Title

Identification of Antimalarial Compounds that Inhibit Apicomplexan AP2 Proteins

Word Count: 11,627

Authors

Timothy Russell[1,2,3,4], Erandi K. De Silva[6], Valerie Crowley[1], Kathryn Shaw-Saliba[7], Namita Dube[1], Gabrielle Josling[1,3,4], Charisse Flerida A. Pasaje[8], Irene Kouskoumvekaki[9], Gianni Panagiotou[10,11], Jacquin C. Niles[8], Marcelo Jacobs-Lorena[7], C. Denise Okafor[1,5], Francisco-Javier Gamo[12], Manuel Llinás[1,2,3,4,5]

1. Department of Biochemistry and Molecular Biology, Pennsylvania State University.
2. Huck Institutes Center for Eukaryotic Gene Regulation (CEGR), Pennsylvania State University.
3. Huck Institutes Center for Malaria Research (CMaR), Pennsylvania State University.
4. Huck Institutes Center for Infectious Disease Dynamics, Pennsylvania State University.
5. Department of Chemistry, Pennsylvania State University.
6. Lewis-Singler Institute for Integrative Genomics, Princeton University.
7. Department of Molecular Biology and Immunology, Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health.
8. Department of Biological Engineering, Massachusetts Institute of Technology.
9. Department of Systems Biology, Technical University of Denmark, Kgs. Lyngby, Denmark.
10. Systems Biology and Bioinformatics, Leibniz Institute for Natural Products Research and Infection Biology, Hans Knöll Institute.
11. Department of Medicine, the University of Hong Kong, Hong Kong SAR, China
12. Global Health Medicines, GlaxoSmithKline, Tres Cantos, Spain.

Abstract

Plasmodium parasites are reliant on the Apicomplexan AP2 (ApiAP2) transcription factor family to regulate gene expression programs. The AP2 DNA binding domains have no homology to the human or mosquito hosts, making them potential drug targets.

Using an *in-silico* screen to dock thousands of small molecules into the crystal structure of the AP2-EXP (Pf3D7_1466400) AP2 domain (PDB:3IGM), we identified compounds that interact with this domain. Four compounds were found to compete for DNA binding with AP2-EXP and at least one additional ApiAP2 protein. Our top ApiAP2 competitor compound perturbs the transcriptome of *P. falciparum* trophozoites and results in a decrease in abundance of $\log_2$ fold change > 2 for 50% (46/93) of AP2-EXP target genes identified. Additionally, two ApiAP2 competitor compounds have anti-*Plasmodium* activity against *P. berghei* mosquito stage parasites. In summary, we describe a novel set of antimalarial compounds that are targeted against the ApiAP2 family of proteins. These compounds may be used for future chemical genetic interrogation of ApiAP2 proteins or serve as starting points for a new class of antimalarial therapeutics.

Introduction

Malaria is a disease that is caused by intracellular parasites of the genus *Plasmodium* and represents a significant health and economic burden worldwide[1]. The most virulent of the human infectious malaria parasites is *Plasmodium falciparum*, which caused more than 200 million cases and resulted in over 600 thousand deaths in 2021[1]. Unfortunately, resistance has been reported for every antimalarial therapeutic deployed to date, necessitating the need for malaria drugs that target new parasite processes[2]. To successfully proliferate, *P. falciparum* parasites must develop through a complex lifecycle that includes intracellular and extracellular stages in the human and *Anopheles* mosquito hosts[3]. To initiate an infection, *Plasmodium* sporozoites are deposited into the

bloodstream of the human host when an infected mosquito takes a blood meal. Sporozoites invade hepatocytes and replicate into daughter merozoites, which egress to invade red blood cells (RBCs). Inside RBCs the parasite replicates asexually over the course of ~48 hours, developing through three major morphological stages referred to as: ring, trophozoite, and schizont. Schizonts rupture at 48 hours post invasion (hpi), releasing 16-32 daughter merozoites into the blood stream. Merozoites reinvade new RBCs and begin replicating again. This iterative invasion and replication, called the intraerythrocytic development cycle (IDC), is responsible for the periodic fevers associated with malaria due to destruction of RBCs[3]. In each round of replication up to 10% of parasites will commit to a sexual development program which results in non-replicating gametocytes[4]. Gametocytes are the only blood stage that are competent to re-establish infection in a mosquito upon ingestion in the blood meal. Male and female gametocytes develop into gametes that fuse to become a zygote, which ultimately develops back into sporozoites, thereby completing the *Plasmodium* lifecycle.

Transcriptomics-based evidence has shown that up to 80% of *P. falciparum* protein coding transcripts are developmentally regulated during the IDC, with peaks of expression occurring at specific IDC stages[5–9]. This pattern has been described as a "just-in-time" cascade due to the observation that for many transcripts, the timing of maximal abundance precedes the developmental stage at which the protein product is required for progression through the IDC[6,9,10]. Transcriptional regulation in *Plasmodium* parasites is driven through a combination of modulation of the chromatin landscape[11–17], the general transcriptional machinery[18,19], and sequence-specific transcription factors[20,21]. Although many of the general transcription factors (GTFs) required for RNA

Polymerase II activity are conserved between *Plasmodium* and model eukaryotes[19,22,23], *Plasmodium* parasites encode seemingly few sequence-specific DNA binding transcription factors.

The major family of sequence specific transcription factors is the 27-member family of Apicomplexan Apetala AP2 (ApiAP2) proteins[21,24,25]. ApiAP2 proteins all contain one to three ~60 amino acid AP2 DNA binding domains that are analogous to the plant APETALA2/Ethylene Responsive Factor (AP2/ERF) domain[22]. Their plant-like origin distinguishes them from all other known transcription factors in humans or mosquitoes[20]. Over half of the ApiAP2 proteins are predicted to be essential for the IDC in *P. falciparum* based on a random mutagenesis screen[26], and in targeted knockout screens only 11/23 ApiAP2 coding sequence disruptions attempted were successful[27–29]. In the rodent malaria model species *P. berghei*[30] and *P. yoelii*[31], 11 ApiAP2 proteins have developmental lethal phenotypes in the sexual blood or mosquito stages. ApiAP2 transcription factors can cause activation or repression of transcription of specific genes, and they are involved in diverse lifecycle processes including invasion, mutually exclusive antigenic gene expression, sexual differentiation, and heat stress tolerance[27,32–40]. Chromatin interacting proteins such as histone deacetylases and methyltransferases have received recent attention as potential drug targets against malaria parasites[41,42]. No efforts to date have focused on targeting the sequence specific DNA binding ApiAP2 transcription factors, which are potential drug targets due to their essentiality for the asexual blood stage and lack of conservation with the host.

In this study we conducted an *in silico* chemical screen against the published crystal structure of the AP2-EXP AP2 domain[43] to predict competitors of DNA binding activity. AP2-EXP is predicted to be essential for the *P. falciparum* asexual blood stage[27,29,33] and its orthologue *P. berghei* AP2-Sp is a master regulator of sporogony[44]. Several compounds identified in the screen can compete for DNA binding activity with the AP2-EXP AP2 domain *in vitro*. Each of these compounds competes DNA binding by at least one additional ApiAP2 protein, suggesting the potential for multi-ApiAP2 inhibition. We also find that modifications to the substitution state of an AP2 domain competitor compound can modulate its ability to compete DNA binding. Our lead compound kills *P. falciparum* parasites in concordance with the stage specific expression of AP2-EXP. We measured AP2-EXP genomic occupancy by ChIP-seq and found high correlation between dysregulated mRNA transcripts in the presence of our lead compound and AP2-EXP bound genes, corroborating the hypothesis that AP2-EXP is inhibited *in vivo*. Finally, we find that two AP2 domain competitor compounds are also active against *Plasmodium* parasite development in the mosquito stage. Overall, our results demonstrate the potential to chemically compete the DNA binding activity of ApiAP2 proteins for the purposes of providing proof-of-concept for a new antimalarial strategy, and interrogation of the function of *P. falciparum* blood stage essential ApiAP2 proteins.

Results

In silico Computational Modeling of AP2-EXP Competitors

To identify competitors of DNA binding by an ApiAP2 protein, we used AP2-EXP (PF3D7_1466400), which is the only ApiAP2 protein whose AP2 domain is structurally

characterized (PDB accession: 3IGM). *In silico* molecular docking was run on AUTODOCK[45] using over ten thousand small molecules from both the Tres Cantos Antimalarial Set (TCAMS)[46] and the Drug Bank[47] against the crystal structure of the AP2-EXP AP2 protein domain (**Fig 1A, Fig S1**). Docking hits were prioritized based on the free energy of interaction and proximity to amino acids necessary for base-specific DNA contacts. In some cases, the compound that generated a high docking score was not available for purchase, so alternative choices with >0.9 Tanimoto similarity score were identified, computationally docked, and used for further evaluation **(All compounds listed in Table S1).** These docking simulations resulted in a final list of nine top scoring predicted competitors of DNA binding by AP2-EXP. Each compound was assigned an alphabetical identifier A-I **(Fig 1B)** for use in this study. The list contains both known antimalarials from the Tres Cantos Antimalarial Set and molecules from the Drug Bank that had not been previously tested for antimalarial activity. To assess antimalarial activity, we assayed *P. falciparum* parasite viability when exposed to each of the putative AP2-EXP competitors and determined that all nine compounds kill asexual *P. falciparum* parasites at micromolar concentrations (range 11-170µM) in a 48-hour asexual stage growth inhibition assay **(Fig 1B, Fig S2)**.

Four Predicted AP2-EXP Competitors Compete DNA Binding *in vitro*

The nine predicted competitors of DNA binding by the AP2-EXP AP2 domain were tested in a competitive electrophoretic mobility shift assay (EMSA). The minimum mass of AP2-EXP required to visualize DNA binding **(Fig S3A)** was mixed with each compound and incubated prior to adding the DNA oligonucleotide **(DNA Oligos in Table S2)**. Competition of DNA binding activity should result in a loss of signal for the

higher molecular weight DNA probe which is bound by AP2-EXP. Four of the nine compounds (A, B, C and I) **(Fig 2A)** were able to effectively compete for DNA binding by AP2-EXP *in vitro*. Three of the four AP2 domain competitors (Compounds A, B and C) have a benzoxazole core moiety, while Compound I is made up of planar rings **(Fig 2B)**.

To assess whether DNA binding competition was specific to the AP2-EXP AP2 domain, we repeated competitive EMSAs with three different purified *P. falciparum* AP2 domains: AP2-I Domain 3 (AP2-I D3)[25,32] **(Fig S3B)**, AP2-HS Domain 1 (AP2-HS D1)[25,36] **(Fig S3C)**, and PfSIP2 Domain 1 (PfSIP2 D1)[35] **(Fig S3D)**, again using the minimum amount of protein needed to visualize binding of each ApiAP2 protein's cognate DNA oligonucleotide **(Table S2)**. AP2-I D3[32] and PfSIP2 D1[35] are essential for the IDC, while *ap2-hs* deletion is tolerated when parasites are grown below physiological temperatures[36]. We found that Compounds A, B, C, and I can compete DNA binding by AP2-I **(Fig S4A)**, Compounds B and I compete DNA binding by AP2-HS D1 **(Fig S4B)** and Compounds B, C and I compete DNA binding by PfSIP2 D1 **(Fig S4C)**.

As a control for compound specificity, we performed competitive EMSAs using the non-*Plasmodium* AP2 domain protein *Arabidopsis thaliana* Ethylene Response Factor 1[48] (AtERF1) **(Fig S3E)** and the High Mobility Group Box (HMGB) domain protein SOX2 **(Fig S3F)**[49]**.** SOX2 is a human encoded transcription factor and has no homology to the AP2 domain. Only Compound I can compete AtERF1 **(Fig S5A)**. SOX2 is competed by

Compound A only **(Fig S5B)**. Overall, Compounds B and C have specificity for Apicomplexan AP2 domains *in vitro*. (EMSA Results Summarized in **Table S3**.)

To test whether the ApiAP2 competing compounds interfere non-specifically with DNA binding proteins by intercalating with DNA, we conducted an ethidium bromide exclusion assay[50]. In this assay, if the molecule of interest can compete away ethidium bromide from intercalating DNA, it will reduce ethidium bromide fluorescence. Compared to the positive control DRAQ5[50], only Compound I was able to effectively displace ethidium bromide **(Fig S6)**. Based on these results, we focused on Compounds B and C due to their specificity for *P. falciparum* AP2 domain proteins and their lack of DNA intercalation.

<u>Structural Analogs of Compound B Differ in DNA Binding Competition Activity</u>

We leveraged the availability of compounds with high structural similarity to Compound B due to its origin in the Tres Cantos Antimalarial Set[46] to determine whether specific features of the chemical structure impact DNA binding competition. We tested the 4 closest available analogs to Compound B from the Tres Cantos Antimalarial Set based on similarity score[46] (Designated as Compounds B-1, B-2, B-3 and B-4) **(Fig 3A)** for competition against AP2-EXP. Compounds B-1 and B-4 compete DNA binding by AP2-EXP **(Fig 3B)** and AP2-I D3 **(Fig S7A)**, while Compounds B-2 and B-3 do not compete DNA binding by either AP2-EXP or AP2-I D3 **(Fig 3B, S7A)**. Compound B-1 is the least effective competitor relative to Compounds B and B-4 **(Fig 3C)**. Non-AP2 domain competitor Compounds B-2 and B-3 each have halogen atoms substituted on the benzene ring in the para and ortho, or para positions of the benzene ring, respectively.

Conversely, Compounds B, B-1, and B-4, have methyl or ethyl groups substituted onto the benzene ring and can compete DNA binding **(Fig 3D)**. Compound B-1 has both a methyl group and a chlorine atom substitution. If the halogen atom substitution decreases DNA binding competition, this mix of substitutions is consistent with its lower efficacy of DNA binding competition relative to Compounds B and B-4**.** Compounds B, B-1 and B-4 have no difference in DNA intercalation ability as measured by ethidium bromide exclusion **(Fig S7B)**. We followed up on these results by running molecular dynamics simulations on AP2-EXP docked with Compounds B,B-1, B-2, B-3, and B-4. While Compound B remains stably docked with AP2-EXP over the course of the simulation, Compound B-3 moves away from the protein and does not re-dock **(Supplemental Movies 1-5)**. Compounds B-1, B-2, and B-4 remain stably associated with AP2-EXP in the simulation **(Supplemental Movies 1-5)**. In summary, substituting bulky, electronegative groups onto Compound B can abolish its ability to compete for DNA binding with AP2-EXP, potentially due to a steric clash.

Compound C Activity Coincides With the Expression of AP2-EXP

To assess whether Compounds B and C are likely to share a common ApiAP2 target, we determined their parasite killing phenotypes against *P. falciparum* during the IDC. Compound H was used as a control since it does not compete AP2-EXP AP2 domain DNA binding **(Fig 2A)**. Each compound was spiked (40µM) into synchronous asexual blood stage *P. falciparum* parasites at 24 hours post invasion and parasite progression was determined at 24 and 48 hours post spike-in **(Fig 4A)**. At 24 hours post spike-in, parasites in the presence of either Compounds B or C had failed to progress past the

mid-trophozoite stage and did not morphologically progress any further by 48 hours post spike-in **(Fig 4B)**. In contrast, parasites spiked with Compound H progressed through one IDC before failing to reinvade in the next cycle **(Fig 4B)**. Compounds B and C therefore may have a common protein target as they exhibit the same morphological impact on the asexual blood stage and similar timing of action. As expected, each compound exposure prevented further parasite growth within 48 hours of spike-in **(Fig S8)**.

Due to its commercial availability, we then continued with Compound C as the lead compound for phenotypic and molecular characterization. To determine the precise timing of Compound C antimalarial activity, synchronous parasites were monitored throughout the 48-hour IDC in the presence of either 40µM Compound C or DMSO control after spiking in Compound C at 6 hpi. While the control parasites progressed normally, the Compound C spiked parasites failed to progress beyond the mid-trophozoite stage at 24-30 hours post invasion **(Fig S9)**. To further narrow down the timing of the anti-*Plasmodium* action of Compound C, 40µM of Compound C was spiked into a parasite culture at 8-hour intervals beginning at 14 hpi. Parasite growth was quantified in the next cycle by counting ring stage parasites at 54 hpi. This growth assay revealed that Compound C has no antimalarial effect after 46 hpi **(Fig 4C).** In support of this result, washing 40µM Compound C out of culture at 8-hour intervals beginning at 14 hpi only rescued parasite growth until 30 hpi, albeit with diminishing effectiveness as the parasites approached 30 hpi **(Fig 4D).** Taken together, these results suggest that

Compound C maximally inhibits of IDC progression of *Plasmodium* parasites between 30-46 hpi.

To validate the stage in the IDC at which AP2-EXP is expressed, we used Selection Linked Integration (SLI)[51] to introduce an endogenous C-terminal tag containing a 2XFK506-Binding Protein (FKBP) dimerization domain and GFP **(Fig S10A)** to create AP2-EXP::GFP. We observed maximal abundance of AP2-EXP::GFP in 30 hpi trophozoites by both Western blot and live fluorescence microscopy **(Fig 4E, F).** Maximal protein abundance of AP2-EXP in mid-trophozoites was independently confirmed in a different parasite line expressing AP2-EXP tagged with 2XHA and the TetR:DOZI mRNA repression aptamer (AP2-EXP::HA) **(Fig S10B, Fig S11)**. AP2-EXP was not detected in maturing Stage III gametocytes **(Fig 4E)**. These results demonstrate that the maximal protein abundance of AP2-EXP at 30hpi coincides with the timing of action for Compound C.

Compound C Disrupts the Parasite Transcriptome with Stage Specificity

Since AP2-EXP is predicted to act as a transcription factor[29,33] we hypothesized that changes in parasite transcript abundance should occur when AP2-EXP DNA binding is competed by Compound C. To test this hypothesis, we cultured parasites in 12µM (0.66xIC_{50}) Compound C or DMSO vehicle and measured comparative RNA abundance by DNA microarray at seven time points during the asexual blood stage. As a quality control for normal IDC progression, we found a high correlation between a set of "control" genes[52,53] with an established 48-hour periodicity with maxima at different

times throughout the 3D7 *P. falciparum* IDC[6] in our DMSO control and Compound C mRNA abundance profiles **(Fig S12A-D)**. Phenotypically, 12µM of Compound C kills less than 50% of parasites in one asexual cycle **(Fig S13A, B)**. To identify perturbations to parasite RNA abundance, we compared all detected transcripts at each time point assayed between the DMSO control and Compound C dosed parasites **(Expression values in Table S4)**. The Spearman Correlation between samples deviated most significantly at 30hpi, (Corr. = -0.16) **(Fig 5A)** as expected from the compound activity measured **(Fig 4B-4D)**.

To follow up on this observation, we quantified differences in transcript level that are unique to the progression from 24 to 30 hpi timepoints for Compound C or DMSO vehicle control populations using Linear Models for Microarray Data (LIMMA[54]) **(Differentially Expressed genes in Table S4)**. Overall, 463 RNA transcripts are decreased in abundance by $\log_2$ fold change >2 between 24-30 hpi in the presence of Compound C. Gene ontology term (GO-term) analysis[55] of the differentially regulated transcripts **(Full GO Terms in Table S4)** revealed an enrichment for genes that encode proteins important for the parasite to invade or modify red blood cells (e.g., RON3, GAP45, RhopH2, RhopH3, MSP1, MSP6). We independently determined differential expression over the entire time course using the RNA time course specific software RNITS[56] and found high overlap with our LIMMA analyses **(Overlaps in Table S4)**. Therefore, the maximal perturbation of the transcriptome **(Fig 5A)**, action of Compound C **(Fig 4B)**, and maximal abundance of AP2-EXP **(Fig 4E, F)** all occur at roughly 30 hpi in the IDC.

AP2-EXP Genomic Occupancy Correlates with Compound C Induced Transcript Dysregulation

We used chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) to determine the genome wide binding occupancy of AP2-EXP. A total of three ChIP-seq samples from highly synchronous trophozoites were collected at 30 hpi using both the AP2-EXP::GFP (2 replicates) and AP2-EXP::HA (1 replicate) parasite lines. Immunoprecipitation of full length AP2-EXP was confirmed by Western blot **(Fig S14A)**, and a co-immunoprecipitation blot confirmed that AP2-EXP interacts with Histone H3 **(Fig S14B)**. Peaks of occupancy were determined for each replicate using the MACS2 peak calling program[57] **(List of Peaks in Table S5)**. Peaks detected in 2/3 ChIP-seq replicates were considered for downstream analysis. In aggregate, AP2-EXP binds 240 genomic loci, corresponding to 101 total target genes **(Genes in Table S5)**. Enrichment of AP2-EXP was well conserved between the GFP and HA tagged parasite lines, as indicated by nearly identical coverage in a metagene plot of conserved AP2-EXP target genes **(Fig 5B)**. A 'no epitope control' ChIP done on a wild type 3D7 parasite line with the anti-GFP antibody detected only 2 peaks, neither of which overlapped with those of AP2-EXP::GFP or AP2-EXP::HA. Both the conserved AP2-EXP peaks and individual replicates were highly enriched for the known AP2-EXP and PbAP2-Sp DNA sequence motif CATGCA[29,58] **(Fig 5C, Fig S15A-C).**

We then determined whether there is overlap between the transcripts dysregulated in the presence of Compound C and AP2-EXP target genes identified by ChIP-seq. Out of the 463 transcripts decreased in abundance by $\log_2$ fold change > 2 in the presence of

Compound C, 46 are AP2-EXP targets **(Fig 5D)**. This represents 50% (46/93) of AP2-EXP target genes detected in the RNA time course **(Fig 5D)**. In general, the transcripts with the greatest decrease in abundance are AP2-EXP targets. We found that 7/14 total transcripts decreased in abundance by $\log_2$ fold change > 5, and 16/33 decreased in abundance by $\log_2$ fold change > 4, are AP2-EXP targets **(Fig 5E, Table S4)**. Using GO-term analysis[55], we found that AP2-EXP targets that are dysregulated at 24-30 hpi have functions related to red blood cell invasion and host remodeling **(Fig 5F, Table S4)**. To assess whether AP2-EXP DNA binding at ChIP-seq defined target promoter regions is impacted by Compound C, parasite cultures were spiked with either 40µM Compound C or DMSO vehicle control at 30 hpi and AP2-EXP occupancy was measured by ChIP-quantitative PCR (ChIP-qPCR). AP2-EXP DNA binding decreases at five specific peaks of occupancy (GAP45, SIP2, RON3, RALP, and AMAI) in the presence of Compound C **(Fig S16)**, supporting a link between AP2-EXP inhibition and Compound C.

We also compared AP2-EXP genomic occupancy (ChIP-seq) with several previously published chromatin datasets in order to further probe its function as a sequence specific transcription factor. AP2-EXP binds chromatin primarily in nucleosome depleted regions[59] compared to randomized control loci **(Fig S17A, B)**. AP2-EXP occupies sites near the Transcription Start Site (TSS) of target genes[60] **(Fig S18)**. A comparison of RNA abundance data[12] to AP2-EXP target genes revealed that many target gene mRNA transcripts are increased in abundance at 24-30hpi, while a subset changes to low abundance **(Fig S19)**. Overall, AP2-EXP genomic occupancy is highly predictive of

enrichment for activating histone marks H3K9ac[17] and H3K4me3[17] as well as the activating chromatin reader Bromodomain Protein 1 (BDP1)[15], again in comparison to randomized control loci **(Fig S20A, B).** Although the known repressing histone markers H3K9me3 and H3K36me2/3[13] do not correlate well with AP2-EXP occupancy, enrichment of the Heterochromatin Protein 1 (HP1)[61] repressing chromatin reader does overlap modestly with AP2-EXP **(Fig S20A, B)**. A GO-term analysis[55] of all AP2-EXP target genes revealed enrichment for invasion and red blood cell modification, **(Table S4)** as was found for the genes that decrease in abundance in the presence of Compound C **(Fig 5F, Table S4)**. In aggregate, these findings suggest that Compound C is able to disrupt the function of AP2-EXP as a sequence specific transcription factor.

ApiAP2 Competitor Compounds B and C are Active Against Mosquito Stage *Plasmodium* Parasites

Elucidation of the role of essential *P. falciparum* ApiAP2 proteins in the IDC is challenging due to the inability to knockout asexual blood stage essential genes from the parasite's haploid genome. Previous attempts to knockout AP2-EXP[27–29,33] failed to produce parasite lines in which the AP2 domain coding sequence was disrupted. We similarly did not recover transgenic parasites after attempting to completely disrupt the coding sequence for AP2-EXP using the targeted gene disruption (pSLI-TGD) system[51] in three independent attempts, suggesting that AP2-EXP is essential for the IDC (not shown). We also attempted to disrupt AP2-EXP protein expression with the conditional knockdown approaches Knock Sideways[51], TetR:DOZI mRNA repression[62], and *glms* ribozyme mediated cleavage[63]. Each of these genetic systems failed to mediate protein

knockdown or mislocalization despite successful integration into the parasite genome **(Fig S10A, B, Fig S21, Fig S22)**. To account for this limitation, we tested the ApiAP2 competitor Compounds B and C against rodent infectious *Plasmodium berghei* parasites due to the well-characterized genetic phenotype of the AP2-EXP orthologue PbAP2-Sp[44,58]. PbAP2-Sp is a master regulator of sporogony and therefore has a phenotype that can be correlated to the antimalarial activity of our lead compounds. We hypothesized that if Compounds B and C can compete DNA binding by the AP2 domain of PbAP2-Sp (89% identity (49/55 amino acids) to the AP2-EXP AP2 domain[55]), they should affect the development of sporozoites in the mosquito stage of *Plasmodium berghei.*

We infected *Anopheles stephensi* mosquitoes with *P. berghei* parasites and allowed the parasites to develop for 10 days to the oocyst stage. At this time, mosquito midguts were injected with the IC_{50} measured for asexual blood stage *P. falciparum* of Compounds B, C or F **(Fig 1B)**. Compound F **(Fig 2A, B)** was used as a non AP2 domain competing control. On day 14 post infection, surviving mosquitoes were dissected and both oocysts and midgut sporozoites were counted **(Fig 6A)**. Mosquitoes injected with Compounds F and C developed comparable oocyst numbers to the DMSO control **(Fig 6B)** while Compound B prevented the development of midgut oocysts entirely **(Fig 6B)**. Compound C treated mosquitoes developed 10-fold fewer midgut sporozoites than the control **(Fig 6C)**. Therefore, consistent with their hypothesized *in vitro* activity against ApiAP2 proteins, Compounds B and C are strong inhibitors of *P. berghei* mosquito stage development into sporozoites *in vivo*.

Discussion

ApiAP2 transcription factors are unique to Apicomplexan parasites due to their plant-like domain architecture[20,22] and many are essential to asexual blood stage development, making them valuable as potential drug targets. It is therefore desirable to discover chemical scaffolds that can target ApiAP2 proteins and to uncover details about their biological function. Additionally, phenotypic characterization has been especially challenging for essential asexual blood stage ApiAP2 proteins as no complete inducible knockdown has been described to date. Here, we used *in silico*, biochemical, and genetic approaches to demonstrate that ApiAP2 proteins can serve as drug targets against *Plasmodium* parasites. We have identified a starting point for chemical-based competition of ApiAP2 proteins and demonstrated a specific effect on parasite development and transcription by Compound C, a benzoxazole moiety compound which competes DNA binding by AP2-EXP, AP2-I D3, and PfSIP2 D1 *in vitro*.

Our results demonstrate the utility of *in-silico* screening to rapidly search a large and diverse set of compounds to target a specific protein domain. This approach allowed us to identify nine potential ApiAP2 competitors from a library of over 10,000 compounds. Compounds B and C both lack off-target activity against AtERF1 and SOX2 DNA binding, indicating selective activity against the AP2 domains of ApiAP2 proteins. Although the mode of DNA binding is shared between AP2-EXP and AtERF1[48], the specific amino acids that mediate DNA binding are not strictly conserved[43]. Since the predicted mode of docking for Compounds A and B would interfere with the AP2-EXP dimerization pocket in addition to DNA binding residues **(Fig S1)**, disruption of

protein:protein interaction may be their mode of action. As AtERF1 is not known to homodimerize, disruption of protein interaction would not impact its DNA binding. Compound I has the greatest magnitude of DNA major groove intercalation ability and does not compete DNA binding by SOX2, consistent with the observation that HMGB family proteins bind DNA *via* the minor groove[64].

We find that the DNA binding competition capacity of Compound B analogs varies *in vitro* based on substitutions to the benzene ring, thereby illuminating a possible determinant of ApiAP2 DNA binding competition. In a molecular dynamics simulation, Compound B remains stably associated with AP2-EXP, while the non ApiAP2 competitor Compound B-3 does not. This suggests that a steric clash between AP2-EXP and the bromine atom on Compound B-3 is responsible for its lack of DNA binding competition. By extension, this may explain the lower efficacy of DNA binding competition observed for Compounds B-1, B-2, and B-3 compared to B and B-4.

ApiAP2 competitor Compounds B and C both cause asexual *P. falciparum* parasites to stop morphologically progressing in the trophozoite stage. Importantly, we observed that more than 50% of parasites successfully reinvade in the presence of 0.66xIC_{50} (12µM) Compound C, which we used for our RNA time course. Therefore, the changes in transcript abundance measured after exposing parasites to Compound C are likely specific to the time points maximally affected, rather than a cumulative effect of completely halting parasite progression in the IDC. Compound C induced growth disruption results in a decrease in the abundance of many transcripts that are AP2-EXP

targets. Affected transcripts are enriched for pathways important for parasites to invade new RBCs and modify the RBC surface.

Our ChIP-seq datasets demonstrate that AP2-EXP preferentially binds the DNA sequence CATGCA *in vivo*. This is the same core DNA motif that was previously shown to be bound by AP2-EXP and PbAP2-Sp *in vitro*[44,65]. Many transcripts that correspond to AP2-EXP target genes are activated, and a small subset are downregulated, at 25-35 hpi in the IDC when AP2-EXP is maximally expressed. Transcripts encoding proteins important for parasite RBC modification and invasion (EBA-181, GAP45, MSP6, EBA-140), and genome integrity (H2A, H2B, H4, DNA Mismatch Repair Protein) are represented among those upregulated and targeted by AP2-EXP. Transcripts downregulated following AP2-EXP occupancy primarily code for proteins of unknown function, or proteins that are expressed outside of the asexual blood stage (*e.g.*, PHISTb, P113). AP2-EXP occupancy is well correlated with the activating chromatin reader PfBDP1[15]. This coincides with results reported for PfAP2-I, which binds many of the same genomic loci as AP2-EXP and BDP1[11,29,32]**.** Together, these results point to transcription activation *via* chromatin modulation as a key role of AP2-EXP in the asexual blood stage. Future studies are needed to understand the precise mechanisms by which AP2-EXP and other ApiAP2 proteins regulate transcription.

New inferences can be made about the function of AP2-EXP based on the description of its genome-wide DNA binding sites and transcript dysregulation in the presence of Compound C. This demonstrates the utility of combining chemical and traditional

genetics because previous attempts to genetically characterize AP2-EXP have not provided a complete picture of its function in the asexual blood stage. The *ap2-exp* AP2 domain coding region has been shown to be essential for the asexual blood stage by both saturating mutagenesis and targeted deletion attempts[26,27,29,33]. Conversely, the coding region beyond the AP2 domain was truncated in two studies[26,33]. Transcriptomic profiling of a truncated AP2-EXP parasite line revealed that many transcripts encoding exported proteins were dysregulated[33]. Unexpectedly, AP2-EXP target genes identified in our study overlap poorly with the differentially regulated transcripts reported as a result of truncation of the AP2-EXP protein **(Table S5)**[33]. There is greater overlap between our study and the AP2-EXP ChIP-seq in trophozoites recently described by Shang *et al*[29], with 50/101 target genes overlapping **(Table S5)**. It is therefore possible that the truncated AP2-EXP retains some critical function that relies on DNA binding. This would explain the apparent essentiality of the AP2 domain, while the full-length protein has other roles beyond DNA binding that impact a different subset of non-asexual blood stage essential genes.

Since Compounds B and C each compete AP2 domain proteins in addition to AP2-EXP *in vitro*, they could potentially be used to assay the function of blood stage essential ApiAP2 proteins in addition to AP2-EXP. A comparison of transcripts downregulated in the presence of Compound C to the DNA binding occupancy of AP2-I[32], for which knockdown has not been reported, revealed that 54/85 AP2-I target genes are decreased in abundance by $\log_2$ fold change >2 at 24-30 hpi. Interestingly, 22/55 AP2-I bound genes that decrease in abundance when parasites are exposed to Compound C

are also AP2-EXP targets, implying the potential for co-regulation of certain gene subsets. This overlap is consistent with our EMSA results and corroborates the hypothesis that Compound C competes both AP2-EXP and AP2-I *in vivo*. Previously, PfSIP2 knockdown has been shown to impact a small subset of genes[35]. Only 3/22 of these are decreased in abundance in the presence of Compound C **(ApiAP2 Target Gene Results Summarized in Table S4)**. This may reflect the cryptic relationship between PfSIP2 DNA binding and transcriptional control, as was noted when PfSIP2 was originally characterized as playing a role in genome integrity and heterochromatin formation[35].

Several ApiAP2 proteins have been implicated in oocyst and sporozoite development in a *P. berghei* ApiAP2 knockout screen[30]. Since Compounds B and C both inhibit mosquito stage development of *P. berghei* with different phenotypes, it is possible that they target different ApiAP2 proteins *in vivo*. The AP2-EXP orthologue PbAP2-Sp acts as the master regulator of infectious sporozoite formation by binding upstream of many genes essential for sporogony[44,58]. Therefore, Compounds B and C should minimally inhibit sporozoite development, which is consistent with our results. Overall, these data support a potential for multi-AP2 domain competition by Compounds B and C *in vivo*.

The IC_{50} values against *P. falciparum* for the ApiAP2 domain competitors identified in this study are all in the micromolar range. Therefore, in order for these compounds to be effective antimalarial drugs, future work will be needed to increase parasite killing potency. Encouragingly, Compounds A, B, and C all have the same core benzoxazole moiety and compete DNA binding by ApiAP2 proteins, but vary widely in their measured

IC_{50}s. This indicates that changes which do not affect ApiAP2 DNA binding competition can change the potency against parasites. Encouragingly, our present results indicate that Compounds B and C can inhibit *P. berghei* parasite development without killing mosquitoes, suggesting that ApiAP2 competitor compounds may be tolerated by the host and could serve as transmission blocking agents[66]. Due to their inclusion in the Tres Cantos Antimalarial Set, Compounds B, B-1, B-2, B-3, and B-4 have all been tested for activity against the human HEPG2 cell line[46]. ApiAP2 competitor Compounds B and B-4 inhibit HEPG2 cell growth by just 4 and 8%, respectively[46]. Therefore, either drug may be prioritized for further development based on their selectivity for *Plasmodium* parasites.

AP2-EXP was originally characterized at the genetic level over a decade ago, and it is the only ApiAP2 protein for which the structure has been solved. Surprisingly, the field still lacks a clear understanding of the biological role of AP2-EXP in the asexual blood stage. Our study has enabled us to make inferences about the role of AP2-EXP and set a proof of principle for targeting the highly unique Apicomplexan AP2 DNA binding proteins as a new antimalarial strategy.

Materials and Methods

Parasite lines

Parasites were grown at 37°C, 5% O2, 7% CO2 using RPMI1640 media supplemented with hypoxanthine and .5% Albumax II (Thermo). The parasite lines used in this study were AP2-EXP::GFP, AP2-EXP::HA, AP2-EXP::*glms*, and wild Type Pf3D7 (Malaria Research and Reference Reagents Repository).

To create AP2-EXP::GFP the C-Terminal region of AP2-EXP was cloned **(Table S2)** and inserted into the plasmid pSLI::2xFKBP[51] for endogenous tagging by single homologous recombination. Red blood cells were pre-loaded with 100µg of plasmid as previously described[67] and Pf3D7 wild type parasites were added at 1% parasitemia. After 24 hours, cultures were maintained in media with 2.5nM WR99210 until parasites were detectable by Giemsa stained slides, at which time they were switched to G418 selection at 400µg/mL as described[51]. Parasites were cloned by limiting dilution and screened for correct genomic integration **(Fig S10A)**. Transgenic parasites with the correct C-Terminal tag were further transfected with pLyn-FRB-mCherry[51] for inducible mislocalization by preloading red blood cells with 100µg plasmid and selection using 1µg/mL blasticidin to create AP2-EXP::GFP + pLyn. All components were confirmed by genotyping PCR. From this point onward, parasites were maintained in media with 2.5nM WR99210. All ChIP and Western blot experiments using AP2-EXP::GFP except Figure S22 were collected using AP2-EXP::GFP without the pLyn mislocalization plasmid.

To create AP2-EXP::HA we used the pSN054 vector system[68]. The right homology region (RHR) was amplified by PCR **(Table S2)**, and the recodonized left homology region (LHR) was synthesized using the BioXP™ 3200 System. Single guide RNA fragments synthesized using the BioXP™ 3200 System (**Table S2)** were cloned into the linearized pSN054 donor vector[68] that incorporates a C-terminal V5 and 2x-hemagglutinin (HA) tags, 10x RNA aptamer array, the TeTR-DOZI expression cassette

containing the blasticidin S-deaminase gene for selection in the parasites, and the reporter gene Renilla luciferase (RLuc). The 3'-end was re-codonized to avoid further modifications of the edited locus in the parasite via removal of the PAM site. Donor vector generation was carried out via Gibson assembly, and the final constructs were sequence verified and further confirmed by restriction digests. Transfections into Cas9- and T7 RNA polymerase-expressing NF54 parasites[69] were carried out by preloading erythrocytes with linearized vectors as described[70]. Cell cultures were maintained in 500 nM anhydrotetracycline (aTc, Sigma-Aldrich 37919), drug selection with 2.5 mg/mL of Blasticidin S was initiated four days after transfection, and recovered parasites were monitored via Giemsa smears and RLuc measurements. Parasites were then cloned by limiting dilution and PCR genotyped **(Fig S10B)**.

To create AP2-EXP::*glms*::HA, the C-terminal coding region of AP2-EXP was amplified **(Table S2)** and cloned into the plasmid pSLI*::*3xHA::*glms*[71] to create pSLI:AP2-EXP*::glms*::HA*.* Red blood cells were pre-loaded with 100µg of plasmid as described[67] and 3D7 wild type parasites were added in at 1% parasitemia. Cultures were maintained on 2.5nM WR99210 until parasites were detectable by Giemsa stained smear, then switched to 400µg/mL G418 as described[51]. Surviving parasites were cloned by limiting dilution, then genotyped **(Figure S21)**. From this point on AP2-EXP::*glms*::HA parasites were maintained in media with 2.5nM WR99210. The human biological samples were sourced ethically and their research use was in accord with the terms of the informed consents under an IRB/EC approved protocol.

Genomic DNA Isolation

Parasite cultures were lysed with .1% Saponin in 1xPBS and collected by centrifugation at 1500 RPM, then resuspended in 1xPBS. Genomic DNA was then isolated using the Qiagen DNeasy nucleic acid isolation kit according to the manufacturer's instructions.

AP2-EXP Knockdown Assays

For AP2-EXP::GFP + pLyn, synchronous parasites were split into two populations. 250nM Rapalog was added to one parasite group for 48 hours as described[51] while the second group was used as a control. Addition of Rapalog did not affect parasite growth (not shown). After 48 hours parasite protein was collected in nuclear and cytosolic fractions and AP2-EXP::GFP, Histone H3, and Pf Aldolase were detected by Western blot.

For AP2-EXP::HA parasites were grown routinely in the presence of 500nM Anhydrotetracycline (Atc). Synchronous parasites were split into two groups. Atc was washed out of one group, while the second group was maintained with Atc as a control. Parasites were maintained for 120 hours before protein harvest. Parasite growth was not affected by removal of Atc (not shown). AP2-EXP::HA and Histone H3 were detected by western blot.

For AP2-EXP::*glms*::HA synchronous parasites were split into two populations. 5mM glucosamine was added to one parasite group for 72 hours as described[63] while the second group was used as a control. Addition of glucosamine did not affect parasite

growth (not shown). After 72 hours parasite protein was harvested for detection of AP2-EXP::*glms*::HA and Histone H3 by western blot.

In silico **docking screen**

The Tres Cantos Antimalarial Set of 13533 small molecules was retrieved from the supplementary material of Gamo *et al*[46] and downloaded in Spatial Data File (SDF) format from BatchEntrez. 4603 small molecule structures were downloaded from DrugBank[47] version 2.5 in SDF format. Using the SDF library, a 3-dimensional MOL2 library was created with the use of BALLOON and the Merck Molecular Force Field. BALLOON was used to calculate up to 20 conformations using a genetic algorithm and to select the conformation with the lowest energy of conformation. The 3-dimensional MOL2 library was then prepared for AutoDock using the python script "ligand_prepare4.py" from AutoDockTools[72]. The prepared ligands were then saved in PDBQT file format. The python script "ligand_prepare4.py" automatically defines the torsions of the molecule.

The homodimerized AP2-EXP AP2 domain was downloaded from the solved crystal structure (PDB ID 3IGM). AutoDockTools was used to remove water molecules, add hydrogens, and calculate charges. Residues Arg-88 and Asn-118 were found to be flexible using FlexPred and set as flexible residues for docking. The flexible residues were saved as flexible residues in a flexible residue file and the remaining molecule was saved in as rigid residues in a rigid residue file as specified in the AutoDock manual[72]. The rigid residues file was used to pre-calculate an energy grid of the macromolecule

using autogrid4[72]. A set of 14 grid map types S, Cl, F,A, Br, N, P, OA, SA, C, I, HD, Ca, NA were calculated for the macromolecule using AutoGrid[72].

From the prepared ligands a docking set was prepared for AutoDock with the macromolecule and pre-calculated energy grids. For each docking set a docking parameter file was prepared, with the Lamarckian Genetic Algorithm (LGA) specified as the search algorithm for AutoDock. To allow for high throughput parallel processing PERL was used on top of AutoDock to manage the dockings. The algorithm was initially set to run with a maximum of 25000 energy evaluations and 20 repeats. Since the calculations are computationally intensive, this setting was used as initial screening of the ligands, to identify a subset of ligands for a more thorough evaluation[72]. The 1000 best hits from the GSK compound evaluation and the DrugBank compound evaluation were selected for a more thorough evaluation. The docking parameters were reconfigured to 250000 energy evaluations and 100 repeats and docking repeated. The docking results were examined with PERL, to provide an automated approach to the interpretation of the docking results. The docking results were evaluated using two different approaches. The top candidates for competition of DNA were selected using a PERL script to filter based on having a geometric center within 10 Å of the location of sense-strand DNA binding. Hits were then filtered based on having a predicted free energy of interaction < -5 kJ/Mol.

IC_{50} Determination

The IC_{50} of each putative ApiAP2 competing compound was determined using a 48-hour growth inhibition assay adapted from[73]. Parasites were seeded at .5% parasitemia and 4% hematocrit on a 96 well plate. Drug or DMSO vehicle control was added in triplicate to each well and parasites were incubated in standard culture conditions for 48 hours. Parasite cultures were then frozen overnight and resuspended in 1:1 volume of Sybr Green Lysis Buffer (20mM Tris-HCl, pH 7.5; 5mM EDTA; 0.08% Triton X-100; 0.008% saponin, 1xPBS). After 1 hour of incubation at room temperature, plates were read at excitation/emission 485/535nm. Growth values were normalized to the vehicle control. IC_{50} values were calculated and growth curves were plotted using GraphPad Prism.

Recombinant Protein Expression

Recombinant AP2 domains AP2-EXP, AP2-I D3, AP2-HS D1, and PfSIP2 D1, were overexpressed and purified from BL21 PlysS *E. coli* as described previously[25], using the same plasmid constructs. Recombinant AtERF1 was purified using the same method after cloning the AtERF1 AP2 domain[48] into the pGex4t-1 overexpression vector **(Table S2)** Briefly, *E. coli* cultures grown to optical density .6 were induced to overexpress recombinant protein by adding 40mM IPTG overnight at room temperature. Bacteria pellets were collected by centrifugation and frozen overnight at –80°C. The pellets were then lysed by sonication in sonication buffer (25mM Tris-Cl pH8.0, 150mM NaCl). GST tagged fusion proteins were purified overnight by incubation with GST affinity beads (Clontech) according to the manufacturer's protocol. Recombinant protein was recovered by elution with 10mM reduced glutathione and quantified by Braford reagent

(Pierce). Input, flowthrough, and eluate fractions were analyzed by SDS-PAGE to ensure recovery of the full- length recombinant protein. Recombinant full-length SOX2 protein was purchased from Abcam (ab169843).

Gel-Shifts

Gel-shifts were done using the Thermo Light Shift EMSA kit with recommended buffer components (10mM Tris pH 7.5, 50mM KCl, 1mM DTT, 6mM MgCl2, 60ng/µL Poly DiDC, 65ng BSA). Recombinant proteins were titrated to empirically determine the minimum mass required for DNA binding and this mass **(Fig S3)** was used for each gel-shift unless otherwise specified. PAGE purified DNA probes with a 5' biotin ligated on the forward DNA strand, along with an unlabeled complementary stand, were purchased from IDT **(Table S2)**. DNA probes were double stranded by heating to 95°C, followed by stepwise cooling in annealing buffer (10mMTtris-Cl pH7.5, 1mM EDTA, 10mM NaCl). Each recombinant protein was incubated in DNA binding mixture plus competitor compound for 15 minutes prior to addition of the cognate double stranded DNA oligonucleotide. Protein, competitor, and DNA oligonucleotide were incubated together for an additional 5 minutes. The mixtures were separated on a .5x TBE polyacrylamide gel, transferred to a nylon membrane (Amersham) at 50 Volts for 30 minutes, and probed using the Light Shift nucleic acid detection module (Thermo) according to the manufacturer's protocol. All gels were imaged using a Bio Rad chemiluminescence imager.

Ethidium Bromide Exclusion Assay

10µM of double stranded DNA (sequence TGCATGCA, purchased from IDT) in .01mM EDTA, 9.4mM NaCl, 2mM HEPES buffer pH 7.9 was incubated for 10 minutes in the presence of 100nM ethidium bromide. DNA with no ethidium bromide was used in parallel as a background control. Fluorescence was measured at excitation/emission 546nm/595nm for each well. Following the baseline reading, each putative AP2 competing compound was added. DRAQ5 DNA dye was used as a positive control for knockdown of ethidium bromide fluorescence by DNA intercalation. Ethidium bromide exclusion assays were performed in technical triplicate using a 96 well plate.

Molecular Dynamics Simulations

Five compounds were prepared for molecular dynamics (MD) simulation: B, B-1, B-2, B-3, and B-4. In preparation for MD, parameters for the compounds were obtained using Antechamber[74–76] with the Generalized Amber Force Field. Starting conformations for each compound bound to AP2-EXP were based on predictions from docking. All complexes were prepared using the tleap module of AmberTools[77] with the protein.ff14SB forcefield[78]. Each complex was solvated in an octahedral box of TIP3P water with a 10 -Å buffer around the protein complex. Na^+ and Cl^- ions were added to neutralize the protein and achieve physiological conditions. All MD minimizations and simulations were performed using Amber with GPU acceleration[79,80]. First, complexes minimized with 5000 steps each of steepest decent and conjugate gradient minimization with 500 kcal/mol·$Å^2$ restraints on all complex atoms. Restraints were reduced to 100 kcal/mol·$Å^2$ and the minimization protocol was repeated. Restraints were then retained

only on the compound for a final minimization step. Following minimization, all complexes were heated from 0 to 300 K using a 100-ps run with constant volume periodic boundaries and 10 kcal/mol·Å^2 restraints on all protein and compound atoms. To equilibrate complexes, 10 ns of MD was performed first with 10 kcal/mol·Å^2, then with 1 kcal/mol·Å^2 restraints on protein and compound atoms using the NPT ensemble. With kcal/mol·Å^2 restraints retained on complexes, 500 ns production simulations were performed. A 2-fs timestep was used and all bonds between heavy atoms and hydrogens were fixed with the SHAKE algorithm[81]. A cut-off distance of 10 Å was used to evaluate long-range electrostatics with Particle Mesh Ewald (PME) and for van der Waals forces. The 'strip' and 'trajout' commands of the CPPTRAJ module[82] were used to remove solvent atoms and extract 50,000 evenly spaced frames from each simulation for analysis.

AP2 Domain Competitor Phenotyping Assays

For Compound B, C, and H spike in phenotyping **(Fig 4B)**, highly synchronous 3D7 wild type *P. falciparum* parasites were spiked with 40µM of each drug at 24 hpi. An equivalent volume of DMSO was used as a control. Parasites were morphologically assessed by Giemsa staining at 24 and 48 hours post spike in (48 and 72 hpi).

For continuous Compound C exposure phenotyping **(Fig S9)**, highly synchronous 3D7 wild type *P. falciparum* parasites were spiked with 40µM Compound C or equivalent DMSO vehicle control at 6 hpi. Parasite progression was then monitored through the IDC by Giemsa staining at 6-hour intervals.

For Compound C spike in and washout assays **(Fig 4C, D)**, 40µM of Compound C was spiked into highly synchronous 3D7 wild type *P. falciparum* parasites at 8-hour intervals starting at 14 hpi. For Compound C washout assays, 40µM of Compound C was washed out of highly synchronous cultures at 8-hour intervals starting at 14 hpi. For each assay an equivalent volume of DMSO vehicle control was used as a control for normal growth.

40µM of compound was used for all phenotyping assays because it was the maximum concentration achievable in culture after resuspending each compound in 100% DMSO. Higher concentrations of compound caused parasite mortality due to the high volume of DMSO required (>1% volume/volume). All assays were performed in triplicate and parasitemia was counted by Giemsa-stained slides. Error bars represent the standard deviation of the mean.

Preparation of RNA for DNA Microarray

DNA microarrays were prepared using the protocol described in[52]. Briefly, RNA was collected from highly synchronous 3D7 wild type *P. falciparum* parasites using TRIzol Reagent in accordance with manufacturer's protocols. After collecting a control timepoint at 6 hpi, parasites were cultured with either 12µM Compound C (0.66xIC_{50}) or DMSO vehicle control. Parasite RNA was then harvested at 12, 18, 24, 30, 42, and 48 hpi. Purified RNA was reverse transcribed using SuperScript II according to the manufacturer's protocol, with dUTP-aminoallyl and a mixture of nonamer and oligo-dT

primers added. cDNA was concentrated using DNA Clean and Concentrator-5 kits. cDNA was labeled using Amersham CyDye Reactive Dyes and re-purified. cDNA for each sample was labeled using Cy5, while Cy3 was used for the reference pool made up of mixed stage asexual cDNA. cDNA was hybridized onto Agilent DNA microarrays[52], washed, and scanned using an Axon 4200A Scanner. Agilent Feature Extraction Software version 11.0.1.1 with the protocol GE2-v5_95_Feb07_nospikein was used to extract signal intensities. Raw signal intensities for all microarrays and the .gal file 3D7v7.1+apimito used for feature extraction are available in **Table S7**.

Analysis of DNA Microarrays

Microarray data were extracted and normalized using the R package LIMMA[54]. Arrays were normalized using robust splines normalization and within arrays parameters selected. Average expression values were calculated per gene by averaging the $\log_2$ Cy5 (cDNA)/Cy3 (Reference Pool) signal intensity across all probes. Correlation plots for the Compound C vs. DMSO control total transcriptome and control genes were made using the $\log_2$ signal intensities for all detected genes using the R package CorrPlot downloaded from: https://github.com/taiyun/corrplot. The LIMMA eBayes function was used on average gene abundance data to determine changes in transcript abundance between 24 and 30hpi that occur differentially in DMSO control vs. Compound C dosed parasites. Volcano plots were made with LIMMA eBayes data using the Enhanced Volcano package in R downloaded from: https://github.com/kevinblighe/EnhancedVolcano. The RNITS[56] R package was used to model differences in gene expression across the entire time course between the DMSO

control and Compound C spike in parasites. RNITS was run using the parameters: center genes, normalize by intensity, and background filter probes. The RNITS model for differential expression was fit at the gene level. All data from LIMMA, eBayes, and RNITS is provided in **Table S4**. Heat maps for control genes and AP2-EXP target genes were made using Java Treeview[83] after means centering and log transformation of transcript abundance values with Cluster3.0[84].

ChIP Sample Preparation

ChIP-seq was performed as described in[38]. Highly synchronous parasites were grown up to 5-10% parasitemia and AP2-EXP ChIP was performed at 30 hpi for all replicates. Parasites were first chemically crosslinked by addition of 1% formaldehyde for 10 minutes at 37°C, followed by 5 minutes of quenching in 125mM Glycine on ice. The supernatant was removed by centrifugation and parasites were resuspended in 1xPBS with .1% saponin for 5 minutes at 4°C to lyse red blood cells. Following lysis, parasites were washed in cold 1xPBS and lysed in Lysis Buffer (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM EDTA pH 8.0, 0.1 mM EGTA pH 8.0, 1 mM DTT, 1x Pierce Protease Inhibitor) on ice for 30 minutes. NP-40 was added to a final concentration of .25% and parasites were homogenized in a douncer homogenizer using 100 strokes per 2X10^9 parasites. Nuclei were separated from the supernatant by centrifugation at 14000 RPM for 10 minutes.

Nuclei were then sonicated in Shearing Buffer (0.1% SDS, 10 mM Tris pH 8.1, 1 mM EDTA, 1X Pierce Protease Inhibitor) using an M220 Covaris Sonicator with the following settings: duty cycle 5%, peak incident power 75 W, cycles per burst 200, temperature

7°C, time 300 seconds. Sonicated material was diluted 1:5 in Dilution Buffer, (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.1, 150 mM NaCl) pre-cleared using Magna ChIP Protein A+G Magnetic Beads (Sigma) for 2 hours, then the Input fraction was separated and set aside. Immunoprecipitation was performed overnight with .5mg/mL 3F10 anti-HA (Sigma) antibody or .1mg/mL Ab290 (Abcam) anti-GFP and immune complexes were collected using Magna ChIP Protein A+G Magnetic Beads. Immunoprecipitated material was washed five times total, once with Low Salt at 4°C (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 150 mM NaCl), High Salt at 4°C (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 500 mM NaCl), Lithium Chloride Buffer at 4°C (0.25 M LiCl, 1% NP-40, 1% Deoxycolate, 1mM EDTA, 10 mM Tris-HCl pH 8.1), and twice at room temperature with 1xTE buffer. Immune complexes were then eluted from the beads using Elution Buffer (1% SDS, .1M NaHCO3). Crosslinking was reversed by overnight incubation with .2M NaCl at 65°C, then RNA and protein were removed by sequential treatments with RNAseA and Proteinase K. DNA was isolated using a Minelute PCR Purification Column (Qiagen) and DNA concentration was measured using a Qbit dsDNA HS assay kit (Life Technologies).

DNA Library Preparation

ChIP-seq DNA libraries were prepared as described in[38] using the NEBNext II DNA library kit (New England Biolabs) according to the manufacturer's instructions. Briefly, DNA ends were blunted, and a single A-tail overhang was added. Next, barcoded DNA adaptors for Illumina sequencing (Bioo Scientific) were ligated to the sample and DNA

was amplified by 11 cycles of PCR using Kapa HiFi DNA Polymerase. DNA libraries were size selected for 250nt fragments and library quality was assessed using an Agilent 2100 Bioanalyzer or TapeStation. Libraries were quantified by Qbit dsDNA HS assay kit and sequenced using a HiSeq 2500 (Replicate GFP1) or NextSeq 550 (Replicates GFP2 and HA3) Illumina sequencer. AMPure XP beads (Beckmann Coulter) were used to size select and purify DNA between NEBNext II library preparation steps.

Whole genome sequencing DNA libraries were prepared using the Illumina Tru-Seq PCR free DNA library kit and sequenced using a HiSeq 2500.

qPCR

ChIP-qPCR samples were collected from 30-35 hpi trophozoites following two hours of Compound C (40μM) or DMSO vehicle control spike in. Primer pairs to be used for ChIP-qPCR were first evaluated to check for 80-110% efficiency using sonicated genomic DNA. RT-qPCR was carried out using Sybr Green Polymerase master mix (Thermo) with the specified primer concentration **(Table S2)**. The Ct was calculated using SDSv1.4 (Applied Biosystems) software, averaged over technical triplicate. The percent of Input per immunoprecipitated DNA fraction was calculated using the delta Ct method. Each assay was performed in biological triplicate, with the exception of the *ama1* primer pair, where n =2. Error bars represent standard deviation. Data was obtained using an Applied Biosystems 7300 Real-Time PCR Machine.

ChIP-seq data analysis

First, reads in each DNA library were trimmed to remove low quality base calls and Illumina adaptor sequences using Trimmomatic[85] with a quality score cutoff of 20. FastQC[86] was used to assess the quality of each DNA library following this step. Reads were then mapped to the *Plasmodium falciparum* genome version 36 downloaded from PlasmoDB using BWA-Mem[87]. Multiply mapped reads were filtered out using Samtools[88]. Filtered and mapped Input and immunoprecipitate bam files were used for peak calling by MACS2[57] with the following parameters: effective genome size 20000000, No Model, q 0.01. Peaks called in each ChIP-seq replicate were overlapped using Bedtools[89] Intersect, and peaks that occur in a minimum of 2/3 replicates were used for downstream analysis. For data visualization using IGV, .bam files were converted to $\log_2$ Immunoprecipitate fraction/Input fraction bigwigs using DeepTools[90] BamCompare. Bedtools[89] ClosestBed was used to correlate regions called as peaks of occupancy by MACS2 to the closest gene. Results were filtered based on 1.5kB proximity to the MACS2 peak and strandedness of the putative target gene. For comparison of ChIP-seq library coverages to Transcription Start Sites and Start Codons, DeepTools PlotHeatmap was used. Matrix files underlying PlotHeatmap were created using DeepTools ComputeMatrix. A four color plot of conserved AP2-EXP DNA binding sites was made using the program cegr-tools four color plot downloaded from https://github.com/seqcode/cegr-tools/tree/master/src/org/seqcode/cegrtools. DNA motifs enriched in AP2-EXP ChIP-seq peaks were determined using the DREME Suite[91]. DNA motifs were sorted from highest to lowest level of conservation to the aggregate AP2-EXP DNA motif using FIMO[92].

For comparison of AP2-EXP peaks with existing post-translational histone marks, chromatin post translational modification datasets[13,17] were downloaded from https://github.com/Daread/plasmodiumExprPrediction[14]. Chromatin reader and nucleosome occupancy datasets[61,93,94] were downloaded from NCBI. ChIP-seq data for BDP1 and HP1 was processed as described above. Nucleosome occupancy data was downloaded from NCBI and processed as described above, with the exception that bam files were turned into bigwig using BamtoBigWig with the --mnase parameter selected. DeepTools ComputeMatrix was used to format the data. DeepTools PlotHeatmap was used to plot coverage of each dataset within 10kB of AP2-EXP peaks. For negative control coverage plots, Bedtools[89] ShuffleBed was used to create random genomic intervals on the same chromosome and the same length as AP2-EXP peaks of occupancy.

For comparison of AP2-EXP peaks identified by Shang *et al*[29] to this study, AP2-EXP peaks of occupancy in the trophozoite stage were downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE184658 . Peak summits from each replicate were expanded to 200 nucleotide windows using Bedtools SlopBed[89]. Conserved peaks in 2/2 replicates were identified using Bedtools intersect[89], then AP2-EXP peaks from this study and Shang *et al* were overlapped using Bedtools Intersect[89].

Western blot

For full parasite lysate Western blots, parasite protein was collected by lysing RBCs with .1% saponin on ice. After lysis, parasite pellets were washed by centrifugation with cold 1xPBS. Parasite pellets were resuspended in 50µL Loading Buffer (50mM Tris-Cl pH 8.0, 20% SDS, 1% Bromophenol Blue) and boiled at 95°C for 10 minutes. For fractionated parasite lysate, red blood cells were saponin lysed and parasite material was washed as described. Fractionated parasite protein was prepared as described in[95]. Briefly, parasites were resuspended in Lysis Buffer (20mM HEPES pH 7.8, 10mM KCl, 1mM EDTA, 1mM DTT, 1x Pierce Protease Inhibitor) and incubated for 15 minutes at 4°C. NP-40 was added to a final concentration 0.65% and nuclear material was pelleted by centrifugation at 2500xg at 4°C for 10 minutes. The cytosolic protein was collected and boiled in 1:1 volume of Loading Buffer at 95°C for 10 minutes. Pelleted nuclei were then resuspended in 1mL Extraction Buffer (20mM HEPES pH 7.8, 800mM KCl, 1mM EDTA, 1mM DTT, 1x protease inhibitor) per 1x10^9 cells for 1 hour at 4°C. Soluble nuclear protein was separated by centrifugation at 2500xg for 10 minutes at 4°C. Nuclear protein was then boiled in 1:1 volume of Loading Buffer at 95°C for 10 minutes.

Protein was run on a 6% SDS PAGE at 100V for 3 hours, then transferred overnight at 4°C and 10 Volts to a nitrocellulose membrane. Membranes were blocked using 5% milk in 1xTBS-.1% tween for 1 hour. Primary antibodies used were: 1/1000 rat ant-HA (Roche 3F10), 1/1000 mouse anti-GFP (Roche), 1/3000 rabbit anti-aldolase conjugated to HRP (Abcam ab38905), or 1/3000 mouse anti-H3 (Abcam ab10799). Membranes were washed three times with 1xTBS-.1% tween between primary and secondary

antibody incubation. Secondary antibody concentrations used were 1/3000 goat anti-rat HRP conjugate (Millipore), 1/3000 goat anti-mouse HRP conjugate, or 1/10,000 (Pierce), goat anti-rabbit HRP conjugate (Millipore). ECL reagent (Pierce) was used to detect HRP signal. Blots were exposed to autoradiography film (VWR) and imaged using an autoradiography developer.

Protein Pulldown

Parasite nuclear protein was isolated as described above for fractionated Western blots. Following the final centrifugation step, the supernatant was collected and diluted by 1:3 in Dilution Buffer (30% glycerol, 20mM HEPES pH 7.8). GFP tagged AP2-EXP was pulled down using Chromotek anti-GFP or mock immunoprecipitated using Chromotek negative control magnetic beads. Beads were washed twice in Wash Buffer prior to use (20mM HEPES pH 7.4, 250mM NaCl, 1mM EDTA, 1mM TCEP, .05% NP-40). Protein and beads were incubated together for 1 hour at 4°C. The beads were then washed twice with Wash Buffer and bound protein was collected in Sample Buffer by boiling at 95°C for 10 minutes.

Fluorescent Microscopy

Samples were prepared by incubating packed infected red blood cells with DRAQ5 dye (Thermo) for 15 minutes. Parasites were then washed in 1xPBS to remove excess dye and immediately placed on a glass slide for imaging. Fluorescent microscopy images were acquired using an Olympus Bx61 fluorescent microscope. All images were processed using SlideBook 5.0.

Mosquito Stage *P. berghei* AP2 Competitor Growth Assay

A female Swiss Webster mouse was inoculated with Plasmodium berghei ANKA 2.34 from frozen stock. Once the parasitemia reached 15%, the blood was harvested by heart puncture, washed twice with 1xPBS and resuspended to 10 mL in 1xPBS. 5 female Swiss Webster mice were infected with 500µL of the resuspended blood. 3-4 days after blood passage, exflagellation of the *P. berghei* gametocytes was assayed. Briefly, a drop of tail vein blood was incubated in RPMI 1640 (Invitrogen) containing 1µM xanthurenic acid (Sigma) for ~12-15 minutes. The mixture was added to a slide and observed under a light microscope at 40x. Ten or more fields were observed. Mice with 0.3-0.7 exflagellations per field were anesthetized and fed on 3-day post emergence *Anopheles stephensi* mosquitoes. Mosquitoes were maintained at 19°C and fed 10% sucrose. Bloodfed mosquitoes were separated 30 hours post-bloodfeeding. 10 days post-bloodfeeding, 5-10 mosquitoes were dissected and oocysts were counted by mercurochrome staining and light microscopy to ensure *P. berghei* oocyst development. The remaining mosquitoes were injected by standard mouth pipette technique with different small molecule inhibitors or a mixture of PBS-DMSO for control. The surviving injected mosquitoes were dissected 14 days post-bloodfeeding. Oocyst numbers were counted as before. Pooled groups of midguts were ground using a pestle, centrifuged at 7,000 rpm for 5 minutes and resuspended in 20µL 1xPBS to release developing mid-gut sporozoites. Sporozoites were counted on a haemocytometer. All in vivo studies were conducted in accordance with the GSK Policy on the Care, Welfare and Treatment of

Laboratory Animals and were reviewed by the ethical review process at the institution where the work was performed.

Data Availability

All sequencing data for AP2-EXP ChIP-seq and whole genome sequencing for transgenic parasite lines AP2-EXP::GFP and AP2-EXP::HA has been deposited in the SRA under **PRJNA818769**.

Raw microarray data for the Compound C and DMSO control RNA time course is available in **Table S7**.

Author Contributions

T.J.R., E.K.D.S., K.S.S., N.D., C.F.A.P., G.P., M.J.L., C.D.O., J.C.N. and M.L. designed experiments. T.J.R., E.K.D.S., V.C., K.S.S., N.D., G.J., C.F.A.P., and I.K. performed the experiments. T.J.R. and N.D. analyzed the data. T.J.R. generated figures. T.J.R. and M.L. wrote the manuscript in collaboration with the other authors

Acknowledgements

This work was funded through NIH/NIAID R01AI076276 (M.L.), R01AI125565 (M.L.), and with support from the Center for Quantitative Biology (P50 GM071508) (M.L.). T.J.R. was supported by NIH T32 Predoctoral Training Grant (5T32GM125592-01) awarded to the Center for Eukaryotic Gene Regulation (CEGR) at The Pennsylvania State University. G.A.J. is a recipient of the Sir Keith Murdoch Fellowship from the

American Australian Association and a Postdoctoral Research Grant from the American Heart Association (16POST26420067).

Conflict of Interest Statement

FJG is a GlaxoSmithKline employee and own shares of the company

REFERENCES

1. WHO. *Word Malaria Report 2021*. *Word Malaria report Geneva: World Health Organization. (2021). Licence: CC* (2021).
2. Ippolito, M. M., Moser, K. A., Kabuya, J.-B. B., Cunningham, C. & Juliano, J. J. Antimalarial Drug Resistance and Implications for the WHO Global Technical Strategy. *Curr. Epidemiol. Reports* **8**, 46–62 (2021).
3. Cowman, A. F., Healer, J., Marapana, D. & Marsh, K. Malaria: Biology and Disease. *Cell* **167**, 610–624 (2016).
4. Josling, G. a & Llinás, M. Sexual development in Plasmodium parasites: knowing when it's time to commit. *Nat. Rev. Microbiol.* **13**, 573–87 (2015).
5. Zanghì, G. *et al.* A Specific PfEMP1 Is Expressed in P. falciparum Sporozoites and Plays a Role in Hepatocyte Infection. *Cell Rep.* **22**, (2018).
6. Bozdech, Z. *et al.* The transcriptome of the intraerythrocytic developmental cycle of Plasmodium falciparum. *PLoS Biol.* **1**, 85–100 (2003).
7. Lu, X. M. *et al.* Nascent RNA sequencing reveals mechanisms of gene regulation in the human malaria parasite Plasmodium falciparum. *Nucleic Acids Res.* **45**, 7825–7840 (2017).
8. Painter, H. J., Carrasquilla, M. & Llinás, M. Capturing *in vivo* RNA transcriptional dynamics from the malaria parasite *Plasmodium falciparum*. *Genome Res.* gr.217356.116 (2017) doi:10.1101/gr.217356.116.
9. Le Roch, K. G. *et al.* Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science (80-.).* **301**, 1503–1508 (2003).
10. Foth, B. J. *et al.* Quantitative time-course profiling of parasite and host cell proteins in the human malaria parasite Plasmodium falciparum. *Mol. Cell. Proteomics* **10**, M110.006411 (2011).
11. Hoeijmakers, W. A. M. *et al.* Epigenetic reader complexes of the human malaria parasite, Plasmodium falciparum. *Nucleic Acids Res.* **47**, 11574–11588 (2019).
12. Toenhake, C. G. *et al.* Chromatin Accessibility-Based Characterization of the Gene Regulatory Network Underlying Plasmodium falciparum Blood-Stage Development. *Cell Host Microbe* **23**, 557-569.e9 (2018).
13. Jiang, L. *et al.* PfSETvs methylation of histone H3K36 represses virulence genes in Plasmodium falciparum. *Nature* **499**, 223–227 (2013).
14. Read, D. F., Cook, K., Lu, Y. Y., Le Roch, K. G. & Noble, W. S. Predicting gene expression in the human malaria parasite Plasmodium falciparum using histone modification, nucleosome positioning, and 3D localization features. *PLoS Comput. Biol.* **15**, 1–23 (2019).
15. Josling, G. A. *et al.* A Plasmodium Falciparum Bromodomain Protein Regulates Invasion Gene

1019 Expression. *Cell Host Microbe* **17**, 741–751 (2015).
1020 16. Pérez-Toledo, K. *et al.* Plasmodium falciparum heterochromatin protein 1 binds to tri-methylated
1021 histone 3 lysine 9 and is linked to mutually exclusive expression of var genes. *Nucleic Acids Res.*
1022 **37**, 2596–2606 (2009).
1023 17. Bártfai, R. *et al.* H2A.Z demarcates intergenic regions of the Plasmodium falciparum epigenome
1024 that are dynamically marked by H3K9ac and H3K4me3. *PLoS Pathog.* **6**, (2010).
1025 18. Ruvalcaba-Salazar, O. K. *et al.* Recombinant and native Plasmodium falciparum TATA-binding-
1026 protein binds to a specific TATA box element in promoter regions. *Mol. Biochem. Parasitol.* **140**,
1027 183–196 (2005).
1028 19. Coulson, R., Hall, N. & Ouzounis, C. Comparative genomics of transcriptional control in the
1029 human malaria parasite Plasmodium falciparum. *Genome Res.* 1–7 (2004)
1030 doi:10.1101/gr.2218604.
1031 20. Jeninga, M., Quinn, J. & Petter, M. ApiAP2 Transcription Factors in Apicomplexan Parasites.
1032 *Pathogens* **8**, 47 (2019).
1033 21. Painter, H. J., Campbell, T. L. & Llinás, M. The Apicomplexan AP2 family: Integral factors
1034 regulating Plasmodium development. *Mol. Biochem. Parasitol.* **176**, 1–7 (2011).
1035 22. Balaji, S., Madan Babu, M., Iyer, L. M. & Aravind, L. Discovery of the principal specific
1036 transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase
1037 DNA binding domains. *Nucleic Acids Res.* **33**, 3994–4006 (2005).
1038 23. Aravind, L., Iyer, L. M., Wellems, T. E. & Miller, L. H. Plasmodium Biology: Genomic Gleanings. *Cell*
1039 **115**, 771–785 (2003).
1040 24. Hollin, T. & Le Roch, K. G. From Genes to Transcripts, a Tightly Regulated Journey in Plasmodium.
1041 *Front. Cell. Infect. Microbiol.* **10**, 1–13 (2020).
1042 25. Campbell, T. L., de Silva, E. K., Olszewski, K. L., Elemento, O. & Llinás, M. Identification and
1043 Genome-Wide Prediction of DNA Binding Specificities for the ApiAP2 family of regulators from
1044 the malaria parasite. *PLoS Pathog.* **6**, (2010).
1045 26. Zhang, M. *et al.* Uncovering the essential genes of the human malaria parasite Plasmodium
1046 falciparum by saturation mutagenesis. *Science (80-.).* **360**, (2018).
1047 27. Shang, X. *et al.* A cascade of transcriptional repression determines sexual commitment and
1048 development in Plasmodium falciparum. *Nucleic Acids Res.* **49**, 9264–9279 (2021).
1049 28. Zhang, M. *et al.* Uncovering the essential genes of the human malaria parasite Plasmodium
1050 falciparum by saturation mutagenesis. *Science (80-.).* **360**, (2018).
1051 29. Shang, X. *et al.* Genome-wide landscape of ApiAP2 transcription factors reveals a
1052 heterochromatin-associated regulatory network during Plasmodium falciparum blood-stage
1053 development. 1–19 (2022).
1054 30. Modrzynska, K. *et al.* A Knockout Screen of ApiAP2 Genes Reveals Networks of Interacting
1055 Transcriptional Regulators Controlling the Plasmodium Life Cycle. *Cell Host Microbe* **21**, 11–22
1056 (2017).
1057 31. Zhang, C. *et al.* Systematic CRISPR-Cas9-Mediated ApiAP2 Genes Reveal Functional Insights into
1058 Parasite Development. *MBio* **8**, 1–17 (2017).
1059 32. Santos, J. M. *et al.* Red Blood Cell Invasion by the Malaria Parasite Is Coordinated by the PfAP2-I
1060 Transcription Factor. *Cell Host Microbe* **21**, 731-741.e10 (2017).
1061 33. Martins, R. M. *et al.* An ApiAP2 member regulates expression of clonally variant genes of the
1062 human malaria parasite Plasmodium falciparum. *Sci. Rep.* **7**, 1–10 (2017).
1063 34. Sierra-Miranda, M. *et al.* PfAP2Tel, harbouring a non-canonical DNA-binding AP2 domain, binds
1064 to Plasmodium falciparum telomeres. *Cell. Microbiol.* (2017) doi:10.1111/cmi.12742.
1065 35. Flueck, C. *et al.* A major role for the Plasmodium falciparum ApiAP2 protein PfSIP2 in
1066 chromosome end biology. *PLoS Pathog.* **6**, (2010).

1067 36. Tintó-font, E., Michel-todó, L., Russell, T. J. & Casas-, N. The PfAP2-HS transcription factor
1068 protects malaria parasites from febrile temperatures. (2021).
1069 37. Kafsack, B. F. C. *et al.* A transcriptional switch underlies commitment to sexual development in
1070 malaria parasites. *Nature* **507**, 248–52 (2014).
1071 38. Josling, G. A. *et al.* Dissecting the role of PfAP2-G in malaria gametocytogenesis. *Nat. Commun.*
1072 **11**, 1–13 (2020).
1073 39. Xu, Y. *et al.* PfAP2-G2 Is Associated to Production and Maturation of Gametocytes in Plasmodium
1074 falciparum via Regulating the Expression of PfMDV-1. *Front. Microbiol.* **11**, 1–11 (2021).
1075 40. Singh, S. *et al.* The PfAP2-G2 transcription factor is a critical regulator of gametocyte maturation.
1076 *Mol. Microbiol.* **115**, 1005–1024 (2021).
1077 41. Coetzee, N. *et al.* Epigenetic inhibitors target multiple stages of Plasmodium falciparum parasites.
1078 *Sci. Rep.* **10**, 1–11 (2020).
1079 42. Vanheer, L. N., Zhang, H., Lin, G. & Kafsack, B. F. C. Activity of Epigenetic Inhibitors against
1080 Plasmodium. 1–6 (2020).
1081 43. Lindner, S. E., De Silva, E. K., Keck, J. L. & Llinás, M. Structural Determinants of DNA Binding by a
1082 P. falciparum ApiAP2 Transcriptional Regulator. *J. Mol. Biol.* **395**, 558–567 (2010).
1083 44. Yuda, M., Iwanaga, S., Shigenobu, S., Kato, T. & Kaneko, I. Transcription factor AP2-Sp and its
1084 target genes in malarial sporozoites. *Mol. Microbiol.* **75**, 854–863 (2010).
1085 45. Morris, G. M. *et al.* Automated docking using a Lamarckian genetic algorithm and an empirical
1086 binding free energy function. *J. Comput. Chem.* **19**, 1639–1662 (1998).
1087 46. Gamo, F. J. *et al.* Thousands of chemical starting points for antimalarial lead identification.
1088 *Nature* **465**, 305–310 (2010).
1089 47. Wishart, D. S. *et al.* DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic*
1090 *Acids Res.* **46**, D1074–D1082 (2018).
1091 48. Allen, M. D., Yamasaki, K., Ohme-Takagi, M., Tateno, M. & Suzuki, M. A novel mode of DNA
1092 recognition by a β-sheet revealed by the solution structure of the GCC-box binding domain in
1093 complex with DNA. *EMBO J.* **17**, 5484–5496 (1998).
1094 49. Tintó-font, E. *et al.* A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1
1095 and Sox-2. *Plant Cell* **9**, 1–6 (2000).
1096 50. Wang, Y., Schellenberg, H., Walhorn, V., Toensing, K. & Anselmetti, D. Binding Mechanism of
1097 Fluorescent Dyes to DNA Characterized by Magnetic Tweezers. *Mater. Today Proc.* **4**, S218–S225
1098 (2017).
1099 51. Birnbaum, J. *et al.* A genetic system to study Plasmodium falciparum protein function. *Nat.*
1100 *Methods* **14**, 450–456 (2017).
1101 52. Kafsack, B. F. C., Painter, H. J. & Llinás, M. New Agilent platform DNA microarrays for
1102 transcriptome analysis of Plasmodium falciparum and Plasmodium berghei for the malaria
1103 research community. *Malar. J.* **11**, 1–9 (2012).
1104 53. Llinás, M., Bozdech, Z., Wong, E. D., Adai, A. T. & DeRisi, J. L. Comparative whole genome
1105 transcriptome analysis of three Plasmodium falciparum strains. *Nucleic Acids Res.* **34**, 1166–1173
1106 (2006).
1107 54. Smyth, G. K., Ritchie, M. & Thorne, N. Linear Models for Microarray Data User ' s Guide.
1108 *Bioinformatics* **20**, 3705–3706 (2011).
1109 55. Bahl, A. *et al.* PlasmoDB: The Plasmodium genome resource. A database integrating experimental
1110 and computational data. *Nucleic Acids Res.* **31**, 212–215 (2003).
1111 56. Sangurdekar, D. P. The Rnits package for normalization and inference of differential expression in
1112 time series microarray data . *Rvignette* 1–13 (2014).
1113 57. Gaspar, J. M. Improved peak-calling with MACS2. *bioRxiv* 1–16 (2018) doi:10.1101/496521.
1114 58. Yuda, M., Kaneko, I., Murata, Y., Iwanaga, S. & Nishi, T. Targetome analysis of malaria sporozoite

1115 transcription factor AP2-Sp reveals its role as a master regulator. (2022).
1116 59. Kensche, P. R. *et al.* The nucleosome landscape of Plasmodium falciparum reveals chromatin
1117 architecture and dynamics of regulatory sequences. *Nucleic Acids Res.* **44**, 2110–2124 (2015).
1118 60. Chappell, L. *et al.* Refining the transcriptome of the human malaria parasite Plasmodium
1119 falciparum using amplification-free RNA-seq. *BMC Genomics* **21**, 1–19 (2020).
1120 61. Fraschka, S. A. *et al.* Comparative Heterochromatin Profiling Reveals Conserved and Unique
1121 Epigenome Signatures Linked to Adaptation and Development of Malaria Parasites. *Cell Host*
1122 *Microbe* **23**, 407-420.e8 (2018).
1123 62. Ganesan, S. M., Falla, A., Goldfless, S. J., Nasamu, A. S. & Niles, J. C. Synthetic RNA-protein
1124 modules integrated with native translation mechanisms to control gene expression in malaria
1125 parasites. *Nat. Commun.* **7**, 1–10 (2016).
1126 63. Prommana, P. *et al.* Inducible Knockdown of Plasmodium Gene Expression Using the glmS
1127 Ribozyme. *PLoS One* **8**, 1–10 (2013).
1128 64. Štros, M. HMGB proteins: Interactions with DNA and chromatin. *Biochim. Biophys. Acta - Gene*
1129 *Regul. Mech.* **1799**, 101–113 (2010).
1130 65. De Silva, E. K. *et al.* Specific DNA-binding by apicomplexan AP2 transcription factors. *Proc. Natl.*
1131 *Acad. Sci.* **105**, 8393–8 (2008).
1132 66. Paton, D. G. *et al.* Exposing Anopheles mosquitoes to antimalarials blocks Plasmodium parasite
1133 transmission. *Nature* **567**, 239–243 (2019).
1134 67. Wu, Y., Kirkman, L. A. & Wellems, T. E. Transformation of Plasmodium falciparum malaria
1135 parasites by homologous integration of plasmids that confer resistance to pyrimethamine. *Proc.*
1136 *Natl. Acad. Sci. U. S. A.* **93**, 1130–4 (1996).
1137 68. Nasamu, A. S. *et al.* An integrated platform for genome engineering and gene expression
1138 perturbation in Plasmodium falciparum. *Sci. Rep.* **11**, 1–15 (2021).
1139 69. Murithi, J. M. *et al.* The Plasmodium falciparum ABC transporter ABCI3 confers parasite strain-
1140 dependent pleiotropic antimalarial drug resistance. *Cell Chem. Biol.* 1–16 (2021)
1141 doi:10.1016/j.chembiol.2021.06.006.
1142 70. Deitsch, K. W., Driskill, C. L. & Wellems, T. E. Transformation of malaria parasites by the
1143 spontaneous uptake and expression of DNA from human erythrocytes. *Nucleic Acids Res.* **29**,
1144 850–853 (2001).
1145 71. Simon, C. S. *et al.* An extended DNA-free intranuclear compartment organizes centrosome
1146 microtubules in malaria parasites. *Life Sci. Alliance* **4**, 1–15 (2021).
1147 72. Huey, R. & Morris, G. M. Using AutoDock 4 with AutoDockTools : A Tutorial. 1–56.
1148 73. Abrahams, L. & Burke, M. w o T w o. 20–36 (2010).
1149 74. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a
1150 general Amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
1151 75. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception
1152 in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260 (2006).
1153 76. Wang, J., W. Wang, P. A. K. and D. A. C. Antechamber: an accessory software package for
1154 molecular mechanical calculations. *J. Am. Chem. Soc* **222**, U403 (2001).
1155 77. Case, D., K. Belfon, I. Ben-Shalom, S. Brozell, D. Cerutti, T. Cheatham III, V. Cruzeiro, T. Darden, R.
1156 Duke, G. Giambasu, M. Gilson, H. Gohlke, A. Goetz, R. Harris, S. Izadi, S. Izmailov, K. Kasavajhala,
1157 A. Kovalenko, R. Krasny, T. Kurtzman, T. Lee, S. L, D. Y. and P. K. AMBER 2020. (2020).
1158 78. Pérez, A. *et al.* Refinement of the AMBER force field for nucleic acids: Improving the description
1159 of α/γ conformers. *Biophys. J.* **92**, 3817–3829 (2007).
1160 79. Götz, A. W. *et al.* Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1.
1161 generalized born. *J. Chem. Theory Comput.* **8**, 1542–1555 (2012).
1162 80. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, W. R. Routine Microsecond Molecular

Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh EwaldNo Title. *J Chem Theory Comput.* **10**, 3878–3888 (2013).
81. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
82. Roe, D. R. & Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).
83. Saldanha, A. J. Java Treeview - Extensible visualization of microarray data. *Bioinformatics* **20**, 3246–3248 (2004).
84. de Hoon, M. J. L., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20**, 1453–1454 (2004).
85. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
86. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc (2010).
87. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. **00**, 1–3 (2013).
88. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
89. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
90. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, 187–191 (2014).
91. Bailey, T. L. DREME: Motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**, 1653–1659 (2011).
92. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
93. Josling, G. A., Selvarajah, S. A., Petter, M. & Duffy, M. F. The role of bromodomain proteins in regulating gene expression. *Genes (Basel).* **3**, 320–343 (2012).
94. Bunnik, E. M. *et al.* DNA-encoded nucleosome occupancy is associated with transcription levels in the human malaria parasite Plasmodium falciparum. *BMC Genomics* **15**, 1–15 (2014).
95. Oehring, S. C. *et al.* Organellar proteomics reveals hundreds of novel nuclear proteins in the malaria parasite Plasmodium falciparum. *Genome Biol* **13**, R108 (2012).

Figure Legends

Figure 1 Nine putative competitors of DNA binding by AP2-EXP were identified by an *in-silico* screen and prioritized for further testing

A) The crystal structure of AP2-EXP (PDB:3IGM) was used as a template to computationally dock thousands of small molecules *in silico* using AutoDock. Results were filtered for

compounds that dock within 10 Angstroms of DNA binding residues with a free energy less than -5kJ/mol. Compounds matching these criteria were sourced and used for further testing.

B) Each compound that was prioritized based on the *in-silico* screen was assigned an identifier (A-I) and tested for anti-*Plasmodium* activity in a 48-hour growth inhibition assay against asexual blood stage *P. falciparum*.

Figure 2 Four of nine predicted AP2-EXP competitors compete AP2-EXP DNA binding in an Electrophoretic Mobility Shift Assay (EMSA)

A) Each putative AP2-EXP competitor was added to an EMSA containing the AP2-EXP AP2 domain with its cognate DNA oligo. DNA binding competition will result in a loss of the shifted probe. DMSO vehicle was used as a control for normal DNA binding by AP2-EXP. 150 fmoles of DNA probe, 125ng of AP2-EXP, and 300μM of each compound were used for each lane.

B) Chemical structures of Compounds A-I. Compounds highlighted in black (A, B, C and I) compete DNA binding by AP2-EXP. Three of the four compounds that compete DNA binding *in vitro* (Compounds A, B, C and D) have a benzoxazole core moiety, denoted by a red box.

Figure 3 Analogues of Compound B compete AP2-EXP DNA binding with differing efficacy

A) The four closest structural analogues to Compound B were sourced through the Tres Cantos Antimalarial Set. These compounds were given the identifiers B-1, B-2, B-3, and B-4. Each compound kills *P. falciparum* parasites at low micromolar (0.16-1.13μM) concentrations as measured by Gamo *et al*[46].

B) Compounds B, B-1, B-2, B-3, and B-4 were added to an EMSA with AP2-EXP. DNA binding competition will result in the loss of the shifted probe. 40 fmoles of DNA probe, 125ng of AP2-EXP, and 300μM of each compound were used for each lane.

C) Compounds B, B-1 and B-4 were titrated at 25, 50, and 150μM into an EMSA with AP2-EXP to determine whether there are differences in DNA binding competition. DMSO vehicle was used as a control for normal DNA binding. 40 fmoles of DNA probe and 125ng of AP2-EXP were used for each lane. The concentration increase for each compound is indicated by the triangle from left to right.

D) Chemical structures of Compound B and B-1, B-2, B-3, and B-4. Compounds that compete DNA binding by AP2-EXP are highlighted in black. Each compound has a different substitution pattern on the right-side benzene ring, denoted by a red box.

Figure 4 AP2 Domain Competing Compounds B and C affect *P. falciparum* at in the mid trophozoite stage, coinciding with the maximal expression of AP2-EXP

A) Schematic for ApiAP2 domain competitor compound phenotyping time course. Highly synchronous asexual *P. falciparum* parasites were spiked with 40μM of AP2 domain competing Compounds B and C, non-ApiAP2 domain competitor Compound H, or DMSO vehicle control at 24 hpi and parasite progression was monitored for 48 hours to determine the parasite killing phenotype.

B) Representative Giemsa-stained microscopy images of *P. falciparum* parasites spiked with Compounds B, C, or H as described in panel A. DMSO vehicle was used as a control for normal progression through the asexual blood stage.

C) Highly synchronous asexual blood stage *P. falciparum* parasites were spiked with 40µM Compound C at eight-hour intervals in the IDC. Each assay was done in triplicate. Error bars represent standard deviation.

D) Highly synchronous asexual stage *P. falciparum* parasites were cultured with 40µM Compound C, which was washed out of the media at eight-hour intervals. Each assay was done in triplicate. Error bars represent standard deviation.

E) AP2-EXP protein expression in the asexual blood stages ring, trophozoite, and schizont, as well as Stage III gametocytes, was probed by Western blot against AP2-EXP endogenously tagged with GFP (AP2-EXP::GFP). The expected molecular weight of AP2-EXP::GFP (147kDa) is indicated by an arrow.

F) Live fluorescent microscopy of AP2-EXP::GFP was used to further probe AP2-EXP protein expression and localization. DRAQ5 was used as a nuclear stain for parasites.

Figure 5 Compound C disrupts the *P. falciparum* transcriptome specifically at 30hpi, with bias towards AP2-EXP targets predicted by ChIP-seq

A) *P. falciparum* parasites were spiked with 12µM (0.66xIC_{50}) Compound C or DMSO vehicle control at 10 hpi. Total RNA was harvested at 6, 12, 18, 24, 30, 42, and 48 hpi for quantification

by DNA microarray. The full transcriptome Spearman Correlation between Compound C and DMSO control spiked parasites was plotted as a correlogram. A * indicates *p-value* < 0.05.

B) Three replicates of AP2-EXP Chip-seq were collected at 30 hpi using two genetically tagged parasite lines (AP2-EXP::GFP and AP2-EXP::HA). A no epitope control was collected using wild type 3D7 parasites and anti-GFP antibodies. Log_2 enrichment of the immunoprecipitate (ChIP) over Input DNA for one replicate of AP2-EXP::GFP, AP2-EXP::HA, and No Epitope Control is plotted relative to the coding sequence start site of AP2-EXP target genes identified in at least 2/3 replicates by the MACS2 peak calling software.

C) CATGA is the most overrepresented DNA motif within the 240 peaks of AP2-EXP occupancy conserved in at least 2/3 ChIP-seq experiments. No significant DNA motif was recovered from the No Epitope Control.

D) Compound C induced changes in transcript abundance at 24-30hpi compared to transcripts bound by AP2-EXP. Of the 463 genes decreased in abundance by log_2 fold change > 2, 46 are AP2-EXP targets.

E) Volcano plot displaying the changes in abundance for AP2-EXP target genes at 24-30 hpi. 46/93 detected AP2-EXP target transcripts decrease in abundance by log_2 fold change > 2 in the presence of Compound C.

F) GO-term analysis of transcripts that are both AP2-EXP targets and decreased in abundance with respect to Compound C (Bonferroni *p-value* cutoff < 0.05).

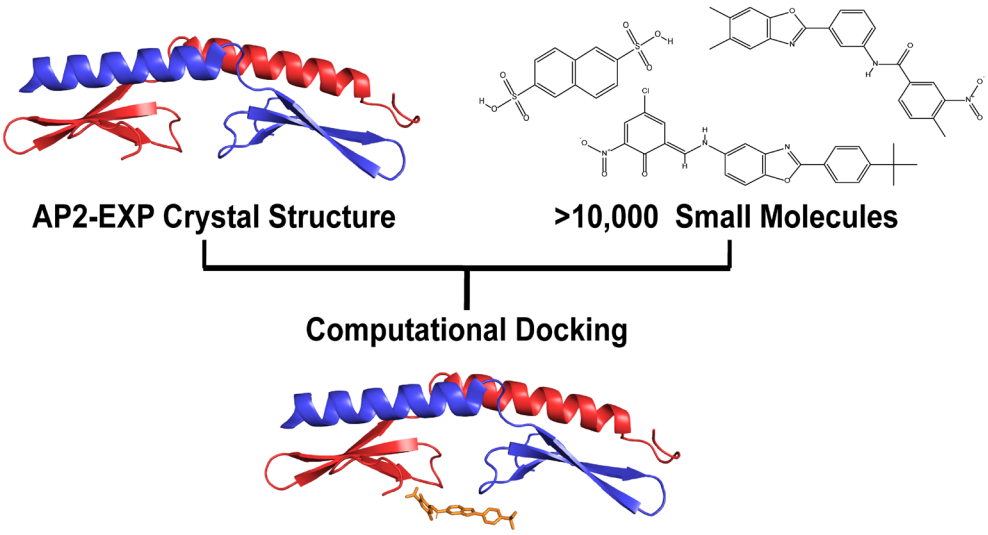
Figure 6 Compounds B and C are active against mosquito stage *Plasmodium berghei* parasites

A) *A. stephensi* mosquitoes were infected with *Plasmodium berghei* parasites. On day 10 post infection, mosquito midguts were injected with the measured IC_{50} of Compounds B, C, F, or DMSO vehicle control. On day 14 post infection mosquitoes were dissected to count oocysts and midgut sporozoites. For Compound C and DMSO vehicle control, each experiment was performed in duplicate. Compounds B and F phenotyping were performed as a single experiment.

B) Midgut oocyst counts per mosquito for *P. berghei* infected mosquitoes on day 14 post infection. Compound identity is indicated in the legend. R1 and R2 correspond to biological replicates one and two, respectively.

C) Midgut sporozoite counts per five infected mosquitoes on day 14 post infection. Compound identity is indicated in the legend. R1 and R2 correspond to biological replicates one and two, respectively.

A



B

Letter Code	PubChem ID	IC_{50} (μM)
A	CID1365835	11
B	CID1365471	80
C	CID5750730	18
D	CID4541005	22
E	CID2343	170
F	CID4437	40
G	CID11390	34
H	CID6914273	40
I	CID25863	125

Figure 1 A-B

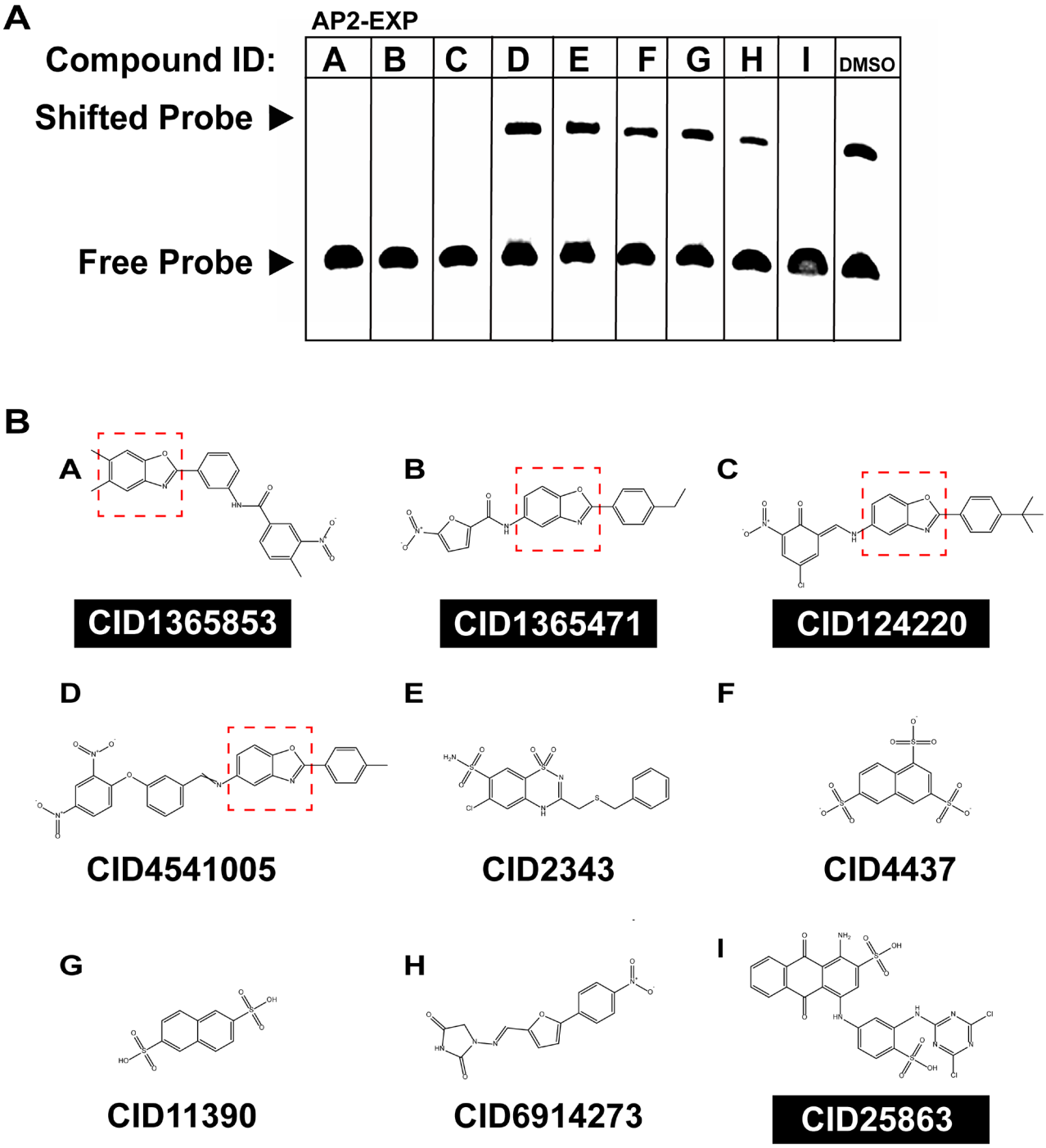

A
AP2-EXP
Compound ID:
A
B
C
D
E
F
G
H
I
DMSO
Shifted Probe
Free Probe
B
A
CID1365853
B
CID1365471
C
CID124220
D
CID4541005
E
CID2343
F
CID4437
G
CID11390
H
CID6914273
I
CID25863

Figure 2 A-B

49

50

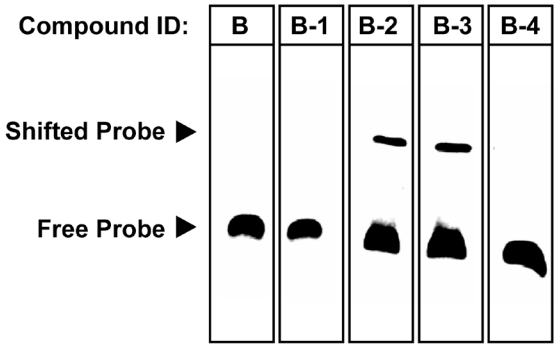
51 **A**

Letter Code	PubChem ID	IC_{50} (μM) (Gamo et al)
B-1	CID1931935	0.16
B-2	CID1104215	1.13
B-3	CID1356045	0.96
B-4	CID15991863	0.87

52

53

54 **B**



55

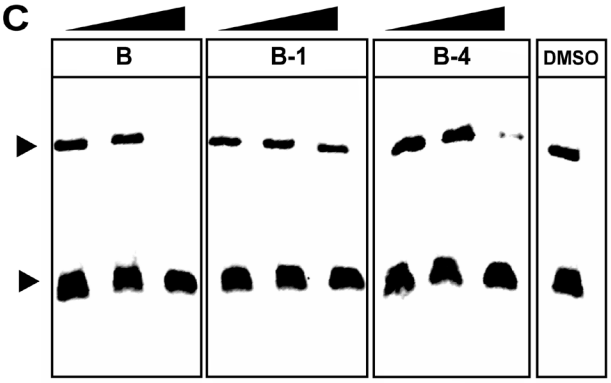
56

57

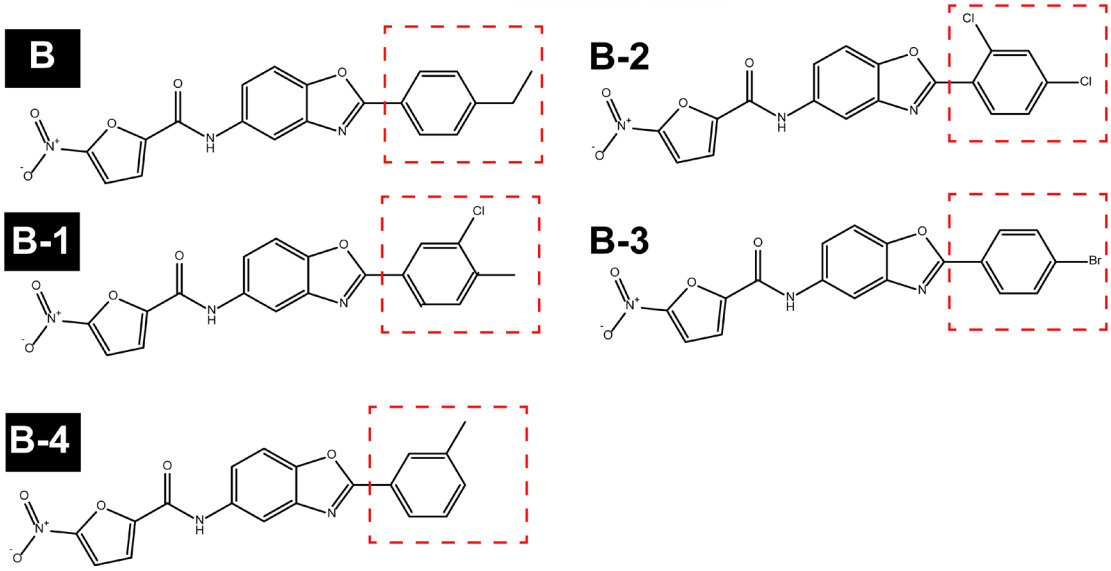
58

59

C



60 **D**



61

62

63

64

65

66

67

68 Figure 3 A-D

69

70

71

72

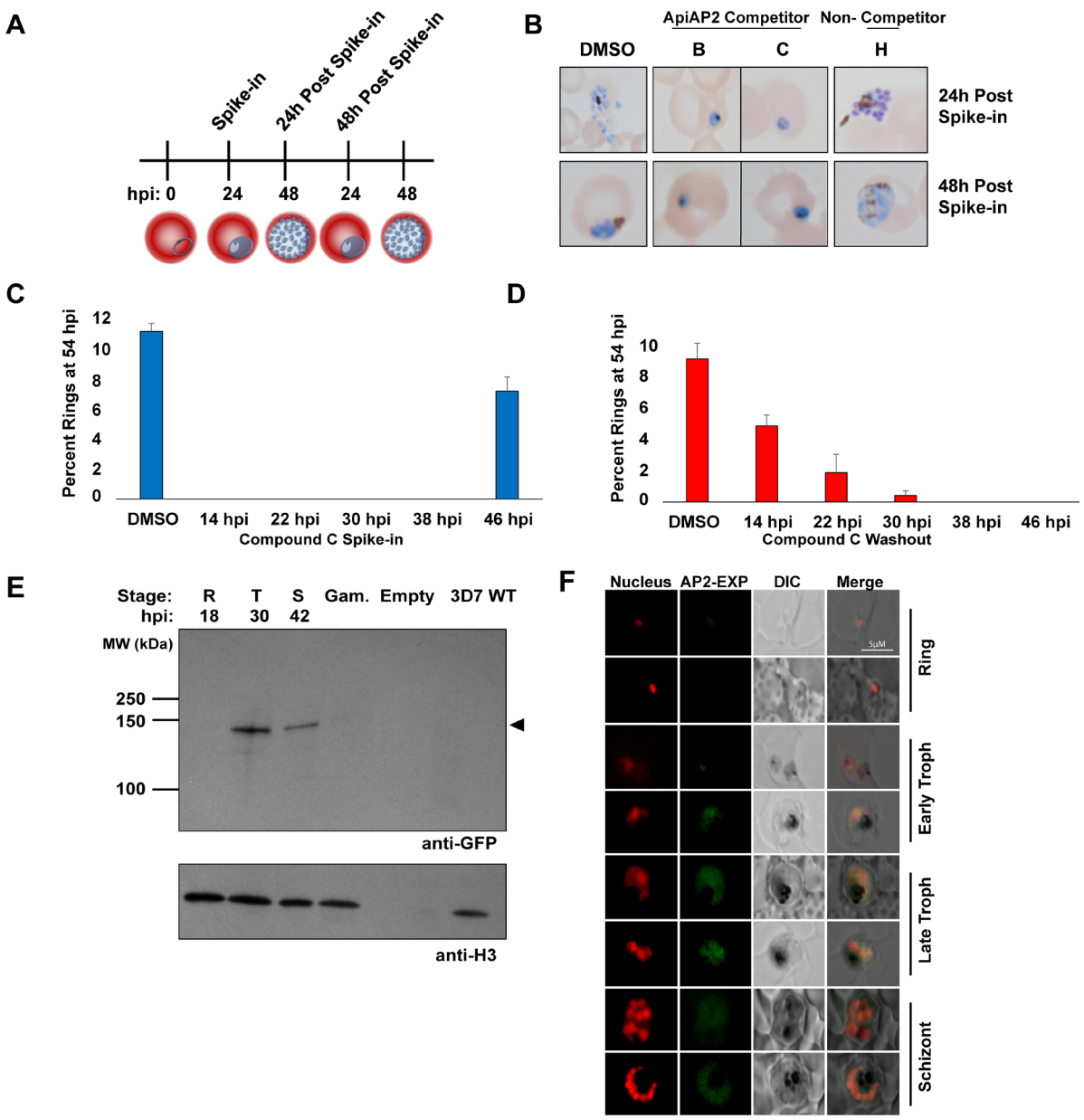
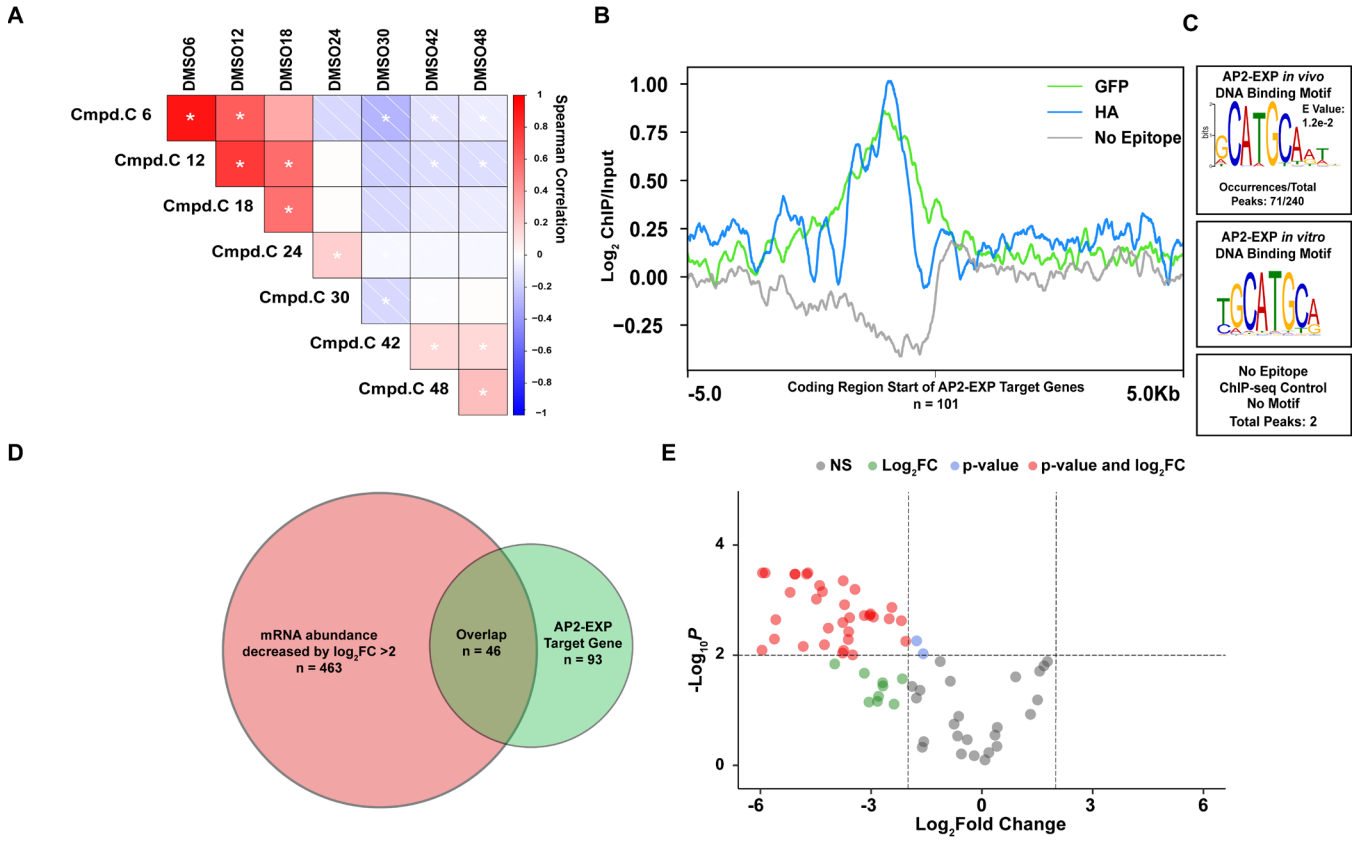

A
Spike-in
24h Post Spike-in
48h Post Spike-in
hpi: 0
24
48
24
48
B
ApiAP2 Competitor
Non- Competitor
DMSO
B
C
H
24h Post Spike-in
48h Post Spike-in
C
Percent Rings at 54 hpi
12
10
8
6
4
2
0
DMSO
14 hpi
22 hpi
30 hpi
38 hpi
46 hpi
Compound C Spike-in
D
Percent Rings at 54 hpi
10
8
6
4
2
0
DMSO
14 hpi
22 hpi
30 hpi
38 hpi
46 hpi
Compound C Washout
E
Stage: R T S Gam. Empty 3D7 WT
hpi: 18 30 42
MW (kDa)
250
150
100
anti-GFP
anti-H3
F
Nucleus
AP2-EXP
DIC
Merge
5μM
Ring
Early Troph
Late Troph
Schizont

Figure 4 A-F



F

AP2-EXP Bound and Decreased by log_2FC >2	Bgd count	Result count	Bonferroni
entry into host	96	14	2.32E-12
movement in host environment	113	14	2.43E-11
biological process involved in interspecies interaction between organisms	409	18	2.03E-07
biological process involved in symbiotic interaction	388	17	7.73E-07
biological process involved in interaction with host	374	14	0.000201045
cell adhesion	70	6	0.004098861
biological adhesion	194	9	0.004708229
pathogenesis	75	6	0.006101223
cell-cell adhesion	62	5	0.028388607

Figure 5A-F

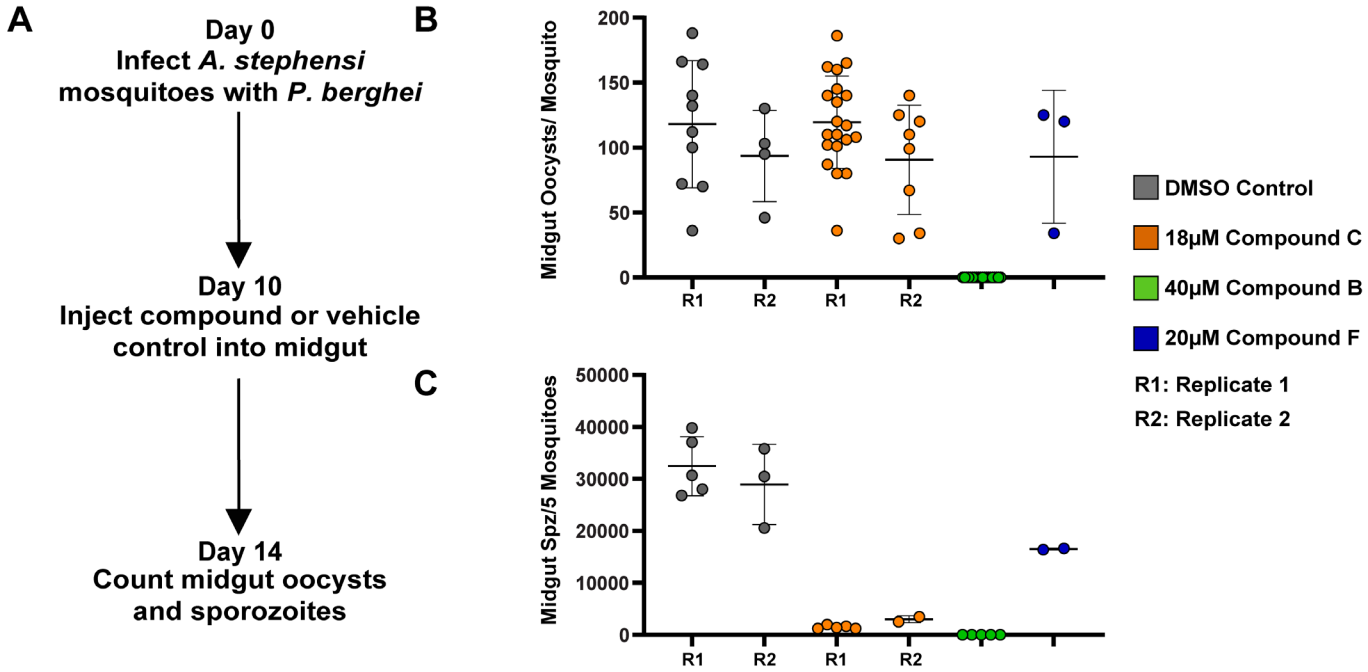

A
Day 0
Infect A. stephensi
mosquitoes with P. berghei
Day 10
Inject compound or vehicle
control into midgut
Day 14
Count midgut oocysts
and sporozoites
B
Midgut Oocysts/ Mosquito
200
150
100
50
0
R1
R2
R1
R2
C
Midgut Spz/5 Mosquitoes
50000
40000
30000
20000
10000
0
R1
R2
R1
R2
DMSO Control
18μM Compound C
40μM Compound B
20μM Compound F
R1: Replicate 1
R2: Replicate 2

Figure 6 A-C