereby avoiding lethality. We could also attempt to tag other Mediator subunits with degron domains to test their functions. Alternate conditional knockdown strategies can be employed such as the plant-derived auxin based degron. This auxin-induced destabilization domain has been used to assess the function of calcineurin in *P. berghei* [200] and has also been tested in *P. falciparum* [201].

Furthermore, it would be interesting to determine the subcellular localization of different Mediator subunits by means of epitope tagging or by raising antibodies against these subunits. The sub-nuclear localization of Med31 and other Mediator subunits could be refined by co-staining with markers of the nucleolus, nuclear pore complex, telomeric heterochromatin, and active *var* loci. This would allow us to validate the expected nuclear localization of this complex as well as imply roles in repression or activation of gene expression. Analysis of Mediator occupancy at ring and trophozoite stages will further provide a comprehensive picture of Mediator association with genomic regions during the IDC. Results from our study show that Mediator enrichment is not correlated with gene expression levels. Therefore, Mediator occupancy needs to be further compared with RNAPII occupancy to better understand the role of the Mediator complex in regulation of gene expression.

7 CONCLUSION

In this study, we have characterized the Mediator complex in *P. falciparum* in order to understand the highly co-ordinated gene expression pattern of the parasite during the IDC. The highly conserved Mediator subunit, Med31 was chosen for the study. For functional analysis of the Med31 subunit, the endogenous gene was successfully tagged with a triple hemagglutinin tag and an *E. coli* DHFR based destabilization domain. However, western blot and growth curve analysis of the transgenic strain showed that protein levels were not regulated by the destabilisation domain thereby rendering the conditional knockdown system ineffective for further analysis of Med31 protein function.

Med31 tagged with a triple-hemagglutinin tag at the C-terminus was expressed episomally in T996 strain of parasites. This transgenic strain was further used to assess Mediator occupancy and its protein-protein interactions. Genome-wide ChIP-Seq analysis of Med31 showed that it occupies the coding region of most genes in the schizont stage of the parasite. For a subset of genes, Mediator enrichment upstream of the ATG is accompanied by a lower level of gene expression in the schizont stage. This observation points towards a repressive role of Mediator in transcription of these genes. Mediator enrichment is also observed downstream of the STOP codon for a subset of genes suggesting that Mediator may have functions beyond transcription initiation and could be involved in transcription termination and mRNA processing events. Peak-based analysis revealed that 407 genes were specifically associated with Mediator. Gene ontology analysis revealed that these genes were particularly enriched in metabolic processes such as selenocysteine biosynthesis, tRNA modifications, inorganic cation transport and copper ion transport. Therefore, Mediator seems to be involved in regulation of specific metabolic pathways in the parasite.

Using affinity precipitation and mass spectrometric analysis of episomally expressed HA-tagged Med31, we have shown the physical presence of the Mediator complex in the parasite for the first time. Our results showed the presence of 11 out of 14 predicted *P. falciparum* Mediator subunits. Mediator complex subunits Med9, Med11 and Med20 were not identified in the pull down experiment. Interestingly, two uncharacterized

proteins; PF3D7_0526800 and PF3D7_1363600 were identified which seem to be strongly associated with the Mediator complex and therefore, could be presumptive Mediator subunits. Parasite strains expressing HA-tagged versions of either PF3D7_0526800 or PF3D7_1363600 episomally were established and the PF3D7_0526800 protein was successfully immunoprecipitated from parasite protein lysates. Co-IP and mass spectrometric analysis of these proteins may help in substantiating their interactions with other Mediator complex subunits. Overall, this study has expanded our understanding of the composition and function of the Mediator complex in *P. falciparum*. A detailed study of the Mediator may further provide insights about modes of gene regulation and help in identifying potential drug targets in the malarial parasite.

8 REFERENCES

- [1] Cox FE. History of the discovery of the malaria parasites and their vectors. *Parasites & vectors* 2010;3:5.
- [2] WHO, World malaria report. World health organization Geneva, Switzerland 2016.
- [3] Carter R, Mendis KN. Evolutionary and Historical Aspects of the Burden of Malaria. *Clinical Microbiology Reviews* 2002;15:564-94.
- [4] Zimmerman PA, Ferreira MU, Howes RE, Mercereau-Puijalon O. Red blood cell polymorphism and susceptibility to *Plasmodium vivax*. *Advances in parasitology* 2013;81:27-76.
- [5] Miller LH, Ackerman HC, Su XZ, Wellems TE. Malaria biology and disease pathogenesis: insights for new treatments. *Nature medicine* 2013;19:156-67.
- [6] Idro R, Jenkins NE, Newton CRJC. Pathogenesis, clinical features, and neurological outcome of cerebral malaria. *The Lancet Neurology* 2005;4:827-40.
- [7] Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. *Nature* 2002;415:673-9.
- [8] Biamonte MA, Wanner J, Le Roch KG. Recent advances in malaria drug discovery. *Bioorg Med Chem Lett* 2013;23:2829-43.
- [9] Prudencio M, Rodriguez A, Mota MM. The silent path to thousands of merozoites: the *Plasmodium* liver stage. *Nature reviews Microbiology* 2006;4:849-56.
- [10] Bannister LH, Hopkins JM, Fowler RE, Krishna S, Mitchell GH. A brief illustrated guide to the ultrastructure of *Plasmodium falciparum* asexual blood stages. *Parasitology today* 2000;16:427-33.
- [11] Cowman AF, Crabb BS. Invasion of red blood cells by malaria parasites. *Cell* 2006;124:755-66.
- [12] Hanssen E, McMillan PJ, Tilley L. Cellular architecture of *Plasmodium falciparum*-infected erythrocytes. *International journal for parasitology* 2010;40:1127-35.
- [13] Hanssen E, Goldie KN, Tilley L. Ultrastructure of the asexual blood stages of *Plasmodium falciparum*. *Methods in cell biology* 2010;96:93-116.
- [14] de Koning-Ward TF, Gilson PR, Boddey JA, Rug M, Smith BJ, Papenfuss AT, Sanders PR, Lundie RJ, Maier AG, Cowman AF, Crabb BS. A newly discovered protein export machine in malaria parasites. *Nature* 2009;459:945-9.
- [15] Taylor DW, Parra M, Chapman GB, Stearns ME, Renner J, Aikawa M, Uni S, Aley SB, Panton LJ, Howard RJ. Localization of *Plasmodium falciparum* histidine-rich protein

1 in the erythrocyte skeleton under knobs. *Molecular and biochemical parasitology* 1987;25:165-74.

[16] Arnot DE, Gull K. The Plasmodium cell-cycle: facts and questions. *Annals of tropical medicine and parasitology* 1998;92:361-5.

[17] Gerald N, Mahajan B, Kumar S. Mitosis in the human malaria parasite *Plasmodium falciparum*. *Eukaryotic cell* 2011;10:474-82.

[18] Francia ME, Striepen B. Cell division in apicomplexan parasites. *Nature reviews Microbiology* 2014;12:125-36.

[19] Abkarian M, Massiera G, Berry L, Roques M, Braun-Breton C. A novel mechanism for egress of malarial parasites from red blood cells. *Blood* 2011;117:4118-24.

[20] Lew Virgilio L. Malaria: Surprising Mechanism of Merozoite Egress Revealed. *Current Biology* 2011;21:R314-R6.

[21] Flannery EL, Chatterjee AK, Winzeler EA. Antimalarial drug discovery - approaches and progress towards new medicines. *Nature reviews Microbiology* 2013;11:849-62.

[22] Srivastava IK, Vaidya AB. A mechanism for the synergistic antimalarial action of atovaquone and proguanil. *Antimicrobial agents and chemotherapy* 1999;43:1334-9.

[23] WHO. World malaria report, World health organization. Geneva, Switzerland 2013.

[24] Arieu F, Witkowski B, Amaratunga C, Beghain J, Langlois AC, Khim N, Kim S, Duru V, Bouchier C, Ma L, Lim P, Leang R, Duong S, Sreng S, Suon S, Chuor CM, Bout DM, Menard S, Rogers WO, Genton B, Fandeur T, Miotto O, Ringwald P, Le Bras J, Berry A, Barale JC, Fairhurst RM, Benoit-Vical F, Mercereau-Puijalon O, Menard D. A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. *Nature* 2014;505:50-5.

[25] Mok S, Ashley EA, Ferreira PE, Zhu L, Lin Z, Yeo T, Chotivanich K, Imwong M, Pukrittayakamee S, Dhorda M, Nguon C, Lim P, Amaratunga C, Suon S, Hien TT, Htut Y, Faiz MA, Onyamboko MA, Mayxay M, Newton PN, Tripura R, Woodrow CJ, Miotto O, Kwiatkowski DP, Nosten F, Day NPJ, Preiser PR, White NJ, Dondorp AM, Fairhurst RM, Bozdech Z. Population transcriptomics of human malaria parasites reveals the mechanism of artemisinin resistance. *Science* 2014.

[26] Doolan DL, Apte SH, Proietti C. Genome-based vaccine design: the promise for malaria and other infectious diseases. *International journal for parasitology* 2014;44:901-13.

[27] Cohen J, Nussenzweig V, Vekemans J, Leach A. From the circumsporozoite protein to the RTS,S/AS candidate vaccine. *Human Vaccines* 2014;6:90-6.

-
- [28] Deng W, Roberts SG. Core promoter elements recognized by transcription factor IIB. *Biochemical Society transactions* 2006;34:1051-3.
- [29] Lee TI, Young RA. Transcription of eukaryotic protein-coding genes. *Annual review of genetics* 2000;34:77-137.
- [30] Buratowski S. Gene expression: Transcription initiation unwrapped. *Nature* 2012;483:286-7.
- [31] Croken MM, Nardelli SC, Kim K. Chromatin modifications, epigenetics, and how protozoan parasites regulate their lives. *Trends in parasitology* 2012;28:202-13.
- [32] Li B, Carey M, Workman JL. The role of chromatin during transcription. *Cell* 2007;128:707-19.
- [33] Vignali M, Hassan AH, Neely KE, Workman JL. ATP-dependent chromatin-remodeling complexes. *Molecular and cellular biology* 2000;20:1899-910.
- [34] Samara NL, Wolberger C. A new chapter in the transcription SAGA. *Current opinion in structural biology* 2011;21:767-74.
- [35] Ansari SA, Morse RH. Mechanisms of Mediator complex action in transcriptional activation. *Cell Mol Life Sci* 2013;70:2743-56.
- [36] Bonnet J, Wang CY, Baptista T, Vincent SD, Hsiao WC, Stierle M, Kao CF, Tora L, Devys D. The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. *Genes Dev* 2014;28:1999-2012.
- [37] Jonkers I, Lis JT. Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol* 2015;16:167-77.
- [38] Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DM, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 2002;419:498-511.
- [39] Batugedara G, Lu XM, Bunnik EM, Le Roch KG. The Role of Chromatin Structure in Gene Regulation of the Human Malaria Parasite. *Trends in parasitology* 2017.
- [40] Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS biology* 2003;1:E5.
-

-
- [41] Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD, De La Vega P, Holder AA, Batalov S, Carucci DJ, Winzeler EA. Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 2003;301:1503-8.
- [42] Foth BJ, Zhang N, Chaal BK, Sze SK, Preiser PR, Bozdech Z. Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite *Plasmodium falciparum*. *Mol Cell Proteomics* 2011;10:M110.006411.
- [43] Le Roch KG, Johnson JR, Florens L, Zhou Y, Santrosyan A, Grainger M, Yan SF, Williamson KC, Holder AA, Carucci DJ, Yates JR, Winzeler EA. Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome research* 2004;14:2308-18.
- [44] Jiang L, Mu J, Zhang Q, Ni T, Srinivasan P, Rayavara K, Yang W, Turner L, Lavstsen T, Theander TG, Peng W, Wei G, Jing Q, Wakabayashi Y, Bansal A, Luo Y, Ribeiro JM, Scherf A, Aravind L, Zhu J, Zhao K, Miller LH. PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*. *Nature* 2013;499:223-7.
- [45] Hoeijmakers WAM, Salcedo-Amaya AM, Smits AH, François KJ, Treeck M, Gilberger TW, Stunnenberg HG, Bártfai R. H2A.Z/H2B.Z double-variant nucleosomes inhabit the AT-rich promoter regions of the *Plasmodium falciparum* genome. *Molecular microbiology* 2013;87:1061-73.
- [46] Petter M, Selvarajah SA, Lee CC, Chin WH, Gupta AP, Bozdech Z, Brown GV, Duffy MF. H2A.Z and H2B.Z double-variant nucleosomes define intergenic regions and dynamically occupy var gene promoters in the malaria parasite *Plasmodium falciparum*. *Molecular microbiology* 2013;87:1167-82.
- [47] Painter HJ, Campbell TL, Llinas M. The Apicomplexan AP2 family: integral factors regulating *Plasmodium* development. *Molecular and biochemical parasitology* 2011;176:1-7.
- [48] Ay F, Bunnik EM, Varoquaux N, Bol SM, Prudhomme J, Vert JP, Noble WS, Le Roch KG. Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome research* 2014;24:974-88.
- [49] Gupta AP, Chin WH, Zhu L, Mok S, Luah YH, Lim EH, Bozdech Z. Dynamic epigenetic regulation of gene expression during the life cycle of malaria parasite *Plasmodium falciparum*. *PLoS pathogens* 2013;9:e1003170.
- [50] Chaal BK, Gupta AP, Wastuwidyaningtyas BD, Luah YH, Bozdech Z. Histone deacetylases play a major role in the transcriptional regulation of the *Plasmodium falciparum* life cycle. *PLoS pathogens* 2010;6:e1000737.
- [51] Josling GA, Petter M, Oehring SC, Gupta AP, Dietz O, Wilson DW, Schubert T, Langst G, Gilson PR, Crabb BS, Moes S, Jenoe P, Lim SW, Brown GV, Bozdech Z, Voss

- TS, Duffy MF. A Plasmodium Falciparum Bromodomain Protein Regulates Invasion Gene Expression. *Cell host & microbe* 2015;17:741-51.
- [52] Bischoff E, Vaquero C. In silico and biological survey of transcription-associated proteins implicated in the transcriptional machinery during the erythrocytic development of Plasmodium falciparum. *BMC genomics* 2010;11:34.
- [53] Coulson RM, Hall N, Ouzounis CA. Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum. *Genome research* 2004;14:1548-54.
- [54] Boschet C, Gissot M, Briquet S, Hamid Z, Claudel-Renard C, Vaquero C. Characterization of PfMyb1 transcription factor during erythrocytic development of 3D7 and F12 Plasmodium falciparum clones. *Molecular and biochemical parasitology* 2004;138:159-63.
- [55] Gissot M, Briquet S, Refour P, Boschet C, Vaquero C. PfMyb1, a Plasmodium falciparum transcription factor, is required for intra-erythrocytic growth and controls key genes for cell cycle regulation. *J Mol Biol* 2005;346:29-42.
- [56] Kun JFJ, Anders RF. A Plasmodium falciparum gene encoding a high mobility group protein box. *Molecular and biochemical parasitology* 1995;71:249-53.
- [57] Briquet S, Boschet C, Gissot M, Tissandie E, Sevilla E, Franetich JF, Thierry I, Hamid Z, Bourgouin C, Vaquero C. High-mobility-group box nuclear factors of Plasmodium falciparum. *Eukaryotic cell* 2006;5:672-82.
- [58] Gissot M, Ting LM, Daly TM, Bergman LW, Sinnis P, Kim K. High mobility group protein HMGB2 is a critical regulator of plasmodium oocyst development. *The Journal of biological chemistry* 2008;283:17030-8.
- [59] Balaji S, Babu MM, Iyer LM, Aravind L. 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* 2005;33:3994-4006.
- [60] Campbell TL, De Silva EK, Olszewski KL, Elemento O, Llinas M. Identification and genome-wide prediction of DNA binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLoS pathogens* 2010;6:e1001165.
- [61] Yuda M, Iwanaga S, Shigenobu S, Mair GR, Janse CJ, Waters AP, Kato T, Kaneko I. Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol Microbiol* 2009;71:1402-14.
- [62] Yuda M, Iwanaga S, Shigenobu S, Kato T, Kaneko I. Transcription factor AP2-Sp and its target genes in malarial sporozoites. *Molecular microbiology* 2010;75:854-63.

-
- [63] Flueck C, Bartfai R, Niederwieser I, Witmer K, Alako BT, Moes S, Bozdech Z, Jenoe P, Stunnenberg HG, Voss TS. A major role for the *Plasmodium falciparum* ApiAP2 protein PfSIP2 in chromosome end biology. *PLoS pathogens* 2010;6:e1000784.
- [64] Kafsack BF, Rovira-Graells N, Clark TG, Bancells C, Crowley VM, Campino SG, Williams AE, Drought LG, Kwiatkowski DP, Baker DA, Cortes A, Llinas M. A transcriptional switch underlies commitment to sexual development in malaria parasites. *Nature* 2014;507:248-52.
- [65] Sinha A, Hughes KR, Modrzynska KK, Otto TD, Pfander C, Dickens NJ, Religa AA, Bushell E, Graham AL, Cameron R, Kafsack BF, Williams AE, Llinas M, Berriman M, Billker O, Waters AP. A cascade of DNA-binding proteins for sexual commitment and development in *Plasmodium*. *Nature* 2014;507:253-7.
- [66] Jackson KE, Habib S, Frugier M, Hoen R, Khan S, Pham JS, Ribas de Pouplana L, Royo M, Santos MA, Sharma A, Ralph SA. Protein translation in *Plasmodium* parasites. *Trends in parasitology* 2011;27:467-76.
- [67] Zhang K, Rathod PK. Divergent regulation of dihydrofolate reductase between malaria parasite and human host. *Science* 2002;296:545-7.
- [68] Bunnik EM, Chung DW, Hamilton M, Ponts N, Saraf A, Prudhomme J, Florens L, Le Roch KG. Polysome profiling reveals translational control of gene expression in the human malaria parasite *Plasmodium falciparum*. *Genome biology* 2013;14:R128.
- [69] Rai R, Zhu L, Chen H, Gupta AP, Sze SK, Zheng J, Ruedl C, Bozdech Z, Featherstone M. Genome-wide analysis in *Plasmodium falciparum* reveals early and late phases of RNA polymerase II occupancy during the infectious cycle. *BMC genomics* 2014;15:959.
- [70] Shock JL, Fischer KF, DeRisi JL. Whole-genome analysis of mRNA decay in *Plasmodium falciparum* reveals a global lengthening of mRNA half-life during the intra-erythrocytic development cycle. *Genome Biol* 2007;8:R134.
- [71] Balu B, Maher SP, Pance A, Chauhan C, Naumov AV, Andrews RM, Ellis PD, Khan SM, Lin JW, Janse CJ, Rayner JC, Adams JH. CCR4-associated factor 1 coordinates the expression of *Plasmodium falciparum* egress and invasion proteins. *Eukaryotic cell* 2011;10:1257-63.
- [72] Goyal M, Banerjee C, Nag S, Bandyopadhyay U. The Alba protein family: Structure and function. *Biochimica et biophysica acta* 2016;1864:570-83.
- [73] Vembar SS, Macpherson CR, Sismeiro O, Coppee JY, Scherf A. The PfAlba1 RNA-binding protein is an important regulator of translational timing in *Plasmodium falciparum* blood stages. *Genome biology* 2015;16:212.
- [74] Zhang Q, Siegel TN, Martins RM, Wang F, Cao J, Gao Q, Cheng X, Jiang L, Hon CC, Scheidig-Benatar C, Sakamoto H, Turner L, Jensen AT, Claes A, Guizetti J,
-

- Malmquist NA, Scherf A. Exonuclease-mediated degradation of nascent RNA silences genes linked to severe malaria. *Nature* 2014;513:431-5.
- [75] Zhang Q, Scherf A. RNase II: A new player enters the game. *RNA Biol* 2015;12:586-9.
- [76] Hughes KR, Philip N, Starnes GL, Taylor S, Waters AP. From cradle to grave: RNA biology in malaria parasites. *Wiley interdisciplinary reviews RNA* 2010;1:287-303.
- [77] Otto TD, Wilinski D, Assefa S, Keane TM, Sarry LR, Bohme U, Lemieux J, Barrell B, Pain A, Berriman M, Newbold C, Llinas M. New insights into the blood-stage transcriptome of *Plasmodium falciparum* using RNA-Seq. *Molecular microbiology* 2010;76:12-24.
- [78] Eshar S, Allemand E, Sebag A, Glaser F, Muchardt C, Mandel-Gutfreund Y, Karni R, Dzikowski R. A novel *Plasmodium falciparum* SR protein is an alternative splicing factor required for the parasites' proliferation in human erythrocytes. *Nucleic acids research* 2012;40:9903-16.
- [79] Eshar S, Altenhofen L, Rabner A, Ross P, Fastman Y, Mandel-Gutfreund Y, Karni R, Llinas M, Dzikowski R. PfSR1 controls alternative splicing and steady-state RNA levels in *Plasmodium falciparum* through preferential recognition of specific RNA motifs. *Molecular microbiology* 2015;96:1283-97.
- [80] Chereji RV, Bharatula V, Elfving N, Blomberg J, Larsson M, Morozov AV, Broach JR, Bjorklund S. Mediator binds to boundaries of chromosomal interaction domains and to proteins involved in DNA looping, RNA metabolism, chromatin remodeling, and actin assembly. *Nucleic acids research* 2017.
- [81] Gill G, Ptashne M. Negative effect of the transcriptional activator GAL4. *Nature* 1988;334:721-4.
- [82] Kornberg RD. Mediator and the mechanism of transcriptional activation. *Trends in biochemical sciences* 2005;30:235-9.
- [83] Kelleher III RJ, Flanagan PM, Kornberg RD. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* 1990;61:1209-15.
- [84] Flanagan PM, Kelleher RJ, 3rd, Sayre MH, Tschochner H, Kornberg RD. A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* 1991;350:436-8.
- [85] Thompson CM, Koleske AJ, Chao DM, Young RA. A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* 1993;73:1361-75.

-
- [86] Nonet ML, Young RA. Intragenic and Extragenic Suppressors of Mutations in the Heptapeptide Repeat Domain of *Saccharomyces Cerevisiae* RNA Polymerase II. *Genetics* 1989;123:715-24.
- [87] Kim Y-J, Björklund S, Li Y, Sayre MH, Kornberg RD. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 1994;77:599-608.
- [88] Asturias FJ, Jiang YW, Myers LC, Gustafsson CM, Kornberg RD. Conserved structures of mediator and RNA polymerase II holoenzyme. *Science* 1999;283:985-7.
- [89] Liu Y, Ranish JA, Aebersold R, Hahn S. Yeast Nuclear Extract Contains Two Major Forms of RNA Polymerase II Mediator Complexes. *Journal of Biological Chemistry* 2001;276:7169-75.
- [90] Dotson MR, Yuan CX, Roeder RG, Myers LC, Gustafsson CM, Jiang YW, Li Y, Kornberg RD, Asturias FJ. Structural organization of yeast and mammalian mediator complexes. *Proceedings of the National Academy of Sciences of the United States of America* 2000;97:14307-10.
- [91] Davis JA, Takagi Y, Kornberg RD, Asturias FA. Structure of the yeast RNA polymerase II holoenzyme: Mediator conformation and polymerase interaction. *Molecular cell* 2002;10:409-15.
- [92] Cai G, Imasaki T, Takagi Y, Asturias FJ. Mediator structural conservation and implications for the regulation mechanism. *Structure* 2009;17:559-67.
- [93] Imasaki T, Calero G, Cai G, Tsai KL, Yamada K, Cardelli F, Erdjument-Bromage H, Tempst P, Berger I, Kornberg GL, Asturias FJ, Kornberg RD, Takagi Y. Architecture of the Mediator head module. *Nature* 2011;475:240-3.
- [94] Koschubs T, Seizl M, Lariviere L, Kurth F, Baumli S, Martin DE, Cramer P. Identification, structure, and functional requirement of the Mediator submodule Med7N/31. *The EMBO journal* 2009;28:69-80.
- [95] Baumli S, Hoepfner S, Cramer P. A conserved mediator hinge revealed in the structure of the MED7.MED21 (Med7.Srb7) heterodimer. *The Journal of biological chemistry* 2005;280:18171-8.
- [96] Lariviere L, Plaschka C, Seizl M, Petrotchenko EV, Wenzek L, Borchers CH, Cramer P. Model of the Mediator middle module based on protein cross-linking. *Nucleic acids research* 2013;41:9266-73.
- [97] Chadick JZ, Asturias FJ. Structure of eukaryotic Mediator complexes. *Trends in biochemical sciences* 2005;30:264-71.
- [98] Wang X, Sun Q, Ding Z, Ji J, Wang J, Kong X, Yang J, Cai G. Redefining the modular organization of the core Mediator complex. *Cell Res* 2014;24:796-808.
-

-
- [99] Tsai KL, Tomomori-Sato C, Sato S, Conaway RC, Conaway JW, Asturias FJ. Subunit architecture and functional modular rearrangements of the transcriptional mediator complex. *Cell* 2014;157:1430-44.
- [100] Robinson PJ, Trnka MJ, Pellarin R, Greenberg CH, Bushnell DA, Davis R, Burlingame AL, Sali A, Kornberg RD. Molecular architecture of the yeast Mediator complex. *Elife* 2015;4.
- [101] Plaschka C, Lariviere L, Wenzek L, Seizl M, Hemann M, Tegunov D, Petrotchenko EV, Borchers CH, Baumeister W, Herzog F, Villa E, Cramer P. Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature* 2015;518:376-80.
- [102] Robinson PJ, Trnka MJ, Bushnell DA, Davis RE, Mattei PJ, Burlingame AL, Kornberg RD. Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell* 2016;166:1411-22 e16.
- [103] Barberis A, Petrascheck M. Transcription Activation in Eukaryotic Cells. eLS: John Wiley & Sons, Ltd; 2001.
- [104] Thompson CM, Young RA. General requirement for RNA polymerase II holoenzymes in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 1995;92:4587-90.
- [105] Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 1998;95:717-28.
- [106] Mittler G, Kremmer E, Timmers HTM. Novel critical role of a human Mediator complex for basal RNA polymerase II transcription. *2001;2:808-13.*
- [107] Takagi Y, Kornberg RD. Mediator as a general transcription factor. *The Journal of biological chemistry* 2006;281:80-9.
- [108] Lacombe T, Poh SL, Barbey R, Kuras L. Mediator is an intrinsic component of the basal RNA polymerase II machinery in vivo. *Nucleic acids research* 2013;41:9651-62.
- [109] Fan X, Chou DM, Struhl K. Activator-specific recruitment of Mediator in vivo. *Nature structural & molecular biology* 2006;13:117-20.
- [110] Lewis BA, Reinberg D. Promoter activation when the ChIPs are down. *Nature structural & molecular biology* 2006;13:96-7.
- [111] Park JM, Werner J, Kim JM, Lis JT, Kim YJ. Mediator, not holoenzyme, is directly recruited to the heat shock promoter by HSF upon heat shock. *Molecular cell* 2001;8:9-19.
- [112] Bhoite LT, Yu Y, Stillman DJ. The Swi5 activator recruits the Mediator complex to the HO promoter without RNA polymerase II. *Genes & Development* 2001;15:2457-69.
-

-
- [113] Bryant GO, Ptashne M. Independent recruitment in vivo by Gal4 of two complexes required for transcription. *Molecular cell* 2003;11:1301-9.
- [114] Takagi Y, Chadick JZ, Davis JA, Asturias FJ. Preponderance of free mediator in the yeast *Saccharomyces cerevisiae*. *The Journal of biological chemistry* 2005;280:31200-7.
- [115] Petrenko N, Jin Y, Wong KH, Struhl K. Evidence that Mediator is essential for Pol II transcription, but is not a required component of the preinitiation complex in vivo. *Elife* 2017;6.
- [116] Kuras L, Borggreffe T, Kornberg RD. Association of the Mediator complex with enhancers of active genes. *Proceedings of the National Academy of Sciences of the United States of America* 2003;100:13887-91.
- [117] Andrau JC, van de Pasch L, Lijnzaad P, Bijma T, Koerkamp MG, van de Peppel J, Werner M, Holstege FC. Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. *Molecular cell* 2006;22:179-92.
- [118] Fan X, Struhl K. Where does mediator bind in vivo? *PloS one* 2009;4:e5029.
- [119] Ansari SA, He Q, Morse RH. Mediator complex association with constitutively transcribed genes in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 2009;106:16734-9.
- [120] Jeronimo C, Robert F. Kin28 regulates the transient association of Mediator with core promoters. *Nature structural & molecular biology* 2014;21:449-55.
- [121] Wong KH, Jin Y, Struhl K. TFIIH phosphorylation of the Pol II CTD stimulates mediator dissociation from the preinitiation complex and promoter escape. *Molecular cell* 2014;54:601-12.
- [122] Grunberg S, Henikoff S, Hahn S, Zentner GE. Mediator binding to UASs is broadly uncoupled from transcription and cooperative with TFIID recruitment to promoters. *The EMBO journal* 2016;35:2435-46.
- [123] Paul E, Zhu ZI, Landsman D, Morse RH. Genome-wide association of mediator and RNA polymerase II in wild-type and mediator mutant yeast. *Molecular and cellular biology* 2015;35:331-42.
- [124] Poss ZC, Ebmeier CC, Taatjes DJ. The Mediator complex and transcription regulation. *Critical reviews in biochemistry and molecular biology* 2013;48:575-608.
- [125] Wang G, Cantin GT, Stevens JL, Berk AJ. Characterization of mediator complexes from HeLa cell nuclear extract. *Molecular and cellular biology* 2001;21:4604-13.

-
- [126] Tsai KL, Sato S, Tomomori-Sato C, Conaway RC, Conaway JW, Asturias FJ. A conserved Mediator-CDK8 kinase module association regulates Mediator-RNA polymerase II interaction. *Nature structural & molecular biology* 2013;20:611-9.
- [127] Petrenko N, Jin Y, Wong KH, Struhl K. Mediator Undergoes a Compositional Change during Transcriptional Activation. *Molecular cell* 2016;64:443-54.
- [128] Jeronimo C, Langelier MF, Bataille AR, Pascal JM, Pugh BF, Robert F. Tail and Kinase Modules Differently Regulate Core Mediator Recruitment and Function In Vivo. *Molecular cell* 2016;64:455-66.
- [129] Gilmour DS. Promoter proximal pausing on genes in metazoans. *Chromosoma* 2009;118:1-10.
- [130] Meyer KD, Lin SC, Bernecky C, Gao Y, Taatjes DJ. p53 activates transcription by directing structural shifts in Mediator. *Nature structural & molecular biology* 2010;17:753-60.
- [131] Knuesel MT, Taatjes DJ. Mediator and post-recruitment regulation of RNA polymerase II. *Transcription* 2011;2:28-31.
- [132] Guglielmi B, Soutourina J, Esnault C, Werner M. TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104:16062-7.
- [133] Nock A, Ascano JM, Barrero MJ, Malik S. Mediator-regulated transcription through the +1 nucleosome. *Molecular cell* 2012;48:837-48.
- [134] Lorch Y, Beve J, Gustafsson CM, Myers LC, Kornberg RD. Mediator-nucleosome interaction. *Molecular cell* 2000;6:197-201.
- [135] Zhu X, Zhang Y, Bjornsdottir G, Liu Z, Quan A, Costanzo M, Davila Lopez M, Westholm JO, Ronne H, Boone C, Gustafsson CM, Myers LC. Histone modifications influence mediator interactions with chromatin. *Nucleic acids research* 2011;39:8342-54.
- [136] Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 2010;467:430-5.
- [137] Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 2013;153:307-19.
- [138] Lai F, Orom UA, Cesaroni M, Beringer M, Taatjes DJ, Blobel GA, Shiekhata R. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* 2013;494:497-501.
-

-
- [139] Huang Y, Li W, Yao X, Lin QJ, Yin JW, Liang Y, Heiner M, Tian B, Hui J, Wang G. Mediator complex regulates alternative mRNA processing via the MED23 subunit. *Molecular cell* 2012;45:459-69.
- [140] Mukundan B, Ansari A. Novel role for mediator complex subunit Srb5/Med18 in termination of transcription. *The Journal of biological chemistry* 2011;286:37053-7.
- [141] Mukundan B, Ansari A. Srb5/Med18-mediated termination of transcription is dependent on gene looping. *The Journal of biological chemistry* 2013;288:11384-94.
- [142] Boubé M, Joulia L, Cribbs DL, Bourbon HM. Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell* 2002;110:143-51.
- [143] Linder T, Gustafsson CM. The Soh1/MED31 protein is an ancient component of *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* Mediator. *The Journal of biological chemistry* 2004;279:49455-9.
- [144] Bourbon HM. Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucleic acids research* 2008;36:3993-4008.
- [145] Soutourina J, Wydau S, Ambroise Y, Boschiero C, Werner M. Direct interaction of RNA polymerase II and mediator required for transcription in vivo. *Science* 2011;331:1451-4.
- [146] Borggreffe T, Yue X. Interactions between subunits of the Mediator complex with gene-specific transcription factors. *Seminars in cell & developmental biology* 2011;22:759-68.
- [147] Stiller JW, Hall BD. Evolution of the RNA polymerase II C-terminal domain. *Proceedings of the National Academy of Sciences* 2002;99:6091-6.
- [148] Furumoto T, Tanaka A, Ito M, Malik S, Hirose Y, Hanaoka F, Ohkuma Y. A kinase subunit of the human mediator complex, CDK8, positively regulates transcriptional activation. *Genes to cells : devoted to molecular & cellular mechanisms* 2007;12:119-32.
- [149] Park JE, Sun Y, Lim SK, Tam JP, Dekker M, Chen H, Sze SK. Dietary phytochemical PEITC restricts tumor development via modulation of epigenetic writers and erasers. *Sci Rep* 2017;7:40569.
- [150] Beck JR, Muralidharan V, Oksman A, Goldberg DE. PTEX component HSP101 mediates export of diverse malaria effectors into host erythrocytes. *Nature* 2014;511:592-5.
- [151] Muralidharan V, Oksman A, Pal P, Lindquist S, Goldberg DE. *Plasmodium falciparum* heat shock protein 110 stabilizes the asparagine repeat-rich parasite proteome during malarial fevers. *Nature communications* 2012;3:1310.
-

-
- [152] Epp C, Raskolnikov D, Deitsch KW. A regulatable transgene expression system for cultured *Plasmodium falciparum* parasites. *Malaria journal* 2008;7:86.
- [153] Wasserman MORaM. Stage-specific expression of the Calmodulin gene in *Plasmodium falciparum*. *J Biochem* 1995;118:6.
- [154] Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* 2005;4:1265-72.
- [155] Esnault C, Ghavi-Helm Y, Brun S, Soutourina J, Van Berkum N, Boschiero C, Holstege F, Werner M. Mediator-dependent recruitment of TFIID modules in preinitiation complex. *Molecular cell* 2008;31:337-46.
- [156] Youn DY, Xiaoli AM, Pessin JE, Yang F. Regulation of metabolism by the Mediator complex. *Biophys Rep* 2016;2:69-77.
- [157] Toth-Petroczy A, Oldfield CJ, Simon I, Takagi Y, Dunker AK, Uversky VN, Fuxreiter M. Malleable machines in transcription regulation: the mediator complex. *PLoS computational biology* 2008;4:e1000243.
- [158] Elemento O, Slonim N, Tavazoie S. A universal framework for regulatory element discovery across all genomes and data types. *Molecular cell* 2007;28:337-50.
- [159] Mehlin C, Boni E, Buckner FS, Engel L, Feist T, Gelb MH, Haji L, Kim D, Liu C, Mueller N, Myler PJ, Reddy JT, Sampson JN, Subramanian E, Van Voorhis WC, Worthey E, Zucker F, Hol WG. Heterologous expression of proteins from *Plasmodium falciparum*: results from 1000 genes. *Molecular and biochemical parasitology* 2006;148:144-60.
- [160] Fan E, Baker D, Fields S, Gelb MH, Buckner FS, Van Voorhis WC, Phizicky E, Dumont M, Mehlin C, Grayhack E, Sullivan M, Verlinde C, Detitta G, Meldrum DR, Merritt EA, Earnest T, Soltis M, Zucker F, Myler PJ, Schoenfeld L, Kim D, Worthey L, Lacount D, Vignali M, Li J, Mondal S, Massey A, Carroll B, Gulde S, Luft J, Desoto L, Holl M, Caruthers J, Bosch J, Robien M, Arakaki T, Holmes M, Le Trong I, Hol WG. Structural genomics of pathogenic protozoa: an overview. *Methods in molecular biology* 2008;426:497-513.
- [161] Birkholtz LM, Blatch G, Coetzer TL, Hoppe HC, Human E, Morris EJ, Ngcete Z, Oldfield L, Roth R, Shonhai A, Stephens L, Louw AI. Heterologous expression of plasmodial proteins for structural studies and functional annotation. *Malaria journal* 2008;7:197.
- [162] Palmer I, Wingfield PT. Preparation and extraction of insoluble (inclusion-body) proteins from *Escherichia coli*. *Current protocols in protein science / editorial board, John E Coligan [et al]* 2004;Chapter 6:Unit 6.3.
-

-
- [163] Singh SM, Panda AK. Solubilization and refolding of bacterial inclusion body proteins. *Journal of bioscience and bioengineering* 2005;99:303-10.
- [164] Webster WA, McFadden GI. From the genome to the phenome: tools to understand the basic biology of *Plasmodium falciparum*. *J Eukaryot Microbiol* 2014;61:655-71.
- [165] de Azevedo MF, Gilson PR, Gabriel HB, Simoes RF, Angrisano F, Baum J, Crabb BS, Wunderlich G. Systematic analysis of FKBP inducible degradation domain tagging strategies for the human malaria parasite *Plasmodium falciparum*. *PloS one* 2012;7:e40981.
- [166] Hallee S, Richard D. Evidence that the Malaria Parasite *Plasmodium falciparum* Putative Rhoptry Protein 2 Localizes to the Golgi Apparatus throughout the Erythrocytic Cycle. *PloS one* 2015;10:e0138626.
- [167] Fan HY, Klein HL. Characterization of mutations that suppress the temperature-sensitive growth of the *hpr1* delta mutant of *Saccharomyces cerevisiae*. *Genetics* 1994;137:945-56.
- [168] Risley MD, Clowes C, Yu M, Mitchell K, Hentges KE. The Mediator complex protein Med31 is required for embryonic growth and cell proliferation during mammalian development. *Dev Biol* 2010;342:146-56.
- [169] Zhu X, Wiren M, Sinha I, Rasmussen NN, Linder T, Holmberg S, Ekwall K, Gustafsson CM. Genome-wide occupancy profile of mediator and the Srb8-11 module reveals interactions with coding regions. *Molecular cell* 2006;22:169-78.
- [170] Teytelman L, Thurtle DM, Rine J, van Oudenaarden A. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proceedings of the National Academy of Sciences of the United States of America* 2013;110:18602-7.
- [171] Bunnik EM, Polishko A, Prudhomme J, Ponts N, Gill SS, Lonardi S, Le Roch KG. DNA-encoded nucleosome occupancy is associated with transcription levels in the human malaria parasite *Plasmodium falciparum*. *BMC genomics* 2014;15:347.
- [172] Lu XM, Bunnik EM, Pokhriyal N, Nasser S, Lonardi S, Le Roch KG. Analysis of nucleosome positioning landscapes enables gene discovery in the human malaria parasite *Plasmodium falciparum*. *BMC genomics* 2015;16:1005.
- [173] Adjalley SH, Chabbert CD, Klaus B, Pelechano V, Steinmetz LM. Landscape and Dynamics of Transcription Initiation in the Malaria Parasite *Plasmodium falciparum*. *Cell Rep* 2016;14:2463-75.
- [174] Anamika K, Gyenis A, Poidevin L, Poch O, Tora L. RNA polymerase II pausing downstream of core histone genes is different from genes producing polyadenylated transcripts. *PloS one* 2012;7:e38769.
-

- [175] Ding N, Zhou H, Esteve PO, Chin HG, Kim S, Xu X, Joseph SM, Friez MJ, Schwartz CE, Pradhan S, Boyer TG. Mediator links epigenetic silencing of neuronal gene expression with x-linked mental retardation. *Molecular cell* 2008;31:347-59.
- [176] Tsutsui T, Fukasawa R, Shinmyozu K, Nakagawa R, Tobe K, Tanaka A, Ohkuma Y. Mediator complex recruits epigenetic regulators via its two cyclin-dependent kinase subunits to repress transcription of immune-response genes. *Journal of Biological Chemistry* 2013.
- [177] Martin RE, Ginsburg H, Kirk K. Membrane transport proteins of the malaria parasite. *Molecular microbiology* 2009;74:519-28.
- [178] Nkrumah LJ, Riegelhaupt PM, Moura P, Johnson DJ, Patel J, Hayton K, Ferdig MT, Wellem TE, Akabas MH, Fidock DA. Probing the multifactorial basis of *Plasmodium falciparum* quinine resistance: evidence for a strain-specific contribution of the sodium-proton exchanger PfNHE. *Molecular and biochemical parasitology* 2009;165:122-31.
- [179] Mourier T, Pain A, Barrell B, Griffiths-Jones S. A selenocysteine tRNA and SECIS element in *Plasmodium falciparum*. *Rna* 2005;11:119-22.
- [180] Lobanov AV, Delgado C, Rahlfs S, Novoselov SV, Kryukov GV, Gromer S, Hatfield DL, Becker K, Gladyshev VN. The *Plasmodium* selenoproteome. *Nucleic acids research* 2006;34:496-505.
- [181] Schneider M, Hellerschmied D, Schubert T, Amlacher S, Vinayachandran V, Reja R, Pugh BF, Clausen T, Kohler A. The Nuclear Pore-Associated TREX-2 Complex Employs Mediator to Regulate Gene Expression. *Cell* 2015;162:1016-28.
- [182] Kubik S, Bruzzone MJ, Shore D. TFIID or not TFIID, a continuing transcriptional SAGA. *The EMBO journal* 2017;36:248-9.
- [183] Liu X, Vorontchikhina M, Wang YL, Faiola F, Martinez E. STAGA Recruits Mediator to the MYC Oncoprotein To Stimulate Transcription and Cell Proliferation. *Molecular and cellular biology* 2008;28:108-21.
- [184] Ansari SA, Ganapathi M, Benschop JJ, Holstege FC, Wade JT, Morse RH. Distinct role of Mediator tail module in regulation of SAGA-dependent, TATA-containing genes in yeast. *The EMBO journal* 2012;31:44-57.
- [185] Chen XF, Lehmann L, Lin JJ, Vashisht A, Schmidt R, Ferrari R, Huang C, McKee R, Mosley A, Plath K, Kurdistani SK, Wohlschlegel J, Carey M. Mediator and SAGA have distinct roles in Pol II preinitiation complex assembly and function. *Cell Rep* 2012;2:1061-7.
- [186] Fan Q, An L, Cui L. *Plasmodium falciparum* histone acetyltransferase, a yeast GCN5 homologue involved in chromatin remodeling. *Eukaryotic cell* 2004;3:264-76.

-
- [187] Fan Q, An L, Cui L. PfADA2, a Plasmodium falciparum homologue of the transcriptional coactivator ADA2 and its in vivo association with the histone acetyltransferase PfGCN5. *Gene* 2004;336:251-61.
- [188] Guglielmi B, van Berkum NL, Klapholz B, Bijma T, Boube M, Boschiero C, Bourbon HM, Holstege FC, Werner M. A high resolution protein interaction map of the yeast Mediator complex. *Nucleic acids research* 2004;32:5379-91.
- [189] Mauldin JP, Lu M, Das S, Park D, Ernst PB, Ravichandran KS. A link between the cytoplasmic engulfment protein Elmo1 and the Mediator complex subunit Med31. *Current biology : CB* 2013;23:162-7.
- [190] East MP, Bowzard JB, Dacks JB, Kahn RA. ELMO Domains, Evolutionary and Functional Characterization of a Novel GTPase-activating Protein (GAP) Domain for Arf Protein Family GTPases. *Journal of Biological Chemistry* 2012;287:39538-53.
- [191] Kim TW, Kwon Y-J, Kim JM, Song Y-H, Kim SN, Kim Y-J. MED16 and MED23 of Mediator are coactivators of lipopolysaccharide- and heat-shock-induced transcriptional activators. *Proceedings of the National Academy of Sciences of the United States of America* 2004;101:12153-8.
- [192] Pesce ER, Acharya P, Tatu U, Nicoll WS, Shonhai A, Hoppe HC, Blatch GL. The Plasmodium falciparum heat shock protein 40, Pfj4, associates with heat shock protein 70 and shows similar heat induction and localisation patterns. *Int J Biochem Cell Biol* 2008;40:2914-26.
- [193] Lariviere L, Geiger S, Hoepfner S, Rother S, Strasser K, Cramer P. Structure and TBP binding of the Mediator head subcomplex Med8-Med18-Med20. *Nature structural & molecular biology* 2006;13:895-901.
- [194] Cevher MA, Shi Y, Li D, Chait BT, Malik S, Roeder RG. Reconstitution of active human core Mediator complex reveals a critical role of the MED14 subunit. *Nature structural & molecular biology* 2014;21:1028-34.
- [195] Luo Z, Lin C, Shilatifard A. The super elongation complex (SEC) family in transcriptional control. *Nat Rev Mol Cell Biol* 2012;13:543-7.
- [196] Takahashi H, Parmely Tari J, Sato S, Tomomori-Sato C, Banks Charles AS, Kong Stephanie E, Szutorisz H, Swanson Selene K, Martin-Brown S, Washburn Michael P, Florens L, Seidel Chris W, Lin C, Smith Edwin R, Shilatifard A, Conaway Ronald C, Conaway Joan W. Human Mediator Subunit MED26 Functions as a Docking Site for Transcription Elongation Factors. *Cell*;146:92-104.
- [197] Backstrom S, Elfving N, Nilsson R, Wingsle G, Bjorklund S. Purification of a plant mediator from Arabidopsis thaliana identifies PFT1 as the Med25 subunit. *Molecular cell* 2007;26:717-29.
-

- [198] Dolan WL, Chapple C. Conservation and Divergence of Mediator Structure and Function: Insights from Plants. *Plant & cell physiology* 2017;58:4-21.
- [199] Levine M, Tjian R. Transcription regulation and animal diversity. *Nature* 2003;424:147-51.
- [200] Philip N, Waters AP. Conditional Degradation of Plasmodium Calcineurin Reveals Functions in Parasite Colonization of both Host and Vector. *Cell host & microbe* 2015;18:122-31.
- [201] Kreidenweiss A, Hopkins AV, Mordmuller B. 2A and the auxin-based degron system facilitate control of protein levels in Plasmodium falciparum. *PloS one* 2013;8:e78661.