



HHS Public Access

Author manuscript

Int J Parasitol. Author manuscript; available in PMC 2015 September 11.

Published in final edited form as:

Int J Parasitol. 2008 August ; 38(10): 1083–1097. doi:10.1016/j.ijpara.2008.01.002.

Histone lysine methyltransferases and demethylases in *Plasmodium falciparum*

Liwang Cui*, Qi Fan, Long Cui, and Jun Miao

Department of Entomology, The Pennsylvania State University, 501 ASI Building, University Park, PA 16802, USA

Abstract

Dynamic histone lysine methylation, regulated by methyltransferases and demethylases, plays fundamental roles in chromatin structure and gene expression in a wide range of eukaryotic organisms. A large number of SET-domain-containing proteins make up the histone lysine methyltransferase (HKMT) family, which catalyzes the methylation of different lysine residues with relatively high substrate specificities. Another large family of Jumonji C (JmjC)-domain-containing histone lysine demethylases (JHDMs) reverses histone lysine methylation with both lysine site and methyl-state specificities. Through bioinformatic analysis, at least nine SET domain-containing genes were found in the malaria parasite *Plasmodium falciparum* and its sibling species. Phylogenetic analysis separated these putative HKMTs into five subfamilies with different putative substrate specificities. Consistent with the phylogenetic subdivision, methyl marks were found on K4, K9 and K36 of histone H3, and K20 of histone H4 by site-specific methyl-lysine antibodies. In addition, most SET domain genes and histone methyl-lysine marks displayed dynamic changes during the parasite asexual erythrocytic cycle, suggesting that they constitute an important epigenetic mechanism of gene regulation in malaria parasites. Furthermore, the malaria parasite and other apicomplexan genomes also encode JmjC-domain-containing proteins that may serve as histone lysine demethylases. Whereas prokaryotic expression of putative active domains of four *P. falciparum* SET proteins did not yield detectable HKMT activity towards recombinant *P. falciparum* histones, two protein domains expressed in vitro in a eukaryotic system showed HKMT activities towards H3 and H4, respectively. With the discovery of these *Plasmodium* SET- and JmjC-domain genes in the malaria parasite genomes, future efforts will be directed towards elucidation of their substrate specificities and functions in various cellular processes of the parasites.

Keywords

Plasmodium falciparum; Histone; Methyltransferase; Demethylase; Phylogeny; Lysine methylation

*Corresponding author. Liwang Cui, Department of Entomology, Penn State University, 501 ASI Building, University Park, PA 16802, USA, Tel.: +1 814 863 7663; fax: +1 814 865 3048. luc2@psu.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Among the diverse histone modifications occurring in eukaryotic cells, histone acetylation is largely linked to active genes, whereas histone lysine methylation is involved in both transcriptional activation and silencing (Li et al., 2007). Histone lysine methylation is regulated by the opposing actions of histone lysine methyltransferases (HKMTs) and histone lysine demethylases. HKMTs belong to two families of proteins: the SET [the *Drosophila* suppressor of variegation – *Su*(var) – *E*(z), the Polycomb-group protein Enhancer of zeste – *E*(z), and Trithorax (TRX) group proteins] domain-containing protein family and non-SET-domain protein DOT1 (disruptor of telomeric silencing-1) (Martin and Zhang, 2005). To date, two families of histone lysine demethylases have also been identified: lysine-specific demethylases 1 (LSD1) and Jumonji C (JmjC)-domain-containing histone demethylases (JHDMs) (Klose et al., 2007). With the finding of a large number of HKMTs and histone demethylases, the significance of histone lysine methylation in regulating various biological processes and pathways in model organisms has begun to be fully appreciated. Recent studies in apicomplexan parasites have also demonstrated the importance of histone lysine methylation in parasite development and pathogenesis (Chookajorn et al., 2007; Cui et al., 2007b; Lopez-Rubio et al., 2007; Sautel et al., 2007).

Plasmodium parasites have a complicated life cycle with two dramatically different hosts, a vertebrate and an invertebrate. In both hosts, the parasites display a tightly regulated transcription program (Bozdech et al., 2003; Le Roch et al., 2003). While basal transcription machinery such as proteins associated with RNA polymerase II is conserved (Callebaut et al., 2005), there is a paucity of recognizable specific transcription factors in the parasite genome (Aravind et al., 2003; Coulson et al., 2004). This notion of transcription factors in malaria parasites, together with the large number of of chromatin-modifying proteins, has led to speculation that epigenetics plays a prominent role in transcription regulation in the parasite. Moreover, this notion can be extended to other apicomplexan parasites such as *Toxoplasma gondii*, suggesting that epigenetics is a conserved phenomenon in the apicomplexan lineage (Sullivan and Hakimi, 2006; Gissot et al., 2007). Of the many enzymes catalyzing the reversible covalent histone modifications, the histone acetyltransferase (HAT) PfGCN5 is a key regulator of gene expression (Fan et al., 2004; Cui et al., 2007b), and represents a potential target for chemotherapy (Cui et al., 2007a). Furthermore, recent studies on antigenic switching in the parasite revealed the participation of the histone deacetylase (HDAC) Sir2 in regulating telomeric silencing and monoallelic expression of the *var* genes (Duraisingh et al., 2005; Freitas-Junior et al., 2005). These studies underline the significance of epigenetic mechanisms in *Plasmodium* gene expression.

Like other eukaryotes, the malaria parasite chromosomes have a typical nucleosomal organization involving both core and variant histones (Cary et al., 1994; Miao et al., 2006). In addition to histone acetylation, tandem mass spectrometry analysis of *Plasmodium falciparum* histones has revealed lysine and arginine methylation (Miao et al., 2006). Some of these methyl marks on histones have been shown to have evolutionarily conserved roles in transcription regulation. In malaria parasites, epigenetic markers have been investigated as a regulatory mechanism of antigenic switching. Di- and tri-methylation of H3K4 (an active gene marker) are found at the *var* promoter that is actively transcribed or poised for

transcription (Lopez-Rubio et al., 2007). In contrast, trimethylation of H3 at K9 (H3K9me3), a heterochromatin marker, has been shown to be associated with the silent *var* genes (Chookajorn et al., 2007; Lopez-Rubio et al., 2007), and is negatively correlated with global gene expression (Cui et al., 2007b). In this survey, we evaluated histone lysine methylation during *P. falciparum* development, performed in silico analysis of potential histone methyltransferases and demethylases in the parasite genomes, and speculated on their enzymatic activities within the context of phylogenetic groupings. These data will serve as the background for future inquiry into the roles of histone lysine methylation in regulating gene expression in malaria parasites.

2. Materials and methods

2.1. Identification of *P. falciparum* SET and JmjC domain proteins

To identify SET and DOT1 family HKMTs and JHDMS in malaria parasites, the *Plasmodium* genomes (<http://PlasmoDB.org>) were searched with various BLAST algorithms using the consensus SET, DOT1 and JmjC motifs as the queries (Jenuwein et al., 1998; Janzen et al., 2006; Klose et al., 2007). The search was extended to cover other available Apicomplexa genomes (<http://www.apidb.org>). A more thorough survey and analysis of the protein architecture were performed using the Simple Modular Architecture Research Tool (SMART) (<http://smart.embl.de>), which uses a Hidden Markov Model (HMM) (Letunic et al., 2006). Alignment of individual proteins as well as signature domains identified by SMART was performed by using Clustal X with visual inspection. A cladogram was generated for *Plasmodium* and apicomplexan SET- and JmjC-domain proteins using the UPGMA method implemented in the MEGA 4 program with pairwise deletions (Tamura et al., 2007). The reliability of the tree topology was assessed by bootstrap analysis.

2.2. Parasite culture

Culture of the *P. falciparum* 3D7 clone and synchronization by two rounds of 5% D-sorbitol treatment were performed as described previously (Trager and Jensen, 1976; Lambros and Vanderberg, 1979; Cui et al., 2007a,b). For time-course studies, parasites were taken at 12, 22, 32 and 42 h post synchronization to represent rings, early trophozoites, late trophozoites and schizonts as determined by microscopy. Infected erythrocytes were treated with 0.05% saponin to lyse the red blood cell membrane, the released parasites were collected by centrifugation and washed twice with cold PBS.

2.3. RNA extraction, cDNA synthesis, PCR and rapid amplification of cDNA ends (RACEs)

Total RNA was extracted from the parasites using Trizol Reagent (Invitrogen, Carlsbad, CA) and treated with RNase-free DNase I (Promega, Madison, WI) to remove contaminating genomic DNA (Cui et al., 2001). cDNA was synthesized from 1 µg of total parasite RNA using oligo-dT primer and Superscript III reverse transcriptase (RT) in a 20 µl reaction. Each cDNA reaction was diluted to 100 µl and 1 µl was used for PCR. To verify the prediction of introns in the SET domain-containing open reading frames (ORFs), primers spanning the SET-domain-containing genes were designed to amplify cDNA for sequencing.

The 5' ends of the SET- and Jmjc-domain genes were determined using the FirstChoice RNA-ligase mediated (RLM)-RACE kit (Ambion, Austin, TX), and the 3' ends were determined by RACE using oligo-dT primer and nested PCR (Fan et al., 2004). To study the temporal expression of the predicted *HKMT* genes during parasite development, real-time quantitative RT-PCR analysis was performed on Opticon DNA Engine II (BioRad, Hercules, CA) using a SYBR Green PCR kit (Roche Applied Science, Indianapolis, IN). Primers used for RT-PCR of SET-domain genes are listed in Table 1. Data analysis and determination of the threshold cycle C_t values were as described previously (Miao et al., 2006). The 2^{-C_t} method was used to calculate the relative expression levels of *HKMT* genes with the constitutively expressed *seryl-tRNA synthetase* gene (PF07_0073) as the internal reference and the C_t value of the ring stage as the calibrator (Livak and Schmittgen, 2001).

2.4. Purification of histones and Western blots

Parasite histones were purified from synchronized parasites using an acid extraction method (Miao et al., 2006). Approximately equal amounts of purified histones from different stages were separated by 18% SDS- PAGE and transferred to nitrocellulose membranes. The histones were blotted using rabbit antibodies specific for methylated H3K4, H3K9, H3K27, H3K36 and H4K20 (Millipore, Billerica, MA) as the primary antibodies and horse radish peroxidase-conjugated anti-rabbit IgG antibodies as secondary. These methylation state-specific antibodies have been determined to be specific for human histones with the respective methylation states. Sequence conservation of *P. falciparum* histones and previous applications of these antibodies also suggest their specificities for the respective methyllysines (Sautel et al., 2007; Lopez-Rubio et al., 2007). Equal loading was monitored using antibodies against histone H4. The Western blots were scanned and the relative intensities of the bands were estimated by using ImageQuant[®] Version 5.0 (Molecular Dynamics, Sunnyvale, CA).

2.5. Protein expression in bacteria

The ORFs of *P. falciparum* canonical histones H2A, H2B, H4 and the variant H2A.Z were cloned at the *NdeI* and *XhoI* sites of pET22b(+) and proteins were expressed in *Escherichia coli* strain BL21. Purification of the C-terminal His-tagged proteins under denaturing conditions and renaturing were performed as previously described for PfH3 (Miao et al., 2006). To express the SET domain proteins, putative active domains of PfSET1 [amino acids (aa) 6,433-6,753], PfSET2 (aa 1,948-2,333), PfSET3 (aa 2,096-2,399) and PfSET8 (aa 882-1,186) were cloned into pGEX-6P-1 at the *BamHI* and *SalI* sites to produce GST fusion proteins. Recombinant proteins and GST were purified from 400 ml of *E. coli* using glutathione Sepharose-4B (Amersham, Piscataway, NJ) and eluted with 50 mM Tris-HCl (pH 8.5), 5 mM MgCl₂, 4 mM DTT, 10% glycerol and 10 mM reduced glutathione.

2.6. Protein expression in an in vitro transcription/translation system

To express the recombinant PfSET domains in a eukaryotic system, we used the TNT[®] SP6 High-Yield Protein Expression System (Promega), which is based on an optimized wheat germ extract with coupled transcription/translation. The cDNA fragments of the

aforementioned PfSET1, 2, 3 and 8 domains fused with N-terminal GST or 6X His tag were cloned into the pF3A WG (BYDV) Flexi vector at *SgfI* and *PmeI* sites. In a 50 µl reaction, 5 µg of plasmid DNA were mixed with 30 µl of transcription/translation master mix and incubated at 25°C for 5 h. Vector DNA was used as a negative control. To detect recombinant protein expression, 5 µl of each reaction was analyzed by SDS-PAGE and Western blot.

2.7. HMKT assays

To detect HMKT activity, 2 µg of each purified recombinant GST fusion protein or GST, or 10 µl of each in vitro transcription/translation reaction were incubated with 4 µg of recombinant *P. falciparum* histone mixture in a 30 µl reaction containing 50 mM Tris-HCl (pH 8.5), 5 mM MgCl₂, 4 mM DTT, 1 µCi ³H-labeled S-adenosyl methionine (SAM) (82 Ci/mmol, Amersham) at 30°C for 1 h. Afterwards, 10 µl of each reaction were spotted on a Whatman P-81 filter for liquid scintillation counting (LSC), and the remainder was analyzed by 18% SDS-PAGE followed by Coomassie blue staining and fluorography (Fan et al., 2004). To determine the target site of the HKMTs, Western blots were performed using methyl-lysine-specific antibodies (Millipore).

3. Results

3.1. Identification of putative HKMTs

Among the known methylated lysines, only H3K79 is methylated by DOT1 (Feng et al., 2002). BLAST search of PlasmoDB with the conserved DOT1 sequence did not identify a DOT1 orthologue in the *Plasmodium* genomes. BLASTP analysis of the *P. falciparum* genome with a conserved SET motif only identified four annotated proteins (PFF1440w, MAP13P1.122, PF08_0012 and PFD0190w) with significant matches to the SET domain consensus (E values < 0.0005). A query of PlasmoDB with the SET domain consensus using HMM identified nine predicted proteins with a recognizable SET domain. Only one protein, PFF1440w, was annotated as a SET-domain protein. Among the nine SET-domain proteins, eight (PFF1440w, MAP13P1.122, PF08_0012, PFD0190w, PFL0690c, PFI0485c, PF13_0293 and PF11_0160) contain a SET domain recognizable by SMART, whereas PFE0400w only partially matched the SET consensus. The SET-domain genes are distributed on eight chromosomes and their sizes range from 178 to 6,753 aa (Table 2). BLASTP analysis of other *Plasmodium* genomes clearly identified the orthologues of all PfSET proteins, demonstrating their conservation and possibly functional importance in *Plasmodium* species. Based on sequence similarities of PFF1440w, MAL13P1.122 and PFD0190w with the *Saccharomyces cerevisiae* SET1 (ScSET1), ScSET2, and human SET8/Pr-SET7 proteins, these genes are named as PfSET1, PfSET2, and PfSET8, respectively (Table 2). Other PfSET genes are numbered consecutively.

Analysis of domain organization of the SET-domain proteins identified additional domains, including the bromodomain, plant homeodomain (PHD), and HDAC-interacting domain (Fig. 1). PFE0400w (PfSET9) has four ankyrin repeats near the C-terminus and located immediately upstream of the SET domain is a Cys-rich motif similar to the C₄-C₂HC consensus found in the MYND (Myeloid translocation protein 8, Nery, and DEAF-1)

domain. The presence of these protein-binding domains suggests that these PfSET proteins may form protein complexes and interact with chromatins.

3.2. The SET domain comparison

SET domain proteins can be divided into at least five subfamilies with varying specificities towards lysines (Martin and Zhang, 2005). CLUSTAL alignment of the PfSET domains with representatives of SET subfamilies is shown in Fig. 2A. Structural analysis of several SET domain HKMTs has determined that the SET domain contains mostly β -sheets with two non-contiguous regions formed by N- and C-terminal sequences referred to as SET-N and SET-C, respectively (Cheng et al., 2005). Four signature motifs are recognized, and the most conserved motifs III and IV form a pseudo knot structure. All PfSETs have a well-conserved SET-C region, where the invariant residues N, H and Y in Motifs III and IV that play catalytic roles are conserved in 7/9 PfSETs. In comparison, the SET-N regions varied considerably in most PfSETs and recognizable Motif I and II were only found in PfSET1-3 and 8 (Fig. 2A). SET-I, which connects SET-N and SET-C, inserts an α -helix into the β -sheets and its variation appears to determine substrate specificity of HKMTs (Xiao et al., 2003b; Couture et al., 2005). SET-I, which can vary greatly in length, is large in three PfSETs (Fig. 2A). In addition to the SET domain, flanking regions such as the pre-SET domain found in H3K9 HKMTs and the C-terminus are required for HKMT activities (Cheng et al., 2005). The pre-SET domain from the H3K9 HKMTs Dim-5 and Clr4 forms a triangular zinc cluster co-ordinated by nine cysteines (Min et al., 2002; Zhang et al., 2002). Only PfSET3 has a canonical pre-SET domain recognized by SMART (Fig. 1, 2B), while PfSET2 has a Cys-rich non-canonical pre-SET domain. The C-terminal flanking domain is involved in forming a narrow channel to accommodate the target lysine side chain (Cheng et al., 2005). This channel is normally formed by a zinc metal center co-ordinated by three conserved Cys residues in the post-SET domain and another Cys in motif III. The post-SET domain contains a CXCX₂₋₄C motif, a common feature in most SET-domain proteins. Consistently, for those PfSET proteins with a post-SET domain, the Cys residue in motif III is absolutely conserved (Fig. 2A). PfSET1-3 have a typical post-SET domain with a CXCX₄C motif, whereas PfSET4, 6, 7 and 9 (PFI0485c, PF13_0293, PF11_0160 and PFE0400w) have a variant post-SET motif (CXCX₂C) that is found in SMYD (SET- and MYND-domain containing) and Suv4-20 family HKMTs (Dillon et al., 2005) (Fig. 2A). For SET-domain proteins that lack the post-SET domain, this channel is formed by packing an α -helix to the active center. Among the two PfSET proteins that do not have a post-SET domain, only the C-terminal sequence of PfSET8 had significant homology to that of human SET8 (Couture et al., 2005) (Fig. 2A).

3.3. Putative substrate specificities of PfSET proteins based on phylogeny

The phylogenetic analysis included 24 *Plasmodium* SET genes and 13 HKMTs with known substrate specificities (Fig. 3). The UPGMA tree supported the existence of at least six distinct subfamilies of SET-domain proteins (Martin and Zhang, 2005), and five were found in *Plasmodium*. The E(z) subfamily members mediating H3K27 methylation in metazoans and plants were not found in protozoa such as *Plasmodium*.

PfSET1 (PFF1440w) showed significant homology to the TRX-related HKMTs (Fig. 3). This subfamily contains the *Drosophila* TRX, plant TRX homologues, MLL1 (mixed-lineage leukemia) in mammals and the *Saccharomyces cerevisiae* SET1 (ScSET1), and they methylate H3K4, a histone mark associated with transcription activation (Alvarez-Venegas and Avramova, 2002; Santos-Rosa et al., 2002). The grouping of PfSET1 and other *Plasmodium* orthologues in the TRX subfamily was supported by a strong bootstrap value. In terms of structural organization, PfSET1 is similar to human MLL1 with C-terminal SET and post-SET domains. Also, PfSET1 possesses four PHD domains, a CXXC domain and a bromodomain, which are normally found in vertebrate proteins of this SET subfamily. The bromodomain and PHD domain have been found to interact with acetylated histones and trimethylated H3K4, respectively (Mellor, 2006; Zhang, 2006). This suggests that PfSET1 may employ these interacting domains to target euchromatic regions that are marked with histone acetylation and H3K4 methylation.

PfSET2 (MAL13P1.122) is grouped with the *S. cerevisiae* SET2 subfamily with moderate bootstrap support (Fig. 3). This subfamily includes ScSET2, *Drosophila* Ash1, and mammalian NSD1 (nuclear-receptor-binding SET domain-containing protein 1). Both ScSET2 and NSD1 possess HKMT activity towards H3K36, while NSD1 also methylates H4K20 (Strahl et al., 2002; Rayasam et al., 2003). Unlike most HKMTs, Ash1 is a multi-catalytic HKMT that methylates K4 and K9 in H3, and K20 in H4 (Beisel et al., 2002). The SET domains of all three members are flanked by a non-canonical pre-SET and a typical post-SET domain. PfSET2 possesses a typical post-SET domain and a Cys-rich region N-terminal of the SET domain (eight Cys within 60 aa), which resembles a non-canonical pre-SET domain. In addition, SET2 members have other domains such as the PHD, AT-hook, PWWP and bromodomain, and PfSET2 has four PHD domains. The similarity in sequence and domain organization between PfSET2 and ScSET2 and the presence of methylated K36 in parasite histone H3 suggest that PfSET2 might be the H3K36 methylase.

PfSET3 (PF08_0012) might be the *Plasmodium* H3K9 methylase with the signature features of the Suv39 subfamily, including the C-terminal location of the SET domain and presence of conserved pre-SET and post-SET domains. PfSET3 has a typical pre-SET domain with 8/9 conserved Cys residues (Fig. 2B). Although PfSET3 and other *Plasmodium* orthologues formed a distinct clade if only the SET-domain sequences were used for phylogenetic analysis (Fig. 3), BLAST analysis clearly showed the highest sequence homology of PfSET3 with the Suv39 subfamily, especially when the pre-SET domain was included in the analysis.

PfSET8 (PF0190w) was highly homologous to human SET8/Pr-SET7 and *T. gondii* SET8 (TgSET8), both of which methylate H4K20 (Fig. 3). The homology not only exists in the SET domain, but also in the C-terminal sequence that forms the α -helix in human SET8 (Couture et al., 2005; Xiao et al., 2005) (Fig. 2A). This suggests that PfSET8 might be the H4K20 methylase in the malaria parasite (Sautel et al., 2007).

The additional five PfSET genes formed three different groups with weak bootstrap supports (Fig. 3). Whereas PfSET5 (PFL0690c) alone formed a distinct clade, PfSET4 (PFI0485c) and PfSET6 (PF13_0293) were related to the SMYD3 HKMT. The SMYD3 protein is a

H3K4-specific methylase, whose activity is enhanced by the heat-shock protein HSP90 (Hamamoto et al., 2004). Although PfSET7 (PF11_0160) and 9 (PFE0400w) formed a different clade, they may be related to SMYD3, since PfSET9 has a recognizable MYND domain. Except for PfSET5, they all have a conserved CXCX₂C post-SET domain. Interestingly, the SMYD3-related SET protein subfamily is more expanded in *T. gondii* with at least nine members (Sautel et al., 2007). So far, SMYD3 is the only member of this subfamily with characterized substrate specificity for H3K4.

3.4. Identification of JmjC-domain containing proteins in Plasmodium and other apicomplexans

BLAST analysis of *Plasmodium* genomes could not clearly identify putative LSD1 homologues, since only the riboflavin-binding region showed sequence similarity to certain *Plasmodium* genes. Search of PlasmoDB using the human JHDM3A/JMJD2A (jumonji domain-containing 2A) and five *S. cerevisiae* JmjC domains identified two JmjC-domain proteins in the *P. falciparum* genome, MAL8P1.111 and PFF0135w, referred to as PfJmjC1 and PfJmjC2, respectively. Orthologues of PfJmjC1 are present in *Plasmodium vivax* and *Plasmodium knowlesi* genomes, but absent in the rodent malaria lineage (Fig. 4, 5). In contrast, PfJmjC2 is present in all analyzed *Plasmodium* lineages. In several other apicomplexan taxa with available genome sequences, PfJmjC1 homologues were found in *Toxoplasma*, *Babesia* and *Theileria*, but not in *Cryptosporidium* (Fig. 4, 5), while PfJmjC2 homologues were identified in *Toxoplasma* and *Theileria*. Phylogenetic analysis of the JmjC domains clearly separated JmjC1 and 2 homologues into two distinct groups (Fig. 5). Among the seven JmjC-domain protein subfamilies, six subfamilies contain additional protein domains (e.g., JmjN, tudor, PHD, zinc finger) (Klose et al., 2007). While JmjC domain alone is sufficient for the demethylase activity in certain JmjC proteins (Tsukada et al., 2006), the JmjN domain, a ~40 aa motif usually found immediately before the JmjC domain, is required for enzymatic activity in JHDM3/JMJD2A even though it does not appear involved in catalysis as viewed from the protein structure (Klose et al., 2006). Phylogenetic analysis using the JmjC domain alone suggests that Apicomplexa JmjC1 proteins were most related to the JHDM3/JMJD2 and JmjC/AT-rich interactive domain (JARID) subfamilies, the only two JHDM subfamilies containing a JmjN domain (Fig. 5). Similarly, Apicomplexa JmjC1 proteins all possess a JmjN domain preceding the JmjC domain but lack other recognizable domains. In this context, the apicomplexan JmjC1 proteins are more related to the JHDM3/JMJD2 subfamily, since the JARID subfamily tends to have additional domains inserted between the JmjN and JmjC domains (Klose et al., 2007). In comparison, the JmjC2 proteins have no additional domains or sequence similarity to any JHDM subfamilies, suggesting that they might be related to the JmjC-domain-only subfamily, which includes numerous orphan members (Klose et al., 2007). Like other chromatin-modifying gene families, the JmjC-domain genes were expanded in *Toxoplasma* with at least seven members (Fig. 4, 5). The five additional TgJmjC proteins are divided into two clades with no close similarity to any characterized JmjC-domain proteins.

JHDMS reverse lysine methylation by an oxidative reaction requiring Fe (II) and α -ketoglutarate (α -KG) as cofactors (Tsukada et al., 2006). Structural analysis of JHDM3/JMJD2A has revealed that the JmjC domain forms eight antiparallel β strands known as the

jelly-roll fold (Chen et al., 2006). All residues involved in Fe (II) binding (HxD/Ex_nH) and α -KG binding (N and K) in the JmjC domain were highly conserved in the Apicomplexa JmjC-domain proteins (Fig. 4). Substrate binding of JHDM3A/JMJD2A is achieved through an unusual hydrogen (CH-O) bond in a methylammonium-binding pocket comprised of conserved G, Y and E/D (Couture et al., 2007; Ng et al., 2007). An S to A substitution in the substrate-binding pocket appears to modulate the methylation-state specificity of the JDJM2 subfamily (Couture et al., 2007; Ng et al., 2007). In apicomplexan JmjC1 proteins, the putative substrate-binding pocket contains the invariant G, Y and E/D residues and the S to A substitution (Fig. 4). Based on the similarity in domain organization and conservation in residues essential for JHDM3A activity, we speculate that apicomplexan JmjC1 proteins might be H3K36 and K9 demethylases with higher activity for H3K9 due to the S-A substitution. Interestingly, other apicomplexan JmjC-domain proteins contain conserved residues involved in F2 (II) and α -KG binding, but not residues in the methyl-lysine substrate-binding pocket in human JHDM3A, implying that they may have different substrate specificities.

3.5. Expression of PfSET and JmjC genes during asexual erythrocytic development

Microarray analysis detected the expression of all PfSET genes during erythrocytic development. To determine their expression levels, we performed real-time quantitative RT-PCR analysis of eight *PfSET* genes (Fig. 6). The results showed that *PfSET1*, 5 and 7 had a relatively constitutive pattern of expression, whereas *PfSET4* and 8 expression increased during the erythrocytic cycle and peaked at the schizont stage. In particular, *PfSET8* had a nearly four-fold increase in schizonts over ring stage. In contrast, *PfSET6* had the highest level of expression in ring and early trophozoite stages. *PfSET2* had similar levels of expression during early development, but its expression was reduced in late trophozoites and elevated again in schizonts. Expression of these seven *PfSET* genes was more or less compatible with the data from microarray analysis (Bozdech et al., 2003; Le Roch et al., 2003). However, RT-PCR analysis detected a more than seven-fold increase in *PfSET3* expression in schizonts, which was quite different from the relatively constitutive expression pattern obtained from microarray analysis (Fig. 6). Microarray analysis showed that PfJmjC1 is expressed in the whole erythrocytic cycle with peak expression in schizonts (Bozdech et al., 2003; Le Roch et al., 2003).

To verify the intron prediction from PlasmoDB, we designed primers across the predicted introns in PfSET genes, amplified the cDNAs from mixed asexual stages and compared those with the genomic sequences. Most of the sequences agreed with the predictions, whereas intron 2 of PfSET1 was 24 bp longer at the 5' end than predicted, thus encoding a protein of 6,753 aa (Table 2). To determine the 5' and 3' end of the SET- and JmjC-domain genes, RACEs were performed using freshly isolated total RNA from asexual stages (Fan et al., 2004). The 5' RLM-RACE results indicated that most of these genes used more than one putative transcriptional initiation site (TIS) (Table 2), consistent with the results from transcriptomic analysis (Watanabe et al., 2002). Whereas the major TIS site(s) revealed a relatively large 5' untranslated region (UTR) for PfSET1, 5, 6, and PfJmjC1 (> 268 bp), the putative 5' UTRs of PfSET3, 4, 7 and PfJmjC2 were relatively short (mostly < 172 bp). At present, these short 5' UTRs could be considered putative, since the actual ATG sites may

lie further downstream from the predicted ATGs and need to be determined experimentally. Similarly, the 3' RACE results identified more than one polyadenylation site clustered within a short region (2-70 bp) for most PfSET genes, giving putative 3' UTRs of 65 – 537 bp (Table 2). The polyadenylation sites of PfSET5 were clustered at two separated regions, while those of PfSET6 were spread in a more extended 120 bp region. In addition to the clustering of major polyadenylation sites, the PfJmjC genes had other minor sites located much further downstream. These results further verified that the *PfSET* and *PfJmjC* genes were expressed in the asexual erythrocytic life cycle of the parasite.

3.5. Enzymatic activity

The recombinant *P. falciparum* core histones and H2A.Z were expressed in *E. coli* and purified to homogeneity as shown in the SDS-PAGE gel (Fig. 7A). To investigate whether PfSET1, 2, 3 and 8 are true HKMTs, we expressed and purified the four PfSET domains from bacteria, but they had no detectable HKMT activities on recombinant *P. falciparum* histones (data not shown). To test whether the lack of activity of bacterially-expressed PfSET domains was due to improper folding, these domains were expressed in vitro using the wheat germ protein expression system. Western blots detected recombinant protein expression (Fig. 7B), but the small amounts of protein expressed did not allow further purification. Therefore, the in vitro translation mix was used directly for HKMT assays. Whereas the vector control reaction had no HKMT activity, two of the four PfSET domains had detectable HKMT activity by LSC and fluorography (Fig. 7C). PfSET1 and 3 domains did not show any detectable HKMT activities. The PfSET8 domain showed strong HKMT activity towards histone H4, whereas the HKMT activity of PfSET2 domain for H3 was weak and visible only after prolonged exposure of the film. Using the current expression system, we only obtained enough PfSET8 for the substrate specificity assay. Consistent with the prediction from phylogenetic analysis, Western blots detected H4K20me1, H4K20me2 and H4K20Me3 with recombinant PfSET8 but not in reactions with control and -SAM (Fig. 7D), indicating PfSET8 indeed methylated H4K20.

3.6. Histone methylation in *P. falciparum*

In core histones, lysine methylation occurs mainly at six positions, i.e. K4, K9, K27, K36, K79 of H3, and H4K20 (Lachner et al., 2003). We previously identified H3K14, H4K5 and H4K16 methylation in *P. falciparum* by tandem mass spectrometry (Miao et al., 2006), but these methylation states have not been characterized in model systems and their functional importance remains unclear. Western blots showed that antibodies specific for H3K4Me3, H3K4Me2, H3K9Me3, H3K36Me3, H4K20Me1 and H4K20Me2 reacted with histones H3 and H4 purified from the parasites (Fig. 8), but not recombinant *P. falciparum* histones expressed in *E. coli* (data not shown). Consistently, immunofluorescence also detected H4K20Me3 in *P. falciparum* schizonts (Sautel et al., 2007). Antibodies against dimethylated H3K27 (H3K27Me2) did not react with parasite histones, consistent with the lack of this histone mark in protozoa. Most antibodies only reacted with one histone, whereas antibodies for H3K4Me2 and H4K20Me also reacted weakly with additional histones (Fig. 8). Despite divergence of the residues flanking the *Plasmodium* H3K36, the anti-H3K36 antibodies did react with PfH3, but it is unknown whether these antibodies could accurately detect the methylation states of H3K36. From Western blots, we did not

observe obvious quantitative changes of H3K36Me3 during development. In comparison, H3K4Me2 and Me3 both displayed peak levels at late trophozoite stage. Similarly, the heterochromatin and silent gene markers H3K9Me3, H4K20Me1 and H4K20Me2 showed an increase during later stages of erythrocytic schizogony. Except for the constitutive H3K36Me3 profile, the methylation patterns at H3K4, H3K9 and H4K20 were congruent with the gene expression profiles of *PfSET1*, 3 and 8, respectively. These results indicate that most histone lysine methylation in *P. falciparum* undergoes dynamic changes during development.

4. Discussion

We have analyzed nine expressed SET-domain genes and two JmjC-domain genes from *P. falciparum* and predicted their enzymatic specificities based on similarities in sequence and domain organization with known family members. The orthologues of PfSET genes and PfJmjC2 are present in all sequenced *Plasmodium* genomes, suggesting the *Plasmodium* ancestor had acquired these genes before speciation and divergence. However, PfJmjC1 orthologues were not found in the sequenced rodent genomes, suggestive of the result of secondary loss of this gene in this branch. The existence of these putative methylase and demethylase homologues in other sequenced Apicomplexa taxa suggests that they may be of functional significance. Interestingly, many histone modification enzymes including HATs, HKMTs, and JHDMs have undergone considerable expansion in closely-related *T. gondii*, and certain paralogues such as TgGCN5 have even acquired unusual substrate specificities and perhaps novel functions (Saksouk et al., 2005). In any case, the large spectrum of SET-domain proteins present in apicomplexan parasites may not merely be the result of gene duplication events. For example, the existence in Apicomplexa of a subfamily of SET proteins phylogenetically related to the mammalian SET8, and absence of this subfamily in the available sequences of other Alveolate taxa such as the ciliates and dinoflagellates suggest that this SET subfamily might have been acquired by the Apicomplexa ancestor via lateral gene transfer from their vertebrate hosts (Sautel et al., 2007).

Unlike other histone modification enzymes, HKMTs and JHDMs possess narrow substrate specificities, often targeting a single lysine within the respective substrates. These enzymes also differ in their preferences for different methylation states (mono-, di-, or tri-methylation) of lysines. In spite of the conserved overall structural features, SET domain HKMTs display considerable structural plasticity even with variations at the active sites, which may contribute to their varying substrate specificities. Even closely-related members of the same subfamily may display subtle differences in enzymatic activity. For instance, the human SET8 is a monomethylase with a preference for nucleosomal histones (Nishioka et al., 2002; Xiao et al., 2005), whereas TgSET8 displays H4K20 mono-, di-, and tri-methylase activity (Sautel et al., 2007). This, together with the complexity of histone lysine modifications, requires the existence of a large number of SET-domain genes. However, the presence in *Plasmodium* genomes of 1-2 JmjC-domain proteins, which may also be substrate lysine-specific, makes it difficult to speculate whether and how JmjC-domain proteins are involved in controlling the dynamics of histone lysine methylation in the parasites. It remains unknown whether i) the JmjC-domain proteins in malaria parasites have broader substrate specificity; ii) there are other yet unidentified histone lysine demethylases;

or iii) methylation at certain lysines is primarily regulated by the dynamics of SET-domain proteins. Based on phylogeny, we predicted that PfSET1, 2, 3 and 8 are homologous to SET subfamilies that methylate histones at H3K4, K36, K9 and H4K20, respectively. This prediction is consistent with the detection of methylation at these four sites in *P. falciparum*. However, we could only substantiate this prediction with confirmed methylase activity of recombinant PfSET8 for H4K20 and recombinant PfSET2 for histone H3. While lack of methylase activity for these recombinant PfSETs expressed in bacteria could be attributed to improper folding of the protein domains, the lack of enzymatic activity of PfSET1 and 3 expressed in a wheat germ extract system suggests that additional protein domains, interacting partners to form complexes, or different substrates (e.g., nucleosomal versus core histones) may be needed for enzymatic activity. For example, although the JmjN domain does not appear to have a catalytic role in the human JMJD2A, it is required for enzymatic activity (Chen et al., 2006). Also, the HAT activity of the *S. cerevisiae* SAS2 MYST family protein for H4K16 absolutely requires an interacting partner (Sutton et al., 2003). Further, it is intriguing that even closely-related species in Apicomplexa may reveal additional differences. Although recombinant PfSET8 and TgSET8 both methylated H4K20, PfSET8 expressed in *E. coli* had no enzymatic activity whereas the bacterially-expressed TgSET8 was active (Sautel et al., 2007). This difference may be due to sequence divergence between PfSET8 and TgSET8, as they only share 35% identity (55% similarity) in the conserved SET domain. Regardless of the predicted enzymatic activities for PfSET1-3, the large size of the proteins and presence of unique domains may indicate functional differences between these enzymes and their counterparts in other model organisms.

As important constituents that make up the complex “histone code”, methylation states of different lysines impart diverse biological functions (Martin and Zhang, 2005). In general, methylation at K4, K36 and K79 of H3 is associated with transcriptionally active regions, whereas methylation at H3K9, H3K27 and H4K20 is associated with transcriptionally silenced chromatin (Martin and Zhang, 2005). This study revealed the dynamically regulated, complex methylation patterns and states of core histones in *P. falciparum*. While functional significance of the entire parasite epigenome has yet to be explored, parallel studies in *T. gondii* suggest that the apicomplexan parasites have a typical organization of hetero- and eu-chromatin and the histone marks play similar roles in gene regulation (Gissot et al., 2007). In *T. gondii*, H3R17 methylation has been attributed to gene activation (Saksouk et al., 2005), whereas the cell cycle-regulated H4K20 methylation is associated with pericentric and telomeric heterochromatin (Sautel et al., 2007). Furthermore, the focal locations of methyl-H4K20 in *P. falciparum* are also reminiscent of telomeric positions, suggesting that SET8 homologues and H4K20 methylation play similar roles in heterochromatic functions in apicomplexan parasites (Sautel et al., 2007). Consistently, findings in *P. falciparum* indicate that H3K9Me3 represents an epigenetic memory mark of silent genes in *P. falciparum* (Chookajorn et al., 2007; Cui et al., 2007b; Lopez-Rubio et al., 2007), whereas H3K4 methylation is associated with active promoters or those poised for transcription (Lopez-Rubio et al., 2007). With the identification of an array of potential HKMTs and JHDMs, an exciting challenge lies ahead, as target specificities and functions of these proteins in shaping the epigenome and orchestrating the development program of the parasites remain to be determined experimentally.

Acknowledgements

We would like acknowledge financial support from NIAID, National Institutes of Health (R01 AI064553).

References

- Alvarez-Venegas R, Avramova Z. SET-domain proteins of the Su(var)3-9, E(z) and trithorax families. *Gene*. 2002; 285:25–37. [PubMed: 12039029]
- Aravind L, Iyer LM, Wellemes TE, Miller LH. *Plasmodium* biology: genomic gleanings. *Cell*. 2003; 115:771–785. [PubMed: 14697197]
- Beisel C, Imhof A, Greene J, Kremmer E, Sauer F. Histone methylation by the *Drosophila* epigenetic transcriptional regulator Ash1. *Nature*. 2002; 419:857–862. [PubMed: 12397363]
- 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:E5. [PubMed: 12929205]
- 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. [PubMed: 16042788]
- Cary C, Lamont D, Dalton JP, Doerig C. *Plasmodium falciparum* chromatin: nucleosomal organisation and histone-like proteins. *Parasitol. Res*. 1994; 80:255–258. [PubMed: 8036241]
- Chen Z, Zang J, Whetstone J, Hong X, Davrazou F, Kutateladze TG, Simpson M, Mao Q, Pan CH, Dai S, Hagman J, Hansen K, Shi Y, Zhang G. Structural insights into histone demethylation by JMJD2 family members. *Cell*. 2006; 125:691–702. [PubMed: 16677698]
- Cheng X, Collins RE, Zhang X. Structural and sequence motifs of protein (histone) methylation enzymes. *Ann. Rev. Biophys. Biomol. Struct*. 2005; 34:267–294. [PubMed: 15869391]
- Chookajorn T, Dzikowski R, Frank M, Li F, Jiwani AZ, Hartl DL, Deitsch KW. Epigenetic memory at malaria virulence genes. *Proc. Natl. Acad. Sci. USA*. 2007; 104:899–902. [PubMed: 17209011]
- Coulson RM, Hall N, Ouzounis CA. Comparative genomics of transcriptional control in the human malaria parasite *Plasmodium falciparum*. *Genome Res*. 2004; 14:1548–1554. [PubMed: 15256513]
- Couture J-F, Collazo E, Brunzelle JS, Trievel RC. Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase. *Genes Dev*. 2005; 19:1455–1465. [PubMed: 15933070]
- Couture J-F, Collazo E, Ortiz-Tello PA, Brunzelle JS, Trievel RC. Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat. Struct. Mol. Biol*. 2007; 14:689–695. [PubMed: 17589523]
- Cui L, Rzomp KA, Fan Q, Martin SK, Williams J. *Plasmodium falciparum*: differential display analysis of gene expression during gametocytogenesis. *Exp. Parasitol*. 2001; 99:244–254. [PubMed: 11888252]
- Cui L, Miao J, Cui L. Cytotoxic effect of curcumin on malaria parasite *Plasmodium falciparum*: inhibition of histone acetylation and generation of reactive oxygen species. *Antimicrob. Agents Chemother*. 2007a; 51:488–494. [PubMed: 17145789]
- Cui L, Miao J, Furuya T, Li X, Su X, Cui L. PfGCN5 mediated histone H3 acetylation plays a key role in gene expression in *Plasmodium falciparum*. *Eukaryot. Cell*. 2007b; 6:1219–1227. [PubMed: 17449656]
- Dillion SC, Zhang X, Trievel RC, Cheng X. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol*. 2005; 6:227. [PubMed: 16086857]
- Duraisingh MT, Voss TS, Marty AJ, Duffy MF, Good RT, Thompson JK, Freitas-Junior LH, Scherf A, Crabb BS, Cowman AF. Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell*. 2005; 121:13–24. [PubMed: 15820675]
- Fan Q, An L, Cui L. *Plasmodium falciparum* histone acetyltransferase, a yeast GCN5 homologue involved in chromatin remodeling. *Eukaryot. Cell*. 2004; 3:264–276. [PubMed: 15075257]
- Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, Zhang Y. Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol*. 2002; 12:1052–1058. [PubMed: 12123582]

- Freitas-Junior LH, Hernandez-Rivas R, Ralph SA, Montiel-Condado D, Ruvalcaba-Salazar OK, Rojas-Meza AP, Mancio-Silva L, Leal-Silvestre RJ, Gontijo AM, Shorte S, Scherf A. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell*. 2005; 121:25–36. [PubMed: 15820676]
- Gissot M, Kelly KA, Ajioka JW, Grealley JM, Kim K. Epigenomic modifications predict active promoters and gene structure in *Toxoplasma gondii*. *PLoS Pathog*. 2007; 3:e77. [PubMed: 17559302]
- Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, Yagyu R, Nakamura Y. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat. Cell Biol*. 2004; 6:731–740. [PubMed: 15235609]
- Janzen CJ, Hake SB, Lowell JE, Cross GA. Selective di- or trimethylation of histone H3 lysine 76 by two DOT1 homologs is important for cell cycle regulation in *Trypanosoma brucei*. *Mol. Cell*. 2006; 23:497–507. [PubMed: 16916638]
- Jenuwein T, Laible G, Dorn R, Reuter G. SET domain proteins modulate chromatin domains in eu- and heterochromatin. *Cell. Mol. Life Sci*. 1998; 54:80–93. [PubMed: 9487389]
- Klose RJ, Kallin EM, Zhang Y. JmjC-domain-containing proteins and histone demethylation. *Nature Rev. Genet*. 2007; 7:715–727. [PubMed: 16983801]
- Klose RJ, Yamane K, Bae Y, Zhang D, Erdjument-Bromage H, Tempst P, Wong J, Zhang Y. The transcriptional repressor JHDM3A demethylates trimethyl histone H3 lysine 9 and lysine 36. *Nature*. 2006; 442:312–316. [PubMed: 16732292]
- Lachner M, O'Sullivan RJ, Jenuwein T. An epigenetic road map for histone lysine methylation. *J. Cell Sci*. 2003; 116:2117–2124. [PubMed: 12730288]
- Lambros C, Vanderberg JP. Synchronization of *Plasmodium falciparum* erythrocytic stages in culture. *J. Parasitol*. 1979; 65:418–420. [PubMed: 383936]
- 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; 301:1503–1508. [PubMed: 12893887]
- Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P. SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res*. 2006; 34:D257–260. [PubMed: 16381859]
- Li B, Carey M, Workman JL. The role of chromatin during transcription. *Cell*. 2007; 128:707–719. [PubMed: 17320508]
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-C(T)}$ method. *Methods*. 2001; 25:402–408. [PubMed: 11846609]
- Lopez-Rubio JJ, Gontijo AM, Nunes MC, Issar N, Rivas RH, Scherf A. 5' flanking region of *var* genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol. Microbiol*. 2007; 66:1296–1305. [PubMed: 18028313]
- Martin C, Zhang Y. The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell. Biol*. 2005; 6:838–849. [PubMed: 16261189]
- Mellor J. It takes a PHD to read the histone code. *Cell*. 2006; 126:22–24. [PubMed: 16839870]
- Miao J, Fan Q, Cui L, Li J, Li J, Cui L. The malaria parasite *Plasmodium falciparum* histones: organization, expression, and acetylation. *Gene*. 2006; 369:53–65. [PubMed: 16410041]
- Min J, Zhang X, Cheng X, Grewal SI, Xu RM. Structure of the SET domain histone lysine methyltransferase Clr4. *Nat. Struct. Biol*. 2002; 9:828–832. [PubMed: 12389037]
- Ng SS, Kavanagh KL, McDonough MA, Butler D, Pilka ES, Lienard BM, Bray JE, Savitsky P, Gileadi O, von Delft F, Rose NR, Offer J, Scheinost JC, Borowski T, Sundstrom M, Schofield CJ, Oppermann U. Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature*. 2007; 448:87–91. [PubMed: 17589501]
- Nishioka K, Rice JC, Sarma K, Erdjument-Bromage H, Werner J, Wang Y, Chuikov S, Valenzuela P, Tempst P, Steward R, Lis JT, Allis CD, Reinberg D. PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol. Cell*. 2002; 9:1201–1213. [PubMed: 12086618]
- Rayasam GV, Wendling O, Angrand PO, Mark M, Niederreither K, Song L, Lerouge T, Hager GL, Chambon P, Losson R. NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *EMBO J*. 2003; 22:3153–3163. [PubMed: 12805229]

- Saksouk N, Bhatti MM, Kieffer S, Smith AT, Musset K, Garin J, Sullivan WJ Jr. Cesbron-Delauw MF, Hakimi MA. Histone-modifying complexes regulate gene expression pertinent to the differentiation of the protozoan parasite *Toxoplasma gondii*. *Mol. Cell. Biol.* 2005; 25:10301–10314. [PubMed: 16287846]
- Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, Schreiber SL, Mellor J, Kouzarides T. Active genes are tri-methylated at K4 of histone H3. *Nature.* 2002; 419:407–411. [PubMed: 12353038]
- Sautel CF, Cannella D, Bastien O, Kieffer S, Aldebert D, Garin J, Tardieux I, Belrhali H, Hakimi MA. SET8-mediated methylations of histone H4 lysine 20 mark silent heterochromatic domains in apicomplexan genomes. *Mol. Cell. Biol.* 2007; 27:5711–5724. [PubMed: 17562855]
- Strahl BD, Grant PA, Briggs SD, Sun ZW, Bone JR, Caldwell JA, Mollah S, Cook RG, Shabanowitz J, Hunt DF, Allis CD. Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Mol. Cell. Biol.* 2002; 22:1298–1306. [PubMed: 11839797]
- Sullivan WJ Jr. Hakimi MA. Histone mediated gene activation in *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* 2006; 148:109–116. [PubMed: 16644030]
- Sutton A, Shia WJ, Band D, Kaufman PD, Osada S, Workman JL, Sternglanz R. Sas4 and Sas5 are required for the histone acetyltransferase activity of Sas2 in the SAS complex. *J Biol Chem.* 2003; 278:16887–16892. [PubMed: 12626510]
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 2007; 24:1596–1599. [PubMed: 17488738]
- Trager W, Jensen JB. Human malaria parasites in continuous culture. *Science.* 1976; 193:673–675. [PubMed: 781840]
- Tsukada Y, Fang J, Erdjument-Bromage H, Warren ME, Borchers CH, Tempst P, Zhang Y. Histone demethylation by a family of JmjC domain-containing proteins. *Nature.* 2006; 439:811–816. [PubMed: 16362057]
- Watanabe J, Sasaki M, Suzuki Y, Sugano S. Analysis of transcriptomes of human malaria parasite *Plasmodium falciparum* using full length enriched library: identification of novel genes and diverse transcription start sites of messenger RNAs. *Gene.* 2002; 291:105–113. [PubMed: 12095684]
- Xiao B, Jing C, Kelly G, Walker PA, Muskett FW, Frenkiel TA, Martin SR, Sarma K, Reinberg D, Gamblin SJ, Wilson JR. Specificity and mechanism of the histone methyltransferase Pr-Set7. *Genes Dev.* 2005; 19:1444–1454. [PubMed: 15933069]
- Xiao B, Wilson JR, Gamblin SJ. SET domains and histone methylation. *Curr. Opin. Struct. Biol.* 2003b; 13:699–705. [PubMed: 14675547]
- Zhang Y. It takes a PHD to interpret histone methylation. *Nature Struct. Mol. Biol.* 2006; 13:572–574. [PubMed: 16826231]
- Zhang X, Tamaru H, Khan SI, Horton JR, Keefe LJ, Selker EU, Cheng X. Structure of the *Neurospora* SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell.* 2002; 111:117–127. [PubMed: 12372305]

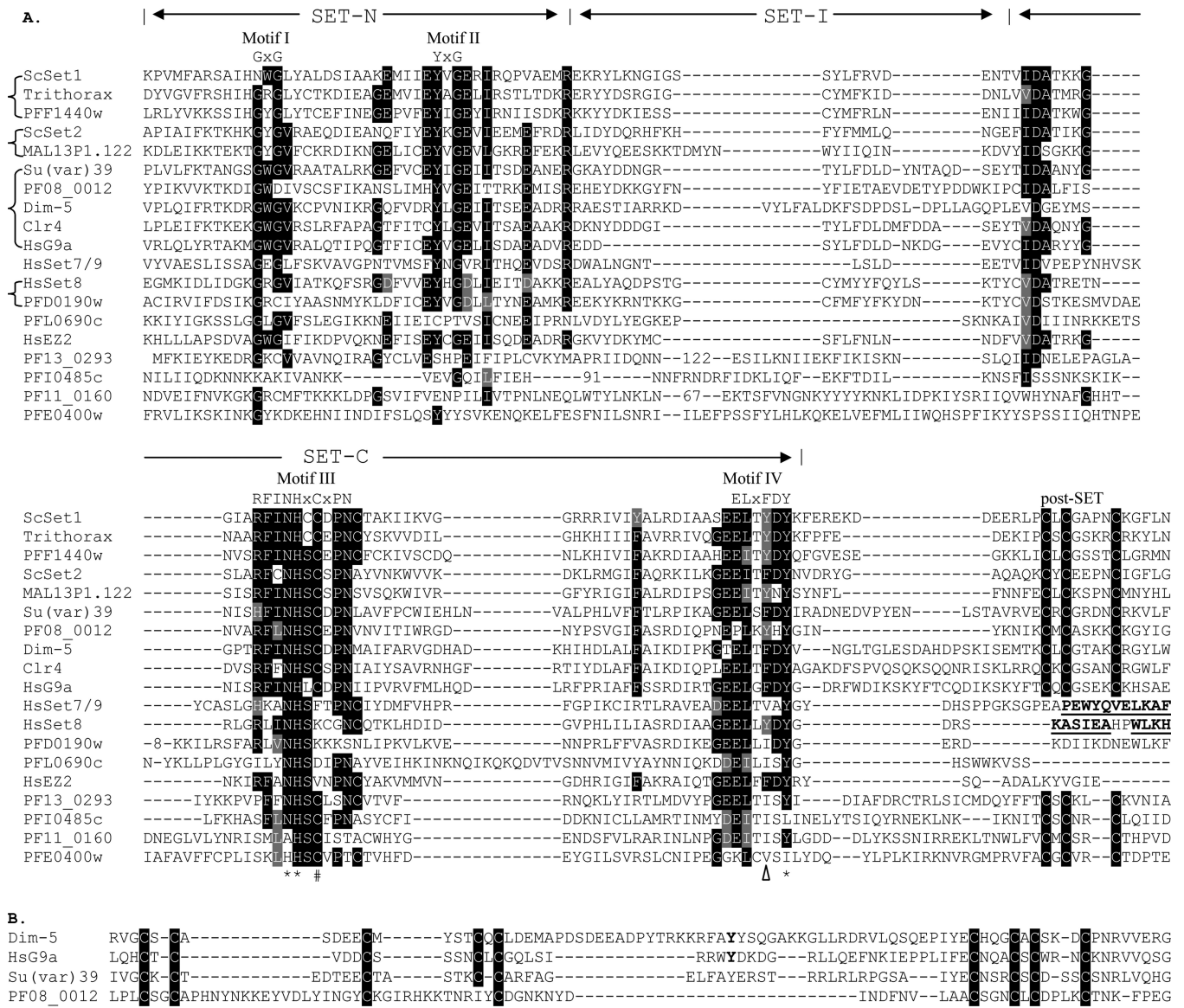
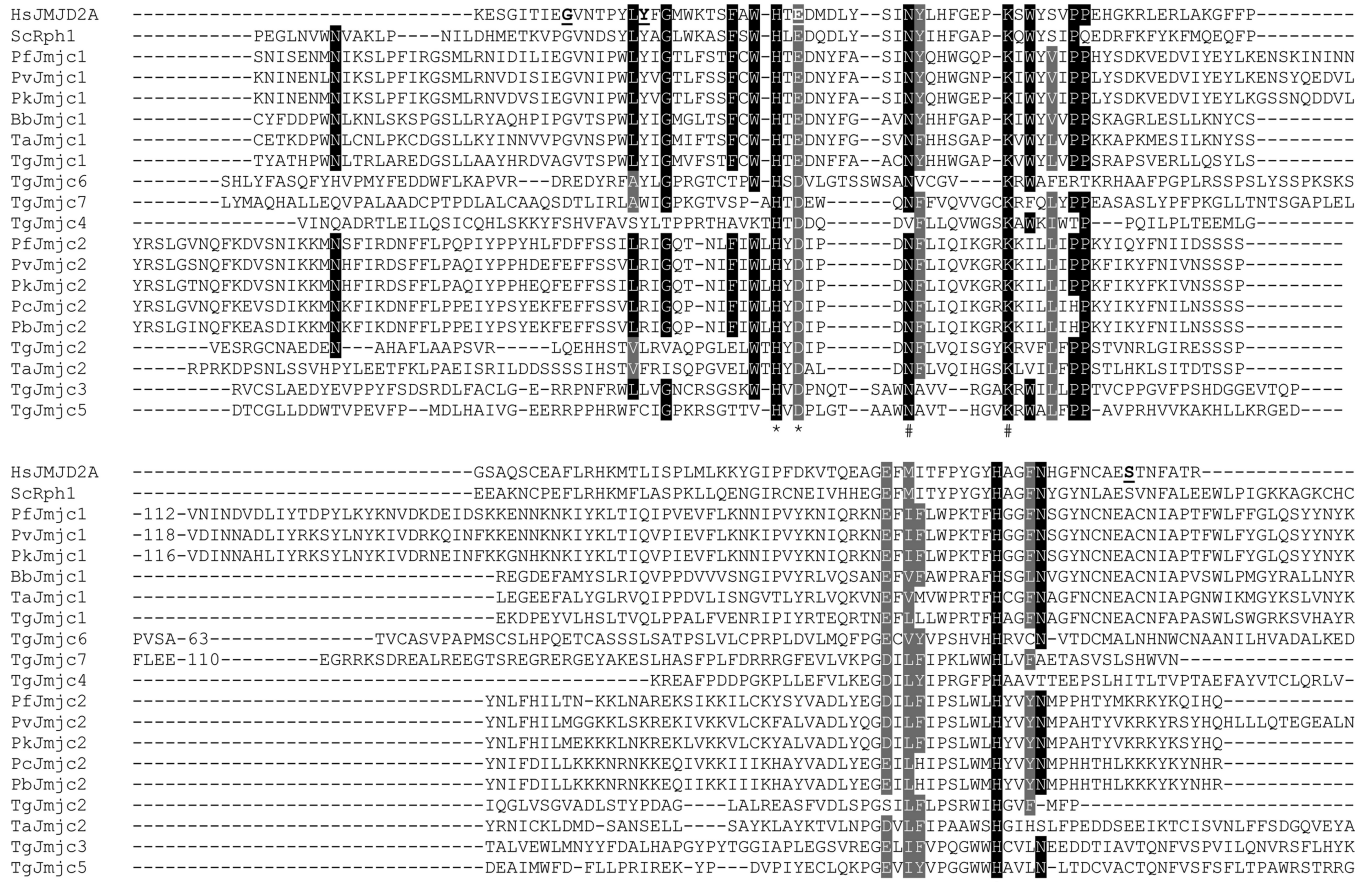


Fig. 2. Sequence comparison of the SET and pre-SET domains. A) Alignment of SET domains of PfSETs with representatives of SET histone lysine methyltransferases. The three regions are indicated as SET-N, SET-I and SET-C. Signature motifs I-IV with the motif consensus sequences are indicated above the alignment. Numbers in the SET-I region indicate deleted residues for optimal alignments. Identical residues are shaded in black and conserved substitutions are shaded in grey. Three residues with catalytic roles are indicated with asterisks (*). The F/Y switch that results in product specificity of human SET7/9 and ScSET1 is indicated with a triangle. The Cys residue in motif III that is involved in the Zn cluster formation with the post-SET domain is indicated with a hash symbol (#). The C-terminal sequences that form an α -helical structure in human SET7/9 and SET8 are in bold and underlined. Representatives of different SET families from model organisms include ScSET1 (GenBank # [EDN62358](#)), ScSET2 ([#P46995](#)), *Drosophila* TRX ([#CAA83515](#)) and

Su(var)39 (**#NP_524357**), *Neurospora crassa* Dim-5 (**#AAL35215**), *Schizosaccharomyces pombe* Clr4 (**#CAA22283**), HsG9a (**#CAA49491**), HsSET7/9 (**#AAI21056**), HsSET8 (**#AAM47033**), and HsEZ2 (**#NP_004447**). Abbreviations: *Saccharomyces cerevisiae* – Sc, *Homo sapiens* – Hs. B) Alignment of the pre-SET domain in PfSET3 and other members of the Su(var)39 subfamily with the Cys residues highlighted.

Fig. 4

**Fig. 4.**

Alignment of Apicomplexa JmjC-domain proteins with two representatives from a major subfamily. Numbers in the sequences indicate residues deleted to optimize the alignment. Residues involved in Fe (II) binding are indicated by an asterisk (*) and residues involved in α -KG binding are indicated by a hash symbol (#). Apicomplexan genes include those from PlasmoDB: PfJmjC1 (MAL8P1.111), PfJmjC2 (PFF0135w), PvJmjC1 (PV123280), PvJmjC2 (Pv113305), PkJmjC1 (PKH_14270), PkJmjC2 (PKH_114710), PbJmjC2 (PB301508.00.0) and PcJmjC2 (PC000228.03.0); from ToxoDB: TgJmjC1 (542.m00225), TgJmjC2 (55.m04820), TgJmjC3 (76.m01561), TgJmjC4 (49.m03238), TgJmjC5 (55.m04927), TgJmjC6 (42.m05839), and TgJmjC7 (28.m00292) and from the GenBank: BbJmjC1 (#XP_001609219), TaJmjC1 (#XP_954636), and TaJmjC2 (#XP_952247), HsJMJD2A/JHDM3 (#KIAA0677) and ScRph1 (#P39956). Abbreviations: *Plasmodium knowlesi* – Pk, *Plasmodium berghei* – Pb, *Plasmodium chabaudi* – Pc, *Toxoplasma gondii* – Tg, *Babesia bovis* – Bb, *Theileria annulata* – Ta.

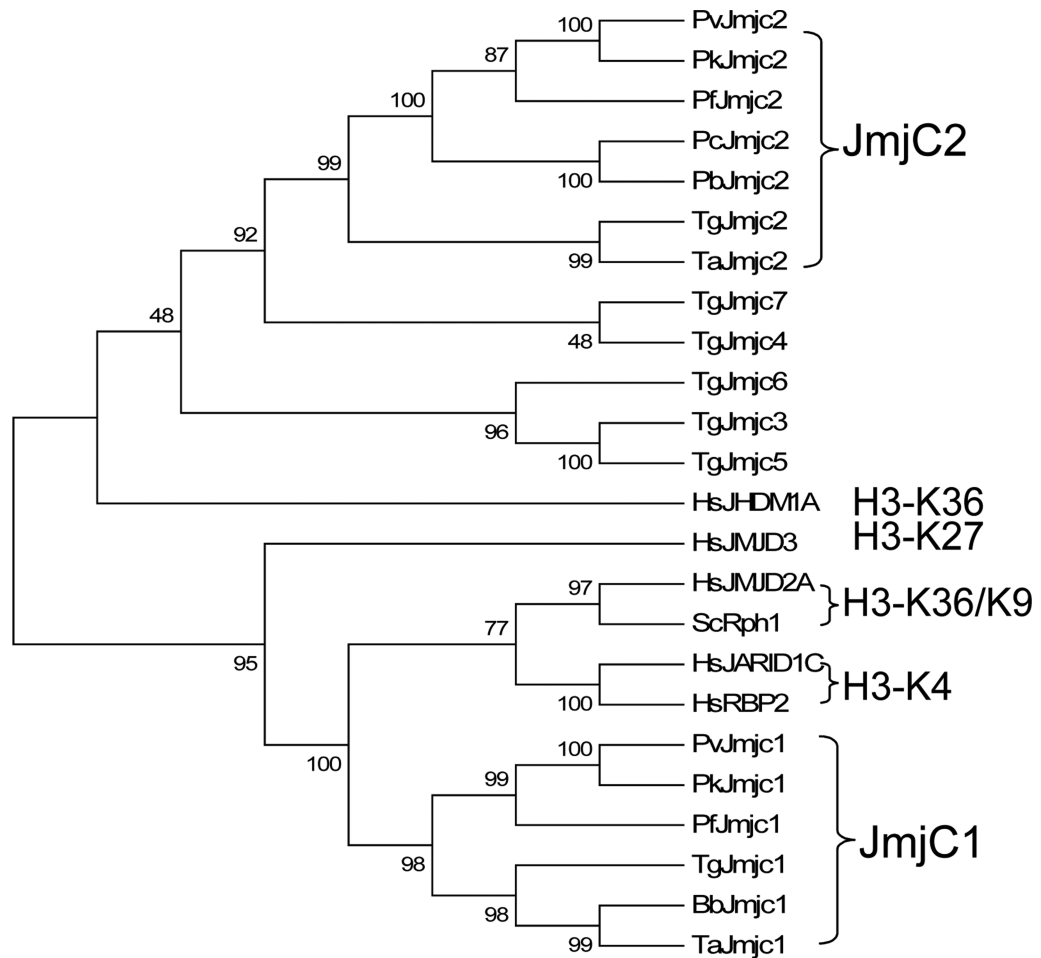


Fig. 5.

Phylogenetic comparison of JmjC-domain proteins. Included in the analysis are the JmjC domains shown in Fig. 4 together with HsJHDM1A (#Q9Y2K7), HsJMJD3 (O15054), HsJARID1C (CAI39837) and HsRBP2 (AAB28544). Shown here is a UPGMA tree with reliability of the tree topology assessed by bootstrap analysis. Substrate specificities are indicated. Abbreviations are as in Fig. 4.

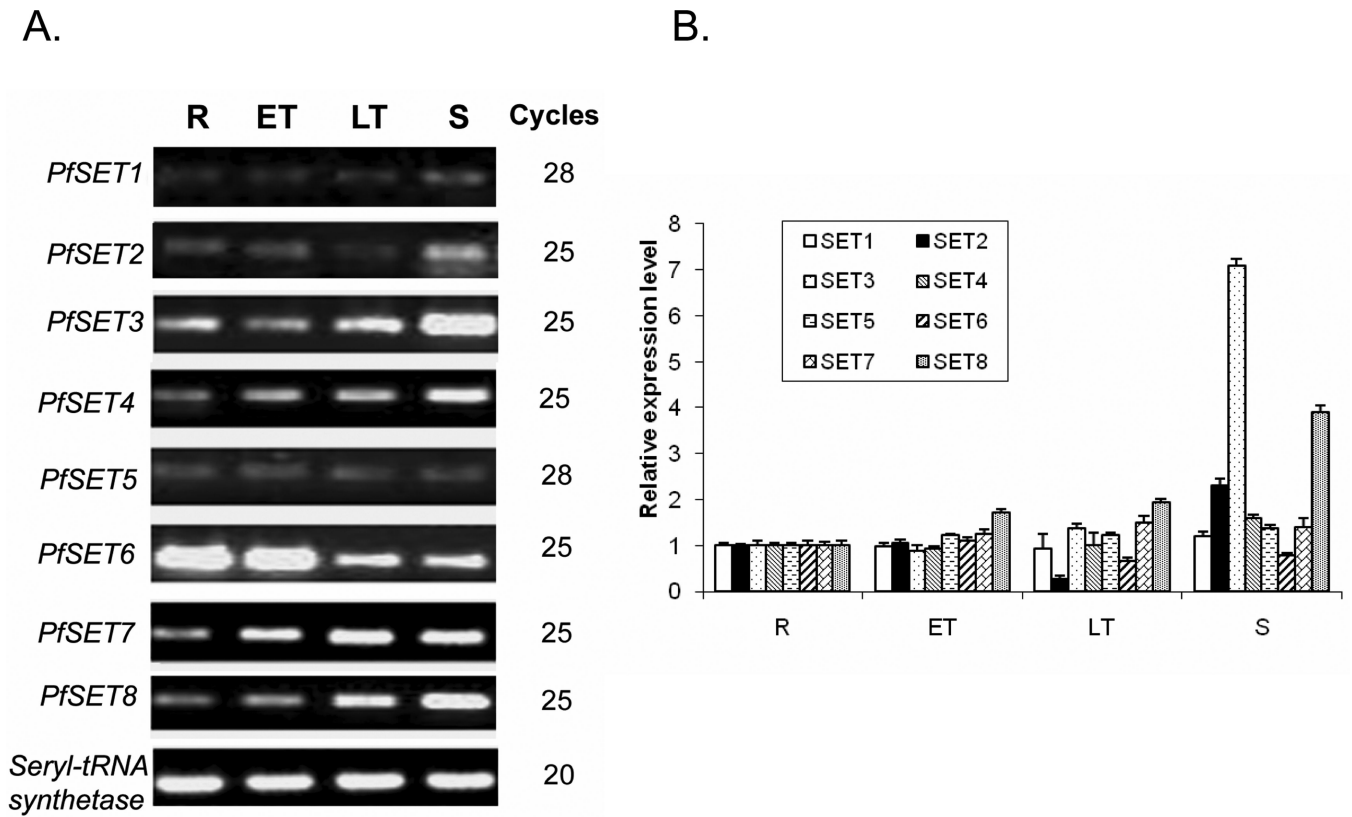


Fig. 6. Expression of *PfSET* genes during asexual erythrocytic development. A) Semi-quantitative reverse transcription (RT)-PCR analysis is illustrated for *PfSET1-8* with seryl-tRNA synthetase as the control. The number of RT-PCR cycles is shown to the right of the panel. B) The relative expression level of each *PfSET* gene quantified by real-time quantitative RT-PCR. R – rings, ET – early trophozoites, LT – late trophozoites, S – schizonts.

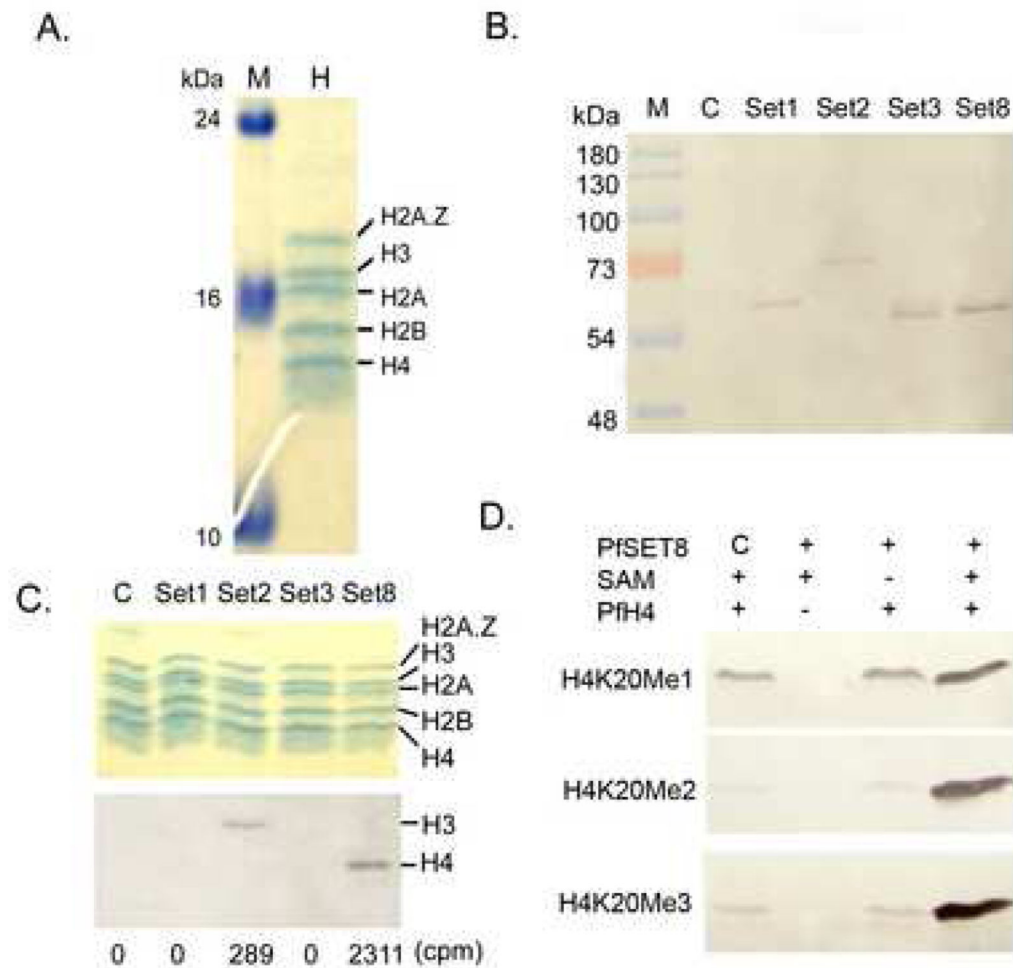


Fig. 7. Expression of putative active domains of SET-domain protein representatives from *Plasmodium falciparum* and their histone lysine methyltransferase activities on recombinant histones. A) Expression and purification of *P. falciparum* histones shown in a Coomassie blue-stained 18% SDS-PAGE gel. M, molecular weight markers in kDa; H, equi-molar mixture of four recombinant histones. B) Western blot analysis of GST-fusion PfSET domains using the wheat germ expression system. C, empty vector control. C) Methylation assays of the in vitro expression products using recombinant *P. falciparum* histones. Upper panel indicates the Coomassie blue-stained gel showing the positions of recombinant histones. Lower panel shows the fluorograph of the same gel after 1-month exposure. Numbers below the panels indicate the liquid scintillation counting data from 10 μ l of the same reactions, where control (C) counting was arbitrarily set as the background value and subtracted from other reactions. D) Western blots of PfH4 methylation by recombinant His-tagged PfSET8 expressed from the wheat germ expression system using α -H4K20Me1, Me2 and Me3 antibodies. C, empty vector control; SAM, S-adenosyl methionine.

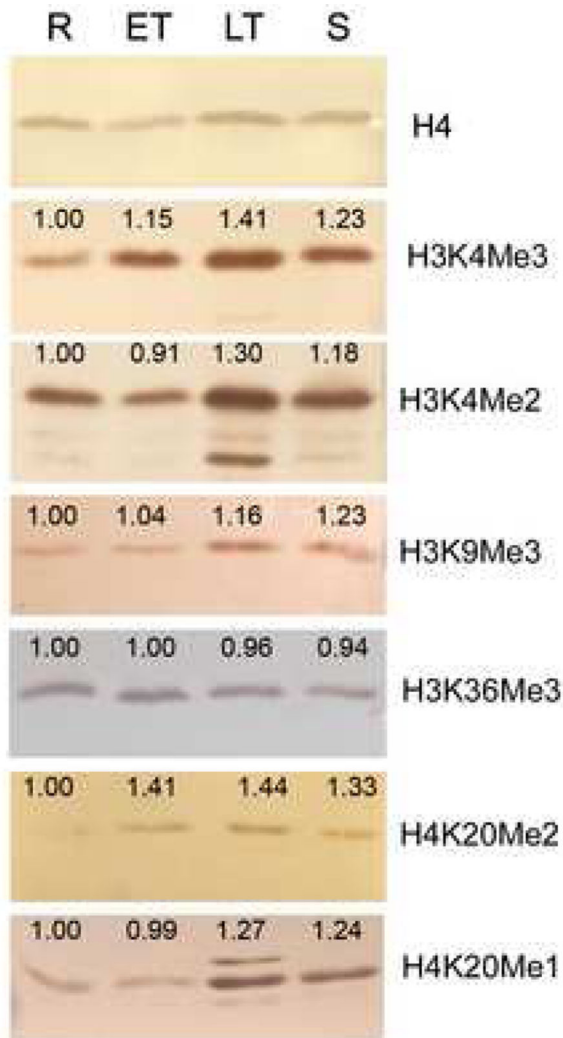


Fig. 8. Dynamic lysine methylation of histones H3 and H4 during asexual erythrocytic development of *Plasmodium falciparum*. Western blots were performed using parasite histones purified from different developmental stages (Abbreviations are as in Fig. 6). Primary antibodies were against conserved mono-, di- and tri-methylated lysine (indicated as Me1, 2 and 3, respectively) at the N-termini of H3 and H4. Antibodies against H4 were used to indicate approximately equal loading. Numbers above the band indicate the ratio of the density of the band relative to that of the ring stage as analyzed by densitometry. The relative amounts of H4 were used as internal references to correct loading differences.

Table 1

Primers used for reverse transcription-PCR analysis of SET-domain genes (F1 and R1) and for expressing PfSET1, 2, 3 and 8 in bacteria (F2 and R2).

Gene ID	Primers	Sequences
PFF1440w	F1	5'-GTCTAGGATGGATGTGTT-3'
	R1	5'-CATAACCCGTGCATGTAT-3'
	F2	5'-CGCGGATCCTCAGAATCTGATGCTACGAAC-3'
	R2	5'-ACGCGTCGACTTAATTCATTCTTCTAGAC-3'
MAL13P1.122	F1	5'-CATGCAATGATATTAGAAG-3'
	R1	5'-CTCCTTTATGTTCATCAT-3'
	F2	5'-CGCGGATCCGAATCCATTCTGATGCTTC-3'
	R2	5'-ACGCGTCGACTTAGGTAGAATAAACTCCATTAAC-3'
PF08_0012	F1	5'-AGTGTATAGATTGTGGT-3'
	R1	5'-GTATATACATAGGTTTATC-3'
	F2	5'-CGCGGATCCGAGATGGCTAAAAGACCAGC-3'
	R2	5'-ACGCGTCGACTCAACCTATGTAACCTTTAC-3'
PFI0485c	F1	5'-GTTTCATGTGTAATCAAGG-3'
	R1	5'-CAAATCGAAACCGAAAGG-3'
PFL0690c	F1	5'-CTGATATACCAAATGCATA-3'
	R1	5'-CCACCAAGAATGTCCATAT-3'
PF130293	F1	5'-GCAGGATATTGCTTGTAG-3'
	R1	5'-CTCATAGTTATAGATGGA-3'
PF110160	F1	5'-ATGGAAACAATATTTAGAA-3'
	R1	5'-CCTCGTCCAGATATGGT-3'
PFD0190w	F1	5'-CTGCTTCAGTTATATGGATTG-3'
	R1	5'-GATATCTTTAAATTCTCG-3'
	F2	5'-CGCGGATCCGACCTTTTTGGAACATCTGC-3'
	R2	5'-ACGCGTCGACTTAAATTTAACCATTCGTTATC-3'

Table 2

A summary of *Plasmodium falciparum* SET- and JmjC-domain proteins.

Gene ID	Name	CL ^a	Length (amino acids)	Family representative	Predicted Site-specificity	5' RLM-RACE ^c	3' RACE ^d
PFF1440w	PiSET1	6	6,753	SET1	H3K4Me1-3	630 (1); 475 (6); 323 (1); 316 (1)	322 (1); 328 (2); 332 (1); 338 (1); 346 (5)
MAL13P1.122	PiSET2	13	2,548	SET2	H3K36Me2, 3	ND	57 (1); 65 (11); 70 (1)
PF08_0012	PiSET3	8	2,399	Suv39	H3K9Me2-3	526 (3); 171 (2); 135 (11); 81 (1)	335 (1); 352 (1); 372 (1); 381 (1); 384 (3); 385 (1); 387 (2); 403 (1); 405 (1)
PFI0485c	PiSET4	9	1,114	SMYD3 ^b	H3K4 ^b	172 (4)	250 (3); 262 (2)
PFL0690c	PiSET5	12	178	Unknown	Unknown	344 (5); 310 (1); 308 (3); 306 (1); 275 (3); 270 (1); 268 (3)	104 (4); 121 (1); 807 (3); 1031 (1); 1034 (1); 1035 (1); 1038 (3); 1043 (1); 1044 (3); 1081 (3); 1104 (2)
PF13_0293	PiSET6	13	509	SMYD3 ^b	H3K4 ^b	494 (1); 492 (1); 453 (6)	455 (2); 499 (1); 508 (1); 512 (2); 518 (1); 529 (2); 536 (3); 537 (1); 543 (1); 574 (1)
PFI10160	PiSET7	11	810	Unknown	Unknown	357 (1); 181 (1); 75 (7); 43 (2); 17 (3); 12 (3)	245 (5); 246 (1); 247 (1); 448 (1); 498 (1)
PF0190w	PiSET8	4	1,186	SET8/Pi-SET7	H4K20Me1-3	688 (3); 683 (1); 661 (1); 659 (4); 468 (1); 380 (3)	285 (4); 286 (2)
PFE0400w	PiSET9	5	1,674	Unknown	Unknown	958 (6)	446 (1); 452 (2); 464 (2)
MAL8P1.111	PiJmjC1	8	1,259	JMJ2A	H3K36, K9	639 (3); 622 (2); 618 (2); 280 (6); 232 (2)	168 (5); 704 (2); 716 (1)
PFF0135w	PiJmjC2	6	646	Unknown	Unknown	118 (8)	140 (2); 151 (1); 181 (4); 190 (1); 191 (6); 205 (3); 247 (1); 251 (1); 274 (1)

The number in parenthesis indicates the number of clones sequenced. ND – not determined.

^aCL – chromosomal location.

^bPhylogenetic similarity of these proteins to the SMYD3 HKMT and potential site-specificity for H3K4.

^cPutative transcription initiation site(s) is shown as nucleotide number upstream from the putative translational start codon(s).

^dPutative polyadenylation site(s) is shown as nucleotide number downstream from the putative stop codon(s).