MICROBIOLOGY

SIP2 is the master transcription factor of *Plasmodium* merozoite formation

Tsubasa Nishi, Izumi Kaneko, Masao Yuda*

Malaria, one of the most serious infectious diseases worldwide, is caused by the proliferation of *Plasmodium* parasites through repeated cycles of intraerythrocytic development. The parasite replicates via schizogony in host erythrocytes, producing multiple progeny merozoites that invade new erythrocytes. Although merozoite formation is the most crucial step in malaria pathogenesis, its molecular mechanism remains unclear. SIP2 is an AP2 transcription factor expressed during schizogony and is particularly conserved among erythrocyte-infecting apicomplexan parasites. Here, we reveal that SIP2 in *Plasmodium berghei* (PbSIP2) functions as the master transcription factor for merozoite formation. Conditional disruption of *pbsip2* resulted in developmental arrest before merozoite formation and notable down-regulation of merozoite-related genes. ChIP-seq showed that PbSIP2 comprehensively activated merozoite-related genes by binding to previously reported cis-regulatory elements of merozoite invasion-related genes, including the bipartite motif (TGCAN$_{4-6}$GTGCA). Collectively, our results indicate that SIP2 is a transcription factor that establishes erythrocyte infectivity and may have an evolutionary origin from the common ancestor of erythrocyte-infecting apicomplexan parasites.

INTRODUCTION

Plasmodium parasites replicate asexually in the blood of vertebrate hosts, which causes malaria, one of the most serious diseases worldwide, with more than 200 million cases and 600,000 deaths every year (*1*). The proliferation of parasites begins with a merozoite invading the host red blood cells (RBCs). Within RBC, parasites grow and begin a unique mode of cell division called schizogony (*2*), wherein the parasites develop into multinucleated cells or schizonts through several asynchronous rounds of DNA replication and nuclear division (karyokinesis) (*3*–*5*). Eventually, multiple merozoites are produced through segmentation (cytokinesis) simultaneously with the final round of karyokinesis (*6*). The progeny merozoites then egress from the RBC and invade a new RBC to continue with intraerythrocytic developmental cycle (IDC). Schizogony, the final step in an IDC, is a crucial process for parasite proliferation in the host blood and is thus strongly related to parasite pathogenesis. However, the molecular mechanisms that regulate schizogony is still largely unknown.

The transcriptional regulation during the *Plasmodium* IDC has been assessed in several time-course transcriptomic studies (*7*–*10*). These studies indicated periodic expression of the majority of genes during IDC. Incidentally, the time-course assay for transposase-accessible chromatin (ATAC) sequencing further showed the transition of active promoter regions along with periodic transcriptomes (*11*). For the schizont stage in particular, those transcriptomic studies demonstrated stage-specific expression of genes essential for major merozoite-specific structures, such as rhoptry and pellicle/inner membrane complex (IMC). Furthermore, ATAC-seq demonstrated the enrichment of several DNA motifs within regulatory regions that are open during schizogony. These motifs include a putative cis-regulatory element upstream of the invasion-related genes, GGTGCA (named PfM18.1), which was detected through an in silico search of DNA motifs in *Plasmodium falciparum* (*12*). In addition, the in silico study identified a bipartite motif comprising GTGCA and GTGCA-like motifs upstream of genes encoding rhoptry proteins and suggested conservation of this bipartite motif in *Plasmodium* spp. These studies indicate the existence of a stage-specific transcription factor that regulates merozoite-related genes at the schizont stage by recognizing these motifs. However, this transcription factors have not been adequately explored, and the molecular mechanisms regulating the transcriptome during schizogony remain unclear.

AP2 transcription factors are the major family of transcription factors found in *Plasmodium* (*13*). Members in this family contain sequence-specific DNA binding domains with three anti-parallel β sheets and one α helix, called AP2 domain (*14*). SIP2 is an AP2 transcription factor expressed during the schizont stage. In *P. falciparum*, SIP2 (PfSIP2; PF3D7_0604100) has been reported to bind to the subtelomeric var. promoter element 2 (SPE2) (*15*) and has been proposed to be involved in chromosome end biology (*16*). However, the primary functions of SIP2 in *Plasmodium* remain to be elucidated. In the present study, we investigated the role of SIP2 in *Plasmodium berghei* (PbSIP2; PBANKA_0102900) during schizont development. We revealed that PbSIP2 plays an essential role in the final step of schizogony, i.e., merozoite formation, by activating the majority of genes important for merozoite-specific structures and abilities. Our data further showed that PbSIP2 was associated with previously reported cis-regulatory motifs of invasion-related genes, including the bipartite motif (*12*).

RESULTS

PbSIP2 is an AP2 family transcription factor that contains tandem AP2 domains conserved in RBC-infecting parasites

PbSIP2 contains two tandem AP2 domains at its N terminus and a putative nuclear localization signal (NLS) on the C-terminal side of the AP2 domains (Fig. 1A). The alignment of amino acid sequences among AP2 domains of PbSIP2 and its orthologs in other *Plasmodium* species showed that the two AP2 domains were conserved by 88 and 69%, respectively (Fig. 1B). In addition, the amino

Department of Medicine, Mie University, Tsu 514-8507, Japan.
*Corresponding author. Email: m-yuda@med.mie-u.ac.jp

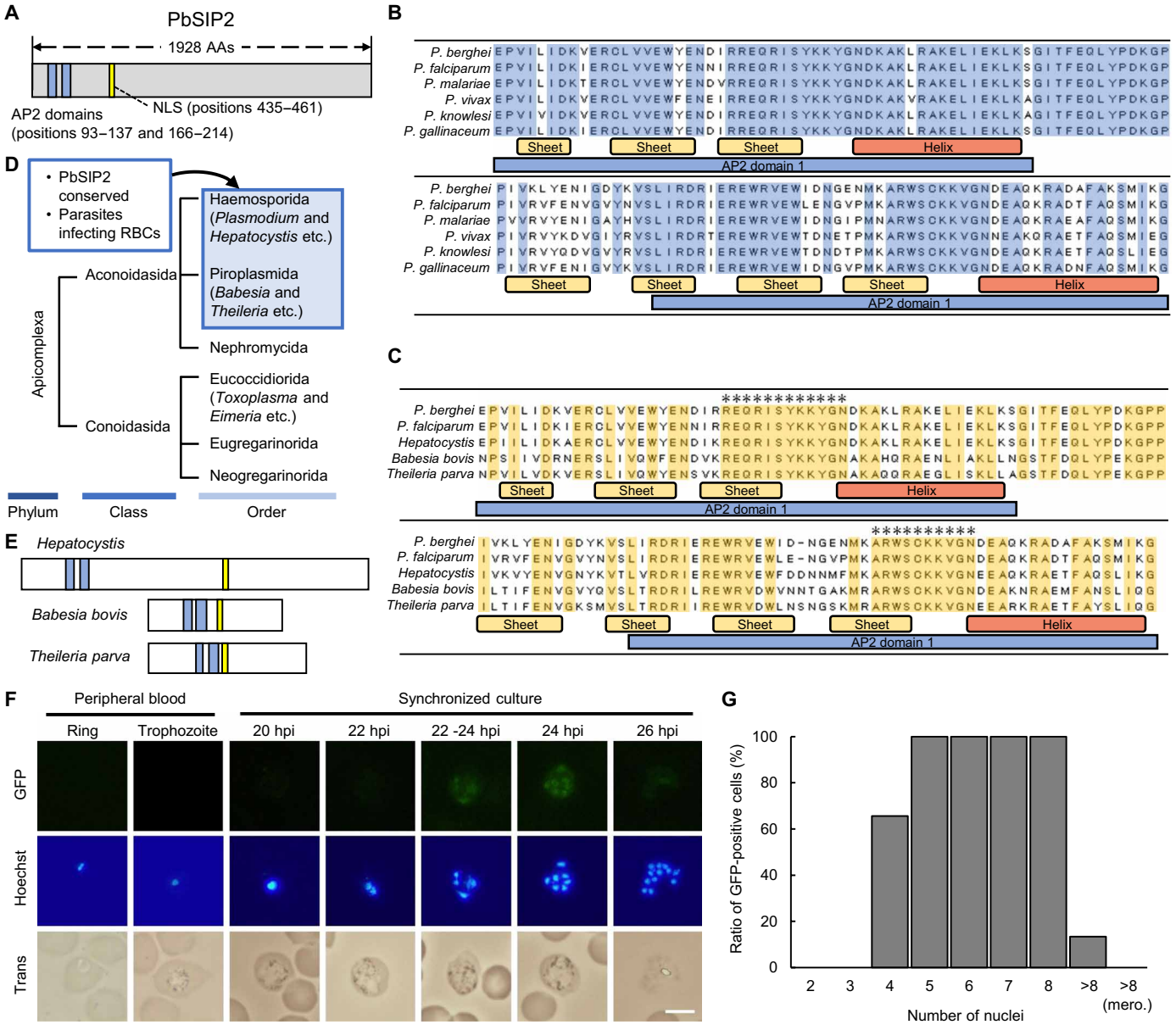

Fig. 1. Sequence features of PbSIP2 and its expression during asexual-blood stage development. (**A**) Schematic illustration of PbSIP2. Blue boxes show the AP2 domains. The NLS was predicted using cNLS Mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi) and indicated using a yellow box. (**B**) Alignment of amino acid (AA) sequences of PbSIP2 orthologs in *Plasmodium* by the ClustalW program in Mega X (*P. berghei*, PbANKA_0102900; *P. falciparum*, PF3D7_0604100; *Plasmodium malariae*, PmUG01_11058700; *Plasmodium vivax*, PVP01_1144800; *Plasmodium knowlesi*, PKNH_1146400; *Plasmodium gallinaceum*, PGAL8A_00132600). Amino acid sequences conserved in all aligned sequences are indicated by blue boxes. Yellow and red bars indicate a β sheet and α helix, respectively. Blue bars show a sequence for an AP2 domain. (**C**) Alignment of amino acid sequences of PbSIP2, PfSIP2, and BLASTP-detected proteins (*Hepatocystis* spp., VWU52144; *Babesia Bovis*, XP_001610230; *Theileria Parva*, XP_764727). Amino acid sequences conserved in all aligned sequences are indicated by yellow boxes. Regions wherein more than 10 amino acids are continuously conserved are indicated with asterisks. (**D**) Taxonomic tree of Apicomplexa. Orders in which PbSIP2 is conserved are indicated in a blue box. (**E**) Schematic illustration of putative SIP2 orthologs in Apicomplexa. Blue boxes show the AP2 domains, and NLSs are indicated using a yellow box. (**F**) Expression of PbSIP2 in asexual blood stage of the PbSIP2::GFP. Rings and trophozoites were detected from the peripheral blood, and schizonts were observed in synchronized cultures. Nuclei were stained with Hoechst 33342. Scale bar, 5 μm. (**G**) Quantification of GFP-positive cells for PbSIP2::GFP with each number of nuclei. Cells were counted by fluorescent microscopic analysis. At least 30 cells were assessed for schizonts with each number of nuclei [>8 (mero.) indicates merozoites]. hpi, hours post-injection.

acid sequences of the linker between the two AP2 domains were also conserved (Fig. 1B).

To further explore the conservation of PbSIP2 in Apicomplexa, we performed a protein-protein BLAST (BLASTP) search using the amino acid sequence of its AP2 domains and the linker between them (positions 87 to 214) as a query. The search detected proteins from *Hepatocystis*, *Babesia*, and *Theileria* species with *E* values < 10^{-50}, but not from other apicomplexan species (Fig. 1C). Notably, *Plasmodium* and these PbSIP2-conserved species belong to the Haemosporida or Piroplasmida, which constitute a group of parasites that infect host RBCs (Fig. 1D) (*17*, *18*). Amino acid sequences of the N- and C-terminal side AP2 domains were conserved with 59 and 62%, respectively, and regions between the third β sheet and α helix were completely conserved for both AP2 domains (Fig. 1C). In addition, the linker lengths between the two AP2 domains were the same (28 amino acids) for PbSIP2 and the BLASTP-detected proteins (Fig. 1C). Furthermore, their overall structures were also similar to those of PbSIP2, as they contained the AP2 domains near their N terminus and a putative NLS toward their C-terminal side (Fig. 1E).

PbSIP2 is expressed during the early schizont development

To determine the stage at which PbSIP2 functions, we generated a parasite line expressing green fluorescent protein (GFP)–fused PbSIP2 (PbSIP2::GFP; fig. S1A) and assessed its expression pattern during the asexual blood stage. In the fluorescence analysis of the peripheral blood, no fluorescent signals were observed in the ring or trophozoite stages of PbSIP2::GFP (Fig. 1F). The analysis was further performed for the schizont stages in culture because schizonts do not appear in the peripheral blood owing to the sequestration of schizont-infected RBCs (*19*, *20*). We synchronized the cell cycle of PbSIP2::GFP by intravenously injecting cultured mature schizonts into mice. The synchronized parasites were cultured again at 6 hours post-injection (hpi), and fluorescent signals were assessed every 2 hours. In the culture, fluorescent signals first appeared in the schizonts with four nuclei (Fig. 1F, gray scale images for each Red-Green-Blue (RGB) channel are shown in fig. S1B). Approximately 65% of the schizonts with four nuclei showed the GFP signal, suggesting that the PbSIP2 expression starts during this period (Fig. 1G). The signals continued until the schizonts had eight nuclei and then faded in most of those with more than eight nuclei (Fig. 1, F and G). *Plasmodium* schizonts undergo several rounds of asynchronous nuclear division and a final round of karyokinesis, which simultaneously occurs with segmentation (cytokinesis), to produce multiple merozoites. Moreover, in *P. berghei*, a median of 12 to 13 merozoites is formed per schizont, with a maximum of 20 (*21*). Thus, these results indicate that PbSIP2 functions from the early rounds of karyokinesis until the beginning of cytokinesis.

Disruption of *pbsip2* affects the late stage of schizont development

To investigate the role of PbSIP2 during schizont development, we performed a conditional knockout of *pbsip2* using a dimerizable Cre recombinase (DiCre) system (*22*, *23*). We first developed a parasite line constitutively expressing Cre59 (Thr^{19}-Asn^{59}) fused with FKBP12 and Cre60 (Asn^{60}-Asp^{343}) fused with FRB, through inserting the expression cassette into the *p230p* locus using the Cas9-expressing parasite, PbCas9 (*24*) (PbDiCre; fig. S2A). Subsequently, two loxP sequences were inserted on 5′ and 3′ sides of *pbsip2*, arranged in the same direction (*pbsip2*-cKO, Fig. 2A and fig. S2B). This parasite will lose the entire open reading frame of *pbsip2* in the presence of rapamycin (Fig. 2A).

We first assessed whether DiCre-mediated recombination in *pbsip2*-cKO occurs quickly enough to disrupt *pbsip2* before PbSIP2 expression begins. The whole blood from mice infected with synchronized *pbsip2*-cKO was split into two cultures at 6 hpi, and the one was treated with 10 nM rapamycin (*pbsip2*-cKO^{Rapa+}) and the other with dimethyl sulfoxide (DMSO) as a control (*pbsip2*-cKO^{Rapa-}) (Fig. 2B). At 18 hpi (12 hours of culture), when no parasites had started nuclear division, the recombination at the *pbsip2* locus was assessed by genotyping polymerase chain reaction (PCR). The assay showed that the *pbsip2* locus was almost completely excised from the genome in *pbsip2*-cKO^{Rapa+}, while no recombination was detected in *pbsip2*-cKO^{Rapa-} (Fig. 2C). Notably, using the parental PbDiCre parasite, we confirmed that schizont development was not affected by rapamycin (fig. S2C and table S1A).

To assess the impact of *pbsip2* disruption on schizont development, we observed *pbsip2*-cKO on Giemsa-stained smears every 2 hours from 18 hpi. In culture, *pbsip2*-cKO^{Rapa+} developed as comparable to *pbsip2*-cKO^{Rapa-} until 24 hpi. At 26 hpi, approximately 20% of *pbsip2*-cKO^{Rapa-} became mature schizonts (number of nuclei > 10), and some have already formed merozoites (Fig. 2, D and E, and table S1A). In contrast, most schizonts of *pbsip2*-cKO^{Rapa+} had less than 10 nuclei at 26 hpi (Fig. 2, D and E, and table S1A). Furthermore, while the number of schizonts with 6 to 10 nuclei increased at 28 hpi, the ratio of mature schizonts were only 1.3 ± 0.6% in *pbsip2*-cKO^{Rapa+}, and no merozoite formation was observed (Fig. 2, D and E, and table S1B). This result suggests that the final round of karyokinesis could not proceed in *pbsip2*-cKO^{Rapa+}.

To investigate whether the final round of DNA replication is also affected by *pbsip2* disruption, we assessed the DNA content in *pbsip2*-cKO by staining the parasites with Hoechst 33342. We first observed Hoechst-stained *pbsip2*-cKO^{Rapa-} and *pbsip2*-cKO^{Rapa+} using fluorescence microscopy at 28 hpi. Intriguingly, >10 Hoechst foci were detected in approximately 40% of *pbsip2*-cKO^{Rapa+} in contrast to those observed following Giemsa staining, in which only 2.2 ± 0.4% of schizonts showed >10 nuclei (Fig. 2F and table S1C). This result indicates that Hoechst staining enabled the visualization of individual haploid chromosome sets segregated in the nucleus, whereas Giemsa staining did not (fig. S2D). Thus, although the ratio of schizonts with >10 Hoechst foci was smaller in *pbsip2*-cKO^{Rapa+} compared with that in *pbsip2*-cKO^{Rapa-} (approximately 60%), DNA replication and chromosome segregation appeared to have been completed in most of the *pbsip2*-cKO^{Rapa+} parasites (Fig. 2F and table S1C). We further assessed the DNA content in *pbsip2*-cKO^{Rapa-} and *pbsip2*-cKO^{Rapa+} using flow cytometry. The analysis showed that, overall, the histograms of DNA content per parasite were not significantly different between *pbsip2*-cKO^{Rapa-} and *pbsip2*-cKO^{Rapa+}, implying that DNA replication proceeded irrespective of whether *pbsip2* was disrupted (Fig. 2G and fig. S2E). These results suggest that while the majority of *pbsip2*-cKO^{Rapa+} completed DNA replication during schizogony to the final rounds, their development is arrested immediately before cytokinesis that accompanies the final round of karyokinesis (Fig. 2H).

PbSIP2 binds to the upstream of genes related to merozoite formation

Next, we performed chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) analysis using PbSIP2::GFP and investigated its binding sites on a genome-wide basis. PbSIP2::GFP

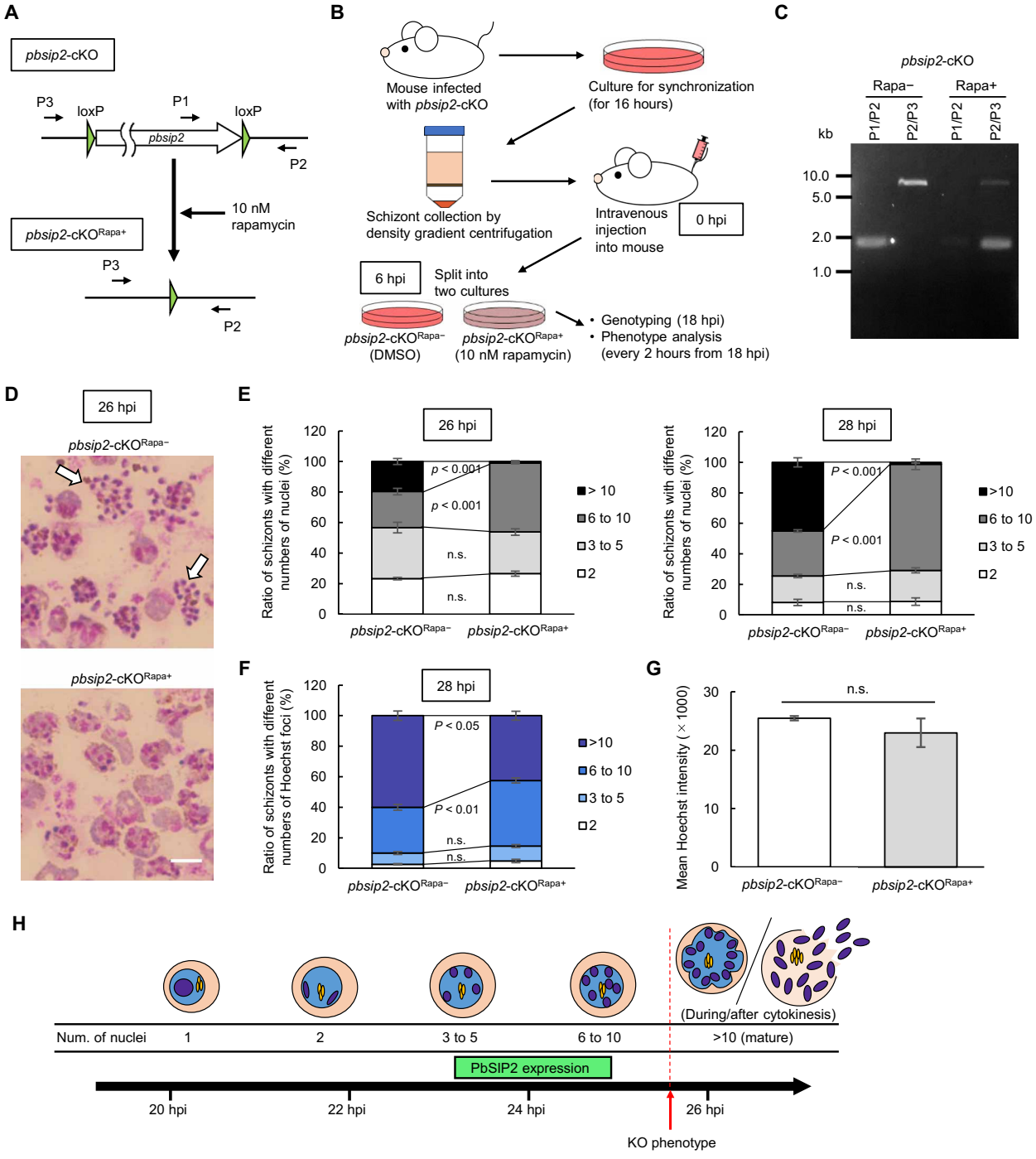

Fig. 2. Conditional knockout of *pbsip2* by DiCre-mediated recombination. (**A**) Schematic illustration of DiCre-mediated recombination at the *pbsip2* locus of *pbsip2*-cKO. Recombination occurs between two loxP sites to excise *pbsip2* from the genome in the presence of rapamycin. P1 to P3 indicate primers used for genotyping PCR. (**B**) Outline of conditional knockout experiments using *pbsip2*-cKO. First, the whole blood from mice infected with *pbsip2*-cKO was cultured. After 16 hours of culture, schizonts were harvested by density gradient centrifugation and injected into mice. At 6 hpi, the whole blood was harvested and split into two cultures: one with 10 nM rapamycin and the other with DMSO. (**C**) A gel image of the genotyping PCR analysis for *pbsip2*-cKO^{Rapa-} and *pbsip2*-cKO^{Rapa+} performed at 16 hpi (a representative image from the first of three replicates). Primers used are illustrated in (A). (**D**) Representative Giemsa-stained images of *pbsip2*-cKO^{Rapa-} and *pbsip2*-cKO^{Rapa+} at 26 hpi. Mature schizonts with merozoites inside are indicated by arrows. Scale bar, 5 μm. (E) Ratio of schizonts with different numbers of nuclei in *pbsip2*-cKO^{Rapa-} and *pbsip2*-cKO^{Rapa+} at 26 hpi (left) and 28 hpi (right). Error bars indicate the SEM value from three biologically independent experiments. The *P* values were calculated by two-tailed Student's *t* test (n.s. stands for not significant, *P* value > 0.05). (**F**) Ratio of schizonts with different numbers of Hoechst foci in *pbsip2*-cKO^{Rapa-} and *pbsip2*-cKO^{Rapa+} at 28 hpi. (**G**) Flow cytometric analysis of *pbsip2*-cKO^{Rapa-} and *pbsip2*-cKO^{Rapa+} at 28 hpi. Mean Hoechst intensities of schizont populations are shown. Error bars indicate the SE from three independent experiments. The *P* value was calculated using a two-tailed Student's *t* test. (**H**) Model of the *P. berghei* schizont development in synchronized cultures. The period for PbSIP2 expression is indicated as a green box. hpi, hours post-injection.

parasites were synchronized as described above, and ChIP-seq was performed at 24 hpi. Two biologically independent experiments were conducted, and their overall peak patterns were highly similar, suggesting high reproducibility of the ChIP-seq experiments (Fig. 3A). Using macs2 (a peak-calling program that genome-widely identifies ChIP-enriched regions), 784 and 804 peaks were identified in experiments 1 and 2, respectively, and 645 peaks (more than 80% of the peaks in experiment 1) overlapped between them (Fig. 3B and data S1, A and B). To investigate the binding motif of PbSIP2, we assessed the enrichment of sequence motifs within 50 base pair (bp) of the summits of the common peaks using Fisher's exact test. The analysis showed enrichment of GTGCA with a *P* value less than 5.0×10^{-324} (the smallest positive real number on the R platform, hereafter referred to as *P* value^{limit}) (Fig. 3C). The GTGCA motif was found within 300 bp of the summit of 556 peaks (86% of the common peaks), and the closer the motif was to the peak summit, more peaks were detected (Fig. 3D). These results indicate that the GTGCA motif could be the binding motif of PbSIP2.

We next examined the PbSIP2 peak locations in the genome and identified target genes. By defining genes containing a peak within 1200 bp of their start codon as PbSIP2 targets, we identified 489 target genes from 645 common peaks (data S1C). Of these, 340 are functionally annotated on PlasmoDB (https://plasmodb.org/plasmo/app). To evaluate the role of PbSIP2, we classified the annotated targets into

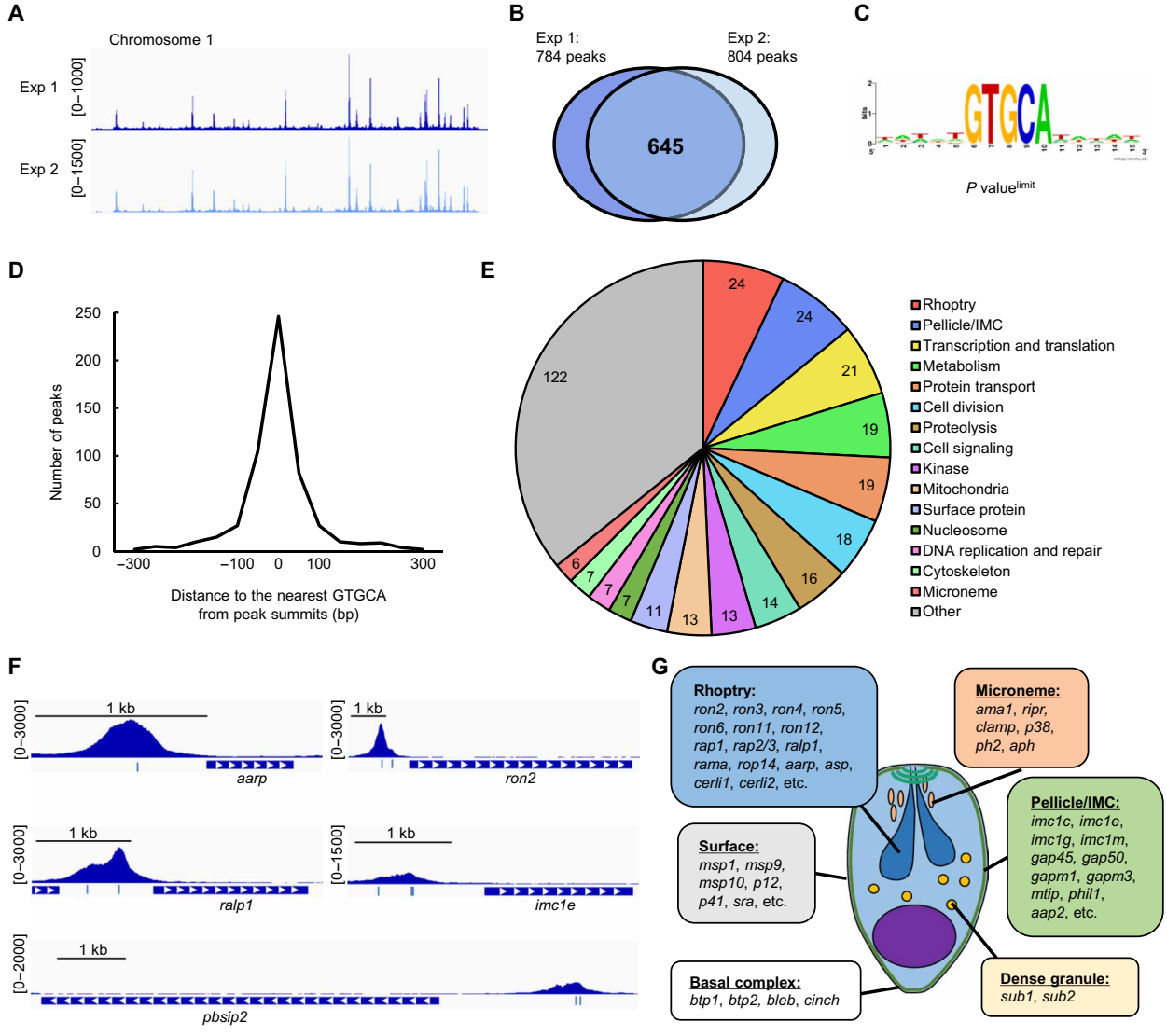

Fig. 3. ChIP-seq analysis of PbSIP2. (**A**) Integrative Genomics Viewer (IGV) images for ChIP-seq experiments 1 and 2 of PbSIP2 on the chromosome 1. Histograms show row read coverage of ChIP data at each base. Scales are indicated in square brackets. (**B**) A Venn diagram showing the number of overlapping peaks between experiment 1 and experiment 2. (**C**) Motif enriched within 50 bp from ChIP-seq peak summits of PbSIP2. The logo was depicted using WebLogo (https://weblogo.berkeley.edu/logo.cgi). *P* value^{limit} is a *P* value less than 5.0×10^{-324} (the smallest positive real number on the R platform). (**D**) Distance between peak summits identified in the ChIP-seq and the nearest GTGCA motifs. (**E**) Classification of functionally annotated target genes of PbSIP2 into 16 characteristic groups. The number of genes categorized into each group is indicated in the graph. (**F**) IGV images showing ChIP-seq peaks upstream of the PbSIP2 target genes. Histograms show row read coverage of ChIP data at each base. Locations of GTGCA motif are indicated as a blue bar. The scales are shown in square brackets. (**G**) Schematic illustration of *P. berghei* merozoite. Characteristic structures of the merozoite and PbSIP2 targets that belong to each structure are shown.

16 characteristic groups (Fig. 3E). The analysis revealed that the two largest groups in the PbSIP2 targets were "rhoptry" and "pellicle/IMC." The group rhoptry contained majority of known rhoptry genes, including the recently identified cytosolically exposed rhoptry leaflet interacting protein genes, *cerli1* and *cerli2* (Fig. 3, F and G) (*25*, *26*). The targets in "pellicle/IMC" included several glideosome-associated protein genes (such as *gap45* and *gap50*) and IMC protein genes (such as *imc1c* and *imc1e*) (Fig. 3, F and G). Moreover, the PbSIP2 targets contained genes related to other merozoite structures, such as microneme [such as *ama1* and *ripr* (*27*)], dense granules (such as *sub1* and *sub2*) merozoite surface (such as *msp1* and *msp9*), and basal complex [such as *btp1* and *bleb* (*28*)] (Fig. 3G). We also performed the Gene Ontology (GO) analysis on the PbSIP2 targets and detected "apical part of cell," "apical complex," and rhoptry as the three most enriched terms (Table 1 and table S2). These results demonstrated that PbSIP2 targets the majority of genes important for merozoite formation, which is consistent with the developmental arrest at the schizont stage observed in the conditional knockout of *pbsip2*. In addition, although not included among the PbSIP2 targets because of the upstream threshold of 1200 bp, *pbsip2* had a PbSIP2 peak approximately 2-kb upstream (Fig. 3F). This result indicates that PbSIP2 could activate its gene through positive transcriptional feedback, which is the characteristic of many transcriptional activators in *Plasmodium* (*29*–*32*).

PbSIP2 functions as a transcriptional activator

Next, we evaluated the role of PbSIP2 in transcriptional regulation during schizont development using high-throughput RNA sequencing (RNA-seq). Conditional knockout of *pbsip2* was performed as described above, and recombination was confirmed at 18 hpi. The *pbsip2*-cKO^{Rapa+} and *pbsip2*-cKO^{Rapa-} parasites were then subjected to RNA-seq analysis at 26 hpi, and the sequence data were compared between them using DESeq2 (data S2). The analysis revealed that in *pbsip2*-cKO^{Rapa+}, 193 genes were significantly down-regulated [$\log_2$(fold change) < -1, *P* value adjusted for multiple testing with the Benjamini-Hochberg procedure (*P* value^{adj}) < 0.05] compared to *pbsip2*-cKO^{Rapa-} [excluding *pbsip2*, which was down-regulated with $\log_2$(fold change) of -4.0 and *P* value^{adj} of 3.9×10^{-231}] (Fig. 4A). These significantly down-regulated genes contained 115 target genes of PbSIP2, showing a significant overlap with a *P* value of 7.9×10^{-66} by Fisher's exact test (Fig. 4B). In contrast, only two genes were significantly up-regulated in *pbsip2*-cKO^{Rapa+} [$\log_2$(fold change) > 1, *P* value^{adj} < 0.05], none of which were targets of PbSIP2 (Fig. 4A). The GO analysis on the significantly down-regulated genes detected apical part of cell, apical complex, and rhoptry as the three most enriched terms same as the analysis on the PbSIP2 targets (Table 2 and table S3). In addition, the two most enriched motifs in the upstream region (300 to 1200 bp from ATG) of the down-regulated genes contained the PbSIP2-binding motif (Fig. 4C). These results indicated that PbSIP2 functions as a transcriptional activator during schizont development. The genes that were significantly down-regulated in *pbsip2*-cKO^{Rapa+} included a considerable number of rhoptry genes (25 genes) (Fig. 4, D and E), 22 of which were also PbSIP2 target genes (Fig. 4B). Furthermore, these rhoptry genes tended to show higher $-\log_2$(fold change) values than the other down-regulated genes; 8 of the 11 genes down-regulated with $\log_2$(fold change) < -4 were a rhoptry gene (Fig. 4E), suggesting that PbSIP2 functions as a strong trans-acting factor for transcriptional activation of rhoptry genes.

To further confirm the down-regulation of rhoptry genes in *pbsip2*-cKO^{Rapa+} at the protein level, we investigated the expression of a rhoptry protein RAP1 in *pbsip2*-cKO. RAP1 fused with mCherry was episomally expressed under the control of its endogenous 5′ and 3′ regions using the centromere plasmid (CenP_RAP1) (fig. S3, left). The expression of RAP1 was observed in *pbsip2*-cKO^{Rapa-} in the form of several mCherry foci around the nuclei (fig. S3, right). However, mCherry-fused RAP1 was not detected in any *pbsip2*-cKO^{Rapa+} schizonts (fig. S3, right), which was consistent with the results of the differential expression analysis. These results suggest that the production of rhoptry proteins was severely decreased in *pbsip2*-cKO^{Rapa+}.

PbSIP2 binds to the GTGCA motif through its tandem AP2 domains

To examine the DNA binding properties of the AP2 domains of PbSIP2, we performed DNA immunoprecipitation followed by high-throughput sequencing (DIP-seq) (*33*) using recombinant PbSIP2 AP2 domains fused with a maltose-binding protein tag on its N-terminal side (MBP::PbSIP2). MBP::PbSIP2 was mixed with fragmented *P. berghei* genomic DNA (approximately 150 to 300 bp). DNA fragments bound by MBP::PbSIP2 were harvested using amylose resin and sequenced by Next-Generation Sequencing (NGS). In the sequence data, we detected 3451 peaks throughout the genome (Fig. 5A and data S3). Within 50 bp from the summit of the DIP peaks,

Table 1. GO analysis for the PbSIP2 targets. Ten most enriched terms are shown. Inf in odds ratio column indicates an infinite number.

Term	*P* value	Odds ratio
Apical part of cell	1.96×10^{-21}	7.95
Apical complex	3.85×10^{-18}	8.92
Rhoptry	1.50×10^{-16}	20.68
Rhoptry neck	1.35×10^{-10}	Inf
Protein binding	8.86×10^{-07}	2.19
Nucleosome	9.38×10^{-07}	61.57
Movement in host	1.85×10^{-06}	4.91
Pellicle	2.69×10^{-06}	4.44
Protein heterodimerization activity	2.88×10^{-06}	17.10
Biological process involved in symbiotic interaction	4.52×10^{-06}	4.30

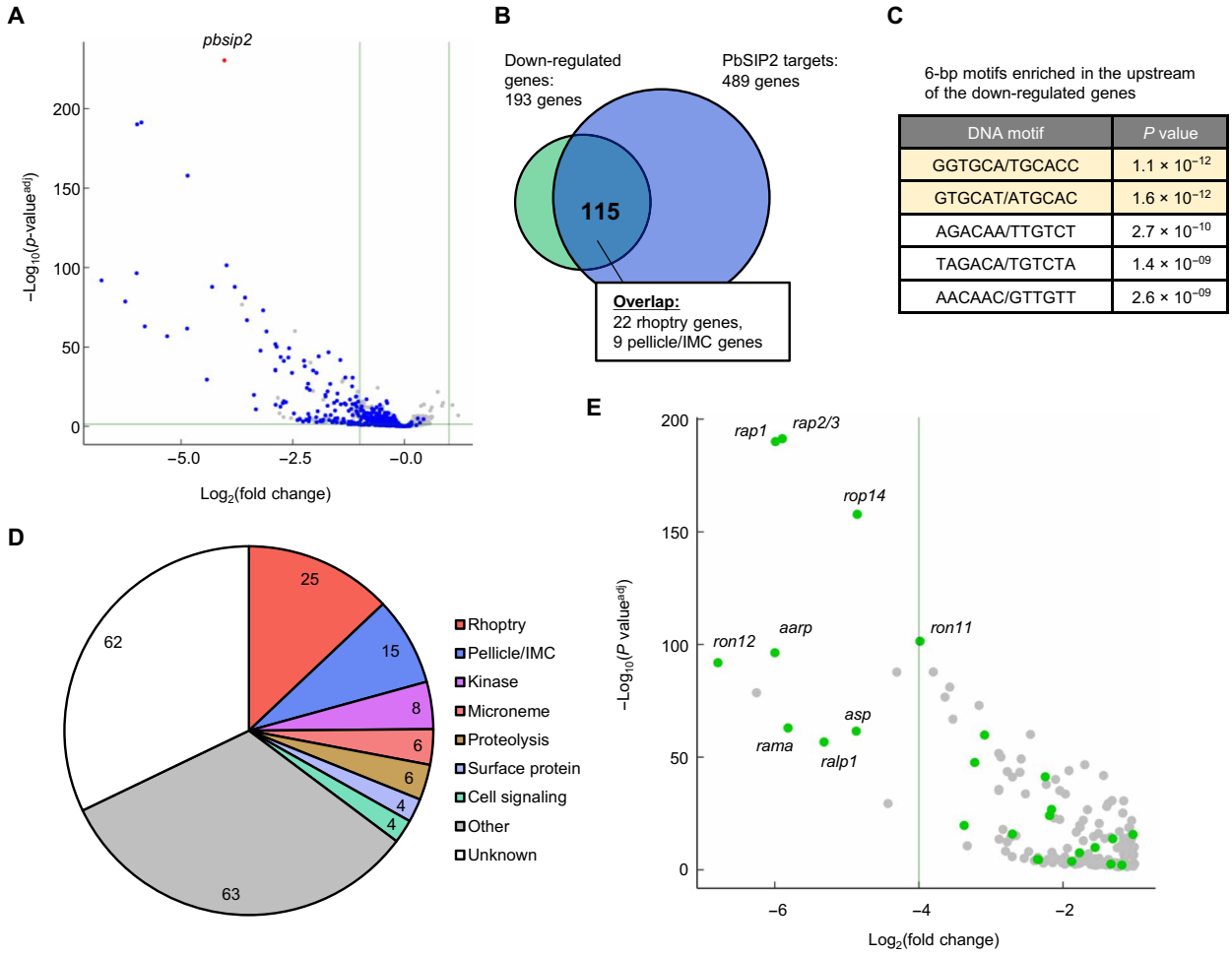

Fig. 4. Differential expression analysis between *pbsip2*-cKO^{Rapa-} and *pbsip2*-cKO^{Rapa+} by RNA-seq. (**A**) Volcano plot showing differential expression of genes between *pbsip2*-cKO^{Rapa-} and *pbsip2*-cKO^{Rapa+}. Blue dots represent target genes of PbSIP2, and a red dot represents *pbsip2*. A horizontal line indicates a *P* value of 0.05 and two vertical lines indicate a $\log_2$(fold change) of 1 and −1. (**B**) A Venn diagram showing an overlap between the genes significantly downregulated in *pbsip2*-cKO^{Rapa+} and the PbSIP2 target genes. (**C**) Six-base motifs enriched in the upstream (300 to 1200 bp from start codons) of genes down-regulated in *pbsip2*-cKO^{Rapa+}. Motifs including GTGCA are indicated in light yellow. (**D**) Classification of the downregulated genes into characteristic groups. (**E**) Section of the volcano plot from (A). Genes significantly down-regulated in *pbsip2*-cKO^{Rapa+} compared with *pbsip2*-cKO^{Rapa-} are shown. Green dots represent a rhoptry gene, and names of some rhoptry genes are shown.

Table 2. GO analysis for the genes significantly downregulated in *pbsip2*-cKO^{Rapa+} compared with *pbsip2*-cKO^{Rapa-}. Ten most enriched terms are shown.

Term	*P* value	Odds ratio
Apical part of cell	6.64×10^{-36}	25.26
Apical complex	4.83×10^{-3}	27.69
Rhoptry	4.36×10^{-26}	62.61
Rhoptry neck	1.78×10^{-14}	Inf
Biological process involved in symbiotic interaction	2.49×10^{-10}	12.37
Biological process involved in interaction with host	2.49×10^{-10}	12.37
Biological process involved in interspecies interaction between organisms	3.59×10^{-10}	11.99
Movement in host	4.04×10^{-09}	11.98
Entry into host	2.25×10^{-07}	11.69
Host cell surface binding	6.35×10^{-07}	33.39

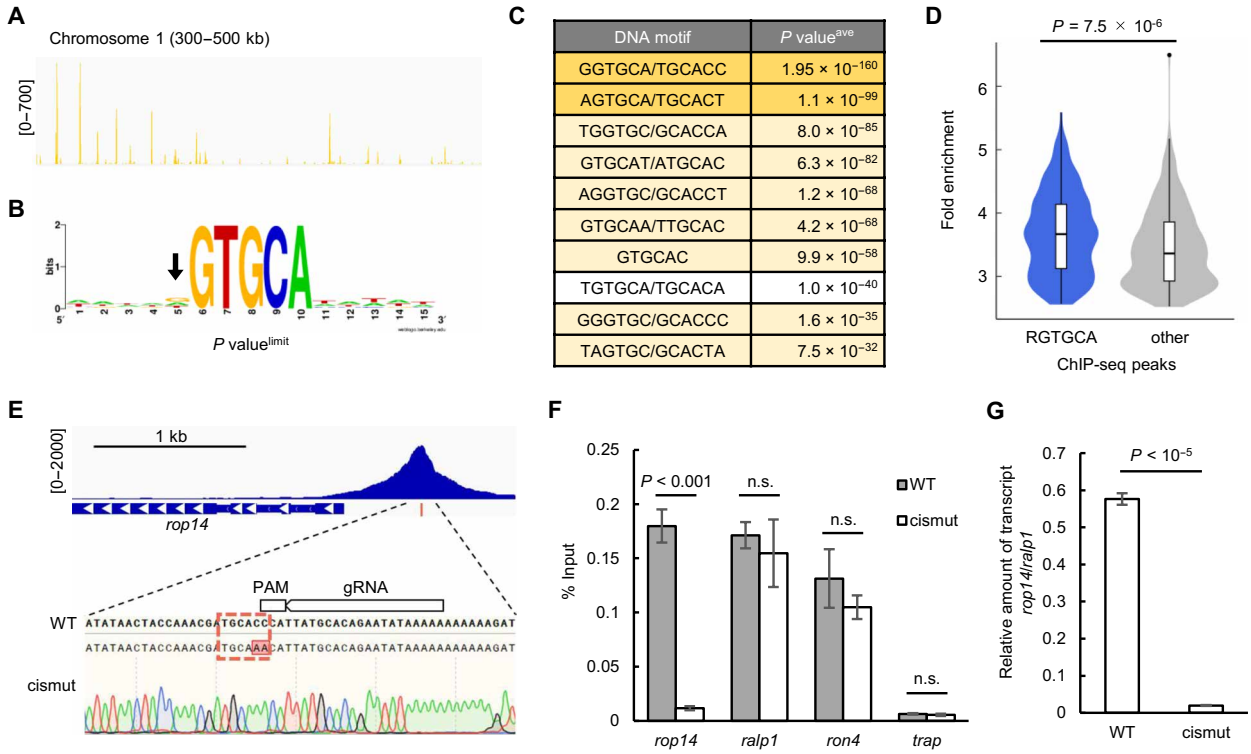

Fig. 5. DNA binding property of PbSIP2 and function of its binding motif as a cis-regulatory element. (**A**) IGV image for DIP-seq using recombinant AP2 domains of PbSIP2 on the chromosome 1. Histograms show row read coverage of IP data at each base. (**B**) A motif enriched within 50 bp from peak summits of DIP-seq. The logo was depicted using WebLogo. *P* value^{limit} is *P* value less than 5.0×10^{-324} (smallest positive real number on the R platform). An arrow indicates slight enrichment of G and A in the logo. (**C**) Six-base motifs enriched in the DIP-seq peaks. RGTGCA motif is indicated in yellow, and its one-base shifted motifs are indicated in light yellow. *P* value^{ave} is the average *P* value derived from the seven separate motif enrichment analyses (the detail is described in Materials and Methods). Motifs were ranked along their *P* value^{ave}. (**D**) Violin plot showing the distribution of fold enrichment values for the ChIP-seq peaks with RGTGCA (blue) and others (gray). The corresponding box plots are shown inside. The *P* value was calculated by two-tailed Student's *t* test. (**E**) IGV image showing the ChIP-seq peak upstream of *rop14* and genomic sequence around RGTGCA in the peak region. Sanger sequence result for PbSIP2::GFP^{Cas9_cismut} is shown under the wild-type sequence. The target region of single guide RNA used for generating PbSIP2::GFP^{Cas9_cismut} is indicated on the sequence. (**F**) ChIP-qPCR analysis of PbSIP2 at the mutated site upstream of *rop14*. Error bars indicate the standard error from three biologically independent experiments. The *P* values were calculated by two-tailed Student's *t* test (i.e., *P* value > 0.01) [*rop14*: PBANKA_0111600, *ralp1*: PBANKA_0619700, *ron4*: PBANKA_0932000, *trap*: PBANKA_1349800]. (**G**) RT-qPCR analysis of *rop14* transcripts between PbSIP2::GFP^{Cas9} and PbSIP2::GFP^{Cas9_cismut}. Error bars and *P* value were generated same as (F).

GTGCA was significantly enriched with a *P* value^{limit} by Fisher's exact test, similar to the ChIP-seq results. This result indicated that PbSIP2 directly binds to genomic DNA through its AP2 domains. In contrast to the ChIP-seq (Fig. 3C), slight enrichment of G and A was observed on the 5′-side of GTGCA when sequence logo was constructed from the sequences around GTGCA detected in the DIP-seq peaks (Fig. 5B). More than 70% of GTGCA motifs in the peak regions were RGTGCA (R = G or A). Consistently, in the motif enrichment analysis for 6-bp motifs, GGTGCA was the most enriched, followed by AGTGCA and motifs one base shifted from GGTGCA (Fig. 5C). Therefore, the AP2 domains of PbSIP2 preferentially bind to RGTGCA. Notably, for the ChIP-seq peaks, those with RGTGCA showed a higher average fold enrichment value than the other peaks that do not contain RGTGCA with a *P* value of 7.5×10^{-6} by two-tailed Student's *t* test (Fig. 5D). Therefore, although enrichment of RGTGCA was not evident in ChIP-seq, PbSIP2 was bound to RGTGCA with high affinity in vivo.

RGTGCA is essential for PbSIP2 binding and transcription of the downstream gene

To assess whether RGTGCA is essential for the DNA binding of PbSIP2 and functions as a cis-regulatory element for the downstream gene, we introduced a mutation into the motif within the ChIP peak upstream of *rop14* (Fig. 5E). Binding of PbSIP2 and *rop14* transcription was assessed by ChIP coupled with quantitative PCR (ChIP-qPCR) and reverse transcription qPCR (RT-qPCR), respectively. ROP14 is a functionally uncharacterized rhoptry protein that localizes to the rhoptry bulbs of merozoites (*34*). *rop14* was selected as the target in this experiment because it is not essential for the asexual blood-stage development in *P. berghei*. We developed parasites expressing GFP-fused PbSIP2 using PbCas9 (PbSIP2::GFP^{Cas9}; fig. S4) and then introduced a mutation upstream of *rop14*, altering GGTGCA to ttTGCA, by the CRISPR-Cas9 system (PbSIP2::GFP^{Cas9_cismut}; Fig. 5E). Using these parasite lines, we examined whether the disruption of the RGTGCA motif affected the binding of PbSIP2 to the genome by ChIP-qPCR analysis. In the PbSIP2::GFP^{Cas9_cismut} parasite, the amount of immunoprecipitated DNA fragments relative to the input DNA (% input value) upstream of *rop14* was more than 15-fold lower than that in PbSIP2::GFP^{Cas9} (Fig. 5F and table S4A). In contrast, upstream of other PbSIP2 targets, *ralp1* and *ron4*, the binding of PbSIP2 was not affected, as the % input values were comparable between the parasites with and without the mutation (Fig. 5F and table S4A). The % input value upstream of the nontarget

gene *trap* was assessed as a negative control and showed no significant change after the introduction of the mutation (Fig. 5F and table S4A).

Next, to investigate the function of the PbSIP2-binding motif as a cis-regulatory element, we compared the transcript levels of *rop14* between PbSIP2::GFP^{Cas9} and PbSIP2::GFP^{Cas9_cismut} by RT-qPCR analysis at 26 hpi. The transcriptional activity of the endogenous *rop14* promoter was assessed as the relative amount of *rop14* transcripts against *ralp1* transcripts. The analysis showed that in PbSIP2::GFP^{Cas9_cismut}, the relative amount of *rop14* transcripts was approximately 30-fold lower than that in PbSIP2::GFP^{Cas9}, revealing significant down-regulation of *rop14* promoter activity after introducing the mutation (Fig. 5G and table S4B). These results demonstrate that RGTGCA is essential for PbSIP2 binding and functions as a cis-regulatory element for activating downstream genes.

PbSIP2 recognize the bipartite motif upstream of rhoptry genes

A previous in silico study reported that two copies of the GTGCA (or GTGCA-like) motif, separated by five or six nucleotides, are conserved upstream of some rhoptry genes in *Plasmodium* species (*12*). In the ChIP-seq of PbSIP2, GTGCA was identified as the binding motif, and most rhoptry genes were detected as targets. Considering these, we explored DNA sequences in the ChIP-seq peaks and detected bipartite motifs, in which GTGCA-like motifs were found on the 5′-side of GTGCA, upstream of some rhoptry genes (Fig. 6A). Because PbSIP2 has two tandem AP2 domains, we hypothesized that PbSIP2 might recognize these bipartite motifs. Accordingly, we predicted the three-dimensional structure of the PbSIP2 AP2 domains using AlphaFold2 to evaluate whether they would have a structure capable of binding to two separate motifs (*35*). In the predicted model, two AP2 domains were aligned in parallel, and the distance between them was 33.11 Å, which is close to the length per turn of B-DNA helix (34 Å) (Fig. 6B). The 5-bp motifs separated by five or six nucleotides appear on the same side over one turn because the B-DNA helix is approximately 10 bp per turn (Fig. 6B). Therefore, the two parallel AP2 domains seemed suitable for binding to the bipartite motifs upstream of rhoptry genes.

To further assess whether PbSIP2 is genome widely associated with the bipartite motifs, we searched for enrichment of motifs within 10 bp on the 5′-side of GTGCA in the ChIP-seq peaks using Fisher's exact test. The analysis detected TGCA as the most enriched motif with a *P* value of 2.4×10^{-54} (Fig. 6C). In addition, motifs one base shifted from the TGCA, such as ATGC and GCAC, were detected with *P* values $<1.0 \times 10^{-10}$ (Fig. 6C). Therefore, the results suggested that PbSIP2 could be associated with the bipartite motif having TGCA on the 5′-side of the GTGCA motif. Given this result, we next searched for peaks with TGCAN$_x$GTGCA (schizont bipartite motif, SBM-N$_x$) and detected 140 peaks containing SBM-N$_{2-8}$ (Fig. 6D). Among these peaks, those containing SBM-N$_{4-6}$ were the majority (111 peaks), with the SBM-N$_5$ peaks being the most abundant (Fig. 6D). In addition, we also performed the search allowing a single mismatch in the 5′ TGCA and detected additional 132 peaks having the SBM-N$_{4-6}$ with one mismatch (Fig. 6E). Thus, nearly 40% of the PbSIP2 peaks are associated with these bipartite motifs. Hereafter, SBM-N$_{4-6}$ and those containing a single mismatch are called full SBM (fSBM) and mSBM, respectively, and altogether called SBM.

The peaks with the SBM tended to have higher intensity compared with the other peaks; i.e., average fold enrichment values of peaks with fSBM and mSBM were significantly higher than that of the other peaks with *P* values of 2.7×10^{-29} and 5.4×10^{-10} by two-tailed Student's *t* test, respectively (Fig. 6F). In addition, the average fold enrichment value of peaks with fSBM was higher than that of peaks with mSBM (*P* value $= 8.7 \times 10^{-6}$), suggesting that the DNA binding of PbSIP2 is stronger without any mismatches in the 5′ TGCA (Fig. 6F). Notably, the peaks with SBM also had a higher average fold enrichment value than those with RGTGCA (*P* value $= 2.6 \times 10^{-4}$) (Fig. 6F). Collectively, these results indicate that PbSIP2 has variations in the DNA binding affinity among its binding motifs, with SBM being stronger than the non-bipartite motifs, including RGTGCA.

We further explored SBMs in the DIP peak regions to evaluate whether SBMs are recognized by the tandem AP2 domains of PbSIP2. The analysis detected 129 and 737 peaks containing fSBM and mSBM (25% of the DIP-seq peaks in total), respectively (Fig. 6G). Furthermore, average fold enrichment values of fSBM and mSBM peaks were higher than that of the other peaks (*P* values $= 1.7 \times 10^{-10}$ and 1.1×10^{-8}, respectively), similar to the ChIP-seq results (Fig. 6H). These results suggest that the variable DNA binding affinities observed for PbSIP2 ChIP-seq were determined by its tandem AP2 domains.

Among the PbSIP2 target genes, 175 genes had a SBM peak upstream. These SBM-associated targets included 15 rhoptry genes and a few other merozoite-related genes, such as *gap45* and *msp10* (Fig. 6I). Rhoptry genes were also detected in genes that were highly downregulated in *pbsip2*-cKO^{Rapa+} (Fig. 4D), suggesting a relationship between genes associated with SBM and high transcription levels induced by PbSIP2. Consistently, among the down-regulated genes, those containing the SBM within 1200 bp from their start codon tended to have higher $-\log_2$(fold change) values compared with the other down-regulated genes (Fig. 6J). Collectively, these results indicate that PbSIP2 acts as a stronger trans-activating factor on the SBM than on the other motifs, as it preferentially binds to the SBM.

DISCUSSION

Schizogony is a crucial step in the intraerythrocytic proliferation of *Plasmodium*. This process proceeds in two major steps: First, the parasites undergo several asynchronous rounds of nuclear division to form a multinucleated cell; second, multiple progeny merozoites are produced through segmentation (cytokinesis) simultaneously with a final round of karyokinesis (*3*–*6*). Merozoite components are expressed during the first step (*36*, *37*), and buds for each progeny merozoite (including rhoptries and apical ends) are already associated with the nuclei by the beginning of segmentation (*6*). Our results showed that PbSIP2 was expressed during the first step and broadly activated merozoite-related genes, including those involved in merozoite morphogenesis, motility, and invasion. Conditional knockout of *pbsip2* caused significant down-regulation of these merozoite-related genes and resulted in developmental arrest immediately before the second step of schizogony. These results suggest that PbSIP2 plays an essential role in schizogony as a master transcription factor that regulates merozoite formation. Meanwhile, most *pbsip2*-knockout parasites could complete DNA replication during schizont development, suggesting that PbSIP2 does not play a major role in DNA replication or nuclear division. Consistently, only seven genes that are related to "DNA replication and repair" were included among

the PbSIP2 targets. Hence, the role of PbSIP2 does not encompass the entire schizogony process, and other transcription factors should regulate the expression of these genes. We consider that further exploration of these factors is required to better understand transcriptional regulation throughout the schizont development.

The ortholog of PbSIP2 in *P. falciparum*, PfSIP2, has been previously identified as a protein that binds to the SPE2 [(T/G)GTGC(A/G)] motif, which exists in tandem on the subtelomeric var gene promoters (*16*). Based on these results, the authors suggested that PfSIP2 may be involved in chromosome end biology, such as chromosomal replication and segregation. Meanwhile, the authors also mentioned the possibility of other roles for PfSIP2 because subtelomeric SPE2 arrays are only detected in the *P. falciparum* genome. Our results showed that PbSIP2 functions as a trans-regulatory factor for merozoite-related genes through its association with motifs similar to SPE2 (GTGCA and SBM). These motifs are analogous to putative cis-regulatory elements upstream of invasion-related genes previously identified in the *P. falciparum* genome (*11*, *12*). Together, we believe that the primary role of SIP2 in *Plasmodium* is the activation of merozoite-related genes through association with these specific cis-regulatory elements and promotion of merozoite formation. In this study, we demonstrated that SIP2 is particularly conserved in parasites belonging to the orders Haemosporida and Piroplasmida. Alignment of their amino acid sequences showed considerably high similarity within the tandem AP2 domains, including the linker regions, suggesting that their roles may also be conserved. Among the apicomplexan parasites, only those in Haemosporida and Piroplasmida produce an invasive stage that can infect host RBCs. Thus, SIP2 may play a conserved role as a master regulator of RBC-invasive stage development in these parasites and may have an evolutionary origin from their common ancestor.

ChIP-seq and DIP-seq analyses revealed that PbSIP2 binds to GTGCA and has a stronger affinity for RGTGCA. Moreover, PbSIP2 was more strongly associated with the SBM. The 3D structure of the tandem AP2 domains of PbSIP2 showed parallel AP2 domains with a distance that almost matched the distance between the two motifs in the SBM on B-DNA. Thus, the DNA binding configuration of PbSIP2 could be considered as follows: One AP2 domain binds to GTGCA, and its binding becomes stronger when the motif is RGTGCA. Furthermore, when the other AP2 domain additionally binds to 5′-side TGCA in the SBM, PbSIP2 can be most strongly associated with DNA. Campbell *et al.* (*38*) previously reported that in the protein-binding microarrays for the AP2 domains of PfSIP2, the N-terminal side AP2 domain alone was capable of binding to the GGTGCA motif. Thus, the N-terminal AP2 may be responsible for binding to GTGCA in the SBM. We propose that due to these different DNA binding properties of the tandem AP2 domains, variations in transcription levels can be established among PbSIP2 targets. The *Plasmodium* genome encodes approximately 30 sequence-specific transcription factor genes, which seems to be a small number for eukaryotic species with a complex life cycle. Considering this, the parasite might use the above mechanism to control variable transcription levels of stage-specific genes with a small number of trans-regulatory factors. In addition, this DNA binding property of PbSIP2 could also be conserved within the RBC-infecting apicomplexan parasites as the amino acid sequences for the linker between two AP2 domains of PbSIP2 were conserved.

During *Plasmodium* IDC, gene transcription is periodically regulated to generate a simple cascade of stage-specific transcriptomes (*7*). AP2 transcription factors are also periodically expressed during IDC (*38*), and some are essential for the asexual blood-stage development. These suggest that the transcriptional activators responsible for stage-specific transcription during IDC could be identified in these AP2 transcription factors. Nevertheless, the involvement of AP2 transcription factors in stage-specific transcription during IDC has barely been elucidated, and most AP2 transcription factors investigated in previous studies have been reported to be involved in the silencing of subtelomeric multigene families (*39*–*42*). As a rare example, PfAP2-I has been shown to be a transcriptional activator of invasion-related genes during the asexual blood-stage development (*43*). However, the activation of invasion-related genes does not seem to be the major role of PfAP2-I during IDC because the expression of PfAP2-I peaks in the trophozoite stage and decreases during the schizont stage. Moreover, the target genes of PfAP2-I include only a limited number of invasion-related genes. We believe that further investigation is required to evaluate other roles of PfAP2-I, such as ChIP-seq analyses at its peak expression period and conditional knockout analyses to assess the impact of disrupting *ap2-i* in parasite asexual development.

Studies on transcriptional regulators in *Plasmodium* have demonstrated surprisingly simple mechanisms of stage-specific transcriptional regulation throughout the life cycle. Notably, these studies have advanced our understanding of the mechanisms regulating sexual development and formation of the two other invasive stages in *Plasmodium*, namely, sporozoites and ookinetes (*30*–*32*, *44*–*48*). Our data indicate that merozoite formation is also regulated by such simple transcriptional regulation through the function of PbSIP2. We believe that further investigation of other AP2 transcription factors, especially those essential for asexual blood-stage development, would enable us to understand the entire *Plasmodium* IDC at the transcriptional level; i.e., a cascade of stage-specific transcription factors and their target genes explain the periodic gene transcription during IDC.

MATERIALS AND METHODS

Ethical statement

All experiments in this study were performed according to the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health to minimize animal suffering and were approved by the Animal Research Ethics Committee of Mie University (permit number 23-29).

Parasite preparation

All the parasites used in this study were inoculated into ddY mice. The transgenic parasites generated in this study were derived from *P. berghei* ANKA strain. Schizont culture was performed at 37°C using RPMI 1640 medium supplemented with 25% fetal bovine serum (Cosmo Bio), penicillin (100 U/ml), and streptomycin (100 μg/ml; Gibco). For cell cycle synchronization, mature schizonts produced by culturing for 16 hours were harvested by density gradient centrifugation using an iodixanol solution (OptiPrep, Serumwerk Bernburg) mixed with tricine solution to adjust the density to 1.077 g/ml. The harvested schizonts were intravenously injected into mice.

Generation of mutant parasites

For tagging PbSIP2 with GFP, the DNA sequences of the two homologous regions were cloned into the *gfp*-fusion vector, which has an *hdhfr* expression cassette next to *gfp* as a pyrimethamine-selectable marker (*49*), to fuse *pbsip2* in frame with *gfp*. The plasmid was linearized

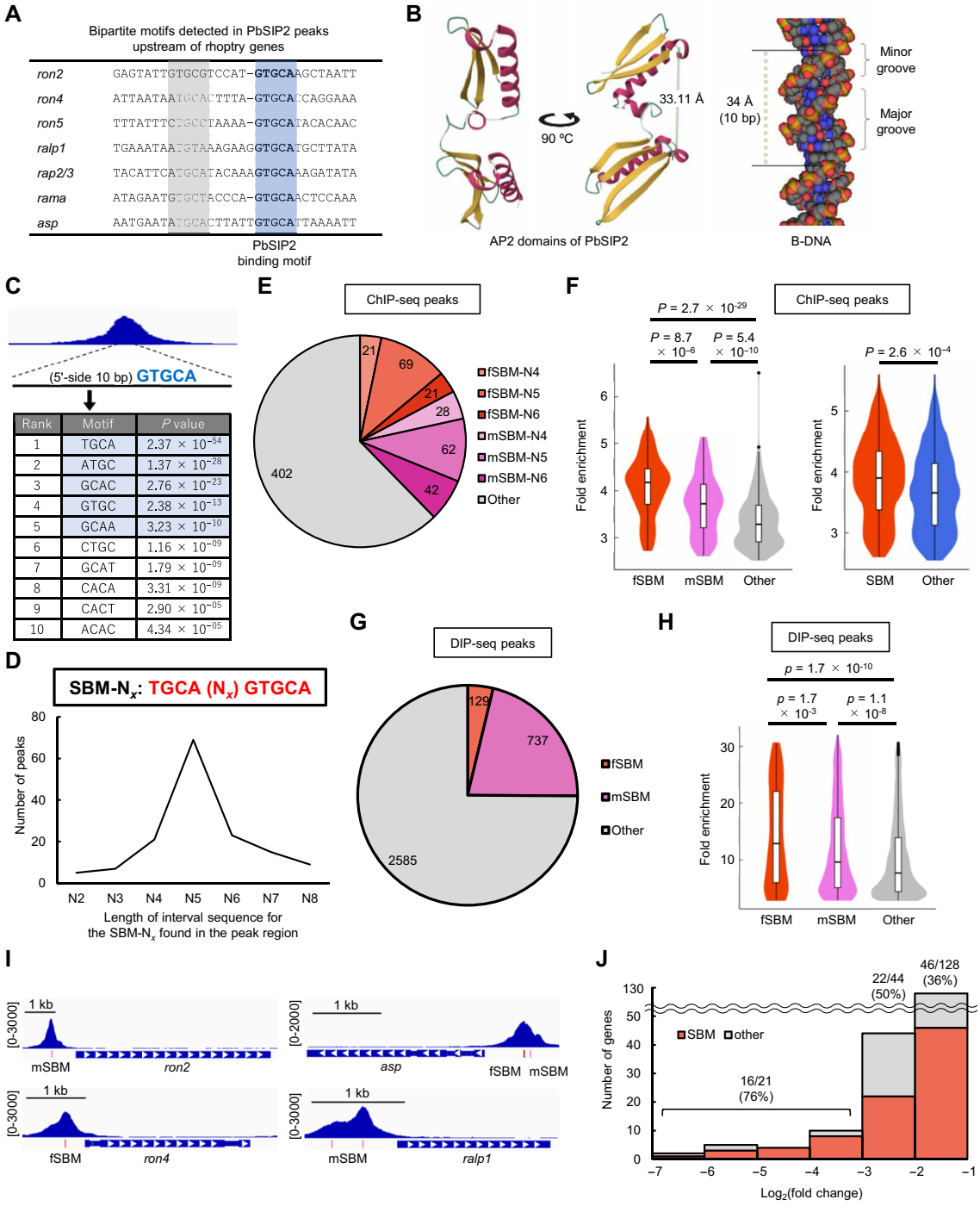

Fig. 6. Association of PbSIP2 with SBM. (**A**) GTGCA-containing bipartite motifs detected in PbSIP2 peaks upstream of rhoptry genes. GTGCA and GTGCA-like motifs are indicated by blue and gray, respectively. (**B**) 3D structure of the PbSIP2 AP2 domains (positions 87 to 214) predicted using AlphaFold2 (*35*). On the right, a B-DNA model depicted using GraphiteLifeExplorer (*50*) is shown. (**C**) Four–base pair motifs enriched within 5′-side 10 bp of GTGCA in the ChIP-seq peak regions. The ChIP-seq data presented in Fig. 3 were used for the analysis. The *P* values were calculated by Fisher's exact test. TGCA motif and its one base–shifted motifs are indicated by blue. (**D**) Number of ChIP-seq peaks containing SBM-N_x with different length of interval sequence in the bipartite motif. (**E**) Pie graph showing the number of ChIP-seq peaks with fSBM (full SBM) and mSBM (SBM with a single mismatch in the 5′ TGCA). (**F**) Violin plot showing distribution of fold enrichment values for the ChIP-seq peaks with fSBM, mSBM, and others (left) and violin plot for those with SBM and RGTGCA (right). The corresponding box plots are shown inside. The *P* values were calculated by two-tailed Student's *t* test. (**G**) Pie graph showing the number of DIP-seq peaks containing fSBM and mSBM. The DIP-seq data presented in Fig. 5 were used for the analysis. (**H**) Violin plot showing distribution of fold enrichment values for the DIP-seq peaks with fSBM, mSBM, and others. (**I**) IGV images showing ChIP-seq peaks on the upstream of PbSIP2 target genes. Red and pink bars indicate location of the fSBM and mSBM, respectively. Scales are indicated in square brackets. (**J**) Histogram showing the number of genes against the $\log_2$(fold change) value in the differential expression analysis between *pbsip2*-cKO^{Rapa-} and *pbsip2*-cKO^{Rapa+}. The percentages of genes associated with SBM are indicated on the graph.

by Xho I and Not I digestion before use in transfection experiments. For gene editing with the CRISPR-Cas9 system using PbCas9 (i.e., loxP-insertion, *gfp*-fusion, and cis-element mutation), donor DNAs were constructed by overlap PCR, cloned into pBluescript KS (+) (Addgene) using the Xho I and Bam HI sites by In-Fusion cloning and then amplified by PCR from the constructed plasmid. The single guide RNA (sgRNA) target sites were designed using CHOPCHOP (https://chopchop.cbu.uib.no/), and sgRNA vectors were constructed as previously described (*24*).

The CenP_RAP1 plasmid was generated from the pCen-GFP plasmid (*46*). The upstream region of *gfp* in pCen-GFP was replaced by 5′ *pbhsp70* using the Kpn I and Nhe I sites, generating the internal control GFP expression cassette. The plasmid was then digested by Sal I, and the expression cassette of RAP1 fused with mCherry, which was generated by the overlap PCR, was cloned at the digested site by In-Fusion cloning.

Transfection was performed using the Amaxa Basic Parasite Nucleofector Kit 2 (Lonza). All transfectants were selected by treatment of mice with pyrimethamine (70 μg/ml) in their drinking water. Recombination was confirmed by PCR and/or Sanger sequencing, and clonal parasites were obtained by limiting dilution. All primers used in this study are listed in data S4.

Fluorescence analysis

Fluorescence analysis was performed using Olympus BX51 microscope with Olympus DP74 camera. Nuclei were stained by incubating infected blood in PBS with Hoechst 33342 (1 ng/ml) for 10 min at 37°C before analysis. Gray scale images were generated using IrfanView.

Flow cytometric analysis was performed using the LSR Fortessa (Becton Dickinson). Nuclei were stained as described above. Cells were gated with forward and side scatter and were assessed for Hoechst (450/50) fluorescence intensity.

Conditional knockout of *pbsip2*-cKO parasites

pbsip2-cKO parasites were cultured for 16 hours, and mature schizonts were harvested by density gradient centrifugation and inoculated into mice. Six hours after inoculation, the whole blood was harvested and split into two cultures. At the beginning of the culture, 1:20,000 volume of rapamycin (Wako) solution (200 μM DMSO stock) was added to one culture, achieving the final concentration of 10 nM (*pbsip2*-cKO^{Rapa+}), and the same volume of DMSO was added to the other (*pbsip2*-cKO^{Rapa-}). For genotyping PCR, genomic DNAs were harvested from 100 μl of each culture. The primers used are listed in data S4.

ChIP-seq and sequencing data analysis

The cell cycle of PbSIP2::GFP parasites was synchronized as described above and cultured from 6 hpi. At 24 hpi, the cultures were passed through a Plasmodipur filter and were immediately fixed in 1% formalin. After fixing for 1 hour at 30°C, RBCs were lysed in ice-cold 150 mM NH_4Cl solution, and then the residual cells were lysed in SDS lysis buffer (50 mM tris-HCl, 1% SDS, and 10 mM EDTA). The cell lysate was sonicated using the Bioruptor (Cosmo Bio) to shear the chromatin and mixed with anti-GFP polyclonal antibodies (ab290, Abcam) conjugated to Protein A Magnetic Beads (Invitrogen). DNA fragments were purified from the immunoprecipitated chromatin and subjected to library construction using the KAPA HyperPrep Kit (Kapa Biosystems). The library was sequenced using the Illumina NextSeq 500. Before immunoprecipitation, DNA fragments were purified from an aliquot of the cell lysate and sequenced to obtain the input sequence data. Two biologically independent experiments were performed for the data analysis.

Sequence data were mapped onto the reference genome sequence of *P. berghei* (v3.0, downloaded from PlasmoDB 46) using Bowtie 2. Reads aligned onto more than two sites were removed from the mapping data, and peaks were called by macs2 callpeak with fold enrichment >2.5 and q value < 0.01 using input sequence data as a control. The parameters for all programs were set to default, unless otherwise indicated. Common peaks in duplicates were defined as those with a distance of less than 150 bp between the peak summits. The enrichment of motifs within 50 bp of the peak summits was analyzed using Fisher's exact test. Genes with peaks 1200-bp upstream from their start codons were identified as target genes. The GO analysis was performed for three aspects—molecular function, cellular component, and biological process—using GSEABase and GOstats programs. All enriched terms from the three analyses were combined and ranked along their P values.

RNA-seq and sequence data analysis

pbsip2-cKO^{Rapa-} and *pbsip2*-cKO^{Rapa+} were prepared as described in the conditional knockout method above. The cultures were passed through a Plasmodipur filter at 26 hpi, and RBCs were lysed in ice-cold 150 mM NH_4Cl solution. Total RNA was extracted from the parasites using ISOGEN II reagent (Nippon Gene), and RNA-seq libraries were prepared using the KAPA mRNA HyperPrep Kit (Kapa Biosystems) for each sample. Libraries were sequenced using the MGI DNBSEQ-G400 (MGI Tech Co. Ltd.). Three biologically independent experiments were performed for each sample. The sequence data were mapped onto the reference *P. berghei* ANKA v3 genome using HISAT2, with the maximum intron length of 1000. The number of reads mapped onto each gene was calculated using featureCounts and compared using DESeq2. Genes in the subtelomeric regions were excluded from the differential expression analysis. The parameters for all programs were set to default, unless otherwise indicated.

Preparation of recombinant protein and DIP-seq analysis

Recombinant proteins were prepared as previously described (*47*). Briefly, the DNA sequence encoding the tandem AP2 domains of PbSIP2 (positions 87 to 214) was cloned into the MBP-fusion vector, pMal-c5X (New England Biolabs), using the Not I and Bam HI sites, and the plasmid was introduced into *Escherichia coli* strain DH5α. The transformed *E. coli* was cultured for 12 hours at 37°C, and then the expression of the MBP-fused protein was induced by isopropyl β-D-thiogalactopyranoside (final concentration of 200 nM) in the culture. Recombinant protein of AP2 domains fused with MBP (MBP::PbSIP2) was purified using amylose resin (New England Biolabs) and recovered in 10 mM maltose solution.

The MBP::PbSIP2 (5 μg) was mixed with *P. berghei* ANKA genomic DNA fragments (2 μg) in 400 μl of binding buffer (100 mM KCl, 2 mM $MgCl_2$, 2 mM tris-HCl, 10 μM $ZnSO_4$, and 10% glycerol) and incubated for 30 min. The recombinant proteins and bound DNA fragments were purified using amylose resin. DNA fragments were subjected to library preparation and NGS using the Illumina NextSeq 500, as described for the ChIP-seq analysis. Genomic DNA fragments before use for DIP were sequenced as the input. The analysis of the sequence data was performed as described for ChIP-seq,

except for the peak calling parameters (fold enrichment >3.0, q value < 0.01). The motif enrichment analyses were separately performed for 7 chromosome sets (chromosomes 1 to 4, 5 to 7, 8 to 9, 10 to 11, 12, 13, and 14) because the analysis using all DIP-seq peaks detected several motifs with P value$^{\text{limit}}$ (Fisher's exact test yields a lower P value when population becomes higher with a constant odds ratio).

ChIP-qPCR and RT-qPCR for reporter experiments

ChIP-qPCR and RT-qPCR were performed using PbSIP2::GFP$^{\text{Cas9}}$ and PbSIP2::GFP$^{\text{Cas9_cismut}}$ to assess the function of the PbSIP2-binding motif as a cis-regulatory motif. For ChIP-qPCR analysis, ChIP experiments were performed at 24 hpi, as described for the ChIP-seq analysis. ChIP and input samples were each prepared using 200 and 10 μl of sonicated cell lysate, respectively. For RT-qPCR analyses, total RNA was harvested at 26 hpi using the ISOGEN II reagent (Nippon Gene), and cDNA was synthesized from the total RNA using the PrimeScript RT reagent Kit with guide DNA Eraser (Takara). Quantification of ChIPed DNA and cDNA was performed by real-time qPCR using the TB Green Fast qPCR Mix (Takara) and Thermal Cycler Dice Real Time System II (Takara). Amplification was performed for 40 cycles, and the cycle threshold (Ct) was detected between 20 and 35 cycles. For ChIP-qPCR analysis, % input values were calculated as $2^{(Ct\text{ChIP}-Ct\text{input})} \times (1/20) \times 100$. Three biologically independent samples were prepared for each experiment and used for analysis. All primers used are listed in data S4.

Supplementary Materials

The PDF file includes:

Figs. S1 to S4
Tables S1 to S4
Legends for data S1 to S4

Other Supplementary Material for this manuscript includes the following:

Data S1 to S4

REFERENCES AND NOTES

1. WHO, *World Malaria Report 2023* (WHO, 2023); www.who.int/publications/i/item/9789240086173.
2. Y. Voß, S. Klaus, J. Guizetti, M. Ganter, Plasmodium schizogony, a chronology of the parasite's cell cycle in the blood stage. *PLOS Pathog.* **19**, e1011157 (2023).
3. M. Read, T. Sherwin, S. P. Holloway, K. Gull, J. E. Hyde, Microtubular organization visualized by immunofluorescence microscopy during erythrocytic schizogony in Plasmodium falciparum and investigation of post-translational modifications of parasite tubulin. *Parasitology* **106**, 223–232 (1993).
4. D. E. Arnot, E. Ronander, D. C. Bengtsson, The progression of the intra-erythrocytic cell cycle of Plasmodium falciparum and the role of the centriolar plaques in asynchronous mitotic division during schizogony. *Int. J. Parasitol.* **41**, 71–80 (2011).
5. S. Klaus, P. Binder, J. Kim, M. Machado, C. Funaya, V. Schaaf, D. Klaschka, A. Kudulyte, M. Cyrklaff, V. Laketa, T. Höfer, J. Guizetti, N. B. Becker, F. Frischknecht, U. S. Schwarz, M. Ganter, Asynchronous nuclear cycles in multinucleated Plasmodium falciparum facilitate rapid proliferation. *Sci. Adv.* **8**, 5362 (2022).
6. R. M. Rudlaff, S. Kraemer, J. Marshman, J. D. Dvorin, Three-dimensional ultrastructure of Plasmodium falciparum throughout cytokinesis. *PLOS Pathog.* **16**, e1008587 (2020).
7. Z. Bozdech, M. Llinás, B. L. Pulliam, E. D. Wong, J. Zhu, J. L. DeRisi, The Transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLOS Biol.* **1**, E5 (2003).
8. X. M. Lu, G. Batugedara, M. Lee, J. Prudhomme, E. M. Bunnik, K. G. Le Roch, Nascent RNA sequencing reveals mechanisms of gene regulation in the human malaria parasite Plasmodium falciparum. *Nucleic Acids Res.* **45**, 7825–7840 (2017).
9. H. J. Painter, N. C. Chung, A. Sebastian, I. Albert, J. D. Storey, M. Llinás, Genome-wide real-time in vivo transcriptional dynamics during Plasmodium falciparum blood-stage development. *Nat. Commun.* **9**, 2656 (2018).
10. L. Chappell, P. Ross, L. Orchard, T. J. Russell, T. D. Otto, M. Berriman, J. C. Rayner, M. Llinás, Refining the transcriptome of the human malaria parasite Plasmodium falciparum using amplification-free RNA-seq. *BMC Genomics* **21**, 1–19 (2020).
11. C. G. Toenhake, S. A. K. Fraschka, M. S. Vijayabaskar, D. R. Westhead, S. J. van Heeringen, R. Bártfai, Chromatin accessibility-based characterization of the gene regulatory network underlying plasmodium falciparum blood-stage development. *Cell Host Microbe* **23**, 557–569.e9 (2018).
12. J. A. Young, J. R. Johnson, C. Benner, S. F. Yan, K. Chen, K. G. Le Roch, Y. Zhou, E. A. Winzeler, In silico discovery of transcription regulatory elements in Plasmodium falciparum. *BMC Genomics* **9**, 70 (2008).
13. S. Balaji, M. M. Babu, L. M. Iyer, L. Aravind, 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.* **33**, 3994–4006 (2005).
14. M. D. Allen, K. Yamasaki, M. Ohme-Takagi, M. Tateno, M. Suzuki, A novel mode of DNA recognition by a β-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J.* **17**, 5484–5496 (1998).
15. T. S. Voss, M. Kaestli, D. Vogel, S. Bopp, H. P. Beck, Identification of nuclear proteins that interact differentially with Plasmodium falciparum var gene promoters. *Mol. Microbiol.* **48**, 1593–1607 (2003).
16. C. Flueck, R. Bartfai, I. Niederwieser, K. Witmer, B. T. F. Alako, S. Moes, Z. Bozdech, P. Jenoe, H. G. Stunnenberg, T. S. Voss, A major role for the Plasmodium falciparum ApiAP2 protein PfSIP2 in chromosome end biology. *PLOS Pathog.* **6**, e1000784 (2010).
17. H. W. Brown, Malaria parasites and other Haemosporidia. *JAMA* **201**, 336–336 (1967).
18. M. Jalovecka, O. Hajdusek, D. Sojka, P. Kopacek, L. Malandrin, The complexity of piroplasms life cycles. *Front. Cell. Infect. Microbiol.* **8**, 393278 (2018).
19. R. L. Jacobs, M. Warren, Sequestration of schizonts in the deep tissues of mice infected with chloroquine-resistant Plasmodium berghei. *Trans. R. Soc. Trop. Med. Hyg.* **61**, 273–275 (1967).
20. B. Franke-Fayard, C. J. Janse, M. Cunha-Rodrigues, J. Ramesar, P. Büscher, I. Que, C. Löwik, P. J. Voshol, M. A. M. Den Boer, S. G. Van Duinen, M. Febbraio, M. M. Mota, A. P. Waters, Murine malaria parasite sequestration: CD36 is the major receptor, but cerebral pathology is unlinked to sequestration. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11468–11473 (2005).
21. L. Mancio-Silva, K. Slavic, M. T. Grilo Ruivo, A. R. Grosso, K. K. Modrzynska, I. M. Vera, J. Sales-Dias, A. R. Gomes, C. R. Macpherson, P. Crozet, M. Adamo, E. Baena-Gonzalez, R. Tewari, M. Llinás, O. Billker, M. M. Mota, Nutrient sensing modulates malaria parasite virulence. *Nature* **547**, 213–216 (2017).
22. N. Jullien, F. Sampieri, A. Enjalbert, J. P. Herman, Regulation of Cre recombinase by ligand-induced complementation of inactive fragments. *Nucleic Acids Res.* **31**, e131 (2003).
23. R. S. Kent, K. K. Modrzynska, R. Cameron, N. Philip, O. Billker, A. P. Waters, Inducible developmental reprogramming redefines commitment to sexual development in the malaria parasite Plasmodium berghei. *Nat. Microbiol.* **3**, 1206–1213 (2018).
24. N. Shinzawa, T. Nishi, F. Hiyoshi, D. Motooka, M. Yuda, S. Iwanaga, Improvement of CRISPR/Cas9 system by transfecting Cas9-expressing Plasmodium berghei with linear donor template. *Commun. Biol.* **3**, 426 (2020).
25. B. Liffner, S. Frölich, G. K. Heinemann, B. Liu, S. A. Ralph, M. W. A. Dixon, T. W. Gilberger, D. W. Wilson, PfCERLI1 is a conserved rhoptry associated protein essential for Plasmodium falciparum merozoite invasion of erythrocytes. *Nat. Commun.* **11**, 1411 (2020).
26. B. Liffner, J. M. Balbin, G. J. Shami, G. Siddiqui, J. Strauss, S. Frölich, G. K. Heinemann, E. M. Edwards, A. Alder, J. S. Wichers, D. J. Creek, L. Tilley, M. W. A. Dixon, T.-W. Gilberger, D. W. Wilson, Cell biological analysis reveals an essential role for Pfcerli2 in erythrocyte invasion by malaria parasites. *Commun. Biol.* **5**, 121 (2022).
27. L. Chen, S. Lopaticki, D. T. Riglar, C. Dekiwadia, A. D. Uboldi, W. H. Tham, M. T. O'Neill, D. Richard, J. Baum, S. A. Ralph, A. F. Cowman, An EGF-like protein forms a Complex with PfRh5 and is required for invasion of human erythrocytes by Plasmodium falciparum. *PLOS Pathog.* **7**, e1002199 (2011).
28. R. L. Clements, A. A. Morano, F. M. Navarro, J. P. Mcgee, E. W. Du, V. A. Streva, S. E. Lindner, J. D. Dvorin, Identification of basal complex protein that is essential for maturation of transmission-stage malaria parasites. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2204167119 (2022).
29. A. Sinha, K. R. Hughes, K. K. Modrzynska, T. D. Otto, C. Pfander, N. J. Dickens, A. A. Religa, E. Bushell, A. L. Graham, R. Cameron, B. F. C. Kafsack, A. E. Williams, M. Llinás, M. Berriman, O. Billker, A. P. Waters, A cascade of DNA-binding proteins for sexual commitment and development in Plasmodium. *Nature* **507**, 253–257 (2014).
30. T. Nishi, I. Kaneko, M. Yuda, Identification of a novel AP2 transcription factor in zygotes with an essential role in Plasmodium ookinete development. *PLOS Pathog.* **18**, e1010510 (2022).
31. M. Yuda, I. Kaneko, Y. Murata, S. Iwanaga, T. Nishi, Targetome analysis of malaria sporozoite transcription factor AP2-Sp reveals its role as a master regulator. *mBio* **14**, e0251622 (2023).

32. R. A. Campelo Morillo, X. Tong, W. Xie, S. Abel, L. M. Orchard, W. Daher, D. J. Patel, M. Llinás, K. G. Le Roch, B. F. C. Kafsack, The transcriptional regulator HDP1 controls expansion of the inner membrane complex during early sexual differentiation of malaria parasites. *Nat. Microbiol.* **7**, 289–299 (2022).
33. A. J. Gossett, J. D. Lieb, DNA Immunoprecipitation (DIP) for the determination of DNA-binding specificity. *CSHarb. Protoc.* **2008**, pdb.prot4972 (2008).
34. E. S. Zuccala, A. M. Gout, C. Dekiwadia, D. S. Marapana, F. Angrisano, L. Turnbull, D. T. Riglar, K. L. Rogers, C. B. Whitchurch, S. A. Ralph, T. P. Speed, J. Baum, Subcompartmentalisation of proteins in the rhoptries correlates with ordered events of erythrocyte invasion by the blood stage malaria parasite. *PLOS ONE* **7**, e46160 (2012).
35. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
36. B. Mahajan, A. Selvapandiyan, N. J. Gerald, V. Majam, H. Zheng, T. Wickramarachchi, J. Tiwari, H. Fujioka, J. K. Moch, N. Kumar, L. Aravind, H. L. Nakhasi, S. Kumar, Centrins, cell cycle regulation proteins in human malaria parasite Plasmodium falciparum. *J. Biol. Chem.* **283**, 31871–31883 (2008).
37. S. Absalon, J. A. Robbins, J. D. Dvorin, An essential malaria protein defines the architecture of blood-stage and transmission-stage parasites. *Nat. Commun.* **7**, 11449 (2016).
38. T. L. Campbell, E. K. de Silva, K. L. Olszewski, O. Elemento, M. Llinás, Identification and genome-wide prediction of DNA Binding specificities for the ApiAP2 family of regulators from the malaria parasite. *PLOS Pathog.* **6**, e1001165 (2010).
39. M. Sierra-Miranda, S. S. Vembar, D. M. Delgadillo, P. A. Ávila-López, A. M. Herrera-Solorio, D. Lozano Amado, M. Vargas, R. Hernandez-Rivas, PfAP2Tel, harbouring a non-canonical DNA-binding AP2 domain, binds to Plasmodium falciparum telomeres. *Cell. Microbiol.* **19**, 10.1111/cmi.12742 (2017).
40. R. M. Martins, C. R. Macpherson, A. Claes, C. Scheidig-Benatar, H. Sakamoto, X. Y. Yam, P. Preiser, S. Goel, M. Wahlgren, O. Sismeiro, J.-Y. Coppée, A. Scherf, An ApiAP2 member regulates expression of clonally variant genes of the human malaria parasite *Plasmodium falciparum*. *Sci. Rep.* **7**, 14042 (2017).
41. E. F. G. Cubillos, I. O. Prata, W. L. Fotoran, L. Ranford-Cartwright, G. Wunderlich, The transcription factor PfAP2-O influences virulence gene transcription and sexual development in Plasmodium falciparum. *Front. Cell. Infect. Microbiol.* **11**, 669088 (2021).
42. A. K. Subudhi, J. L. Green, R. Satyam, R. P. Salunke, T. Lenz, M. Shuaib, I. Isaioglou, S. Abel, M. Gupta, L. Esau, T. Mourier, R. Nugmanova, S. Mfarrej, R. Shivapurkar, Z. Stead, F. Ben Rached, Y. Ostwal, R. Sougrat, A. Dada, A. F. Kadamany, W. Fischle, J. Merzaban, E. Knuepfer, D. J. P. Ferguson, I. Gupta, K. G. Le Roch, A. A. Holder, A. Pain, DNA-binding protein PfAP2-P regulates parasite pathogenesis during malaria parasite blood stages. *Nat. Microbiol.* **8**, 2154–2169 (2023).
43. J. M. Santos, G. Josling, P. Ross, P. Joshi, L. Orchard, T. Campbell, A. Schieler, I. M. Cristea, M. Llinás, Red blood cell invasion by the malaria parasite is coordinated by the PfAP2-I transcription factor. *Cell Host Microbe* **21**, 731–741.e10 (2017).
44. I. Kaneko, S. Iwanaga, T. Kato, I. Kobayashi, M. Yuda, Genome-wide identification of the target genes of AP2-O, a Plasmodium AP2-family transcription factor. *PLOS Pathog.* **11**, e1004905 (2015).
45. Y. Murata, T. Nishi, I. Kaneko, S. Iwanaga, M. Yuda, Coordinated regulation of gene expression in Plasmodium female gametocytes by two transcription factors. *eLife* **12**, RP88317 (2024).
46. M. Yuda, I. Kaneko, S. Iwanaga, Y. Murata, T. Kato, Female-specific gene regulation in malaria parasites by an AP2-family transcription factor. *Mol. Microbiol.* **113**, 40–51 (2020).
47. M. Yuda, I. Kaneko, Y. Murata, S. Iwanaga, T. Nishi, Mechanisms of triggering malaria gametocytogenesis by AP2-G. *Parasitol. Int.* **84**, 102403 (2021).
48. G. A. Josling, T. J. Russell, J. Venezia, L. Orchard, R. van Biljon, H. J. Painter, M. Llinás, Dissecting the role of PfAP2-G in malaria gametocytogenesis. *Nat. Commun.* **11**, 1503 (2020).
49. M. Yuda, S. Iwanaga, S. Shigenobu, G. R. Mair, C. J. Janse, A. P. Waters, T. Kato, I. Kaneko, Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. *Mol. Microbiol.* **71**, 1402–1414 (2009).
50. S. Hornus, B. Lévy, D. Larivière, E. Fourmentin, Easy DNA modeling and more with GraphiteLifeExplorer. *PLOS ONE* **8**, e53609 (2013).

Acknowledgments: We thank A. Noro (Mie University) for the technical assistance. **Funding:** This work was supported by the following research funds: Grant-in-Aid for Scientific Research (B) 23 K27400 (to M.Y. and T.N.), Grant-in-Aid for Scientific Research (C) 24 K10187 (to T.N.), and Grant-in-Aid for Scientific Research (C) 23 K06515 (to I.K.) **Author contributions:** Conceptualization: T.N. and M.Y. Methodology: T.N. and M.Y. Investigation: T.N., I.K., and M.Y. Formal analysis: T.N. and I.K. Visualization: T.N. Data curation: T.N. and M.Y. Resources: T.N. and M.Y. Validation: T.N. and M.Y. Supervision: M.Y. Project administration: M.Y. Funding acquisition: T.N., I.K., and M.Y. Writing—original draft: T.N. and M.Y. Writing—review and editing: T.N. and M.Y. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All FASTQ files for ChIP-seq, RNA-seq, and DIP-seq experiments are available from the Gene Expression Omnibus database (accession numbers GSE266502, GSE266503, and GSE266869).

Submitted 18 August 2024
Accepted 19 February 2025
Published 21 March 2025
10.1126/sciadv.ads5458