International Journal of Biomedical and Health Sciences
Vol. 9, No. 3, September 30, 2013
Printed in Nigeria

0794-4748/2013 $5.00 + 0.00

http://www.asopah.org

IJBHS 2012153/9303

Control of Gene Expression in *Plasmodium*§

Mauro Ferreira de Azevedo[1] and Hernando A. del Portillo[2].

[1]Departamento de Parasitologia, Instituto de Ciências Biomédicas, Universidade de São Paulo. Av. Lineu Prestes 1374, São Paulo, SP 05508-900, Brazil
[2]Present Address: Barcelona Centre for International Health Research (CRESIB), Hospital Clinic/IDIBAPS, Universitat de Barcelona, Roselló 132, 4a planta, 08036, Barcelona, Spain. Phone: 34 93 2275706; Fax: 34 93 4515272 (hernandoa.delportillo@cresib.cat)

Introduction

Malaria parasites have more than 10 stages of cellular differentiation and invade at least four types of cells in two different hosts with a considerable variation in temperature between them. All of this complex biology depends on the efficient control of gene expression, about which our knowledge still has many shortcomings. Although this parasite has some general mechanisms in common with yeast and higher eukaryotes, many aspects of its genetic regulation seem to be specific to this genus: (i) during the asexual blood stages, the parasites seem to turn on a rigid, viral-like program of early, middle, and late genes expressed as a cascade of continuous events; (ii) it seems likely that malaria parasites have acquired unique and yet-to-be-described transcription factors; (iii) antisense transcription has been described in about 10% of the coding genome, clearly indicating as-yet-undefined, post-transcriptional control mechanisms; and (iv) control gene expression of the var subtelomeric multigene family involves a gene-specific cross-talk between intron and exon, as well as epigenetic mechanisms to control allelic exclusion. Here, we review our present knowledge on control of gene expression in malaria parasites and illustrate the importance of bioinformatics in advancing our knowledge in this area, with illustrative examples on promoters, transcription factors, and the transcriptome analysis of the intraerythrocytic developmental cycle.

General Aspects

Malaria is the most important parasitic disease in the world. Each year, 300 to 500 million people are infected, and more than 1 million people, mostly children under 5 years of age, die. More than 40% of the world's population live in high-risk areas including sub-Saharan Africa, South and South-East Asia, Indonesia, and South and Central America (Figure 1) (1).

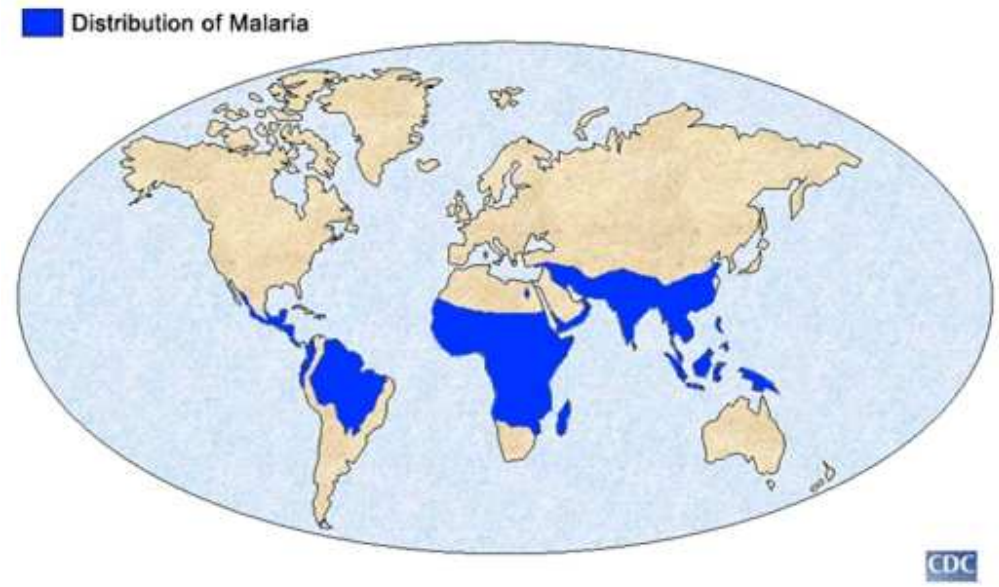

Figure 1**: Geographical distribution of human malaria worldwide. Obtained from CDC http://www.cdc.gov/malaria/distribution_epi/distribution.htm.**

The causative agent is a parasite of the phylum Apicomplexa, genus *Plasmodium*. Although there are more than 100 species of this genus, only *P. falciparum, P. malariae, P. ovale,* and *P. vivax* infect humans. More than 99% of infections are caused by *P. falciparum* and *P. vivax.*
The geographic distribution of these two malaria species is very different. *P. falciparum* is the species most prevalent in sub-Saharan Africa, where the highest number of deaths occurs. *P. vivax* is widely distributed and is the most frequently found species in the majority of regions, except Africa. A genetic characteristic of the Negro population is one of the reasons that *P. vivax* is less prevalent in Africa. A mutation in the gene encoding the "Duffy" antigen confers the Duffy-negative phenotype to this population. *P. vivax* depends on this receptor to invade the erythrocyte and is incapable of infecting Duffy-negative red blood cells (2, 3). *P. falciparum*, however, has alternative mechanisms for invasion, independent of binding to the "Duffy" protein (4).

Life Cycle

The life cycles of the different malaria species are very complex (Figure 2). An infected female *Anopheles* mosquito bites the vertebrate host, injecting parasite forms, known as sporozoites, into the dermis. Intense movement in the dermis allows some of the parasites to encounter blood and lymphatic vessels, thereby causing a decrease in movement. The parasites that enter the lymphatic system invade the closest lymph node and transform into a stage similar to the hepatic stage. However, they are unlikely to complete the process of division or reach cell maturity.

This extraerythrocytic stage was described recently by Amino *et al.* (2006) (5), but its importance in establishing infection and the hosts' immune response is still unknown. Parasites that enter the blood vessels are carried in the blood and reach the hepatocytes, invading these cells. Here, they are modified to a more rounded form, which initiates the process of asexual reproduction, forming the hepatic schizont, which now contains thousands of merozoites that are liberated into the host circulation. In *P. vivax*, after the sporozoites have invaded the hepatocytes, they can differentiate into a dormant stage called a hypnozoite. These can lie dormant for months and re-initiate replication when the original disease has already been cured, leading to another infection, known as relapse. In this way, patients cured of the disease can develop a new infection. Independent of this particular life stage of *P. vivax*, once merozoites are released into the bloodstream, they quickly invade red blood cells, initiating the asexual intraerythrocytic cycle, which causes the disease pathology.

The parasite forms a parasitophorous vacuole, where it develops into the first stage, known as the ring stage. This feeds on the contents of the red blood cell and transforms into a trophozoite. This stage begins the process of cell division known as schizogony and transforms into the schizont. This contains the new merozoites that rupture the red blood cell and are released into the bloodstream to infect new cells, continuing the disease cycle. Some parasites do not begin the process of cell division and instead, differentiate into the sexual form of the parasite, the male and female gametocytes that will be ingested by a mosquito during its blood meal. Once in the digestive tract of the insect, the red blood cells are digested, releasing the male and female gametes. The male gametes (microgametes) undergo three rounds of mitosis, giving rise to eight cells, a phenomenon know as exflagellation, and are transformed into individual male gametes. Fertilization of the gametes occurs forming a zygote, the only diploid stage of the parasite. This differentiates into the ookinete , an amoeboid form able to cross the peritrophic membrane, and lodges in the internal wall of the digestive tract between the epithelium and the basal lamina. It forms a protective envelope and transforms into an oocyst. This undergoes the first reducing meiosis, followed by many stages of mitosis, releasing haploid sporozoites into the haemolymph of the insect. Sporozoites migrate, and some of them actively penetrate the salivary gland of the mosquito until a new blood meal is done, and another individual may then be infected.

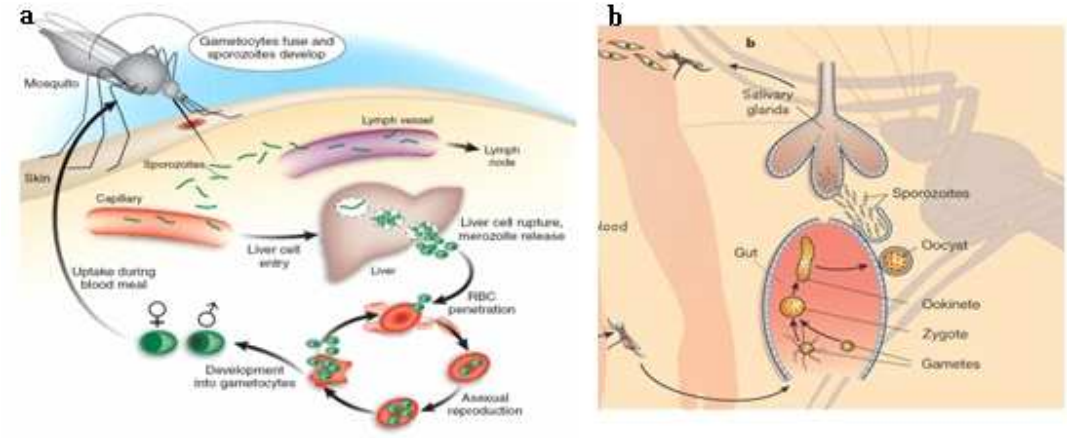

Figure 2: Life cycle. a. The life cycle in the vertebrate host, demonstrating the modifications proposed by Amino *et al.* (2006) (5). Extracted from Jones *et al.* (2006) (104). b. Life cycle in the mosquito. (http://www.sanger.ac.uk/PostGenomics/plasmodium/presentations/plasmodium_lifecycle.shtml)

Plasmodium falciparum* and *Plasmodium vivax

P. falciparum and *P. vivax* are the most important human malaria species. The pathology caused by infection by each of these parasites is very distinct. Although *P. vivax* infects only reticulocytes, *P. falciparum* infects young and mature erythrocytes indiscriminately, producing much higher parasitemias. *P. falciparum* has a cytoadhering phenotype where mature forms of the parasite adhere to the endothelium, escaping clearance by the spleen. Because of this phenotype, mostly ring stages are observed on examination of peripheral blood. *P. vivax* does not have a similar phenotype, and it is unknown how clearance by the spleen is avoided. Recently, it has been speculated that *P. vivax* parasites cytoadhere to the barrier cells of the spleen, escaping spleen macrophage clearance and installing themselves into a reticulocyte-rich area (6). These characteristics mean that *P. falciparum* malaria is much more lethal than *P. vivax*, which rarely kills its host but is also a chronic and debilitating disease.

The types of study that can be carried out using these two parasites have different limitations. *P. falciparum* can be continuously cultured easily *in vitro*, allowing application of methods such as: *in vitro* drug testing; inhibition of invasion using specific antibody serum; and transfection and genetic manipulation. All of these methods are severely limited in *P. vivax* because of the absence of a simple, reproducible method of *in vitro* continuous culture. In fact, this limitation is reflected in the large amount of data that have been gathered from the genome, transcriptome, and proteome of *P. falciparum* as compared with *P. vivax*.

Genome

In the last two decades, considerable advances in the techniques of DNA sequencing and partial or complete sequences of hundreds of genomes have been published (http://www.ncbi.nih.gov/entrez/query.fcgi?db=genomeprj). Although the genome probably contains most of the information necessary for the development of an organism, the function and regulation of expression of predicted genes are not immediately apparent after publication of DNA sequence data. This information requires much more time and resources than the initial genome projects and also depends on advances in post-genome technology, such as bioinformatics and reverse genetics, that permit the formation and functional testing of models for biological problems.

The study of malaria has benefited from the advance of genome projects, *Plasmodium* spp. being one of the parasites with the greatest number of species and strains whose genomes have been or are being sequenced. The genomes of *P. falciparum* and *P. yoelli* (a species that infects rodents) have been sequenced, and their first drafts were published 4 years ago (7, 8). Partial sequences of the *P. chabaudi* and *P. berghei* genomes, and comparison with the other two genomes of the genus, were published recently (9). The genome of *P. vivax* is in the final stages of annotation and should be published in the Spring/Summer of 2007 (http://www.tigr.org/tdb/e2k1/pva1/). As well as these, total or partial sequencing projects of the genomes of *P. knowlesi, P. gallinaceum*, and *P. reichenowi* are under way. The sequencing data and annotation of these genomes, as well as information about gene expression and protein interactions, are available in databases such as GenBank, GeneDB, and PlasmoDB, the last being the official database of the *Plasmodium* spp. genomes.

The *P. falciparum* genome project was undertaken as a consortium of the sequencing centers The Institute for Genomic Research (TIGR), the Sanger Center, and Stanford University. The strategy used involved separating the 14 chromosomes by pulsed field electrophoresis and then sequencing each one using the "shotgun" method. This entails cloning smaller parts of the chromosomes into plasmid vectors and then sequencing these clones using vector-specific oligonucleotides. In some cases, larger clones, constructed by using yeast artificial chromosomes (YACs), were also used to help with mapping. Chromosomes 1, 3, 4, 5-9, and 13 were sequenced by the Sanger Center; chromosomes 2, 10, 11, and 14 by TIGR; and chromosome 12 by Stanford University.

The haploid genome of *P. falciparum* has more than 23 million base pairs (Mb) and a total AT content of 80.62%. The AT content in the coding regions is 76.23%, increasing to almost 90% in introns and intergenic regions. At the last data update, 5,440 genes had been annotated, 5,374 of which encode proteins. Of these, more than 50% contain introns, and 63% were annotated as hypothetical proteins without known function (http://www.sanger.ac.uk/Projects/P_falciparum/genome.overview.shtml). Of the RNA genes, 45 transfer RNA (tRNA) and 20 ribosomal (rRNA) genes were identified (7).

The genome of *P. vivax* has yet to be published, but preliminary data indicate an AT content of about 55% and a total size of about 30 Mbp. The subtelomeric regions—where many of the genes are involved in virulence and/or antigenic variation, such as the *vir* gene family (10), are found—have a higher AT content of about 80%. In the central chromosome regions, synteny with the *P. falciparum* genome is high, as is the similarity between the genes of the two species.

Analysis of the genomes of the two rodent species of malaria, *P. berghei* and *P. chabaudi*, together with the data published previously for *P. yoelli* and *P. falciparum*, demonstrates that about 80% of the rodent malaria genes have orthologs in *P. falciparum*, possibly forming a group of universal plasmodial genes (9). Most of these genes are localized in the central regions of the chromosomes, with the subtelomeric regions being much more diverse.

The data reviewed above show that the genome sequences of the most important malaria parasites are readily available for mining and comparisons. Of general interest, comparisons between syntenic chromosome regions of some species of *Plasmodium* can be performed using tools available at the new PlasmoDB site, and several programs such as Artemis allow genome visualizations and annotations.

Transcriptome

Some years before the *P. falciparum* genome was published, the first large-scale studies of the transcriptome were begun using "microarray" technology. In spite of working with limited data from the pre-genomic era, these studies were capable of identifying many differences in abundance of transcription between trophozoites and gametocytes (11) and between the asexual intraerythrocytic stages (12).

As the sequencing data available for *Plasmodium* spp. increased, new studies on transcriptomes appeared in much greater scale. The transcriptome, analyzed by microarray for 48 hours during the intraerythrocytic life cycle of *P. falciparum*, with resolution of 1 hour, revealed that the parasite has ridged regulation, which seems unique in eukaryotes (13). Indeed, most of the genes are transcribed during this phase of the life cycle, and about three-quarters of these are expressed only once (see below). Another study (14) analyzed the transcriptome of six stages of the intraerythrocytic cycle, as well as merozoites, sporozoites, and gametocytes, using

microarray. By using methods with less temporal resolution but analyzing transcripts from a greater variety of stages, this study found that less than one-half the parasites' genes (49%) are regulated at the transcription level.

Analysis of transcripts from less abundant parasite stages, such as sporozoites and hepatic forms, has also been undertaken. Subtractive cDNA libraries, obtained from sporozoite and merozoite RNA, were used to identify 25 sporozoite-specific transcripts possibly involved in invasion (15). Preliminary data from transcriptomes of hepatic stages of the parasite indicate an absence of specific transcripts and a change from the expression profile of sporozoites to that of blood-stage forms (16-18). The use of different methodology to isolate the parasites in these three studies leads to results that are difficult to compare. This may reflect on the presence of different profiles, depending on the time of infection of the hepatocyte and consequent maturity of the parasites.

An increase in the sensitivity of the microarray technique has permitted analysis of the transcriptome of parasites isolated from patients. In these studies, it was demonstrated that the pattern expression of most genes is equivalent to parasites maintained in culture. However, some genes behave differently, perhaps reflecting differences in the biology of the parasites in different conditions. Among these are the genes that encode surface antigens and proteins involved in sexual differentiation that are overexpressed *in vivo* (19, 20). It is possible that these differences in expression represent a greater need for the parasite to escape the host immune response and that *in vivo* conditions favor greater production of gametocytes for infecting mosquitoes.

The advent of microarray technology thus faces a computational challenge to analyze this large amount of data because, for instance, the single experiment on the transcriptome of the intraerythrocytic developmental cycle (IDC) at 1-hour resolution generated close to 350,000 data. Indeed, many computational, statistical, and bioinformatics algorithms have been developed to analyze microarray data (21).

Proteome

An increase in the sensitivity of mass spectrometry techniques and improving methods for cultivation and purification of parasites have allowed considerable advances in studies of the *Plasmodium* spp. proteome. Data from the *P. falciparum* proteome (22) and *P. berghei* (9) confirm most of the microarray data from Bozdech *et al.* (2003) (13). More than one-half of the proteins in the parasite are expressed in only one stage, and for the genes of known function, there is a high correlation between the genes expressed and metabolic and cellular activities of the parasite at that stage. Some stage-specific genes belong to multigene families whose members are expressed at different stages in a strategically specific manner, such as involvement in invasion, indicating that the parasite might use a limited repertoire of strategies to survive.

Genes that were thought to be specific to the intraerythrocytic cycle, such as the genes *var*, *rifin*, *stevor*, and the super family *pir*, which include the *vir* genes of *P. vivax*, are transcribed and often translated in parasite stages in the mosquito (9, 22). It is possible that some members of these families have an important function for survival of these stages, or that the control of expression is simply more relaxed.

The proteome of less abundant stages, such as the gametocytes and sporozoites, has also been determined, demonstrating many differences between the repertoire of proteins in these stages and the asexual blood forms (23, 24). Purification of male and female gametocytes was

performed by fluorescence activated cell sorting (FACS) of parasites that express the *gfp* gene controlled by active promoters in each of the stages (25). The data produced in this study demonstrate a very distinct repertoire in the gametocytes compared with the other parasite stages. The male gametocytes have the most specialized proteome, perhaps reflecting the intense differentiation process that occurs during exflagellation.

Organelles specific to this phylum, such as rhoptries, are essential for invasion. Rhoptries have been purified, and the proteins have been identified by mass spectrometry. A study by Bradley *et al.* (2005) (26) identified some of the proteins present in the rhoptries of *T. gondii*. They have orthologous genes in *P. falciparum*, indicating the possibility of common mechanisms of invasion. A similar study was done on *P. falciparum* (27).

The data reviewed above indicate that malaria parasites are able to express mostly non-redundant protein repertoires at each particular stage, reinforcing the idea of a tight control of gene expression. On the other hand, data from the transcriptome of the IDC at 1-hour resolution indicates a simple program for controlling gene expression during the intraerythroctic developmental cycle where more than 75% of the genes of the parasites are transcribed. Understanding how the parasite controls gene expression at each differentiation stage is essential to the development of alternative strategies and might shed light into unique mechanisms acquired during the evolution of malaria parasites.

Control of Gene Expression

Malaria parasites have more than 10 stages of cellular differentiation, precisely regulate differentiation and replication, and invade at least four types of cells in two different hosts with a considerable variation in temperature between them. All of this complex biology depends on an efficient control of gene expression—and what we know of gene expression control still has many shortcomings. Although this parasite has some general mechanisms in common with yeast and higher eukaryotes, many aspects of their genetic regulation seem to be specific to this genus.

Chromatin

In nucleated organisms, DNA is compartmentalized in the nucleus, and this provides the first barrier to gene activation. Transcription factors are produced in the cytoplasm and must enter the nucleus to exercise their functions. Inside the nucleus, the DNA is packed and compartmentalized into chromatin that has regions with different levels of packaging. Heterochromatin is highly compact, and thus it is difficult to transcribe genes in these regions. In contrast, euchromatin is more relaxed, and genes are more easily transcribed. There is movement in the compaction of chromatin, allowing genes in a region of heterochromatin to be changed to a region of euchromatin and activated. One of the first steps in activation of transcription is the unpackaging of the DNA, allowing access of the proteins needed to initiate transcription. This process frequently involves the acetylation of histones in the promoter region, making them negatively charged and decreasing their affinity for the DNA..

In *Plasmodium*, the chromatin DNA is structured in the form of nucleosomes of approximately 155 bp (28, 29). The genes that code for each of the structural histones (H2A, H2B, H3, and H4) and also for four histone variants are present in the genome (30). The gene that codes for histone H1 has never been found in any of the Apicomplexa organisms, indicating that they may have a different mechanism from other eukaryotes for organizing their

nucleosomes (31). It has also been observed that this protein is very divergent among other protozoans and higher eukaryotes, with the absence of the amino-terminal and globular domains in the more primitive organisms (32). It is possible that the Apicomplexa H1 histone has a very distinct structure, making it difficult to be predicted. Proteins responsible for nuclesome assembly have also been identified (33).

It has been observed that many parasite promoters are regulated in a similar manner in the chromosomal and episomal context. This might indicate that the organization of chromatin does not have a great influence on gene regulation. However, Horrocks *et al.* (2002) (34) demonstrated that episomal plasmids are organized in the form of nucleosomes and that the structure and maintainance of the pattern of gene regulation of the promoters depend on the parasites passing through the S-phase of mitosis. More evidence, such as the presence of proteins involved in the acetylation of histones (such as the ADA2-GCN5 complex, which is conserved in eukaryotes), suggests an important role for the regulation of the DNA structure in the parasite (35). The deacetylation of histones represses transcription. Orthologs to histone deacytelases such as HDAC (36) and Sir2 (37), responsible for repression of gene expression in yeast, have been identified in *P. falciparum*. Sir2 appears to be involved in the formation of heterochromatin and silencing of some types of *var* genes (38, 39).

Promoters

The promoters of *Plasmodium* spp. have a bi-partite structure with a core promoter followed by regulatory elements (40, 41). Transcription is monocystronic (42), and the genes that encode proteins are transcribed by RNA polymerase II (43). Data from the transcription analyses demonstrate the great importance of transcriptional control. Many promoters characterized by transient transfection experiments express the reporter gene with the same temporal pattern as the endogenous genes. Some, such as the dihydrofolate reductase (*dhfr*) and the elongation factor alpha one (*ef*1-α) promoters, are active during all of the intraerythrocytic cycle. Others are active during specific stages, such as merozoite surface protein 2 (*msp*2), apical membrane antigen 1 (*ama*1), *Pfs*16 and *Pfs*25 that are active during early schizogony, in mature schizonts, in gametocytes, and in gametes, respectively (44-46).

Plasmodium spp. promoters have distinct characteristics, because the transcription machinery of the parasite does not recognize the promoters of other eukaryotes. Yet, the mechanism of transcriptional regulation appears to be conserved within the genus because promoters from different species are functional in heterologous transfection experiments (41, 47-49). Unexpectedly, however, a recent report on the functional analysis of *P. vivax* promoters demonstrated that unlike all other *Plasmodium* spp., promoters of *P. vivax* are poorly or not recognized by the *P. falciparum* transcription machinery (50). These results suggest the existence of *P. vivax*-specific transcription regulatory elements.

Binding sites for various eukaryotic transcription factors are found in the sequence of different *Plasmodium* spp. promoters. Functional analyses detecting DNA–protein interactions, either by transient transfection experiments with reporter genes or by electrophoretic mobility shift assays (EMSAs), were unable to prove whether these sites were recognized by parasite proteins. On the contrary, it was observed that the important sequences involved in promoter activation do not demonstrate similarities with the motifs involved in the regulation of other organisms (46, 51-57). A few functional motifs shared with other eukaryotes have been characterized, such as the “TATA box” (58) responsible for recruitment of the core transcription

complex and the long adenine- and thymidine-rich sequences (poly (dT) poly (dA)) (59) that change the conformation of chromatin, making it more accessible to transcription factors.
The data reviewed above clearly demonstrate the scarce knowledge on promoters gathered in malaria through functional analysis, mainly because of the large technical difficulties of such analyses. Bioinformatics approaches are now accelerating discovery of this key aspect of control of gene expression in malaria parasites (see "Post-Transcriptional Control", below).

Basal Transcription Complex

Initiation of transcription is much more complex in eukaryotes than in bacteria. Instead of a single type of RNA polymerase, able to recognize promoters via the sigma subunit, three different classes of polymerase transcribe different types of genes. RNA polymerase I transcribes ribosomal RNA genes, type II transcribes genes encoding proteins, and type III transcribes transfer RNAs and other small RNAs. These enzymes act as a complex of 8-14 subunits of up to 500 kilodaltons (kD). Recognition of promoter regions is not carried out directly by RNA polymerase but depends on the help of specific proteins, known as transcription factors. These recognize specific sites and bind to DNA and, together with the RNA polymerase, form the core complex for transcription.

P. falciparum and other species from the genus contain nuclear genes that code for polypeptide chains of the three types of polymerase. The largest of these polypeptides has variable regions longer than those found in other eukaryotes (60) and contains a repeat motif rich in serine (61). The carboxyl-terminal domain (CTD) of polymerase II has a very divergent amino-acid sequence compared with other eukaryotes. However, a structure similar to the hepta-peptide repeat, SPTSPSY, is present, indicating that the conserved mechanism of phosphorylation of this region is used for initiation of transcription (62, 63). The mitochondrial RNA polymerase has similarities with the RNA polymerase of the bacteriophage T3/T7 (64) and is encoded by nuclear genes in many organisms. The parasite also contains another RNA polymerase with bacterial characteristics, that of the Apicoplast . The gene encoding one of its polypeptide chains is found in the genome of this organelle (65).

For each of the RNA polymerases to be able to transcribe a gene, it is necessary that transcription factors bind to the promoter region of the DNA. Of the three types of RNA polymerase, pol II needs the greatest number of transcription factors. In the region of the promoter closest to the site of initiation of transcription, the core promoter elements are found, such as the “TATA box”, the “CAAT box”, transcription initiators (INR), and some others. The *trans* elements of the core transcription complex bind to this region of the promoter. These elements are composed principally of TFIID, TFIIA, TFIIB, TFIIF, RNA polymerase, TFIIE, TFIIH, and TFIIJ. Of note, malaria parasites seem to contain much lower transcription factors than crown eukaryotes, including factors associated with the basal transcription complex, which has been mostly conserved throughout evolution (see below).

Capping

After initiation of transcription, when the recently transcribed RNA reaches a size of about 20-30 nucleotides, it undergoes its first modification, known as “capping”. This process occurs using three basic reactions consisting of: 1) hydrolysis of the triphosphate at the 5' extremity by a RNA 5' triphosphatase, leaving a diphosphate; 2) addition of a guanidine monophosphate group

(GMP) in a 5'-to-3' direction to this extremity by the action of a guanyltransferase; and 3) methylation of the last guanine at position N7 by a methyltransferase.

The mechanism of "capping" is different in animals and yeast. Animals have the triphosphatase and the guanyltransferase encoded by a single gene, giving rise to a bifunctional protein, whereas in yeast, three different proteins encoded by individual genes are used. The "capping" is related to protection and stability of the RNA, transport to the cytoplasm, and translation. In *Plasmodium* spp. and other protozoa such as *Trypanosoma brucei* and *Giardia lamblia*, the "capping" mechanism is similar to that of yeast, with the first two enzymes being encoded by different genes (66, 67). However, the methyltransferase has not yet been characterized and may have distinct characteristics from the groups already studied.

Splicing

At the end of transcription, and in some cases, before this, the messenger RNA (mRNA) should have its introns removed by "splicing", leaving an open reading frame (ORF) for translation. In many eukaryotes, this is an important regulation point. During the splicing of some genes, part of the exons can be removed with the introns, modifying the reading frame and allowing translation of different isoforms of the protein. This is a regulated process, and the transcripts can be processed in a stage- or tissue-specific form, depending on the needs of the cell.

In *P. falciparum*, this mechanism of regulation appears to be present. Knapp *et al*. (1991) (68) demonstrated that a mRNA from a gene with an unknown function is processed into at least three different transcripts. Species of this genus appear to have certain conserved loci in relation to gene organization and structure, including the number and position of introns and splicing alternatives. This indicates the presence of common control mechanisms (69, 70). In some cases, the alternative splicing may not alter the reading frame. This occurs in the B7 gene, where transcript processing in the asexual and sexual blood stages gives rise to 5' untranslated regions (5'UTR) of different sizes (71). It is possible that this influences the efficiency of translation or the mRNA stability. Although these few examples demonstrate that alternative splicing occurs in *Plasmodium*, the extent of such phenomenon in controlling gene expression in the different developmental stages remains to be determined.

Termination

The end of transcription in eukaryotes depends on the interaction of various proteins with the recently synthesized mRNA and polymerase II and is accompanied by a second modification in the mRNA, polyadenylation. In the 3' untranslated regions (3'UTR) of the genes, conserved sequences are used as binding sites for a protein complex that cleaves the RNA, 10-30 nucleotides from this site, and adds a ribonucleotide tail of adenines (poly A). It is not yet known how the polymerase, which continues to transcribe the gene, stops transcription and is released from the DNA. Recently, a model was proposed where the yeast protein Rat1 and its cofactor Rai1 bind directly to the 5' extremity of the RNA that continues being transcribed after cleavage of the mRNA and is degraded by the activity of 5'-3' exonuclease until it meets with the RNA polymerase, which, by an unknown method, releases the DNA (72).

As in many eukaryotes, the end of transcription in *Plasmodium* spp. occurs at multiple sites, and the poly A tail is added by cleavage and polyadenylation (73). The elements responsible for

this process are not well characterized. Canonical sites of polyadenylation of the type ATTAAA and AATAAA are found in multiple copies in the 3'UTR of various genes, but it is unknown whether these are recognized by the transcription machinery of the parasite in the same manner as other eukaryotes. In some cases, it is probable that the AATAA pentamer is recognized as the site of polyadenylation, which can occur 1-30 nucleotides downstream of this site (74).

For the *P. gallinaceum* gene that encodes for the gametocyte antigen Pgs28, it was demonstrated that the poly A tail is added around 20 nucleotides after the fifth canonic polyadenylation site (AATAAA/ATTAAA) and that, before this, there is an important U-rich element (75). This 3'UTR was used in transient transfection experiments in human cells, being recognized by our own transcription machinery, but in a different manner than that occurring in the parasite in relation to the polyadenylation site and the important *cis* elements for this process (76). Transient transfection experiments demonstrated that the presence of a 3'UTR from the parasite genus is necessary for expression of the reporter genes. Distinct mechanisms for regulation of the termination of transcription and processing of mRNA must be involved, because the parasite does not recognize 3'UTRs from other eukaryotes (77, 78).

Transcription Factors

In comparison to other organisms, *Plasmodium* spp. appears to have a distinct and perhaps more limited repertoire of proteins involved in gene regulation. Thus, a recent study, undertaken by Coulson *et al.* (2004) (79), identified fewer transcription factors than those of crown eukaryotes using different algorithms and bioinformatics (see below). The initial annotation of the genome did not identify certain components of the core transcription complex, such as TFIIA, the beta subunits of TFIIE and TFIIF, and other factors associated with TFIID and TFIIH. On the basis of these data and analysis of the *T. gondii* genome, it was initially proposed that the core transcription complex in Apicomplexa organisms is simpler than that of other eukaryote groups (31). Improvements in bioinformatic techniques have since then led to the identification of more factors (80) and proteins with complex DNA-binding domains. These data indicate that *P. falciparum*, and perhaps other Apicomplexa, have more genes involved in regulation of transcription than was previously predicted (81). The difficulty in annotating transcription factors in *Plasmodium* spp. is probably attributable to divergence in the sequence of its domains compared with known groups.

Of the proteins that form the core transcription complex, only the “TATA binding protein” (TBP) has been functionally characterized and has an amino acid sequence very divergent from the TBP of other organisms. However, the structure is similar and is capable of binding to the conserved sequences TATAA and TGTAA (58, 82). Another *trans* element that has been characterized is a protein relating to the Myb factors. This is regulated transcriptionally during the intraerythrocytic cycle and can bind conserved *cis* elements present in different promoters of the parasite (83).

The low number of transcription factors present in the parasite is contrasted by the high numbers of *cis* elements present in each promoter (84). In yeast for example, most of the promoters are regulated by one or two *cis* elements. Considering that the combination of these two factors (*cis* and *trans*) determines the diversity of regulation in the cell, it is possible that this is maintained in the parasite using a few transcription factors that bind in combination to various *cis* elements in the promoters.

Post-Transcriptional Control

The 3'UTR has a role in the regulation of gene expression in many eukaryotes. In animals, they determine the temporal expression of numerous genes, particularly during embryonic development. Their secondary structure and interaction with other proteins influence the processing, transport, stability, and efficiency of translation of the transcript. In *Plasmodium* spp., a great deal of evidence indicates the existence of important post-transcriptional control, particularly during the passage from the vertebrate host cycle to its development in the mosquito.

The first evidence of post-transcriptional control came from studies of the *P. berghei* gene *Pbs21*, which encodes a protein localized in the membrane of female gametocytes, zygotes, and oocytes. As well as being detected in these parasite stages, transcripts of this gene are also present in female gametocytes, where they are not translated, demonstrating post-transcriptional control (85). Orthologs of this gene have been found in *P. falciparum* and other parasites of this genus. Other genes that undergo post-transcriptional control are those that encode dihydrofolate reductase and thymidylate synthetase (*dhfr-ts*). In animals, these two enzymes are encoded by independent genes, contrary to the mechanism in protozoans. Each of these proteins can bind its corresponding mRNA and inhibit its translation. In the presence of substrates or competitive inhibitors, these proteins lose the capacity to bind to RNA, and translation increases. In *P. falciparum*, the mRNA binding site is not located close to the active site of the enzyme, which therefore has the same affinity for the mRNA in its active and inactive forms. This characteristic of post-transcriptional control impedes the production of more proteins in a situation of greater necessity, for example in the presence of inhibitors, imposing a much lower tolerance of antifolate drugs than in human cells (86).

Protein families involved in RNA binding are highly represented in the *P. falciparum* genome (79). Among these are the Puf proteins that bind the nano-responsive elements (NRE), present in the 3'UTR of genes of various eukaryotes and inhibit RNA translation. Two genes that encode members of this gene family have been found in the *P. falciparum* genome and as orthologs in other species of the genus. They are overexpressed in gametocytes (87), and functional analysis demonstrates that both proteins are capable of binding to *Drosophila* mRNA that contains the NRE sequence in its 3'UTR, indicating its possible role in post-transcriptional control of genes involved in the development of sexual stages of the parasite (88).

When the transcriptome and the proteome of *P. berghei* were compared, new gametocyte-transcribed genes that have their translation inhibited were found (9). Analysis of the 3'UTR of these genes revealed a conserved motif of 47 nucleotides that contain repetitions of the NRE sequence. These data reinforce the importance of post-transcriptional control in gametocytes and assign the role of recognizing specific elements in the 3'UTR of genes that should have their translation inhibited to the Puf proteins.

Direct evidence of the importance of translational control on gene expression in malaria has been recently and elegantly demonstrated in gametocytes in *P. berghei* (89). Using reverse genetics, the gene encoding the DOZI helicase of *Plasmodium berghei* was knocked-out, and its disruption inhibited the formation of the ribonucleoprotein complexes, diverting at least 370 transcripts for degradation.

Another mechanism of post-transcriptional control is “antisense” RNA, which involves the *cis* or *trans* synthesis of an RNA complementary to the mRNA of some genes. This RNA binds to the corresponding mRNA silencing its translation, “splicing”, or diminishing its stability.

Antisense RNA has been found in all the large groups of organisms and has been most studied in bacteria (90).

Some evidence points to the existence of an antisense silencing mechanism in *Plasmodium* spp. Analysis of the transcriptome of intraerythrocytic stages of the parasite by "serial analysis of gene expression" (SAGE) revealed that more than 10% of the genes have antisense transcripts (91, 92). Later, it was determined that they are transcribed by an RNA polymerase sensitive to α-amanitina, a characteristic typical of RNA polymerase II (93). The antisense RNA mechanism is functional in transfection experiments for silencing of specific genes, such as *Cla*9 (94).

Non-Coding RNAs

The pattern of gene expression used by the *Plasmodium* genus differs in many aspects to most eukaryotes. Genes that are normally transcribed constitutively, such as histones, actin, and calmodulin, are transcribed in a stage-specific manner. Probably the most surprising difference in regulation of a group of genes in the parasite, compared with other eukaryotes, is the ribosomal RNA (rRNA) genes. Instead of the standard configuration of three subunits (18S, 5S, and 28S) repeated dozens or hundreds of times in sequence, it has around 20 genes, distributed on different chromosomes, often with some of the subunits missing. In addition, two or three structurally distinct forms of rRNA are present, and their expression is regulated. Type A rRNA is transcribed in the blood stages of the parasite, and type S is predominantly transcribed in the mosquito stages of the parasite (95, 96). The exact function of this type of mechanism is unknown, but knockout of one of the S-type genes in *P. berghei* inhibits development in the mosquito in a dose-dependent manner (97). The A-type sequence is more similar to that of other eukaryotes, whereas type S is more divergent, appearing similar to the bacterial ribosome in some regions (98). This type of mechanism of expression of ribosomes is unique, because organisms such as plants and bacteria regulate the type of ribosomes by altering the protein component.

Another group of non-coding RNA also has a peculiar structure in *Plasmodium*. Unlike most major eukaryotic organisms that have redundancy in the most commonly used codons, *P. falciparum* has a tRNA for almost all codons as single copy (7). Transfection experiments with genes containing a much lower AT content than the parasites demonstrate that these are translated without apparent necessity for optimization of the codons. These data may indicate that the parasite is capable of translating GC- and AT-rich codons with the same efficiency, unlike most organisms studied.

As has happened with practically all the genome, transcriptome, and proteome projects, the data generated by the studies in *Plasmodium* spp. have generated many more questions than answers as well as new lines of research. Among these, the use of computational tools to advance our scarce knowledge on control of gene expression is a brand new field that promises to rapidly advance our scarce knowledge in this area. A few illustrative examples on promoters, transcription factors, and transcriptomes will demonstrate its importance.

Bioinformatics and Gene Expression in *Plasmodium*

Promoter Elements

Although functional/structural analyses of a few promoter regions from *Plasmodium* have revealed that they posses the common bipartite structure of promoters from other eukaryotes,

analysis of intergenic regions has failed to identify canonical sequences such as CCATT and SP1 boxes. The laboratory of Dyann Wirth, as reported by Militello and coworkers (57), was the first to use a bioinformatics approach to gain insight into predicted regulatory elements in intergenic regions of *P. falciparum*. To do so, they analyzed the upstream intergenic regions of the heat shock protein genes as a model using the TRANSFAC (99) database on the Biobase homepage, which contains data on transcription factors, their binding sites, and regulated genes. TRANSFAC uses Match, a web-based tool to search for transcription factors by weight matrix search, and Patch, another web-based tool that identifies transcription factors by pattern matching. TRANSFAC thus allows making predictions of poorly characterized gene promoters. Interestingly, except for TATA boxes, whose prediction is questionable in malaria because of the high AT content of intergenic regions (close to 90%), non-other, common elements present in crown eukaryotes were identified in these *Plasmodium* upstream regions, suggesting species-specific regulatory elements. To further look for such elements, all intergenic regions of the *hsp* genes were analyzed using AlignAce to find common repeated elements to which a particular value was given (100, 101). Strikingly, a high-score, G-rich palindromic element was found in the intergenic regions of the *hsp* genes (Figure 3a). To determine whether such element was conserved in other *Plasmodium* species, the coding and upstream intergenic regions of the *hsp*86 gene of *P. yoelli*, *P. berghei*, and *P. vivax* were retrieved using TBLASTN and BLASTN, with the *P. falciparum hsp*86 amino acid sequence used as a query. The genomic sequence of the hsp86 locus of these four *Plasmodium* species were aligned, and significantly, the same G-rich element was observed in the upstream regions in these species, strongly suggesting that it represented an important regulatory element for *hsp* gene expression maintained through evolution (Figure 3b).

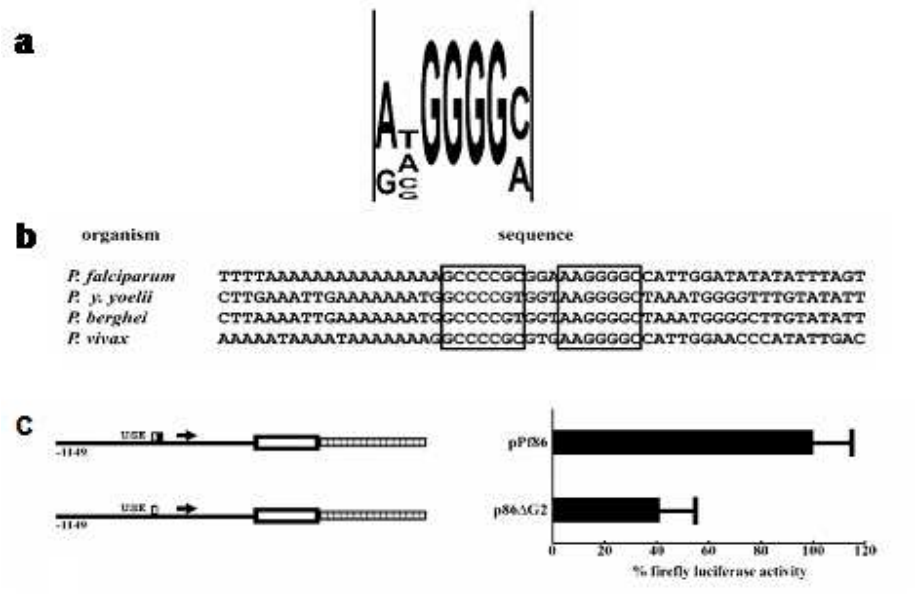

Figure 3: A functional G-rich palindromic element is conserved among the intergenic regions of *hsp* genes in *Plasmodium*. a. G-rich element. b. G-rich palindromic element in different malaria species. c. A reporter plasmid containing the intergenic region of the *hsp* gene of *P. falciparum* and driving the expression of the luciferase reporter gene is functional. Data and figures obtained with permission from Dyanne Wirth.

To prove these computational findings, recombinant plasmids containing the luciferase reporter gene under the control of the *P. falciparum hsp*86 upstream intergenic region, as well as mutated versions lacking the G-rich element, proved this one is required for gene expression of the *hsp*86 gene (Figure 3c). Similar approaches are now being used to analyze other promoter regions in *P. falciparum*, such as *var* gene promoters. Most relevant, because of the high AT content of intergenic regions as well as a putative role of long stretches of polypyrimidines (i.e., low complexity regions) in basal transcription (59), the development of new algorithms to solve this intrinsic problem of promoter predictions in malaria parasites represents an interesting computational challenge.

Transcription Factors

Controlled gene expression in higher eukaryotes is achieved mainly through a plethora of Transcription Associated Proteins (TAP), which are mainly taxon specific. Strikingly, the first annotated complete genome draft from *P. falciparum* failed to predict several TAPs commonly described in other eukaryotes (7). This suggested that either malaria parasites contain significantly lower TAPs than other eukaryotes, or that because of the high AT content of the genome, their prediction required a more sophisticated computational analysis. To address this issue, Coulson and coworkers (2004) (79) devised two different computational approaches to globally search for *P. falciparum* TAPs. In the first approach, they searched the TRANSFAC and SWISS-PROT databases for sequences retrieved from "transcription" in their keyword or gene description fields. Next, sequences were linked to their corresponding Protein Families (Pfam) database annotations; in this way, they obtained 51 profile-Hidden Markov Models (HMMs) of protein domains involved in transcription. HMMER version 2.2 was then used to search for *P. falciparum* sequences using the 51 HMMs. In addition, seven other databases from crown eukaryotic organisms, which included *Schizosaccharomyces pombe, Saccharomyces cerevisiae, Arabidopsis thaliana, Caenorhabditis elegans, Anopheles gambiae, Drosophila melanogaster*, and *Homo sapiens*, were similarly searched with these HMMs. Interestingly, only 17 models had matches with *P. falciparum* sequences corresponding to 69 unique proteins. In contrast, a similar analysis revealed 40 models matching *Arabidopsis thaliana* sequences.

In the second approach, they queried the *P. falciparum* genome as well as the genome of the same crown eukaryotes genomes against 17,054 TAPs. Clustering of TAPs and homologs by similarity revealed 84 clusters containing at least one *P. falciparum* gene and one TAP pertaining to 95 genes, of which only 8 were also identified by the HMMs. Together, these approaches thus revealed a total of 156 malarial TAPs, overall representing 3% of the genome. This is in striking contrast to the other higher eukaryotes similarly analyzed, which revealed an average of 10% of their genomes corresponding to TAPs. The complete list of TAPs as well as information on PlasmoDB ID, gene description, and expression at different developmental stages of malaria parasites are available from the Supplementary Material (http://www.ebi.ac.uk/research/cgg/projects/transcription/plasmodium).

Lack of predictions of the TFIID complex, essential for the formation of the transcription preinitiation complex at promoter regions of Pol II genes, encouraged a search for new computational methods to look for such factors. Analysis of *P. falciparum* protein domains revealed that regions of low complexity were rarely found within globular hydrophobic domains (80). Thus, a new computational, two-dimensional program called Hydrophobic Cluster Analysis

(HCA), which is not sensitive to gaps, was generated and used to search for globular domains of *P. falciparum* having similarity to domains conserved in protein sequences representing different subunits of the basal transcription factors and cofactors of humans and yeast (80). Importantly, 33 putative transcription factors, of which only 10 had been predicted using HMMs, were identified using HCA (Table 1). Thus, the combination of profile-based searches along with HCA promises to be a general method to identify TAPs in a genome scale for malaria parasites.

Table 1: General transcription factors predicted in *Plasmodium falciparum*.

Table 1: General transcription factors predicted in *Plasmodium falciparum*

Factors	*H. sapiens*	*S. cerevisiae*	*P. falciparum*	Nuclear signals prediction		Expression pattern	
				NLS	NES	Micro-array	Proteomic data
TFIIA α	P52655	P32773 (TCA1)	MAL7P1.78 +	-	-	G	-
TFIIA β							
TFIIA γ	P52657	P32774 (TCA2)	PFL2435w *+	-	-	T,Sc	-
			PFI1630 *	+	-	G	-
TFIIB	Q00403	P29055	PFA0525w	-	-	All stages	-
TFIID TBP	P20226	P13393	PFB0305w	-	-	R	-
TFIID TAF1	P21675 (TAF250)	P46677 (TAF145)	PFL1645w	+	-	S,LT	S,G
TFIID TAF2	*gi:4507347* (TAF150)	P23255 (TAF150)	MAL7P1.124	+	+	R,T	S
TFIID TAF5	Q15542 (TAF100)	P38129 (TAF90)	?				
TFIID TAF7	Q15545 (TAF55)	Q05021 (TAF67)	PFI1425w	+	+	R,S	-
TFIID TAF14	P42568 (ENL/AF-9)	P35189 (TAF30)	?				
TFIID TAF4	O00268 (TAF135)	P50105 (TAF48)	?				
TFIID TAF12	Q16514 (TAF20)	Q03761 (TAF68/61)	?				
TFIID TAF6	P49848 (TAF80)	P53040 (TAF60)	?				
TFIID TAF9	Q16594 (TAF31)	Q05027 (TAF17)	?				
TFIID TAF11	Q15544 (TAF28)	Q04226 (TAF40)	?				
TFIID TAF13	Q15543 (TAF18)	P11747 (TAF19)	?				
TFIID TAF3	*gi:13374079* (TAF140)	Q12297 (TAF47)	?				
TFIID TAF8	*gi:31323620* (TAF43)	Q03750 (TAF65)	?				
TFIID TAF10	Q12962 (TAF30)	Q12030 (TAF25)	PFE1110w	-	+	R, Sc	-
TFIIE α	P29083	P36100	MAL7P1.86 +	+	+	Sc	-
TFIIE β	P29084	P36145	MAL13P1.360 *	+	-	ND	-
TFIIF α	P35269 (RAP74)	P41895 (Tfg1)	?				
TFIIF β	P13984 (RAP30)	P41896(Tfg2)	PFI1 0458 *	-	+	R,G	-
TFIIH core p62/TFB1	P32780 (p62)	P32776 (TFB1)	MAL3P7.42 *+ (Chr3.phat 258)	+	+	R,T	-
TFIIH core p52/TFB2	Q92759 (p52)	*gi:6325135* (TFB2)	PFL2125c	+	+	R,T,Sc	-
TFIIH core p44/SSL1	Q13888 (p44)	Q04673 (SSL1)	MAL13P1.76	+	+	R,T	-
TFIIH core p34/TFB4	Q13889 (p34)	*Gi:6325313*	PFI3_0279	-	+	T	-
TFIIH core TFB5	*Gi:55665883*	*Gi:13129164*	PFI4 0398	-	-	R, T, G	-
TFIIH core XPB/SSL2-RAD25	P19447 (XPB)	Q00578 (SSL2/RAD25)	PFI0_0369	+	-	G	S
TFIIH XPD/RAD3	P18074 (XPD)	P06839 (RAD3)	PFI1650w	+	+	R,T,G	G
TFIIH CAK MAT1/TFB3	P51948 (MAT1)	*Gi:6320668* (TFB3)	PFB0610c	+	-	R,T	-
TFIIH CAK Cdk7/KIN28	P50613 (CDK7)	P06242 (KIN28)	? $				
TFIIH CAK Cyclin H/ CCL1	P51946 (cyclin H)	P37366 (CCL1)	? $				

Factors	*H. sapiens*	*S. cerevisiae*	*P. falciparum*	Nuclear signals predictio n			
				NLS	NES	Microarray	Proteomi c data
TFIIAα	P52655	P32773 (TCA1)	MAL7P1.70+	–	–	G	–
TFIIAβ							
TFIIAγ	P52657	P32774 (TCA2	PFL2435w *+	–	–	T,Sc	–

			PF11630*	+	–	G	–
TFIIB	Q00403	P29055	PFA0525w		–	All Stages	–
TFIID TBP	P20226	P13393	PFE0305w		–	R	–
TFIIDTAF1	P21675 (TAF250)	P46677 (TAF145)	PFL1645w	+	–	S,LT	S,G
TFIIDTAF2	Gt4507347 (TAF150)	P23255 (TAF150)	MAL7P1.134	+	+	R,T	S
TFIIDTAF5	Q15542 (TAF100)	P20129 (TAP90)					
TFIID TAF7	Q15545 (TAF55)	Q05021 (TAF67)	PFI1425w	+	+	R,S	–
TFIID TAF14	P42568 (ENL/AF-9)	P35189 (TAF30)					
TFIID		P50105 (TAF48)					
TFIID		Q03761 (TAF68/61)					
TFIID		P53040 (TAF60)					
TFIID		Q05027 (TAF17)					
TFIID		Q04226 (TAF40)					
TFIID		P11747 (TAF19)					
TFIID		Q12297 (TAF47)					
TFIID		Q00750 (TAF65)					
TFIID		Q12000 (TAF25)	PFE1110w	–	+	R,Sc	–
TFIID		P36100	MAL7P1.86+	+	+	Sc	–
TFIID		P36145	MAL13P1.360*	+	–	ND	–
TFIID		P41895 (Tfg1)					–
TFIID		P41896 (Tfg2)	PF11 0458*	–	+	R,G	–
		P32776 (TFB1)	MAL3P7.42*+	+	+	R,T	–
			(Chr3 ph 25B)	+	–		–
			PFL2125c	+	+	R,T,Sc	–
			MAL13P1.76	+	+	R,T	–
			PF13_0279	–	+	T	–
			PF14_0398	–	–	R,T,G	–
			PF10_0369	+	–	G	S
			PF11650w	+	+	R,T,G	G
						R,T	
			PFB0610c	+	–		
			?$				
TFIIH CAK			?$				
CCLI							

The general transcription factors which were identified in this report are underlined and shown in bold; * and + indicate similarities which were identified in this report after assessing PSI-BLAST marginal similarities at the sequence 2D level or after extending the comparison at the 2D level outside the limits primarily defined by PSI-BLAST, respectively. The Swiss-Prot accession numbers are given for the human and yeast sequences. When a Swiss-Prot identifier is not available, the GenBank identifier (gi) is indicated in italics instead. The references of the human TAF3 and TAF8 sequences can be found in 50 and 100, respectively. $: see text for comments on these putative homologues of Cdk7/KIN28 and cyclin H/CCL1. The presence of nuclear localization sequences (NLS 102), nuclear export sequences (NES 103), and the expression of these predicted general transcription factors using microarray and proteomics on different parasite stages (9–11, 89) are indicated. G: gametocyte, T: trophozoite, LT: late trophozoite, Sc: schizonte, R: ring, (-): absent, (+): present, (?): cannot be found. Table and legend reproduced with permission from Isabelle Callebaut (80).

Several other important aspects on control of gene expression of malaria parasites were revealed by these analyses. First, unlike crown eukaryotes, malaria parasites do not contain large paralogous families of TAPs; in fact, this lack of gene duplications is not limited to TAPs but to other internal genes, in contrast to subtelomeric genes, where malaria have clustered multigene families involved in immune escape. Second, malaria parasites contain homologs of all 12 subunits of RNA polymerase II, TBP protein, and TFIIB, indicating that the basal transcription machinery is similar to that of crown eukaryotes. Yet, most components of the TFIID complex, as well as other factors essential for forming the initiation complex of RNA Pol II genes, were not predicted (see below). Third, TAPs involved in RNA stability and translation, such as the CCCH-type Zn finger, were overrepresented in the genome of *P. falciparum*, reinforcing the idea

of an important post-transcriptional control of gene expression. Last, TAPs involved in acetylation/deacetylation were identified, indicating that epigenetic mechanisms can also be involved in control of gene expression in malaria. Recent elegant evidence has indeed demonstrated epigenetic control of *var* gene expression (38, 39).
In summary, using regular similarity analysis (7), HMMs (79), and profile-based searches/HCA (80), there are still about 40% of general transcription factors found in crown eukaryotes that are not predicted in the genome of *P. falciparum*. Thus, either these factors have evolved to a point that none of these computational methods is able to predict them, or basal transcription is different in malaria parasites. If so, these differences can be rationally exploited for alternative control strategies against malaria.

Transcriptome Analysis of the Intraerythrocytic Developmental Cycle

Pathology in malaria is mostly associated with the intraerythrocytic developmental cycle (IDC), where sequestration of infected red blood cells in the deep capillaries of internal organs, as well as synchronous rupture of infected red blood cells and release of merozoites, can induce obstruction of blood flow, edema, cerebral malaria, and anemia, among others. Interestingly, during these stages, malaria parasites have a cyclic developmental program in which, upon entrance of merozoites inside red blood cells, the parasites differentiate into a stage known as "ring", which is surrounded by a parasitophorous vacuole. After a period of 12-18 hours, the parasites enter into a highly metabolic growth stage known as a trophozoite, which lasts for about 16 h, after which DNA synthesis starts in a stage known as the schizont. After nuclear divisions inside a syncithium, cytoplasmic membranes are synthesized, red blood cells burst, and free individual merozoites are liberated into the circulation to invade new red blood cells (Figure 2). In a seminal work carried out at the laboratory of Joe DeRisi at the University of California, San Francisco, the entire transcriptome analysis of the asexual blood stages at 1-hour scale resolution was undertaken (13). To do so, this group had previously developed a web-based program called ArrayOligoSelector that allowed them to construct a 70-base-long oligonucleotide microarray representing the entire *P. falciparum* genome as it had been released since 2000 and annotated as 2002 (7). This program takes into account different parameters including uniqueness, user-defined elements to be avoided, self-binding, complexity, and GC content and is freely available for downloading. A total of 7,462 oligonucleotides representing 5,409 annotated open reading frames were printed into this first-generation microarray. Next, they grew parasites into a 5-liter culture bioreactor, and after a double round of synchronization by sorbitol, they removed two 5-ml aliquots at each hour for 48 hours. One of the aliquots was pooled with all others, making a reference pool labeled with Cy-3, and the other was labeled with Cy-5 and used to analyze the expression profile at each particular hour of the IDC. Thus, at 1-hour scale resolution, they were able to determine those genes that were expressed higher than the reference pool for the entire IDC. Unexpectedly, analysis of the results revealed that close to 85% of the genes were expressed during the IDC, and that of those close to 75% were expressed only once in a periodic manner, having a single maximum and a single minimum of 1-1.5 hours. In sharp contrast, a similar analysis of the expression profiles of the yeast cell cycle revealed that only 3% of the genes were expressed periodically (102). The periodic expression allowed them to use Fourier Analysis to order expression profiles of the IDC into a phaseogram according to their time and phase of expression (Figure 4). Results demonstrated that malaria parasites have a

cascade of expression lacking clear boundaries throughout the IDC, and that genes expressed at each particular phase could be grouped into genes with similar biological functions.

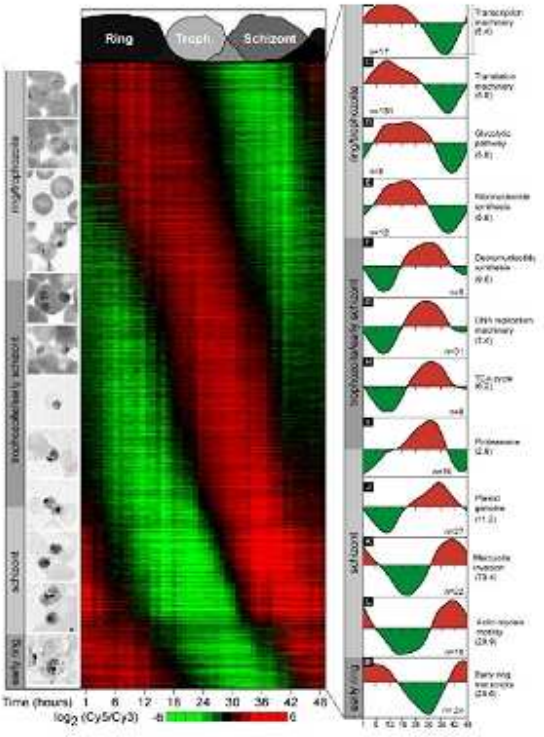

Figure 4: Overview of the *P. falciparum* IDC transcriptome (13). (A) A phaseogram of the IDC transcriptome was created by ordering the transcriptional profiles for 2,712 genes by phase of expression along the y-axis. The characteristic stages of intraerythrocytic parasite morphology are shown on the left, aligned with the corresponding phase of peak gene expression. (B–M) The temporal ordering of biochemical processes and functions is shown on the right. Esach graph corresponds to the average expression profile for the genes in each set, and the eman peak-to-through amplitude is shown in parentheses. Figure and legend reproduced with permission from Joe DeRisi.

A striking example of this similar phase/similar function was with genes expressed during the late schizogony and involved in merozoite invasion. Thus, during the merozoite stages, the parasites have to expose a series of proteins at their surface that will allow them to specifically interact and enter new red blood cells, and these surface antigens have become prime candidates for the development of vaccines. Strikingly, several merozoite surface proteins—AMA1, EBA, and RESA, all of which have experimental evidence supporting their roles as potential vaccine candidates—were expressed at the same phase, along with many other coding genes, the functions of which are unknown and which expand the list of putative vaccine candidates. Similar results and lists were generated for the entire phaseogram. Because close to 60% of the *P. falciparum* genome represents hypothetical proteins of unknown function, this analysis and lists of genes have potentially and significantly advanced malaria research by assigning function

of hypothetical proteins by phase of expression during the IDC; yet, formal proof-of-principle of any one particular hypothetical protein expressed at a particular phase and from which we can assign a similar function as the ones annotated is still lacking. This seminal paper thus illustrates the power of mathematical-computational analysis to advance our biological knowledge of the asexual blood stages of malaria parasites. Moreover, these data indicate that there is a tight, simple, virus-like control of gene expression during the IDC in spite of its complex cycle involving three differentiation stages. It is important to recall, however, that to obtain such data and interpretation, the authors used only 2,714 ORFs of the 5,440 predicted genes. Thus, close to 50% of the entire coding genome of *P. falciparum* was represented or was not periodically expressed (i.e., not amenable to Fourier analysis) on this particular array.

An alternative systems biology approach to create probabilistic genetics networks (PGNs) based on the entire transcriptome of the IDC from these same data was reported by Barrera and co-workers (103). Irrespective of periodically or non-periodically expressed genes, they built a Markov model, which was able to create PGNs based on the interactions on individual or double-genes throughout the IDC. Validation of the model was shown by creating a PGN of glycolysis and by demonstrating that genes in this PGN were biologically sound, and that key enzymes, such as isomerase and which had oligonucleotides representing them and not giving a periodic signal, were included into this PGN. Similar to the Fourier analysis, formal proof of these PGNs still awaits further experimentation. Regardless, this different analysis of the same set of data clearly exemplifies the importance of computational biology, statistics, and mathematics in advancing biologically meaningful results from data on transcriptomes.

Concluding Remarks

As reviewed in this chapter, control of gene expression of malaria parasites seems unique among crown eukaryotes. Although significant advances in our knowledge of this fundamental aspect of the biology of malaria parasites have been achieved through functional assays, they remain technically difficult. Bioinformatics and computational approaches are rapidly advancing our knowledge in this area. Indeed, as data on genomes, transcriptomes, and proteomes of several *Plasmodium* species are expanding, new challenges will include crossing borders of all these “omes” to integrate a view of the parasites as a dynamic system. An integrative approach involving biology, mathematics, statistics, and computational biology to develop rational approaches for alternative control strategies will be needed.

References

1. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI. The global distribution of clinical episodes of Plasmodium falciparum malaria. Nature. 2005;434(7030):214–217.
2. Miller LH, Mason SJ, Clyde DF, McGinniss MH. The resistance factor to Plasmodium vivax in blacks. The Duffy-blood-group genotype, FyFy. N Engl J Med. 1976;295(6):302–304.
3. Barnwell JW, Wertheimer SP. Plasmodium vivax: merozoite antigens, the Duffy blood group, and erythrocyte invasion. Prog Clin Biol Res. 1989;313:1–11.
4. Baum J, Maier AG, Good RT, Simpson KM, Cowman AF. Invasion by P. falciparum merozoites suggests a hierarchy of molecular interactions. PLoS Pathog. 2005;1(4):e37.
5. Amino R, Thiberge S, Martin B, Celli S, Shorte S, Frischknecht F, Menard R. Quantitative imaging of Plasmodium transmission from mosquito to mammal. Nat Med. 2006;12(2):220–224.

6. del Portillo HA, Lanzer M, Rodriguez-Malaga S, Zavala F, Fernandez-Becerra C. Variant genes and the spleen in Plasmodium vivax malaria. Int J Parasitol. 2004;34(13-14):1547–1554.
7. 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, Mungal 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(6906):498–511.
8. Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC. et al. Genome sequence and comparative analysis of the model rodent malaria parasite Plasmodium yoelii yoelii. Nature. 2002;419(6906):512–519.
9. Hall N, Karras M, Raine JD, Carlton JM, Kooij TW, Berriman M. et al. A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. Science. 2005;307(5706):82–86.
10. del Portillo HA, Fernandez-Becerra C, Bowman S, Oliver K, Preuss M, Sanchez CP. et al. A superfamily of variant genes encoded in the subtelomeric region of Plasmodium vivax. Nature. 2001;410(6830):839–842.
11. Hayward RE, Derisi JL, Alfadhli S, Kaslow DC, Brown PO, Rathod PK. Shotgun DNA microarrays and stage-specific gene expression in Plasmodium falciparum malaria. Mol Microbiol. 2000;35(1):6–14.
12. Ben Mamoun C, Gluzman IY, Hott C, MacMillan SK, Amarakone AS, Anderson DL. et al. Co-ordinated programme of gene expression during asexual intraerythrocytic development of the human malaria parasite Plasmodium falciparum revealed by microarray analysis. Mol Microbiol. 2001;39(1):26–36.
13. Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu J, DeRisi JL. The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. PLoS Biol. 2003;1(1):E5.
14. Le Roch KG, Zhou Y, Blair PL, Grainger M, Moch JK, Haynes JD. et al. Discovery of gene function by expression profiling of the malaria parasite life cycle. Science. 2003;301(5639):1503–1508.
15. Kaiser K, Matuschewski K, Camargo N, Ross J, Kappe SH. Differential transcriptome profiling identifies Plasmodium genes encoding pre-erythrocytic stage-specific proteins. Mol Microbiol. 2004;51(5):1221–1232.
16. Gruner AC, Hez-Deroubaix S, Snounou G, Hall N, Bouchier C, Letourneur F. et al. Insights into the P. y. yoelii hepatic stage transcriptome reveal complex transcriptional patterns. Mol Biochem Parasitol. 2005;142(2):184–192.
17. Sacci JB, Azad AF. Gene expression analysis during liver stage development of Plasmodium. Int J Parasitol. 2002;32(13):1551–1557.
18. Wang Q, Brown S, Roos DS, Nussenzweig V, Bhanot P. Transcriptome of axenic liver stages of Plasmodium yoelii. Mol Biochem Parasitol. 2004;137(1):161–168.
19. Daily JP, Le Roch KG, Sarr O, Fang X, Zhou Y, Ndir O. et al. In vivo transcriptional profiling of Plasmodium falciparum. Malar J. 2004;3:30.
20. Daily JP, Le Roch KG, Sarr O, Ndiaye D, Lukens A, Zhou Y. et al. In vivo transcriptome of Plasmodium falciparum reveals overexpression of transcripts that encode surface proteins. J Infect Dis. 2005;191(7):1196–1203.
21. Quackenbush J. Computational analysis of microarray data. Nat Rev Genet. 2001;2:418–427.
22. Florens L, Washburn MP, Raine JD, Anthony RM, Grainger M, Haynes JD. et al. A proteomic view of the Plasmodium falciparum life cycle. Nature. 2002;419(6906):520–526.
23. Blair PL, Carucci DJ. Functional proteome and expression analysis of sporozoites and hepatic stages of malaria development. Curr Top Microbiol Immunol. 2005;295:417–438.
24. Lasonder E, Ishihama Y, Andersen JS, Vermunt AM, Pain A, Sauerwein RW. et al. Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. Nature. 2002;419(6906):537–542.
25. Khan SM, Franke-Fayard B, Mair GR, Lasonder E, Janse CJ, Mann M. et al. Proteome analysis of separated male and female gametocytes reveals novel sex-specific Plasmodium biology. Cell. 2005;121(5):675–687.
26. Bradley PJ, Ward C, Cheng SJ, Alexander DL, Coller S, Coombs GH. et al. Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in Toxoplasma gondii. J Biol Chem. 2005;280(40):34245–34258.
27. Sam-Yellowe TY, Florens L, Wang T, Raine JD, Carucci DJ, Sinden R. et al. Proteome analysis of rhoptry-enriched fractions isolated from Plasmodium merozoites. J Proteome Res. 2004;3(5):995–1001.
28. Cary C, Lamont D, Dalton JP, Doerig C. Plasmodium falciparum chromatin: nucleosomal organisation and histone-like proteins. Parasitol Res. 1994;80(3):255–258.
29. Lanzer M, de Bruin D, Wertheimer SP, Ravetch JV. Transcriptional and nucleosomal characterization of a subtelomeric gene cluster flanking a site of chromosomal rearrangements in Plasmodium falciparum. Nucleic Acids Res. 1994;22(20):4176–4182.

30. Miao J, Fan Q, Cui L, Li J. The malaria parasite Plasmodium falciparum histones: Organization, expression, and acetylation. Gene. 2006;369:53–65.
31. Meissner M, Soldati D. The transcription machinery and the molecular toolbox to control gene expression in Toxoplasma gondii and other protozoan parasites. Microbes Infect. 2005;7(13):1376–1384.
32. Kasinsky HE, Lewis JD, Dacks JB, Ausio J. Origin of H1 linker histones. FASEB J. 2001;15(1):34–42.
33. Chandra BR, Olivieri A, Silvestrini F, Alano P, Sharma A. Biochemical characterization of the two nucleosome assembly proteins from Plasmodium falciparum. Mol Biochem Parasitol. 2005;142(2):237–247.
34. Horrocks P, Pinches R, Kriek N, Newbold C. Stage-specific promoter activity from stably maintained episomes in Plasmodium falciparum. Int J Parasitol. 2002;32(10):1203–1206.
35. 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(2):251–261.
36. Joshi MB, Lin DT, Chiang PH, Goldman ND, Fujioka H, Aikawa M. et al. Molecular cloning and nuclear localization of a histone deacetylase homologue in Plasmodium falciparum. Mol Biochem Parasitol. 1999;99(1):11–19.
37. Scherf A, Figueiredo LM, Freitas-Junior LH. Plasmodium telomeres: a pathogen's perspective. Curr Opin Microbiol. 2001;4(4):409–414.
38. Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, Thompson JK. et al. Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in Plasmodium falciparum. Cell. 2005;121(1):13–24.
39. Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, Rojas-Meza AP. et al. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. Cell. 2005;121(1):25–36.
40. Wu Y, Sifri CD, Lei HH, Su XZ, Wellems TE. Transfection of Plasmodium falciparum within human red blood cells. Proc Natl Acad Sci U S A. 1995;92(4):973–977.
41. Crabb BS, Cowman AF. Characterization of promoters and stable transfection by homologous and nonhomologous recombination in Plasmodium falciparum. Proc Natl Acad Sci U S A. 1996;93(14):7289–7294.
42. Lanzer M, Wertheimer SP, de Bruin D, Ravetch JV. Chromatin structure determines the sites of chromosome breakages in Plasmodium falciparum. Nucleic Acids Res. 1994;22(15):3099–3103.
43. Lanzer M, de Bruin D, Ravetch JV. Transcription mapping of a 100 kb locus of Plasmodium falciparum identifies an intergenic region in which transcription terminates and reinitiates. EMBO J. 1992;11(5):1949–1955.
44. Wickham ME, Thompson JK, Cowman AF. Characterisation of the merozoite surface protein-2 promoter using stable and transient transfection in Plasmodium falciparum. Mol Biochem Parasitol. 2003;129(2):147–156.
45. de Koning-Ward TF, Speranca MA, Waters AP, Janse CJ. Analysis of stage specificity of promoters in Plasmodium berghei using luciferase as a reporter. Mol Biochem Parasitol. 1999;100(1):141–146.
46. Dechering KJ, Kaan AM, Mbacham W, Wirth DF, Eling W, Konings RN. et al. Isolation and functional characterization of two distinct sexual-stage-specific promoters of the human malaria parasite Plasmodium falciparum. Mol Cell Biol. 1999;19(2):967–978.
47. van der Wel AM, Tomas AM, Kocken CH, Malhotra P, Janse CJ, Waters AP. et al. Transfection of the primate malaria parasite Plasmodium knowlesi using entirely heterologous constructs. J Exp Med. 1997;185(8):1499–1503.
48. Mota MM, Thathy V, Nussenzweig RS, Nussenzweig V. Gene targeting in the rodent malaria parasite Plasmodium yoelii. Mol Biochem Parasitol. 2001;113(2):271–278.
49. Fernandez-Becerra C, de Azevedo MF, Yamamoto MM, del Portillo HA. Plasmodium falciparum: new vector with bi-directional promoter activity to stably express transgenes. Exp Parasitol. 2003;103(1-2):88–91.
50. Ferreira de Azevedo M, del Portillo HA. Mal J. 2007;6:20.
51. Horrocks P, Lanzer M. Mutational analysis identifies a five base pair cis-acting sequence essential for GBP130 promoter activity in Plasmodium falciparum. Mol Biochem Parasitol. 1999;99(1):77–87.
52. Mbacham WF, Chow CS, Daily J, Golightly LM, Wirth DF. Deletion analysis of the 5' flanking sequence of the Plasmodium gallinaceum sexual stage specific gene pgs28 suggests a bipartite arrangement of cis-control elements. Mol Biochem Parasitol. 2001;113(1):183–187.
53. Porter ME. The DNA polymerase delta promoter from Plasmodium falciparum contains an unusually long 5' untranslated region and intrinsic DNA curvature. Mol Biochem Parasitol. 2001;114(2):249–255.

54. Osta M, Gannoun-Zaki L, Bonnefoy S, Roy C, Vial HJ. A 24 bp cis-acting element essential for the transcriptional activity of Plasmodium falciparum CDP-diacylglycerol synthase gene promoter. Mol Biochem Parasitol. 2002;121(1):87–98.
55. Voss TS, Kaestli M, Vogel D, Bopp S, Beck HP. Identification of nuclear proteins that interact differentially with Plasmodium falciparum var gene promoters. Mol Microbiol. 2003;48(6):1593–1607.
56. Chow CS, Wirth DF. Linker scanning mutagenesis of the Plasmodium gallinaceum sexual stage specific gene pgs28 reveals a novel downstream cis-control element. Mol Biochem Parasitol. 2003;129(2):199–208.
57. Militello KT, Dodge M, Bethke L, Wirth DF. Identification of regulatory elements in the Plasmodium falciparum genome. Mol Biochem Parasitol. 2004;134(1):75–88.
58. Ruvalcaba-Salazar OK, del Carmen Ramirez-Estudillo M, Montiel-Condado D, Recillas-Targa F, Vargas M, Hernandez-Rivas R. Recombinant and native Plasmodium falciparum TATA-binding-protein binds to a specific TATA box element in promoter regions. Mol Biochem Parasitol. 2005;140(2):183–196.
59. Polson HE, Blackman MJ. A role for poly(dA)poly(dT) tracts in directing activity of the Plasmodium falciparum calmodulin gene promoter. Mol Biochem Parasitol. 2005;141(2):179–189.
60. Bzik DJ. The structure and role of RNA polymerases in Plasmodium. Parasitol Today. 1991;7(8):211–214.
61. Fox BA, Li WB, Tanaka M, Inselburg J, Bzik DJ. Molecular characterization of the largest subunit of Plasmodium falciparum RNA polymerase I. Mol Biochem Parasitol. 1993;61(1):37–48.
62. Li WB, Bzik DJ, Gu HM, Tanaka M, Fox BA, Inselburg J. An enlarged largest subunit of Plasmodium falciparum RNA polymerase II defines conserved and variable RNA polymerase domains. Nucleic Acids Res. 1989;17(23):9621–9636.
63. Giesecke H, Barale JC, Langsley G, Cornelissen AW. The C-terminal domain of RNA polymerase II of the malaria parasite Plasmodium berghei. Biochem Biophys Res Commun. 1991;180(3):1350–1355.
64. Li J, Maga JA, Cermakian N, Cedergren R, Feagin JE. Identification and characterization of a Plasmodium falciparum RNA polymerase gene with similarity to mitochondrial RNA polymerases. Mol Biochem Parasitol. 2001;113(2):261–269.
65. Gardner MJ, Goldman N, Barnett P, Moore PW, Rangachari K, Strath M. et al. Phylogenetic analysis of the rpoB gene from the plastid-like DNA of Plasmodium falciparum. Mol Biochem Parasitol. 1994;66(2):221–231.
66. Ho CK, Shuman S. A yeast-like mRNA capping apparatus in Plasmodium falciparum. Proc Natl Acad Sci U S A. 2001;98(6):3050–3055.
67. Hausmann S, Altura MA, Witmer M, Singer SM, Elmendorf HG, Shuman S. Yeast-like mRNA capping apparatus in Giardia lamblia. J Biol Chem. 2005;280(13):12077–12086.
68. Knapp B, Nau U, Hundt E, Kupper HA. Demonstration of alternative splicing of a pre-mRNA expressed in the blood stage form of Plasmodium falciparum. J Biol Chem. 1991;266(11):7148–7154.
69. van Lin LH, Pace T, Janse CJ, Birago C, Ramesar J, Picci L. et al. Interspecies conservation of gene order and intron-exon structure in a genomic locus of high gene density and complexity in Plasmodium. Nucleic Acids Res. 2001;29(10):2059–2068.
70. Singh N, Preiser P, Renia L, Balu B, Barnwell J, Blair P. et al. Conservation and developmental control of alternative splicing in maebl among malaria parasites. J Mol Biol. 2004;343(3):589–599.
71. Pace T, Birago C, Janse CJ, Picci L, Ponzi M. Developmental regulation of a Plasmodium gene involves the generation of stage-specific 5' untranslated sequences. Mol Biochem Parasitol. 1998;97(1-2):45–53.
72. Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF. et al. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. Nature. 2004;432(7016):517–522.
73. Ruvolo V, Altszuler R, Levitt A. The transcript encoding the circumsporozoite antigen of Plasmodium berghei utilizes heterogeneous polyadenylation sites. Mol Biochem Parasitol. 1993;57(1):137–150.
74. Lanzer M, Wertheimer SP, de Bruin D, Ravetch JV. Plasmodium: control of gene expression in malaria parasites. Exp Parasitol. 1993;77(1):121–128.
75. Golightly LM, Mbacham W, Daily J, Wirth DF. 3' UTR elements enhance expression of Pgs28, an ookinete protein of Plasmodium gallinaceum. Mol Biochem Parasitol. 2000;105(1):61–70.
76. Shue P, Brown SV, Cann H, Singer EF, Appleby S, Golightly LM. The 3' UTR elements of P. gallinaceum protein Pgs28 are functionally distinct from those of human cells. Mol Biochem Parasitol. 2004;137(2):355–359.
77. Wu Y, Kirkman LA, Wellems TE. Transformation of Plasmodium falciparum malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. Proc Natl Acad Sci U S A. 1996;93(3):1130–1134.
78. Goonewardene R, Daily J, Kaslow D, Sullivan TJ, Duffy P, Carter R. et al. Transfection of the malaria parasite and expression of firefly luciferase. Proc Natl Acad Sci U S A. 1993;90(11):5234–5236.

79. Coulson RM, Hall N, Ouzounis CA. Comparative genomics of transcriptional control in the human malaria parasite Plasmodium falciparum. Genome Res. 2004;14(8):1548–1554.
80. Callebaut I, Prat K, Meurice E, Mornon JP, Tomavo S. Prediction of the general transcription factors associated with RNA polymerase II in Plasmodium falciparum: conserved features and differences relative to other eukaryotes. BMC Genomics. 2005;6:100.
81. 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 Res. 2005;33(13):3994–4006.
82. McAndrew MB, Read M, Sims PF, Hyde JE. Characterisation of the gene encoding an unusually divergent TATA-binding protein (TBP) from the extremely A+T-rich human malaria parasite Plasmodium falciparum. Gene. 1993;124(2):165–171.
83. 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. Mol Biochem Parasitol. 2004;138(1):159–163.
84. van Noort V, Huynen MA. Combinatorial gene regulation in Plasmodium falciparum. Trends Genet. 2006;22(2):73–78.
85. Paton MG, Barker GC, Matsuoka H, Ramesar J, Janse CJ, Waters AP. et al. Structure and expression of a post-transcriptionally regulated malaria gene encoding a surface protein from the sexual stages of Plasmodium berghei. Mol Biochem Parasitol. 1993;59(2):263–275.
86. Zhang K, Rathod PK. Divergent regulation of dihydrofolate reductase between malaria parasite and human host. Science. 2002;296(5567):545–547.
87. Cui L, Fan Q, Li J. The malaria parasite Plasmodium falciparum encodes members of the Puf RNA-binding protein family with conserved RNA binding activity. Nucleic Acids Res. 2002;30(21):4607–4617.
88. Fan Q, Li J, Kariuki M, Cui L. Characterization of PfPuf2, member of the Puf family RNA-binding proteins from the malaria parasite Plasmodium falciparum. DNA Cell Biol. 2004;23(11):753–760.
89. Mair GR, BraksGarver LS, Wiegant JC, Hall N, Dirks RW, Khan SM, Dimopoulos G, Janse CJ, Waters AP. Regulation of sexual development of Plasmodium by translational repression. Science. 2006;313:667–669.
90. Brantl S. Antisense-RNA regulation and RNA interference. Biochim Biophys Acta. 2002;1575(1-3):15–25.
91. Gunasekera AM, Patankar S, Schug J, Eisen G, Kissinger J, Roos D. et al. Widespread distribution of antisense transcripts in the Plasmodium falciparum genome. Mol Biochem Parasitol. 2004;136(1):35–42.
92. Patankar S, Munasinghe A, Shoaibi A, Cummings LM, Wirth DF. Serial analysis of gene expression in Plasmodium falciparum reveals the global expression profile of erythrocytic stages and the presence of anti-sense transcripts in the malarial parasite. Mol Biol Cell. 2001;12(10):3114–3125.
93. Militello KT, Patel V, Chessler AD, Fisher JK, Kasper JM, Gunasekera A. et al. RNA polymerase II synthesizes antisense RNA in Plasmodium falciparum. RNA. 2005;11(4):365–370.
94. Gardiner DL, Holt DC, Thomas EA, Kemp DJ, Trenholme KR. Inhibition of Plasmodium falciparum clag9 gene function by antisense RNA. Mol Biochem Parasitol. 2000;110(1):33–41.
95. Gunderson JH, Sogin ML, Wollett G, Hollingdale M, de la Cruz VF, Waters AP. et al. Structurally distinct, stage-specific ribosomes occur in Plasmodium. Science. 1987;238(4829):933–937.
96. Waters AP, Syin C, McCutchan TF. Developmental regulation of stage-specific ribosome populations in Plasmodium. Nature. 1989;342(6248):438–440.
97. van Spaendonk RM, Ramesar J, van Wigcheren A, Eling W, Beetsma AL, van Gemert GJ. et al. Functional equivalence of structurally distinct ribosomes in the malaria parasite, Plasmodium berghei. J Biol Chem. 2001;276(25):22638–22647.
98. Rogers MJ, Gutell RR, Damberger SH, Li J, McConkey GA, Waters AP. et al. Structural features of the large subunit rRNA expressed in Plasmodium falciparum sporozoites that distinguish it from the asexually expressed subunit rRNA. RNA. 1996;2(2):134–145.
99. Wingender E, Dietze P, Karas H, Knuppel R. TRANSFAC: a database on transcription factors and their DNA binding sites. Nucleic Acids Res. 1996;24(1)
100. Roth FP, Hughes JD, Estep PW, Church GM. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nat Biotechnol. 1998;16(10):939–945.
101. Hughes JD, Estep PW, Tavazoie S, Church GM. Computational identification of cis-regulatory elements associated with groups of functionally related genes in Saccharomyces cerevisiae. J Mol Biol. 2000;296(5):1205–14.

102. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B. Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol Biol Cell. 1998;9:3273–3297.
103. Deleted in press.
104. Jones MK, Good MF. Malaria parasites up close. Nat Med. 2006;12(2):170–171.