Cell

Resource

Supergenomic Network Compression and the Discovery of EXP1 as a Glutathione Transferase Inhibited by Artesunate

Andreas Martin Lisewski,[1,2,13,*] Joel P. Quiros,[3,13] Caroline L. Ng,[8] Anbu Karani Adikesavan,[1,4] Kazutoyo Miura,[10] Nagireddy Putluri,[4,5] Richard T. Eastman,[8,10] Daniel Scanfeld,[8] Sam J. Regenbogen,[7] Lindsey Altenhofen,[11,12] Manuel Llinás,[11,12] Arun Sreekumar,[4,5,6] Carole Long,[10] David A. Fidock,[8,9] and Olivier Lichtarge[1,2,5,7,*]

[1]Department of Molecular and Human Genetics
[2]Computational and Integrative Biomedical Research Center
[3]Integrative Molecular and Biomedical Sciences Graduate Program
[4]Department of Molecular and Cell Biology
[5]Verna and Marrs McLean Department of Biochemistry and Alkek Center for Molecular Discovery
[6]Department of Biochemistry and Molecular Biology
[7]Department of Pharmacology
Baylor College of Medicine, Houston, TX 77030, USA
[8]Department of Microbiology and Immunology
[9]Division of Infectious Diseases, Department of Medicine
Columbia University Medical Center, New York, NY 10032, USA
[10]Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD 20852, USA
[11]Department of Molecular Biology and Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA
[12]Department of Biochemistry and Molecular Biology and Center for Infectious Disease Dynamics, The Pennsylvania State University, State College, PA 16802, USA
[13]Co-first author
*Correspondence: lisewski@bcm.edu (A.M.L.), lichtarge@bcm.edu (O.L.)
http://dx.doi.org/10.1016/j.cell.2014.07.011

SUMMARY

A central problem in biology is to identify gene function. One approach is to infer function in large supergenomic networks of interactions and ancestral relationships among genes; however, their analysis can be computationally prohibitive. We show here that these biological networks are compressible. They can be shrunk dramatically by eliminating redundant evolutionary relationships, and this process is efficient because in these networks the number of compressible elements rises linearly rather than exponentially as in other complex networks. Compression enables global network analysis to computationally harness hundreds of interconnected genomes and to produce functional predictions. As a demonstration, we show that the essential, but functionally uncharacterized *Plasmodium falciparum* antigen EXP1 is a membrane glutathione S-transferase. EXP1 efficiently degrades cytotoxic hematin, is potently inhibited by artesunate, and is associated with artesunate metabolism and susceptibility in drug-pressured malaria parasites. These data implicate EXP1 in the mode of action of a frontline antimalarial drug.

INTRODUCTION

The biological functions of most genes are unknown (Erdin et al., 2011) and therefore require novel methods of identification (Radivojac et al., 2013). Increasingly, these methods rely on computational network analysis (Sharan et al., 2007). Such networks are composed of protein or gene nodes connected by intrinsic links, which indicate common evolutionary origins across species, and contextual links, which indicate interactions or biological correlations among genes and proteins within a genome. The function of a protein node can then be inferred either through local network analysis that transfers annotations from the nodes it directly connects to or through global analysis that optimizes some relatedness measure over the entire network (Sharan et al., 2007; Vazquez et al., 2003). Although local network analyses are computationally relatively inexpensive, they are also of limited value in sparsely annotated network regions (Erdin et al., 2011) because they cannot reach for information beyond an immediate neighborhood (Chua et al., 2006). Unfortunately, such areas of very sparsely annotated genome regions include genomes of disease-causing agents (Ideker and Sharan, 2008). For example, in the human malarial parasite *Plasmodium falciparum*, the low sensitivity of current annotation methods leaves most genes without any known biological function (Aurrecoechea et al., 2009; Gardner et al., 2002). By contrast, global network approaches can be more sensitive, but their computational

demands typically restrict them to smaller networks that can encompass single proteomes with only several thousand nodes (Erdin et al., 2011; Vazquez et al., 2003).

In this study, we apply a known global network-based function prediction method, termed graph-based information diffusion (GID; Venner et al., 2010), over so-called supergenomic networks that comprise all the genes from hundreds of genomes. To achieve this, we propose a general network compression scheme that dramatically reduces the number of network links by eliminating redundancies within and between network cliques. In the context of supergenomic networks, these cliques consist of clusters of orthologous groups of proteins (COGs) (Tatusov et al., 1997). Critically, network compression does not significantly perturb global network analysis: GID is much more efficient on a compressed network but still accurately reproduces GID outputs on an uncompressed network. Thus network compression opens, in principle, the known gene and protein space to global network-based function prediction.

As a validation of this approach, we tested a GID-based prediction in *P. falciparum* 3D7 parasites of the biological function of exported protein 1 (EXP1), also referred to as exported antigen 5.1 (Ag5.1) or circumsporozoite-related antigen/protein (CRA) (Hope et al., 1984; Simmons et al., 1987). Even though the biological role of EXP1 has not been characterized, several lines of evidence suggest that this small 17 kDa polypeptide is important to malaria pathogenesis. Failure to disrupt this *exp1* gene, which is well conserved among *Plasmodium* species (Simmons et al., 1987), suggests its essentiality for the parasite (Maier et al., 2008). This gene is also one of the most abundantly transcribed loci (PF11_0224/PF3D7_1121600) during the ring and early trophozoite stages (Bozdech et al., 2003; Le Roch et al., 2004), which is the asexual parasite's initial growth phase in erythrocytes. The protein product is mainly exported to the parasitophorous vacuolar membrane (PVM) and to cytosolic compartments in infected red blood cells (RBCs) (Simmons et al., 1987), where it forms homomers in the membrane (Spielmann et al., 2006). EXP1 triggers an antigenic immune response in humans (Simossis et al., 2005) and has been explored as a malaria vaccine candidate (Caspers et al., 1991; Meraldi et al., 2004). We demonstrate, as predicted by GID, that EXP1 is a glutathione S-transferase (GST) that conjugates glutathione onto hematin—the main cytotoxin released during malarial blood stage infection. This activity is unique among known membrane GSTs but is nonetheless consistent with their ability to protect cells against xenobiotic and oxidative stress (Morgenstern et al., 2011). We further show that EXP1 is potently inhibited by the current antimalarial drug of choice, artesunate (ART). This soluble artemisinin derivative is currently recommended by the World Health Organization as the frontline treatment of severe falciparum malaria (World Health Organization (WHO), 2011) even though its future efficacy is uncertain due to emerging parasite resistance to artemisinins (Ariey et al., 2014). Our identification of EXP1 as a *P. falciparum* membrane-bound GST suggests previously unresolved modes of hematin metabolism and ART-mediated stress response in parasitized RBCs.

RESULTS

Supergenomic Networks of Evolutionary Relationships Are Compressible

Network compression exploits the COG cliques present in supergenomic networks (see Experimental Procedures) in two steps. First, *intra-clique compression* replaces each COG clique with a star graph (Figure 1A). All the edges among the members of a clique are removed, and instead every node becomes connected by a single edge to a new core node for the clique, weighted by the size of the clique, n_{cog} (Figures 1A and 1B). In a realistic supergenomic network, however, still more compression is required. For example, over 373 proteomes from different species, the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database (von Mering et al., 2007), yielded a supergenomic network with $n \approx 1.51 \times 10^6$ protein nodes connected by $n_c \approx 3.86 \times 10^7$ contextual links and up to $n_i \approx 1.93 \times 10^{11}$ intrinsic links (Figure 1B and Experimental Procedures). The latter included 5.41×10^8 intra-COG links from 33,929 COGs. Intra-clique compression reduces these nearly 400-fold, to 1.38×10^6, at the cost of adding ~2% (33,929) new core nodes to build the star graphs. The network structure, however, is still dominated by $\sim 10^{11}$ inter-COG links and remains beyond practical computational capabilities.

To further reduce computational cost, the second essential step is *inter-clique compression*, which additionally compresses evolutionary relationships *between* COGs. All the links between a pair of COGs are replaced by a single edge connecting their core nodes, weighted by average sequence similarity between COGs (Figure 1C). The number of inter-COG links now falls almost 10^5-fold from 1.93×10^{11} to 6.81×10^6. After full compression, the total number of remaining intrinsic links is 8.19×10^6, which is of the same order of magnitude as that of contextual links.

Critically, this compression perturbs GID only mildly: in extensive simulations on random networks with nodes assigned random initial conditions, the relative error between GID on the full network and on the compressed network remained below ~10% when the original uncompressed graph had an edge density below ~20% (Figures S1A and S1B; see specific example in Figures 1D and 1E), i.e., when the graph contained less than ~20% of all possible edges (see Extended Experimental Procedures). Thus network compression perturbs GID minimally when the underlying network is sparse. These and additional numerical data (Figures S1C–S1E and Extended Experimental Procedures) suggest that network compression was essential to practically enable GID on large supergenomic networks with a loss of accuracy that is tolerable and adjustable.

Theoretical Reason for Compressibility

Network compression was computationally efficient because in supergenomic networks, the number of maximal cliques has a linear relation to the number of network nodes (Figure S1F). Given that the computational time to find all maximal cliques in a network is bounded linearly by the number of nodes per single maximal clique (Eppstein et al., 2010), the observation that the total maximal clique number rises linearly and not exponentially with network size ensures that a complete list of compressible cliques can be computed efficiently, i.e., in nonexponential,

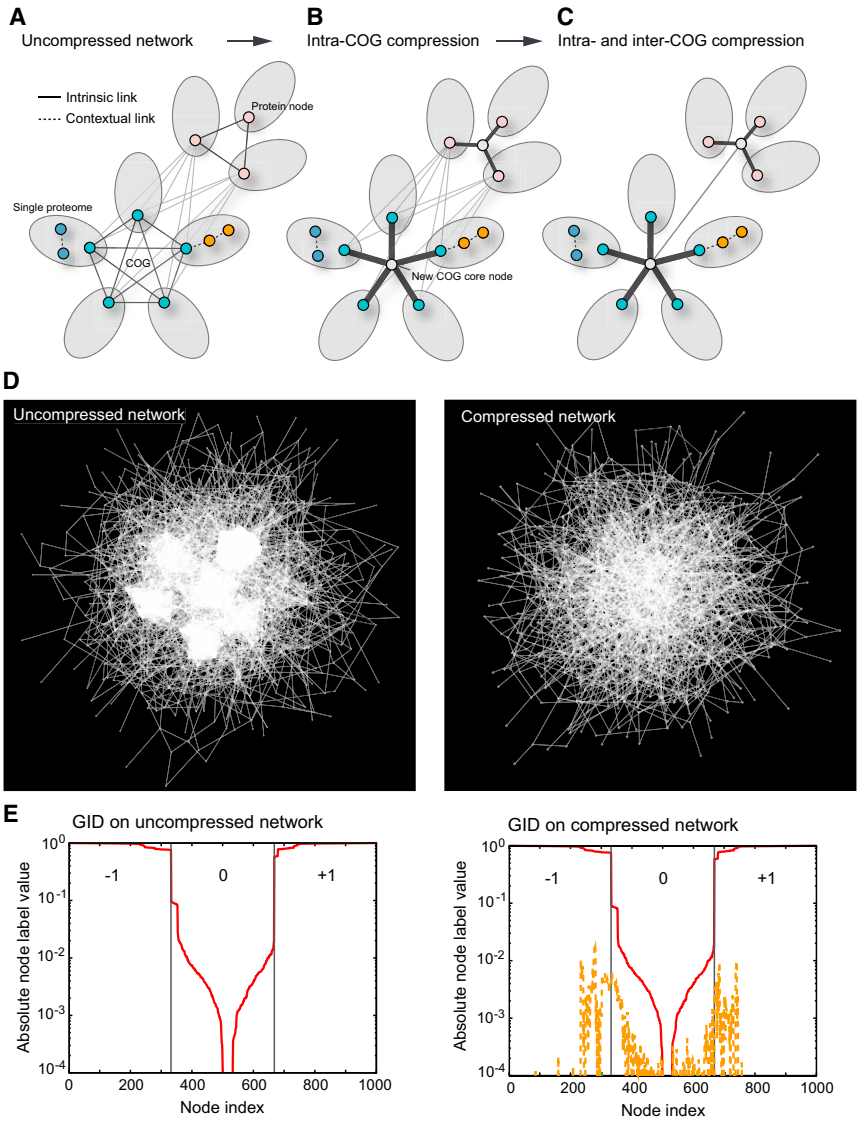

Figure 1. Compression of Supergenomic Networks

(A) Supergenomic network with multiple genomes/proteomes (gray ovals) and links depicting intrinsic and contextual gene associations.

(B) COG cliques of size n_{cog} are transformed in intra-COG compression into weighted star graphs. For both cliques and star graphs, every node is connected to the new node with a link of weight w, and the resulting link weight at each clique or star graph member is $w\ n_{cog}$.

(C) Inter-COG compression additionally replaces all sequence similarity links between members in two different COGs with a single link. Contextual links are not compressed.

(D) Intra-clique network compression example of an n = 1,000 node network with six cliques. The edge density, i.e., the ratio of the actual number of edges over the maximum edge number $n(n-1)/2$, of the uncompressed network is 0.013, and the compressed network has only one-third the number of links from the original network.

(E) GID output comparison in logarithmic scale between the full, uncompressed, and compressed networks as shown in (D). GID input labeling in the first third of indexed nodes was "−1," second was "0," and third was "+1" (gray curve). The GID output is shown in red, and the absolute error (vector norm difference between uncompressed and compressed GID output) in yellow. Despite the loss of more than 60% of its links, the GID output on the compressed networks had only 0.4% average relative error, defined as the absolute error divided by the vector norm of the GID output of the uncompressed network. Due to the logarithmic scale, negative class label values are represented in absolute values.

linear time (Figure S1G, black graph). In contrast, other common complex networks such as Erdős-Rényi random graphs have a number of maximal cliques that are exponential in the number of nodes (Figure S1G, red graph). Consequently, in larger networks, their detection quickly becomes computationally prohibitive. Supergenomic networks are therefore biological examples of networks where clique detection is computationally efficient.

To explain the efficiency of compression in supergenomic networks, we modeled evolutionary relationships between genes by *preferential attachment* among cliques (Figure S1H): beginning with a small number of a few given cliques, a new clique was added to the network by linking it to an existing clique with a probability defined by the connectivity of that clique. This scale-free network generation model (Wang et al., 2009) is a natural generalization of both Yule's model of biological speciation (Yule, 1925) and the Barabasi-Albert model of preferential attachment for scale-free networks (Barabasi and Albert, 1999; Newman, 2005). It predicts that for large clique sizes $k >> 1$, their distribution follows a power law $\sim k^{-\gamma}$ with $\gamma = 3$. In agreement with this prediction, we measured $\gamma = 2.94 \pm 0.03$ for the clique (COG) sizes in the supergenomic network (Figure S1I), which implies that its structure was consistent with preferential attachment. Because preferential attachment necessarily produces scale-free networks, in which all maximal cliques can be localized efficiently in computer time that scales linearly with network size (Eppstein et al., 2010), the model offers a rationale for the compressibility of supergenomic networks.

Functional Matching on Compressed Supergenomic Networks

In order to benchmark the performance of functional predictions, we next tested the ability to recover the activities of test proteins after hiding their Gene Ontology (GO) Molecular Function information from the network (*leave-one-out* experiments). The GID output was ranked by statistical Z scores. These scores were insensitive to compression error (Figure 2A) and were based on a normal approximation for the global distribution of reciprocal GID output log-ratios (Figure 2B, example in Figure 2C, and Extended Experimental Procedures). For comparison, we used

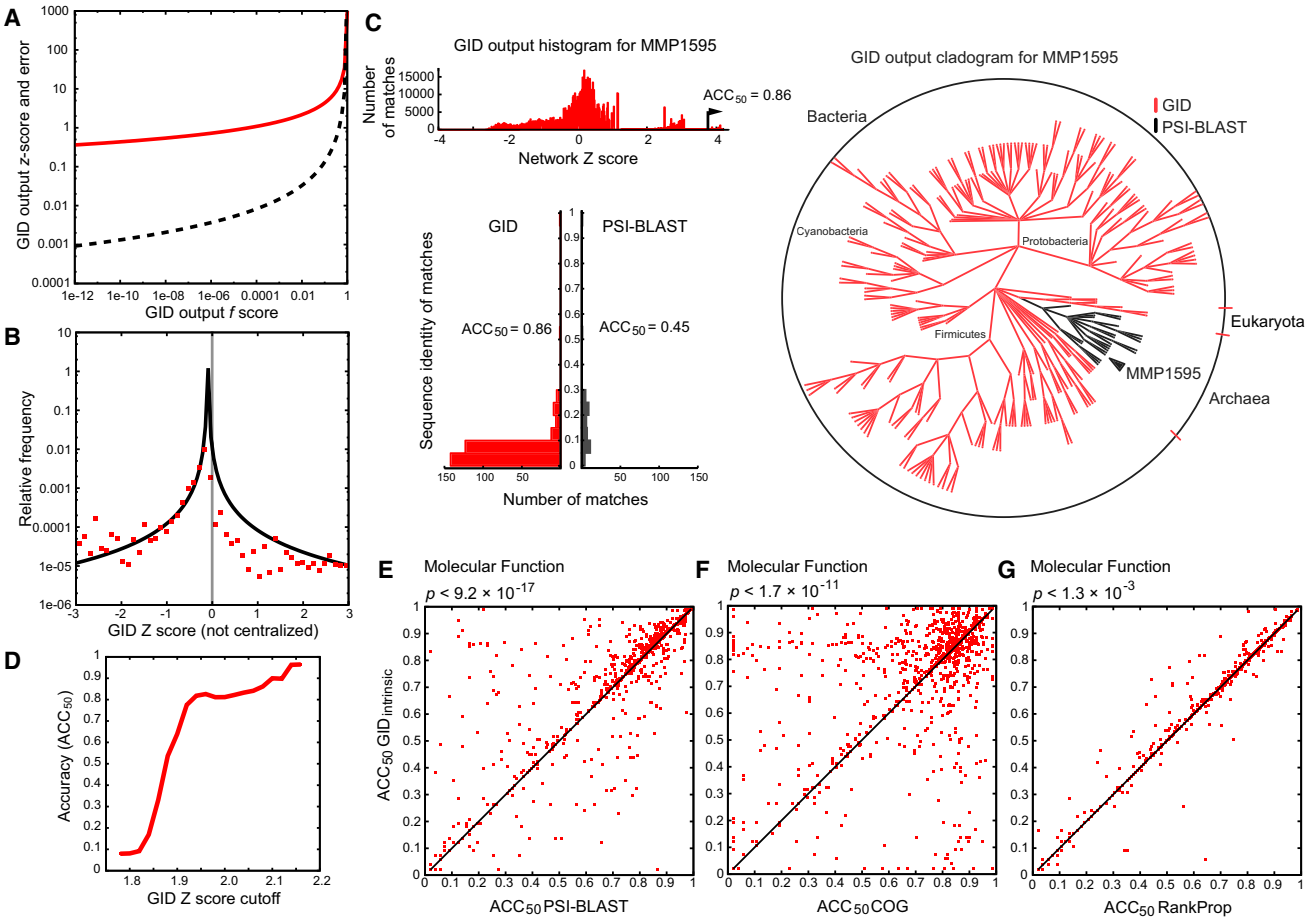

Figure 2. Functional Matching on a Supergenomic Network

(A) Transformation of GID output *f* scores (absolute values) into output *Z* scores (red solid line); error in output *Z* scores due to a given GID compression error in *f* scores (dashed black line).

(B) GID output *Z* scores follow approximately a shifted Gaussian distribution (represented here on a reciprocal log scale). The shift follows from the bias in the functional input labels to negative values.

(C) Example output for one leave-one-out test (MMP1595, a *Methanococcus maripaludis* queuine tRNA-ribosyltransferase; EC 2.4.2.29). The histogram shows the distribution of output *Z* scores over the entire network, where statistically significant matches are above $Z = 2$. For a fixed number of 50 false matches, GID is more sensitive than PSI-BLAST: it detects 307 true positives (*TP*) against 42 with PSI-BLAST, or $ACC_{50} = 0.86$ against $ACC_{50} = 0.45$, where ACC_{50} is the accuracy defined as $TP/(TP + 50)$. Shown are corresponding histogram of sequence identities to the query sequence and a dendrogram of true matches across species for PSI-BLAST (black) and GID (red).

(D) Accuracy as a function of minimum *Z* scores cut-off.

(E) GID versus PSI-BLAST scatter plot of accuracy values for the test set of 1,000 fully annotated (EC numbers) enzyme sequences.

(F and G) GID versus COG (F) and GID versus RankProp (G) scatter plots for the same test set.

two local and one global algorithm for sequence similarity networks. The local methods were PSI-BLAST sequence search (Altschul et al., 1997), where all links from the full uncompressed network were allowed for matching, and the ranking of network COGs was by their average sequence similarity to the query protein (COG method). Because neither method takes into account contextual information, GID was limited to intrinsic evolutionary links for fair comparison. A set of 1,000 enzymes from the curated SwissProt collection of the UniProt database (UniProt Consortium, 2010) defined the test set of the leave-one-out experiments. Over these experiments, we measured success by the average number of true positive predictions made prior to reaching 50 false predictions, or ACC_{50} accuracy values (Gribskov and Robinson, 1996). We found that GID *Z* scores above $Z \sim 2$ were reliable indicators of functional accuracy (Figure 2D) giving accuracies that were 9% greater for GID than for PSI-BLAST ($p < 9.2 \times 10^{-17}$, Wilcoxon signed rank test; Figure 2E) and 21% more than with COGs ($p < 1.7 \times 10^{-11}$; Figure 2F). Our results showed that local algorithms on the full, uncompressed network (such as PSI-BLAST) yielded lower prediction accuracies than GID on the compressed network.

We also benchmarked GID against an established global algorithm, RankProp (Melvin et al., 2009). The iterative flow algorithm underlying RankProp was computationally intractable on the

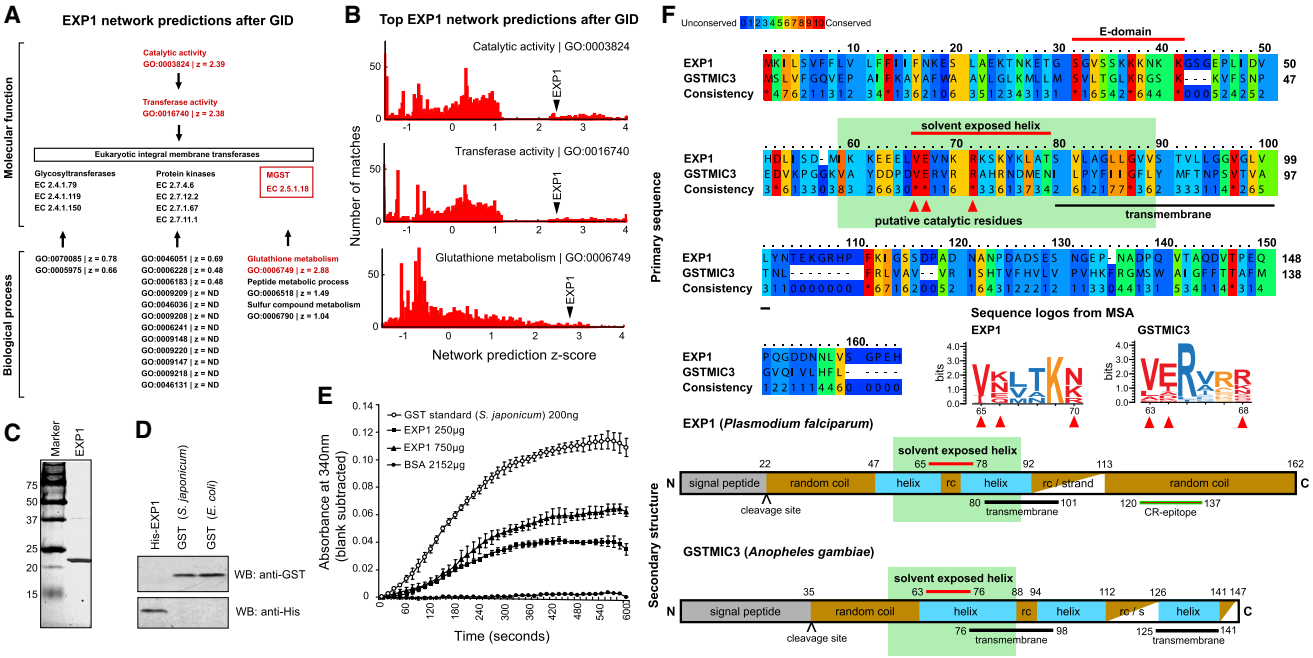

Figure 3. Prediction and Experimental Confirmation of EXP1 Function

(A) GID analysis of EXP1 performed for all GO terms that were enriched significantly in the three possible clusters.

(B) GID predictions of molecular function and biological process in *Z* score histograms for all *P. falciparum* genes extracted from the network.

(C) Purified EXP1 at apparent molecular weight near 23 kDa, which is higher than the predicted 17 kDa and a known anomaly of this antigen (Hope et al., 1985).

(D) No EXP1 antibody interference with both expression host GST from *Escherichia coli* and positive control GST from *Schistosoma japonicum* (*Sj*GST).

(E) GST activity of EXP1 toward CDNB; negative control with bovine serum albumin (BSA). Mean values and standard error bars are from three experimental results.

(F) Extended homology sequence alignment of *P. falciparum* EXP1 and *Anopheles gambiae* microsomal glutathione S-transferase GSTMIC3 indicates a putative shared site (positions 59–89 with 46% similarity, green box); this region includes a charged helical stretch (red bar). Sequence logos from multiple sequence alignments indicate similar amino acid composition at the putative catalytic residue positions. Secondary structure comparison between EXP1 and GSTMIC3 reveals further commonalities. *E-domain* indicates putative MGST oligomerization domain (Holm et al., 2006).

uncompressed supergenomic network, so RankProp was also applied to the intrinsic compressed network. In that test (Figure 2G), GID detected 5% more true matches than RankProp ($p < 1.3 \times 10^{-3}$). This further confirmed that compression is an essential step toward a global analysis of large networks, and that GID can improve function detection in comparison with other global methods.

Application to the *P. falciparum* Genome: The Molecular Function of EXP1

EXP1 Enzymatic Function Prediction and Validation

To predict novel molecular functions over many genomes, we used *competitive GID* (Venner et al., 2010) on the compressed supergenomic network: at any tested level of the GO hierarchy, the most likely molecular function was determined by the GO term with the highest significant *Z* score. Given the prior distribution of input functional labels (Figure S3A), we produced 92,892 significant predictions (with $Z > 2$) for 1,889 Molecular Function enzyme commission (EC) numbers in archaea (4% of predictions), bacteria (48%), and eukaryota (48%; see Table S1). *Bacillus anthracis* and *P. falciparum*, causative agents of anthrax and malaria, had the smallest numbers of genes annotated with GO Molecular Function terms compared to their total numbers of genes in the prokaryotic and eukaryotic categories, respectively (Figure S3B). This last case confirmed earlier observations that the majority of the *P. falciparum* 3D7 genome is difficult to annotate with biological function (Ochoa et al., 2011). To rank the importance of gene targets among the 305 *P. falciparum* functional predictions made with GID, we reviewed their number of references in the scientific literature (Table S2). This analysis highlighted EXP1 as the most cited *P. falciparum* gene with no known molecular function and for which we had a functional prediction.

When applied to EXP1, competitive GID predicted "catalytic activity" ($Z = 2.39$) as the only significant term at the highest GO level. Out of six main enzymatic classes, "transferase activity" ($Z = 2.38$) most strongly associated with EXP1 (Figures 3A and 3B and Table S2). No prediction from the transferase subclasses achieved a significant *Z* score, but three main functional clusters in eukaryotic integral membrane transferases emerged: glycosyltransferases (EC 2.4.1), protein kinases (EC 2.7), and microsomal glutathione S-transferases (GST, EC 2.5.1.18; Figure 3A). A search in the GO Biological Process category produced "glutathione metabolism" ($Z = 2.88$; Figures 3A and 3B). To identify the network sources of these predictions, we ordered all other protein nodes by their *Z* score distance to EXP1 (Table S3): the

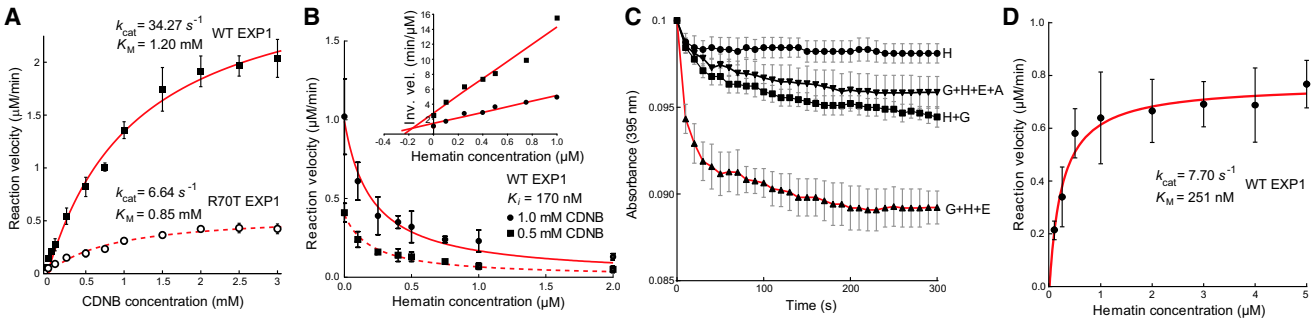

Figure 4. CDNB and Hematin Are Substrates of EXP1

(A) Reaction kinetics and Michaelis-Menten analysis for WT EXP1 and R70T EXP1.

(B) Hematin inhibits the GST activity of EXP1 competitively (inset Dixon plot) with a K_i of 170 nM.

(C) Spectrometry-monitored kinetics of the spontaneous degradation of hematin (H) in GSH solution (H+G), EXP1-mediated enzymatic degradation of hematin (G+H+E), and its inhibition after adding ART (10 nM, G+H+E+A).

(D) Michaelis-Menten analysis for EXP1-mediated degradation of hematin; curve (solid red line), least-square fits to Michaelis-Menten equation.

Data points represent mean values with standard error bars from three measurements.

closest was a type 3 membrane GST from *Saccharomyces cerevisiae* (with 10% sequence identity to EXP1), leading to the prediction that EXP1 has a function akin to that of microsomal glutathione S-transferases (MGSTs).

Guided by this prediction, we tested EXP1 in vitro for GST activity, after purifying the full-length protein expressed in *Escherichia coli* (Figures 3C and 3D). EXP1 was active in a detergent-solubilized form, and standard GST assay spectrophotometric measurements of EXP1 conjugating the thiol group of glutathione toward the benzene ring of the cytotoxic benzene derivative 1-chloro-2,4-dinitrobenzene (CDNB) yielded a specific activity of 8.72 ± 3.73 nmol min^{-1} mg^{-1} (Figures 3E and S3C). This activity of EXP1 was comparable to that of the human microsomal GST (MGST1, Figures S3D–S3F) and the unrelated cytosolic 27 kDa GST from *P. falciparum* (*Pf*GST, PF11_0187). Heterologous bacterial expression showed that no other *P. falciparum* proteins were required for the observed activity of EXP1. This GST activity of EXP1 establishes a new apicomplexan member of the widespread superfamily of membrane-associated proteins in eicosanoid and glutathione metabolism (MAPEG; Jakobsson et al., 1999). No MAPEG members have previously been described in the *Plasmodium* genus, although there is likeness in quaternary structure: both EXP1 and eukaryotic MGST form membrane homo-oligomers (Holm et al., 2006; Spielmann et al., 2006).

EXP1 Putative Catalytic Site

Even though EXP1 lacks substantial homology to other MGSTs, there is a suggestive sequence identity, detected at 9% using extended homology alignments (Simossis et al., 2005), and global secondary structure similarity to a type 3 MGST from the malaria mosquito vector *Anopheles gambiae* (GSTMIC3). An Evolutionary Trace analysis (Mihalek et al., 2004) identified Arg70 as well as the nearby Glu66 (Figure 3F) as being among the most important residues in the MAPEG superfamily (top 8% of evolutionarily traced residue positions). Indeed, a substitution of this arginine to a threonine is sufficient for a loss of GST activity in leukotriene C4 synthase (Lam et al., 1997). We therefore predicted that Arg70 in *P. falciparum* EXP1 might be catalytically important for GST activity.

To test this hypothesis, we expressed and purified an R70T mutant of EXP1 (Figures S3E and S3F). Enzyme kinetics confirmed that the reaction was approximately five times faster for wild-type (WT) compared to the R70T mutant (Figure 4A) and gave rate constants k_{cat} = 34.27 s^{-1} for WT EXP1 and k_{cat} = 6.64 s^{-1} for R70T EXP1; the Michaelis-Menten constants were K_M = 1.20 mM and K_M = 0.85 mM, respectively. Our data suggest that R70 is important to the transferase mechanism although not critically involved in substrate binding. WT EXP1 reaction kinetics were in good agreement with other eukaryotic MGSTs (Andersson et al., 1995).

Hematin Is an EXP1 Substrate

During malarial infection, the principal cytotoxic compound that is released during parasite-mediated hemoglobin catabolism is iron-bound heme, or its hydroxide, hematin. Given that hematin weakly and uncompetitively inhibits *Pf*GST through binding to a preformed *Pf*GST-GSH complex with a K_i value of 3 μM (Hiller et al., 2006), we asked whether hematin might inhibit EXP1. We found that inhibition was almost 20 times stronger, with K_i = 170 nM, and was competitive (Figure 4B, inset)—i.e., hematin competed with the substrate CDNB for binding to the EXP1 active site. As a control, spectrophotometry around 400 nm showed that recombinant EXP1 purified from *Escherichia coli* carried no detectable residual heme or hematin, and that only the addition of exogenous hematin produced a prominent Soret peak at 395 nm that is characteristic of hematin (Figure S4A). This result opened the possibility that hematin itself is a substrate of EXP1, which to our knowledge would be a unique property of this malarial GST.

Reduced GSH spontaneously degrades heme by forming a GSH-heme complex involving a nucleophillic attack in which heme's porphyrin ring is disrupted leading to the release of Fe^{3+} (Atamna and Ginsburg, 1995). This process, which can be spectrophotometrically monitored near 395 nm (Atamna and Ginsburg, 1995; Ginsburg et al., 1998), has been proposed as the underlying mechanism of GSH-mediated heme/hematin degradation in *P. falciparum*-infected RBCs (Famin et al., 1999; Ginsburg et al., 1998). We confirmed this reaction in vitro where,

in the presence of reduced GSH, loss of free hematin was monitored by a decrease of the Soret peak at 395 nm (Figures S4B and S4C) and by mass spectrometry-based measurements of the production of GSH-hematin complexes (Figures S4D–S4F). Motivated by these observations, we asked whether the addition of EXP1 enzymatically accelerates hematin degradation. Two experiments demonstrate that EXP1 actively facilitates the conjugation of reduced GSH to hematin. First, spectrophotometry showed a 4-fold higher rate of decrease in 395 nm absorbance over the first ~30 s of reaction time in the presence of EXP1 (G+H+E), compared to the spontaneous degradation of hematin (H+G, Figure 4C). Second, mass spectrometry identified the GSH-hematin product in which addition of EXP1 increased product yield over the spontaneous reaction >14-fold (Figures S4G and S4H). A Michaelis-Menten analysis (Figure 4D) measured k_{cat} = 7.7 s^{-1} and a K_M = 251 nM. Remarkably, the high specificity constant k_{cat}/K_M ~3 × 10^7 $s^{-1}M^{-1}$ placed EXP1 near other exceptionally efficient enzymes such as fumarase or catalase and thus not far from the universal diffusion limit of ~10^8 $s^{-1}M^{-1}$. This activity toward hematin was unique among all tested WT (*Sj*GST, human MGST1) and mutant (R70T EXP1) GSTs (Figure S4I).

In line with the functional formation of MGST homotrimers (Holm et al., 2006), an inhibition analysis with R70T EXP1 (Figure S4J) indicated that several WT EXP1 units constitute one functional catalytic unit. In addition, we bacterially expressed, purified, and tested the function of the uncharacterized *Plasmodium yoelii* EXP1 ortholog HEP17 (17 kDa hepatocyte erythrocyte protein). HEP17 shares only 44% sequence identity with EXP1 such that, and consistent with the sequence logo in Figure 3F (inset), the proposed EXP1 catalytic residue R70 changes to N in HEP17, and E66 changes to K. Despite this variation, HEP17 matched the GST activity of EXP1 both toward CDNB and toward hematin (Figures S4K and S4L). These data provide evidence that hematin is a substrate of EXP1, which itself belongs to this *Plasmodium* family of membrane GSTs that share hematin as a substrate.

ART Is a Potent Inhibitor of EXP1

Studies on artemisinin mode of action mostly agree that hemoglobin uptake, digestion, and heme/hematin release lead to cleavage of artemisinin's endoperoxide bridge. The resulting activated drug then causes oxidative damage to parasite membranes (Klonis et al., 2011). EXP1 primarily localizes to the PVM, and a construct of GFP fused to the EXP1 signal sequence has been observed in spherical/cytostomal PVM invaginations that appear to contain RBC cytosol and hemoglobin (Grüring et al., 2011). Endocytic vesicles have been reported to pinch off from these invaginations to form acidified compartments during ring stages, i.e., before the formation of a main food vacuole (FV) (Abu Bakar et al., 2010). Using differential interference contrast and immunofluorescence microscopy, our observations suggested the presence of similar peripheral PVM invaginations and vesicles, which in part colocalized with foci of EXP1 expression during trophozoite (Figure 5A) and late ring stages (Figure 5B). Such peripheral EXP1 foci were also observed in three distinct parasite lines (ART-sensitive 3D7, ART-sensitive Dd2, and ART-tolerant 3b1—see below and Experimental Procedures), in two different drug-exposure conditions (with and without ART), and at three different time points after RBC invasion (12, 18, and 30 hr; Figures 5C, 5D, and S5A–S5D). These invaginations and vesicles are potential sites of early hemoglobin degradation and hematin release (Abu Bakar et al., 2010), which could supply the peroxide bridge activator during ring stages (Klonis et al., 2013b). Because of this temporal and spatial profile of EXP1 expression, and the high affinity of EXP1 to its substrate hematin (Figures 4B, 4D, and S4G), we hypothesized that ART might inhibit hematin degradation catalyzed by EXP1. To test this hypothesis in vitro, we added ART to our EXP1-hematin reactions. These assays revealed competitive inhibition (Figure S5F) with a half-maximal inhibitory concentration (IC_{50}) of 2.05 nM (Figures 4C and 5E). Nonspecific inhibition was excluded through stoichiometry controls (Habig et al., 2009): after reducing enzyme concentration 10-fold, from 100 nM to 10 nM, potency changed by only ~25% and remained within error margins (Figure 5F). A control through liquid chromatography-mass spectrometry (LC-MS) confirmed an ART-dependent suppression of EXP1-mediated GSH-hematin adduct formation down to spontaneous levels (Figure 5F, inset and S5E). Consistent with the peroxide bridge activation mechanism, the reaction displayed strong hematin dependence (Figure 5G): inhibition of EXP1 activity toward CDNB, in the absence of iron-bound hematin, was uncompetitive (Figure S5G) and 100-fold less potent with an average IC_{50} of 184 nM. In a negative control, we tested the standard cytosolic *Sj*GST from *Schistosoma japonicum,* which showed 6-fold lower reaction velocities that were close to the background noise level (Figures 5E and S4I). Because EXP1 units form membrane homo-oligomers, we further tested for cooperativity. A Hill equation analysis resulted in a calculated IC_{50} value of 1.21 nM and in a Hill coefficient n_H = 0.56, which indicated lack of cooperativity: one EXP1 unit appeared to bind one ART molecule (Figure 5G, inset). Our results provide evidence of a potent, hematin-dependent inhibition of EXP1 activity by ART.

As a control for drug specificity, we tested the antimalarial naphthoquinone atovaquone that targets the cytochrome bc_1 complex in the mitochondrial electron transport chain (ETC) (Fry and Pudney, 1992). Even though eukaryotic MGSTs have a structural similarity with the ETC enzyme cytochrome *c* oxidase (Holm et al., 2006), atovaquone did not inhibit EXP1-mediated hematin degradation (Figure 5H). This suggests that EXP1 may be targeted by endoperoxides but not by naphthoquinones.

EXP1 in Parasite Drug Response

We then asked whether EXP1 might be involved in drug action of ART, whose cellular effectors may extend beyond the genes and genomic regions that recent studies associated with emerging ART resistance (Ariey et al., 2014; Cheeseman et al., 2012; Park et al., 2012; Takala-Harrison et al., 2013). Specifically, we tested a model (Mukanganyama et al., 2001) of reductive metabolism of ART, mediated by a hypothetical GST that could catalyze the formation of GSH-ART complexes via GSH binding to the drug's lactone ring to form a hydroperoxide moiety. We addressed this model through EXP1 in the dihydroartemisinin-pressured ART-tolerant parasite line 3b1 (see Experimental Procedures), compared to the parental line Dd2 and to the reference drug-sensitive line 3D7.

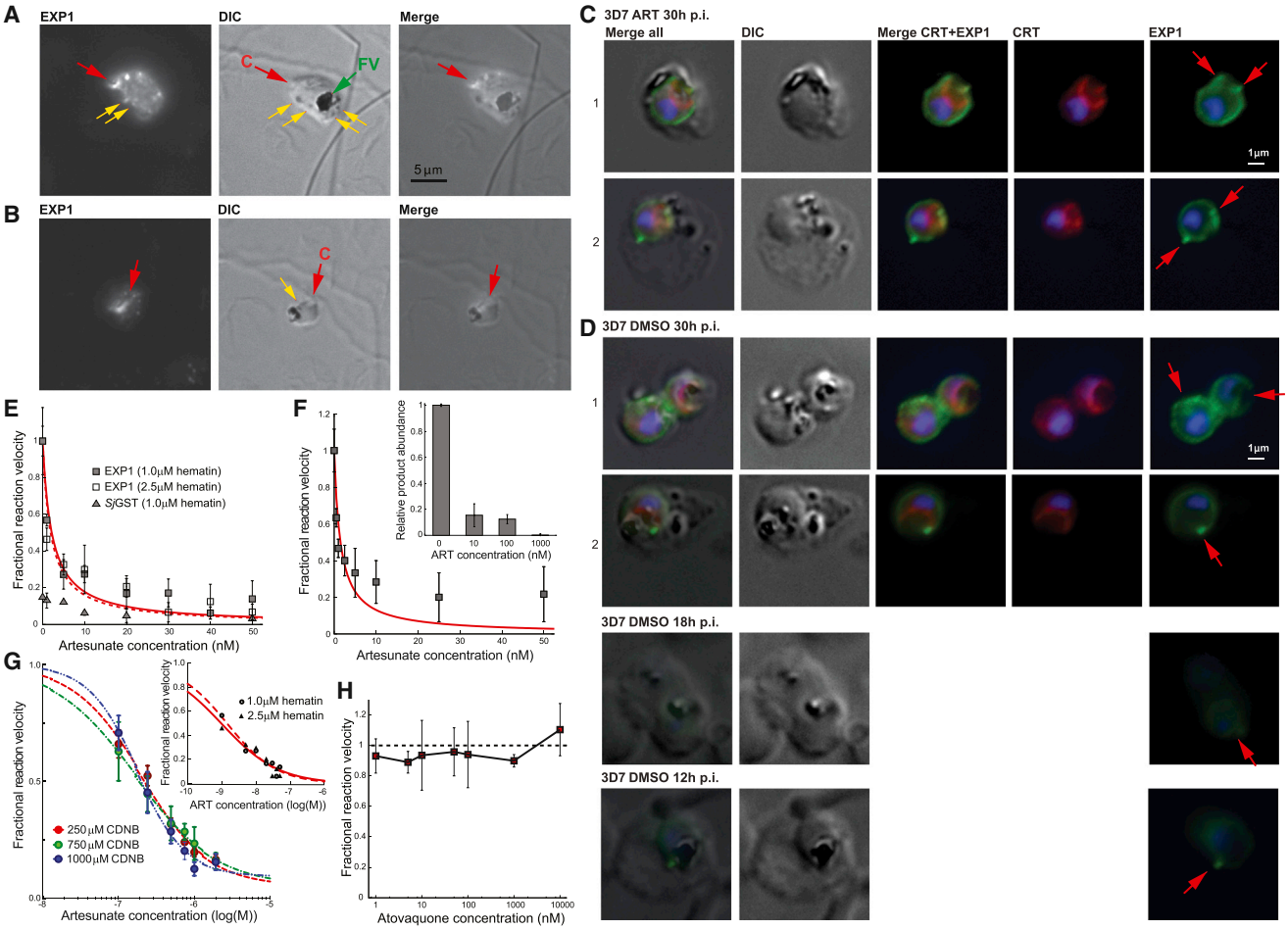

Figure 5. ART Is a Potent Inhibitor of EXP1

(A) Immunofluorescence microscopy of 3D7 trophozoite parasites labeled with anti-EXP1 antibodies (EXP1) shows evidence of peripheral expression of EXP1 that colocalizes with an apparent cytostomic invagination (*C*, red arrow; DIC panel). A fully formed trophozoite stage food vacuole (FV, green arrow) presents a large accumulation of hemozoin, whereas small darker spots (yellow arrows) might be vesicles that have pinched off from the invagination.

(B) A late ring stage parasite displays a prominent spherical invagination that also partly corresponds to a convexly shaped EXP1 expression signal; small vesicles (yellow arrow) may become components of an internal pre-FV compartment (dark irregular structure).

(C) Trophozoite stage 3D7 parasites exposed to ART indicate peripheral EXP1 expression (row C1) or enhanced expression foci at the PVM (row C2).

(D) Control trophozoites (DMSO) displayed similar morphology in EXP1 expression patterns to the drug-exposed parasites. Antibodies to the FV marker CRT (*P. falciparum* chloroquine resistance marker) indicate that EXP1 expression was largely independent of FV location.

(E) ART inhibits EXP1 GST activity toward hematin.

(F) Stoichiometry control of ART-mediated inhibition of EXP1 activity toward hematin at a reduced enzyme concentration of 10 nM. Inset: ART-dependent inhibition of GSH-hematin product formation.

(G) Inhibition of EXP1 activity toward CDNB through ART at three increasing CDNB concentrations. Inset: Hill equation analysis for inhibition of EXP1-mediated hematin degradation through ART.

(H) Atovaquone does not inhibit EXP1 GST activity toward hematin. Dashed line represents fractional reaction velocity at zero concentration of atovaquone (in absolute physical units 0.74 ± 0.06 µM min^{-1}).

Mean values and standard error bars from three experimental results.

In vitro data from LC-MS showed that EXP1 can catalyze the production of GSH-ART adducts in an ART concentration-dependent manner (Figure 6A). A control with *Sj*GST, which belongs to the same GST family as *Pf*GST (PF14_0187), yielded GSH-ART levels indistinguishable from their spontaneous formation. These results recall earlier findings that *Pf*GST does not actively conjugate GSH to ART and is unlikely to modulate parasite susceptibility to artemisinins (Deponte and Becker, 2005). LC-MS analysis with extracts of cultured parasites yielded GSH-ART levels that, in ART-tolerant 3b1 parasites, were estimated to be 3.76 ± 0.31 times higher after ART exposure than the background signal without drug challenge (Figures 6B and S6), whereas in 3D7 parasites, those levels were low and almost indistinguishable from post-drug exposure levels (0.93 ± 0.09 ratio). Similarly, in Dd2 parasites, GSH-ART measurements after

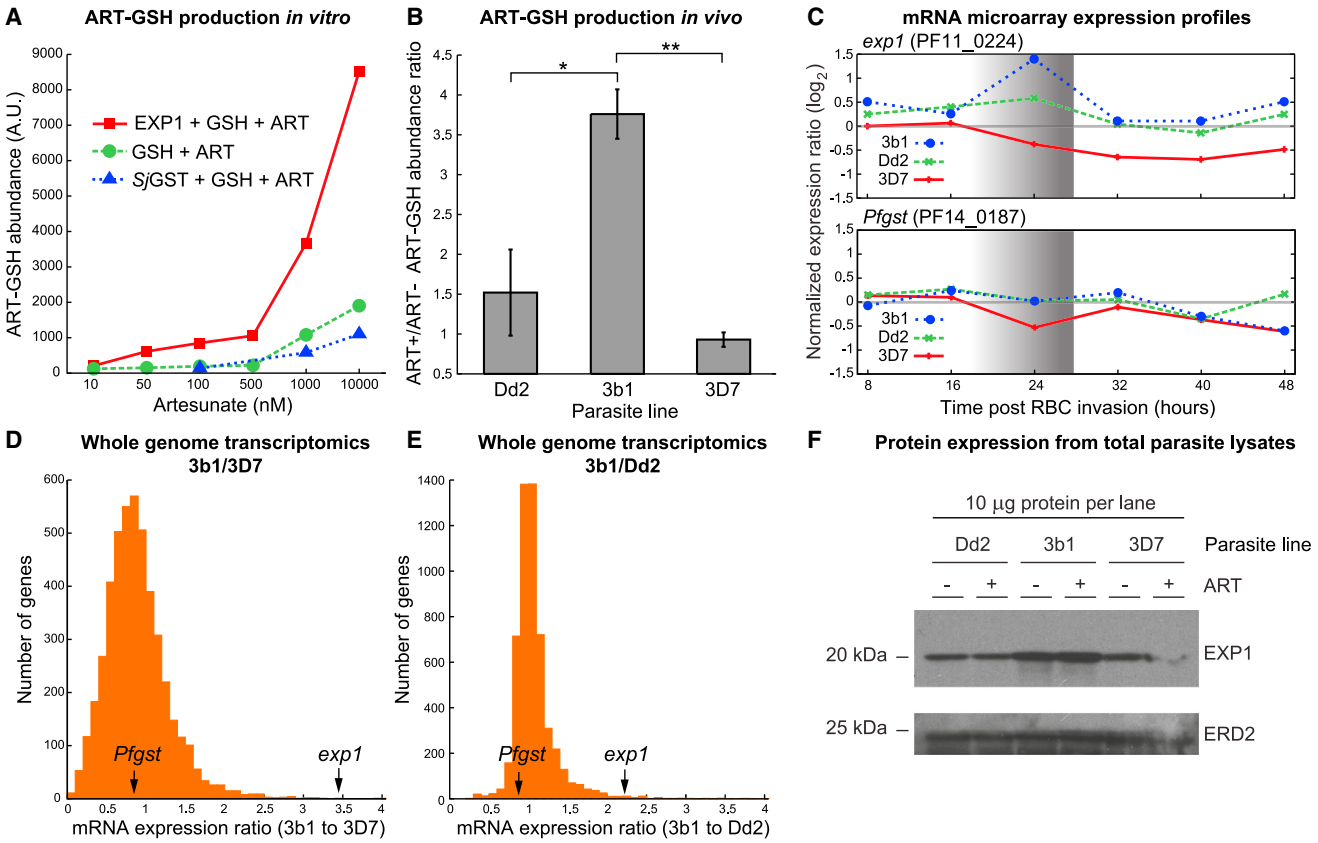

Figure 6. EXP1 Is Associated with ART Metabolism and Modulation of ART Susceptibility

(A) In vitro LC-MS of GSH-ART product formation (monitored at *m/z* 692.0) with increasing concentrations of ART.

(B) Abundance ratios of GSH-ART products from cultured parasite samples challenged with ART (ART^+) against unchallenged parasites (ART^-); error bars represent minimum and maximum ratios from two independent samples per condition. Significance levels *$p < 0.05$ and **$p < 0.01$ are from Student's t test.

(C) Blood stage RNA microarray expression temporal profiles for *exp1* and *Pfgst* at six time points between 8 and 48 hr post-RBC invasion. Shaded area indicates late-ring/trophozoite transition.

(D) Distribution of transcript-level fold changes between the ART-resistant line 3b1 and the ART-sensitive line 3D7 from whole-genome mRNA expression arrays over 5,446 *P. falciparum* genes (see Extended Experimental Procedures).

(E) Distribution of transcript-level fold changes between the ART-tolerant line 3b1 and the ART-sensitive line Dd2 (the parental line from which 3b1 was derived) from whole-genome mRNA expression arrays over 5,446 *P. falciparum* genes (see Extended Experimental Procedures).

(F) Western immunoblotting shows evidence of elevated EXP1 levels in the ART-tolerant 3b1 line compared to Dd2 and 3D7. Loading controls employed antibodies to the ER lumen protein-retaining receptor ERD2 (PF13_0280).

ART challenge were only 1.52 ± 0.54 times higher than those without drug exposure.

DNA sequencing of the *exp1* locus, including 1.6 kb upstream and 0.9 kb downstream of the coding region, revealed no difference between Dd2 and 3b1 (data not shown), and quantitative PCR indicated no *exp1* copy number change in 3b1 compared to Dd2 (Extended Experimental Procedures). mRNA microarray temporal profiles over six time points in the intra-erythrocytic parasite stages showed the strongest transcriptional upregulation of *exp1* in line 3b1 near the late-ring/trophozoite transition (Figure 6C), whereas this upregulation was less pronounced in Dd2 by a factor of two. Relative to 3D7, *exp1* mRNA levels in 3b1 parasites were the third most highly upregulated among all 5,446 genes tested (3.5-fold increase; $p < 6 \times 10^{-5}$; Figure 6C; Table S4; our unpublished data). In comparison to Dd2, *exp1* mRNA levels in 3b1 were also upregulated by a factor of 2.2 ($p < 0.012$; Figure 6D; Table S4). Remarkably, *Pfgst* (PF14_0187) did not present higher transcript levels (expression ratio 0.8 in 3b1/3D7 and 0.8 in 3b1/Dd2; Figures 6C–6E). To increase microarray mRNA signal, 3b1 and Dd2 parasite samples were pretreated with terminator exonuclease, and we note that lower *exp1* expression ratios were observed without this treatment (expression ratio 1.3 in 3b1/3D7 and 1.2 in 3b1/Dd2). EXP1 protein levels measured through immunoblotting with EXP1 antibodies in parasites challenged for 6 hr with ART suggested elevated expression in 3b1, followed by EXP1 levels in the unchallenged line 3b1 (Figure 6F). This evidence of upregulation was not observed in the control lines Dd2 and 3D7 (Figure 6F). Together, these results suggest that EXP1 is a GST with ART as a substrate and support a model in which EXP1 may be part of a reductive metabolic pathway of ART by conjugating it to reduced glutathione.

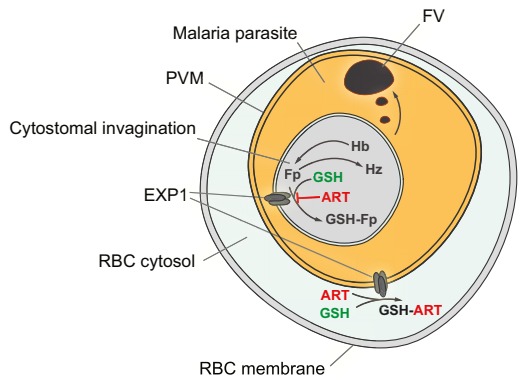

Figure 7. Model of EXP1 in the Asexual Intraerythrocytic *P. falciparum*

EXP1 resides in homotrimeric complexes mainly in the PVM and inside cytostomal invaginations of the PVM where, through GST activity, it may protect membranes from hematin (ferriprotoporphyrin IX, Fp), which can be released from internalized red blood cell (RBC) cytosol and catabolized hemoglobin (Hb). GSH-Fp complex formation may be an alternative to the detoxification pathway of Fp into hemozoin (Hz) through biomineralization, which eventually leads to hemozoin (Hz) deposits in the food vacuole (FV). In the presence of heme/hematin, EXP1 activity can be potently inhibited by ART. EXP1 may also degrade ART by conjugating it to reduced glutathione (GSH) in drug-tolerant parasites.

DISCUSSION

Compression of supergenomic networks may be the first positive evidence to address the hypothesis that large biological systems, when represented through evolutionarily related information sequences, are algorithmically compressible (Oexle, 1995). Critically, increased computational efficiency through compression did not prove costly in terms of functional sensitivity and accuracy: the malarial antigen EXP1 case study, in which the network prediction preceded and guided the experimental validation, illustrates the value of global network information.

The discovery of GST activity for EXP1, which we now also refer to as *Pf*MGST (*P. falciparum* membrane glutathione S-transferase) and, for HEP17, *Py*MGST (*Plasmodium yoelii* membrane glutathione S-transferase), points to EXP1 as being a key component of glutathione metabolism in *P. falciparum*. Heme/hematin release through hemoglobin degradation, and separately antimalarial drug action, both confer strong oxidative insults on the host cell and on the parasite (Becker et al., 2004). Thus, relative to the classical hemozoin formation pathway, EXP1-catalyzed hematin degradation may constitute a novel GSH-mediated detoxification pathway within the malaria parasite (Figure 7). Given that no other known essential protein activity or recently discussed drug target candidates (Cardi et al., 2010; Chugh et al., 2013) are inhibited by ARTs at comparably low concentrations, and the mounting but indirect evidence that these drugs interfere with the glutathione-redox cycle (O'Neill et al., 2010), we propose that EXP1 inhibition might be an important component of the mode of action of ART.

A prominent feature of *P. falciparum* lines selected for resistance to artemisinins appears to be ring-stage quiescence (Witkowski et al., 2010): a temporary arrest of development and hemoglobin uptake and heme/hematin release until toxic ART levels have been reduced sufficiently to resume growth (Klonis et al., 2013a). Sensitive parasites appear to also be able to enter the dormant stage, albeit at a reduced frequency compared to in vitro-derived resistant lines, suggesting that resistant lines have multiple means to withstand ART exposure (Teuscher et al., 2012). Prior studies with ART-resistant *P. yoelii* parasites have observed an upregulation of GSH-mediated detoxification pathways (Witkowski et al., 2012). During ring-stage quiescence, EXP1 might directly contribute to the degradation of ARTs by conjugating them to reduced GSH, thus lowering the drugs' oxidative potential and allowing a stronger parasite population to recrudesce (Figure 7). Our data support further studies on EXP1 and GSH-mediated oxidative stress responses in ART action and *P. falciparum* susceptibility.

EXPERIMENTAL PROCEDURES

Supergenomic Network Data and Compressible Cliques

Large-scale protein network data came from the Search Tools for Interacting Genes (STRING) database version 7.1. This had a total of 1.513 million protein nodes of which 33% belonged to eukaryota, 63% to bacteria, and 4% to archaea. Nodes were connected with intrinsic evolutionary and with contextual species-specific links, including six types of protein-protein associations: immunoprecipitation, yeast two-hybrid, coexpression, conserved genomic neighborhood, phylogenetic co-occurrence, and literature co-occurrence. Contextual links between nodes had weights given by a Bayes classifier of evidence scores from STRING and normalized between 0 and 1, such that 1 indicated the highest confidence in a functional association. In total, the network had $n_c = 38{,}573{,}579 = 3.86 \times 10^7$ contextual links. Intrinsic links were binary: 1 indicated that two proteins were orthologs or that they were paralogs that shared orthology to a third protein, and 0 indicated that no evolutionary relationship was detected. Orthology was established in a pair if its two protein sequences had symmetrical best hits in sequence identity based on the Smith-Waterman algorithm among all given lineages; mutually orthologous proteins from at least three different species gave rise to COGs, which were computed using the automated eggNOG algorithm (Jensen et al., 2008). COGs were characterized as *maximal cliques* (Mohseni-Zadeh et al., 2004). A clique is *maximal* if it is not a subset of any other clique, and here we always mean a maximal clique with n_{cog} nodes when simply referring to a clique. Paralogous genes from single species were grouped into single units (Jensen et al., 2008), and COGs were maximal n_{cog}-cliques when detected through best reciprocal sequence hits or became so through mergers of maximal cliques of n_{cog} nodes that have $(n_{cog} - 1)$ nodes in common (Falls et al., 2004; Montague and Hutchison, 2000). Empirically (Mohseni-Zadeh et al., 2004), mergers affect only a few (often less than 1%) of the originally detected cliques, which allowed us directly to model COGs as maximal cliques in the sequence matching network. The total number of intrinsic links within COGs is the sum of n_{cog} $(n_{cog} - 1)/2$ over all 33,929 COGs present in the network. This gave $540{,}563{,}390 \approx 5.41 \times 10^8$ intra-COG links. Additionally, any pair of cliques with significant sequence similarity results in at most $n_{cog1} \times n_{cog2}$ *inter*-COG links, and overall such pairs produced a maximum of $n_i \approx 1.93 \times 10^{11}$ links.

Cloning, Expression, and Purification of *P. falciparum* EXP1

The *P. falciparum exp1* cDNA was PCR amplified using the plasmid DNA clone pHRPExGFP (kindly deposited by Dr. Kasturi Haldar at the Malaria Research and Reference Reagent Resource Center, MR4, Manassas, VA, USA). The recombinant *exp1* product was sequence verified and cloned between NdeI and HindIII restriction sites in the Kan^R pET28a (+) vector from Novagen (New Canaan, CT, USA) to enable *E. coli* expression of N-terminal His-tagged EXP1 protein (OL617, pET28a-His-EXP1). The EXP1 R70T mutant (OL618, pET28a-His-EXP1-R70T) was generated using the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA, USA). The His-tagged

EXP1 WT and R70T proteins from the total lysate were purified using Ni-NTA agarose (QIAGEN, Valencia, CA, USA) affinity column chromatography. Recombinant proteins were aliquoted and stored at −80°C in GST assay buffer (pH 7.3) containing 2% glycerol and 0.1% Triton X-100.

EXP1 GST Enzyme Assay

Purified EXP1 was preincubated in GST assay buffer (pH 6.5) with 0.1% Triton X-100 and 2 mM reduced GSH on ice for 90 min, followed by the addition of CDNB. Absorbance of the reaction was monitored at 340 nm, every 15 s for a period of 10 min. GST activity toward hematin was assayed by preincubating 100 nM WT protein with 0.1% Triton X-100, 2 mM GSH in pH 6.5 assay buffer for 30 min on ice, followed by the addition of hematin to initiate hematin degradation. Absorbance was monitored at 395 nm every second for 3 min.

EXP1 ART Inhibition Assay

ART and atovaquone were dissolved in 100% ethanol and added to 100 nM WT EXP1 hematin reactions to determine hematin degradation inhibition. WT EXP1 was preincubated in GST assay buffer (pH 6.5) with 0.1% Triton X-100 and ART on ice for 30 min followed by the addition of 2 mM GSH and hematin. The reaction was monitored at 395 nm every second for 3 min.

Generation of DHA/ART-Tolerant Parasites

3b1 parasites were selected from the Dd2 strain by culturing asexual blood stage parasites (~10^8 on average) in the presence of the artemisinin derivative dihydroartemisinin (DHA) at sub-IC_{50} concentrations (2.8 nM, as compared to a parental IC_{50} value of 6.4 nM) for 55 days, followed by progressive increases in drug concentration (up to 28 nM) over the next 200 days. Acquired tolerance to DHA and to ART was evidenced as an increased frequency of recrudescence over a 30 day period following parasite exposure to high concentrations of DHA/ART for 4 days (up to 112 nM; our unpublished data).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2014.07.011.

AUTHOR CONTRIBUTIONS

A.M.L., C.L.N., A.K.A., N.P., and A.S. conceived, designed, and conducted experiments, analyzed data, produced figures, and wrote the text. J.P.Q. conceived, designed, and conducted experiments, analyzed data, and produced figures. C.L. and K.M. designed and conducted experiments, analyzed data, produced figures, and edited the text. R.T.E. conducted experiments and edited the text. D.S. analyzed data. S.J.R. analyzed data and produced figures. L.A. conducted experiments and analyzed data. M.L., D.A.F., and O.L. conceived and designed experiments, analyzed data, and edited the text.

ACKNOWLEDGMENTS

Financial support came from NIH GM066099 and GM079656 and NSF CCF-0905536, DBI-1062455 (to O.L.) and from NIH AI50234 and AI109023 (to D.A.F.). N.P. and A.S. were supported by Alkek CMD Grants. K.M., R.E., and C.L. were supported by the Divisions of Intramural Research at the National Institute of Allergy and Infectious Diseases, National Institutes of Health. M.L. is funded through a Burroughs Wellcome Fund Investigators in Pathogenesis of Infectious Disease Grant and an NIH Director's New Innovators award (1DP2OD001315-01) with generous support from the Centre for Quantitative Biology (P50 GM071508). The authors gratefully acknowledge comments and help from Dr. Theodore Wensel, Dr. Christophe Herman, Dr. Timothy Palzkill, Dr. Santha Kumar, and Dr. David Marciano.

Received: February 12, 2013
Revised: January 2, 2014
Accepted: July 7, 2014
Published: August 14, 2014

REFERENCES

Abu Bakar, N., Klonis, N., Hanssen, E., Chan, C., and Tilley, L. (2010). Digestive-vacuole genesis and endocytic processes in the early intraerythrocytic stages of *Plasmodium falciparum*. J. Cell Sci. *123*, 441–450.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Andersson, C., Piemonte, F., Mosialou, E., Weinander, R., Sun, T.H., Lundqvist, G., Adang, A.E., and Morgenstern, R. (1995). Kinetic studies on rat liver microsomal glutathione transferase: consequences of activation. Biochim. Biophys. Acta *1247*, 277–283.

Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., et al. (2014). A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria. Nature *505*, 50–55.

Atamna, H., and Ginsburg, H. (1995). Heme degradation in the presence of glutathione. A proposed mechanism to account for the high levels of non-heme iron found in the membranes of hemoglobinopathic red blood cells. J. Biol. Chem. *270*, 24876–24883.

Aurrecoechea, C., Brestelli, J., Brunk, B.P., Dommer, J., Fischer, S., Gajria, B., Gao, X., Gingle, A., Grant, G., Harb, O.S., et al. (2009). PlasmoDB: a functional genomic database for malaria parasites. Nucleic Acids Res. *37* (Database issue), D539–D543.

Barabasi, A.L., and Albert, R. (1999). Emergence of scaling in random networks. Science *286*, 509–512.

Becker, K., Tilley, L., Vennerstrom, J.L., Roberts, D., Rogerson, S., and Ginsburg, H. (2004). Oxidative stress in malaria parasite-infected erythrocytes: host-parasite interactions. Int. J. Parasitol. *34*, 163–189.

Bozdech, Z., Llinás, M., Pulliam, B.L., Wong, E.D., Zhu, J., and DeRisi, J.L. (2003). The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. PLoS Biol. *1*, E5.

Cardi, D., Pozza, A., Arnou, B., Marchal, E., Clausen, J.D., Andersen, J.P., Krishna, S., Møller, J.V., le Maire, M., and Jaxel, C. (2010). Purified E255L mutant SERCA1a and purified PfATP6 are sensitive to SERCA-type inhibitors but insensitive to artemisinins. J. Biol. Chem. *285*, 26406–26416.

Caspers, P., Etlinger, H., Matile, H., Pink, J.R., Stüber, D., and Takács, B. (1991). A *Plasmodium falciparum* malaria vaccine candidate which contains epitopes from the circumsporozoite protein and a blood stage antigen, 5.1. Mol. Biochem. Parasitol. *47*, 143–150.

Cheeseman, I.H., Miller, B.A., Nair, S., Nkhoma, S., Tan, A., Tan, J.C., Al Saai, S., Phyo, A.P., Moo, C.L., Lwin, K.M., et al. (2012). A major genome region underlying artemisinin resistance in malaria. Science *336*, 79–82.

Chua, H.N., Sung, W.K., and Wong, L. (2006). Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions. Bioinformatics *22*, 1623–1630.

Chugh, M., Sundararaman, V., Kumar, S., Reddy, V.S., Siddiqui, W.A., Stuart, K.D., and Malhotra, P. (2013). Protein complex directs hemoglobin-to-hemozoin formation in *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA *110*, 5392–5397.

Deponte, M., and Becker, K. (2005). Glutathione S-transferase from malarial parasites: structural and functional aspects. Methods Enzymol. *401*, 241–253.

Eppstein, D., Loffler, M., and Strash, D. (2010). Listing all maximal cliques in sparse graphs in near-optimal time. Proceedings of the 21st International Symposium on Algorithms and Computation (ISAAC 2010) *6506*, 403–413.

Erdin, S., Lisewski, A.M., and Lichtarge, O. (2011). Protein function prediction: towards integration of similarity metrics. Curr. Opin. Struct. Biol. *21*, 180–188.

Famin, O., Krugliak, M., and Ginsburg, H. (1999). Kinetics of inhibition of glutathione-mediated degradation of ferriprotoporphyrin IX by antimalarial drugs. Biochem. Pharmacol. *58*, 59–68.

Fry, M., and Pudney, M. (1992). Site of action of the antimalarial hydroxynaphthoquinone, 2-[trans-4-(4′-chlorophenyl) cyclohexyl]-3-hydroxy-1,4-naphthoquinone (566C80). Biochem. Pharmacol. *43*, 1545–1553.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M., Pain, A., Nelson, K.E., Bowman, S., et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature *419*, 498–511.

Ginsburg, H., Famin, O., Zhang, J., and Krugliak, M. (1998). Inhibition of glutathione-dependent degradation of heme by chloroquine and amodiaquine as a possible basis for their antimalarial mode of action. Biochem. Pharmacol. *56*, 1305–1313.

Gribskov, M., and Robinson, N.L. (1996). Use of receiver operating characteristic (ROC) analysis to evaluate sequence matching. Comput. Chem. *20*, 25–33.

Grüring, C., Heiber, A., Kruse, F., Ungefehr, J., Gilberger, T.W., and Spielmann, T. (2011). Development and host cell modifications of *Plasmodium falciparum* blood stages in four dimensions. Nat Commun *2*, 165.

Habig, M., Blechschmidt, A., Dressler, S., Hess, B., Patel, V., Billich, A., Ostermeier, C., Beer, D., and Klumpp, M. (2009). Efficient elimination of nonstoichiometric enzyme inhibitors from HTS hit lists. J. Biomol. Screen. *14*, 679–689.

Hiller, N., Fritz-Wolf, K., Deponte, M., Wende, W., Zimmermann, H., and Becker, K. (2006). *Plasmodium falciparum* glutathione S-transferase—structural and mechanistic studies on ligand binding and enzyme inhibition. Protein Sci. *15*, 281–289.

Holm, P.J., Bhakat, P., Jegerschöld, C., Gyobu, N., Mitsuoka, K., Fujiyoshi, Y., Morgenstern, R., and Hebert, H. (2006). Structural basis for detoxification and oxidative stress protection in membranes. J. Mol. Biol. *360*, 934–945.

Hope, I.A., Hall, R., Simmons, D.L., Hyde, J.E., and Scaife, J.G. (1984). Evidence for immunological cross-reaction between sporozoites and blood stages of a human malaria parasite. Nature *308*, 191–194.

Hope, I.A., Mackay, M., Hyde, J.E., Goman, M., and Scaife, J. (1985). The gene for an exported antigen of the malaria parasite *Plasmodium falciparum* cloned and expressed in *Escherichia coli*. Nucleic Acids Res. *13*, 369–379.

Ideker, T., and Sharan, R. (2008). Protein networks in disease. Genome Res. *18*, 644–652.

Jakobsson, P.J., Morgenstern, R., Mancini, J., Ford-Hutchinson, A., and Persson, B. (1999). Common structural features of MAPEG — a widespread superfamily of membrane associated proteins with highly divergent functions in eicosanoid and glutathione metabolism. Protein Sci. *8*, 689–692.

Klonis, N., Crespo-Ortiz, M.P., Bottova, I., Abu-Bakar, N., Kenny, S., Rosenthal, P.J., and Tilley, L. (2011). Artemisinin activity against *Plasmodium falciparum* requires hemoglobin uptake and digestion. Proc. Natl. Acad. Sci. USA *108*, 11405–11410.

Klonis, N., Creek, D.J., and Tilley, L. (2013a). Iron and heme metabolism in *Plasmodium falciparum* and the mechanism of action of artemisinins. Curr. Opin. Microbiol. *16*, 722–727.

Klonis, N., Xie, S.C., McCaw, J.M., Crespo-Ortiz, M.P., Zaloumis, S.G., Simpson, J.A., and Tilley, L. (2013b). Altered temporal response of malaria parasites determines differential sensitivity to artemisinin. Proc. Natl. Acad. Sci. USA *110*, 5157–5162.

Lam, B.K., Penrose, J.F., Xu, K., Baldasaro, M.H., and Austen, K.F. (1997). Site-directed mutagenesis of human leukotriene C4 synthase. J. Biol. Chem. *272*, 13923–13928.

Le Roch, K.G., Johnson, J.R., Florens, L., Zhou, Y., Santrosyan, A., Grainger, M., Yan, S.F., Williamson, K.C., Holder, A.A., Carucci, D.J., et al. (2004). Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. Genome Res. *14*, 2308–2318.

Maier, A.G., Rug, M., O'Neill, M.T., Brown, M., Chakravorty, S., Szestak, T., Chesson, J., Wu, Y., Hughes, K., Coppel, R.L., et al. (2008). Exported proteins required for virulence and rigidity of *Plasmodium falciparum*-infected human erythrocytes. Cell *134*, 48–61.

Melvin, I., Weston, J., Leslie, C., and Noble, W.S. (2009). RANKPROP: a web server for protein remote homology detection. Bioinformatics *25*, 121–122.

Meraldi, V., Nebié, I., Tiono, A.B., Diallo, D., Sanogo, E., Theisen, M., Druilhe, P., Corradin, G., Moret, R., and Sirima, B.S. (2004). Natural antibody response to *Plasmodium falciparum* Exp-1, MSP-3 and GLURP long synthetic peptides and association with protection. Parasite Immunol. *26*, 265–272.

Mihalek, I., Res, I., and Lichtarge, O. (2004). A family of evolution-entropy hybrid methods for ranking protein residues by importance. J. Mol. Biol. *336*, 1265–1282.

Morgenstern, R., Zhang, J., and Johansson, K. (2011). Microsomal glutathione transferase 1: mechanism and functional roles. Drug Metab. Rev. *43*, 300–306.

Mukanganyama, S., Naik, Y.S., Widersten, M., Mannervik, B., and Hasler, J.A. (2001). Proposed reductive metabolism of artemisinin by glutathione transferases *in vitro*. Free Radic. Res. *35*, 427–434.

Newman, M.E.J. (2005). Power laws, Pareto distributions and Zipf's law. Contemp. Phys. *46*, 323–351.

O'Neill, P.M., Barton, V.E., and Ward, S.A. (2010). The molecular mechanism of action of artemisinin—the debate continues. Molecules *15*, 1705–1721.

Ochoa, A., Llinás, M., and Singh, M. (2011). Using context to improve protein domain identification. BMC Bioinformatics *12*, 90.

Oexle, K. (1995). Data compressibility, physical entropy, and evolutionary a priori relation between observer and object. Phys. Rev. E *51*, 2651–2653.

Park, D.J., Lukens, A.K., Neafsey, D.E., Schaffner, S.F., Chang, H.H., Valim, C., Ribacke, U., Van Tyne, D., Galinsky, K., Galligan, M., et al. (2012). Sequence-based association and selection scans identify drug resistance loci in the *Plasmodium falciparum* malaria parasite. Proc. Natl. Acad. Sci. USA *109*, 13052–13057.

Radivojac, P., Clark, W.T., Oron, T.R., Schnoes, A.M., Wittkop, T., Sokolov, A., Graim, K., Funk, C., Verspoor, K., Ben-Hur, A., et al. (2013). A large-scale evaluation of computational protein function prediction. Nat. Methods *10*, 221–227.

Sharan, R., Ulitsky, I., and Shamir, R. (2007). Network-based prediction of protein function. Mol. Syst. Biol. *3*, 88.

Simmons, D., Woollett, G., Bergin-Cartwright, M., Kay, D., and Scaife, J. (1987). A malaria protein exported into a new compartment within the host erythrocyte. EMBO J. *6*, 485–491.

Simossis, V.A., Kleinjung, J., and Heringa, J. (2005). Homology-extended sequence alignment. Nucleic Acids Res. *33*, 816–824.

Spielmann, T., Gardiner, D.L., Beck, H.P., Trenholme, K.R., and Kemp, D.J. (2006). Organization of ETRAMPs and EXP-1 at the parasite-host cell interface of malaria parasites. Mol. Microbiol. *59*, 779–794.

Takala-Harrison, S., Clark, T.G., Jacob, C.G., Cummings, M.P., Miotto, O., Dondorp, A.M., Fukuda, M.M., Nosten, F., Noedl, H., Imwong, M., et al. (2013). Genetic loci associated with delayed clearance of *Plasmodium falciparum* following artemisinin treatment in Southeast Asia. Proc. Natl. Acad. Sci. USA *110*, 240–245.

Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997). A genomic perspective on protein families. Science *278*, 631–637.

Teuscher, F., Chen, N., Kyle, D.E., Gatton, M.L., and Cheng, Q. (2012). Phenotypic changes in artemisinin-resistant *Plasmodium falciparum* lines in vitro: evidence for decreased sensitivity to dormancy and growth inhibition. Antimicrob. Agents Chemother. *56*, 428–431.

UniProt Consortium (2010). The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res. *38* (Database issue), D142–D148.

Vazquez, A., Flammini, A., Maritan, A., and Vespignani, A. (2003). Global protein function prediction from protein-protein interaction networks. Nat. Biotechnol. *21*, 697–700.

Venner, E., Lisewski, A.M., Erdin, S., Ward, R.M., Amin, S.R., and Lichtarge, O. (2010). Accurate protein structure annotation through competitive diffusion of enzymatic functions over a network of local evolutionary similarities. PLoS ONE *5*, e14286.

von Mering, C., Jensen, L.J., Kuhn, M., Chaffron, S., Doerks, T., Krüger, B., Snel, B., and Bork, P. (2007). STRING 7—recent developments in the integration and prediction of protein interactions. Nucleic Acids Res. *35* (Database issue), D358–D362.

Wang, B., Yang, X.-h., and Wang, W.-l. (2009). A novel scale-free network model based on clique growth. J Cent South Univ Technol *16*, 474–477.

Witkowski, B., Lelièvre, J., Barragán, M.J., Laurent, V., Su, X.Z., Berry, A., and Benoit-Vical, F. (2010). Increased tolerance to artemisinin in *Plasmodium falciparum* is mediated by a quiescence mechanism. Antimicrob. Agents Chemother. *54*, 1872–1877.

Witkowski, B., Lelièvre, J., Nicolau-Travers, M.L., Iriart, X., Njomnang Soh, P., Bousejra-Elgarah, F., Meunier, B., Berry, A., and Benoit-Vical, F. (2012). Evidence for the contribution of the hemozoin synthesis pathway of the murine *Plasmodium yoelii* to the resistance to artemisinin-related drugs. PLoS ONE *7*, e32620.

World Health Organization (WHO) (2011). Section 8.4.1. In Guidelines for the Treatment of Malaria, 2nd edition – Rev. 1 (World Health Organization).

Yule, G.U. (1925). A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis. Philos. Trans. R. Soc. Lond. B *213*, 21–87.

EXTENDED EXPERIMENTAL PROCEDURES

GID

Given n nodes that each depicts one protein, some carry non-zero labels y_i that assign functional classes. If nodes i and j were linked a weighted adjacency matrix $W = \{w_{ij}\}_{1 \le i,j \le n}$ was set to $w_{ij} > 0$, otherwise $w_{ij} = 0$. Inferring function for the remaining $n - p$ nodes is based on a quadratic minimization problem (Tsuda et al., 2005) $H = \sum_i (f_i - y_i)^2 + \gamma/2 \sum_{i,j} w_{ij}(f_i - f_j)^2$ where the functional label y_i is set to either 1 if node i has the function, to –1 if it does not have that function, or 0 if there is no evidence either way (unlabeled node). The solution $f = \{f_1, \ldots, f_n\}$ is given by $f = (1 + \gamma L)^{-1} y$, where $L = D - W$ is the Laplacian matrix, with the weight matrix and $D = diag(d_i)$, $d_i = \sum_j w_{ij}$. The parameter γ was set to the value $1/\|L\|$, with the maximum matrix norm $\|L\| = \max\{d_i\}_{1 \le i \le n}$, which ensured convexity of the cost function (Lisewski and Lichtarge, 2010). A Z score was then calculated among all nodes $i \in \Omega$, where Ω is the set of all nodes labeled with $y_i = 0$ at the input and which had a non-zero output value f_i; then $Z_i = (-\log^{-1}(f_i) - \log^{-1}(f))/\mathrm{std}(\log^{-1}(f))$, where $\log^{-1}(f)$ was the mean of the reciprocal logarithm of the output f among all, and is the corresponding standard variation. For negative output f_i the same procedure was applied to the absolute values $|f_i|$ and the Z score obtained a negative sign. If an output value $|f_i|$ was below the fixed numerical precision (10^{-8}) in the solution, then a Z score was not determined (*ND*). All calculations were carried out using the *Matlab* computer program (version 7.1, MathWorks Inc., Natick, MA, USA). A *Matlab* program code (file gid_supergenomic.m) for GID on a supergenomic network is deposited at the Internet address http://mammoth.bcm.tmc.edu/GID/lib.tgz. Statistical p values from the Wilcoxon signed rank test were calculated with the *R* computer program (v2.12) through the wilcox.test function. The protein functional information necessary for the input labels in GID was retrieved from Gene Ontology Annotation (http://www.ebi.ac.uk/GOA/). Gene Ontology terms enrichment analysis was performed using the AmiGO internet page (http://amigo.geneontology.org/). Calculated enrichment p_e values were taken directly from the AmiGO output page; significance cut-off was chosen at $p_e = 10^{-6}$.

Error and Computational Cost of Network Compression

To estimate the relative error caused by full compression, the combination of intra- and inter-clique compression was tested on an extensive sample of random model networks (Figure S1B). The relative error in the final state of each node after GID remained less than 10% if the original graph had an edge density of less than ~20%. In the 373 genomes supergenomic network with an edge density of 16% these data specifically suggest a relative error of ~7%. This error could be further reduced by making the network more sparse; for example, setting a stricter similarity cutoff, from 10% to 40%, lowered the number of links between COGs and the network link density to 5%, which yielded an estimated GID error of ~2% (Figure S1B). For GID on the entire supergenomic network with over 1.5 million nodes these simulations extrapolate to ~600 s runtime (Figures S1C and S1E, blue data points and fitted curve) with a comparatively modest memory demand of 1.73 GB (Figures S1D and S1E, blue data points and fitted curve). This is consistent with observations: GID on the compressed supergenomic network averaged 842 ± 85 s of single CPU time per run and required 1.53 GB of computer memory. In sharp contrast, an uncompressed network would require an impractical computer memory of 65.2 TB (Figures S1D and S1E, red data points and fitted curve extrapolation) and a computer time of ~44 hr per run (Figures S1C and S1E, red data points and fitted curve extrapolation).

Integrating Intrinsic and Contextual Network Data

To further increase the functional coverage of the GID predictions (Table S2), the Laplacian matrix of the contextual network was added to the existing Laplacian matrix of the compressed intrinsic network. The two Laplacian kernel matrices, L_i and L_c, were added together in a linear superposition, i.e., $L = \alpha_i L_i + \alpha_c L_c$, where the relative weights were kept equal ($\alpha_i = \alpha_c$). This step is equivalent to adding together the corresponding weighted adjacency matrices: $\alpha_i W_i + \alpha_c W_c$. This was feasible because contextual and intrinsic network links had separate link sets that were individually relevant for predicting gene function (Figure S2A), and therefore a further optimization of relative weights was not necessary (Mostafavi et al., 2008; Tsuda et al., 2005). A validation test on the same Molecular Function test set as before showed that GID on the integrated network was more accurate (Figure S2A). Higher accuracies were also observed in a series of leave-one-out tests for the other two Gene Ontology categories: Biological Process (Figure S2B) and Cellular Component (Figure S2C). In these experiments accuracies from intrinsic network data were significantly higher than those obtained with contextual network links; nevertheless the combination of both network data sources always resulted in significant accuracy gains (Figures S2A–S2C). GID thus enabled the large-scale integration of contextual and intrinsic network data by improving the prediction of Gene Ontology terms that describe protein function.

Cloning, Expression, and Purification of *P. falciparum* EXP1

The *P. falciparum exp1* cDNA was amplified by polymerase chain reaction using the plasmid DNA clone pHRPExGFP (kindly deposited by Dr. Kasturi Haldar at the Malaria Research and Reference Reagent Resource Center, MR4, Manassas, VA, USA). The amplified *exp1* product was sequence verified and cloned between NdeI and HindIII restriction sites in the Kan^R pET28a (+) vector from Novagen (New Canaan, CT, USA) to enable expression of N-terminal His-tagged EXP1 protein (OL617, pET28a-His-EXP1). The EXP1 R70T mutant (OL618, pET28a-His-EXP1-R70T) was generated using the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA, USA). Both strands of the WT *exp1* and R70T *exp1* ORF were sequence verified. The *exp1* plasmid was transformed into BL21 (DE3) *E. coli* cells. 1L cultures in LB broth were grown at 37°C to an OD600 of 0.5, and protein expression was induced with 0.1 mM IPTG. Induced cultures were grown for 14 hr at 30°C. The cells were harvested by centrifugation and stored

at −80°C, until further processing. Cell pellets were thawed and lysed with bacterial lysis reagent (*Bugbuster Master Mix*, Novagen, New Canaan, CT, USA) with additional detergent (Octyl thioglucoside, 60 mM final concentration, Thermo Scientific), to extract the membrane-bound proteins. The His-tagged EXP1 from the total lysate was purified using Ni-NTA agarose (QIAGEN, Valencia, CA, USA) affinity column chromatography. The EXP1 protein was determined to be greater than 90% pure using silver-stained SDS-PAGE gels. For the western blot analyses of His-tagged Exp1 expression and its cross-reactivity with GST antibody, the bacterially expressed recombinant His-tagged EXP1 protein along with standard GST proteins (*E. coli* and *S. japonicum*) were resolved by SDS-PAGE using 12% gels and immunoblotted with anti-His antibody (0.1 μg/ml) and anti-GST antibody (1:5,000). Chemiluminescence was detected by western blotting and autoradiographed. Recombinant EXP1 WT and R70T proteins were aliquoted and stored in GST assay buffer (pH 7.3) containing 20% glycerol and 10% Triton X-100 at −80°C. Protein estimations were done by the bicinchoninic acid method. *P. yoelii* HEP17 cDNA was cloned into pET28a (GenScript, Piscataway, NJ, USA). *hep17* cDNA was cloned into pET28a HEP17 expression and purification were performed using the same protocol as was done for EXP1. Purified recombinant human MGST1 protein expressed in *E. coli* was purchased from Bioclone Inc. (San Diego, CA, USA).

Glutathione S-Transferase Enzyme Assay

Glutathione S-transferase activity in the purified recombinant WT EXP1 and its R70T mutant protein toward 1-chloro-2,4-dinitrobenzene (CDNB) was assayed using the GST assay kit (Novagen, New Canaan, CT, USA). Assays were conducted with buffer at pH 6.5 to minimize the nonenzymatic reaction. Protein, either purified EXP1 or BSA (negative control) or *Sj*GST (positive control) was preincubated with 0.1% Triton X-100 and 2 mM reduced glutathione on ice for 90 min, followed by the addition of CDNB to initiate the enzymatic reaction. The absorbance of the reaction was monitored at 340 nm, every 15 s for a period of 10 min. Assays were repeated at least three times independently for each sample, and the GST-specific activity was calculated as per the kit protocol. *P. yoelii* HEP17 GST activity was similarly assayed.

EXP1 Hematin Inhibition Assay

Using the GST kit, WT EXP1 was assayed with increasing amounts of hematin (Sigma-Aldrich, St. Louis, MO, USA) dissolved in 0.2 M NaOH for 5 min before addition of CDNB. Absorbance was recorded at 340 nm every 15 s for a period of 5 min.

EXP1 Hematin Degradation Assay

GST activity toward hematin was assayed by preincubating 100 nM WT protein with 0.1% Triton X-100, 2 mM GSH in pH 6.5 assay buffer for 30 min on ice followed by the addition of hematin to initiate hematin degradation. Absorbance was monitored at 395 nm every second for 3 min. *P. yoelii* HEP17 hematin degradation was similarly assayed.

EXP1 Drug Inhibition Assay

Varying amounts of ART and atovaquone were dissolved in 100% ethanol was added to 100 nM WT EXP1 hematin reactions to determine hematin degradation inhibition. WT EXP1 was preincubated in GST assay buffer (pH 6.5) with 0.1% Triton X-100 and ART on ice for 30 min followed by the addition of 2 mM GSH and hematin. The reaction was monitored at 395 nm every second for 3 min. Assays were repeated at least three times independently, and GST activity was monitored by the loss of hematin over time. Reaction kinetics and IC_{50} values were calculated and plotted using *Prism* software v6.0 (GraphPad, La Jolla, CA, USA).

Parasite In Vitro Culture Maintenance

P. falciparum asexual blood stage parasite strains Dd2, 3b1, and 3D7 were cultured in human erythrocytes as described (Fidock et al., 1998).

Sequence and Copy Number Analysis

exp1 sequence analysis was performed on the full-length coding region plus 1.6 kb and 0.9 kb of 5′ and 3′ UTR respectively, amplified using two overlapping sets of primers: 5′-AAGTGTATAATGCTAGCTTTGGG + 5′-GAAAATGATAAAGAAAAGAGCAAG' and 5′-TCTTCTTCTCTTATAGTTTGTAG + 5′-TTATTGTACTACATATGTATATGC. Sequencing with internal primers was performed on the direct PCR products as well as three independent colonies of each PCR fragment. This analysis showed no sequence differences between 3b1 and Dd2. To assess *exp1* copy number in Dd2 and 3b1 we performed quantitative PCR using genomic DNA input amounts of 5.0, 1.67, and 0.56 ng, each run in triplicate. The amplification was done using the iQ SYBR Green Supermix (Bio-Rad, Hercules, CA) using a DNA Engine thermal cycler (Bio-Rad). Cycling parameters were 95°C for 3 min, 95°C for 30 s, 55°C for 30 s and 62°C for 30 s for 40 times. *exp1* primers were 5′-CTTGCCACTTCAGTACTTGCAGG and 5′-ACCTCTGGTGTAACATCTTGAGC. The internal reference was the β-tubulin gene (PF10_0084; PF3D7_1008700), amplified using 5′-TGATGTGCGCAAGTGATCC and 5′-TCCTTTGTGGA CATTCTTCCTC. Exp1 copy number was normalized to β-tubulin using the ΔC(t) method. Data from three independent repeats yielded mean ± SEM ratios of 3b1/Dd2 of 1.03 ± 0.21 for *exp1*, consistent with no change in copy number.

Hematin-GSH Product Assay with LC-MS

P. falciparum strains Dd2, 3b1, and 3D7 were synchronized with 5% sorbitol and allowed to recover for one round of replication. In the second cycle of replication post-sorbitol treatment, parasites were treated for 6 hr with 3 × IC_{50} ART at 26–28 hr

post-invasion. Parasites were released from the RBC by saponin lysis and parasite extracts were prepared as described (Koncarevic et al., 2007). The formation of hematin-GSH product ion with mass-to-charge ratio *m/z* 923 was optimized with low fragmentation voltage; the collision induced dissociation mass spectra (CID) were recorded by selecting the parent adduct ion *m/z* 923 display and the *m/z* 615 ions that formed after loss of reduced glutathione. Samples were examined in positive ionization modes using an electrospray ionization source. Single reaction monitoring (SRM) experiments were performed using an electrospray ionization triple quadrupole mass spectrometer (*QQQ*, Agilent Technologies, Santa Clara, CA, USA). Chromatographic separation of the hematin-GSH product was performed using reverse phase (RP) separation in-line with *QQQ* mass spectrometers (Agilent Technologies). The RP separation was carried out using an analytical Zorbax Eclipse XDB-C18 column (50 × 4.6 mm id.; 1.8 μm, or for SRM, Agilent Technologies) that was used for SRM analysis. Amounts of eluted GSH-hematin product were calculated in the presence or absence of EXP1; in both cases their reaction times were controlled for and solvent blank readings were subtracted.

Immunofluorescence Assays

P. falciparum strains Dd2, 3b1, and 3D7 were cultured and synchronized as mentioned above. Cultures were harvested at ~12 hr post-invasion and 18 hr post-invasion, as determined by Giemsa stain. Alternatively, cultures were treated with DMSO or 3 × IC_{50} ART at 24 hr post-invasion for 6 hr. Harvested cultures were washed once in 1 × PBS, then allowed to adhere to poly-L-lysine-treated coverslips. Cells were fixed with 4% formaldehyde and 0.0075% glutaraldehyde, permeabilized with 0.1% Triton X-100, blocked with 3% BSA, incubated with mouse anti-*Pf*EXP1 antibodies (a kind gift from Dr. Joao Aguiar, Naval Medical Research Center, Silver Spring, MD, USA) followed by goat anti-mouse Alexa 488 (Molecular Probes, Eugene, OR, USA) antibodies, or rabbit anti-*Pf*CRT (Fidock et al., 2000) followed by goat anti-rabbit Alexa 596 antibodies (Molecular Probes) and Hoechst staining. Coverslips were mounted on glass slides with ProLong Gold Antifade Reagents (Invitrogen, Carlsbad, CA, USA) and visualized on a Nikon Eclipse T*i* microscope. Images were taken with unsynchronized *P. falciparum* 3D7 parasites using monoclonal antibody (5.1 mAb) that was kindly provided by Dr. David Cavanagh at the European Malaria Reagent Repository (www.malariaresearch.eu), and originally provided by Dr. Jana McBride (McBride et al., 1982).

Sample Preparation for Transcriptomics

3b1 and Dd2 parasites were synchronized by two consecutive sorbitol treatments timed 12 hr apart for three or more successive generations and allowed to recover for another parasite asexual blood stage cycle before initiating time point samplings. For each time course, cultures were progressively expanded to a volume of 250–300 ml at 3% hematocrit with 8%–10% of tightly synchronous populations of late *P. falciparum* schizonts. Parasites were inoculated in a bioreactor (Applikon, Foster City, CA, USA) in the presence of fresh erythrocytes and allowed to reinvade for the next 2–3 hr at 8% hematocrit. Gas and temperature conditions were monitored with a Bio Controller unit ADI 1030. The first time point of the time series was determined by the peak of invasion after which cultures were diluted to a final volume of 1,000 ml and 2% hematocrit. At least 90% of the parasites were in the early ring stage and in general, cultures reached a parasitemia of 10%–20%. Equal parasite samples were harvested every 8 hr throughout the 48 hr intra-erythrocytic life cycle. To increase microarray signal-to-noise ratios, samples were pretreated with Terminator 5′-Phosphate-Dependent Exonuclease (Epicenter, Madison, WI, USA) to degrade any RNA that does not have a 5′-triphosphate, 5′-cap or 5′-hydroxyl group and thus does not contribute to mRNA. Pellets of parasitized erythrocytes were directly stored in Trizol for subsequent RNA extraction and cDNA synthesis (Bozdech et al., 2003; Kyes et al., 2000) using 125U of Superscript II reverse transcriptase (Invitrogen). Each reaction was concentrated on a Zymo DNA clean and concentrator-5 column (Zymo Research, Irvine, CA, USA) and labeled with Cy5 dye (GE Healthcare, Piscataway, NJ, USA). The reference pool consisted of Cy3-coupled cDNA samples prepared from Dd2 RNAs representing all developmental stages at 8 hr intervals of the intra-erythrocytic life cycle. Equal amounts of labeled samples from each time point and reference pool were subjected to array hybridization for 16–18 hr at 65°C on a 60-mer *P. falciparum* microarray (Agilent Technologies, AMADID 029872) as previously described (Painter et al., 2013). Data were acquired using an Agilent G2505B scanner and analyzed with Agilent Feature Extractor v9.5.3.1 (Agilent Technologies, Santa Clara, CA, USA).

Transcriptomics

P. falciparum transcriptional profiles were reconstituted using the Fast Fourier Transform method to sort genes based on their expression peaks. To correct for variations in developmental stage speed between parasite lines, dynamic time warping (Aach and Church, 2001) was applied to determine the optimal post-invasion time point (corresponding to the first sampling time point) along the intra-erythrocytic life cycle, using the Pearson correlation coefficient between each pair of imputed points. Each parasite line was aligned using the high-resolution 3D7 intra-erythrocytic transcriptome data as reference (Bozdech et al., 2003; Llinás et al., 2006). The alignments were confirmed by applying principal component analysis. To assess gene expression-level differences between parasite lines, the area under the curve (AUC) of absolute gene expression plots was computed and the fold change was calculated as the difference between AUCs in pairwise comparisons. To determine significant variation between parasite lines, a cut-off of > 1.5-fold difference was assigned. Each time-course transcriptome was aligned to the 3D7 reference data, imputed and for each individual gene, expression fold change was normalized using the gene fold change averaged across all pairwise comparisons, with +1 corresponding to one standard deviation above the mean fold change.

Western Blot

P. falciparum asexual blood stage parasite strains cultured Dd2, 3b1, and 3D7 parasites were synchronized with 5% sorbitol, and harvested 48 hr post-synchronization at 90%–100% rings. Parasites were released from erythrocytes with 0.025% saponin, washed once with 1 × PBS, and then resuspended in ice-cold RIPA lysis buffer (RIPA buffer (Boston BioProducts, Boston, MA, USA) supplemented with 10 μg/ml pepstatin A, 2 mM orthophenantroline, 2 mM EDTA, pH 8.0, and 2× Complete EDTA-free Protease Inhibitor cocktail (Roche Applied Science, Indianapolis, IN, USA). Extracts were subjected to five freeze/thaw cycles. Total protein content was determined using the BioRad Protein Assay. Ten micrograms of protein of each sample was loaded on a pre-cast Criterion XT Bis-Tris Gel (BioRad). Separated polypeptides were transferred onto a PVDF membrane (Millipore), and subsequently incubated with monoclonal mouse anti-*Pf*EXP1 antibodies and polyclonal rabbit anti-*Pf*ERD2 antibodies (MR4), followed by incubations with the respective HRP-conjugated secondary antibodies. Polypeptides were visualized with chemiluminescence.

SUPPLEMENTAL REFERENCES

Aach, J., and Church, G.M. (2001). Aligning gene expression time series with time warping algorithms. Bioinformatics *17*, 495–508.

Falls, C., Powell, B., and Snoeyink, J. (2004). Computing high-stringency COGs using Turan-type graphs. CiteSeerX. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.107.8589.

Fidock, D.A., Nomura, T., and Wellems, T.E. (1998). Cycloguanil and its parent compound proguanil demonstrate distinct activities against *Plasmodium falciparum* malaria parasites transformed with human dihydrofolate reductase. Mol. Pharmacol. *54*, 1140–1147.

Fidock, D.A., Nomura, T., Talley, A.K., Cooper, R.A., Dzekunov, S.M., Ferdig, M.T., Ursos, L.M., Sidhu, A.B., Naudé, B., Deitsch, K.W., et al. (2000). Mutations in the *P. falciparum* digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. Mol. Cell *6*, 861–871.

Jensen, L.J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., and Bork, P. (2008). eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Res. *36* (Database issue), D250–D254.

Koncarevic, S., Bogumil, R., and Becker, K. (2007). SELDI-TOF-MS analysis of chloroquine resistant and sensitive *Plasmodium falciparum* strains. Proteomics *7*, 711–721.

Kyes, S., Pinches, R., and Newbold, C. (2000). A simple RNA analysis method shows *var* and *rif* multigene family expression patterns in *Plasmodium falciparum*. Mol. Biochem. Parasitol. *105*, 311–315.

Lisewski, A.M., and Lichtarge, O. (2010). Untangling complex networks: risk minimization in financial markets through accessible spin glass ground states. Physica A *389*, 3250–3253.

Llinás, M., Bozdech, Z., Wong, E.D., Adai, A.T., and DeRisi, J.L. (2006). Comparative whole genome transcriptome analysis of three *Plasmodium falciparum* strains. Nucleic Acids Res. *34*, 1166–1173.

McBride, J.S., Walliker, D., and Morgan, G. (1982). Antigenic diversity in the human malaria parasite *Plasmodium falciparum*. Science *217*, 254–257.

Mohseni-Zadeh, S., Brézellec, P., and Risler, J. (2004). Cluster-C, an algorithm for the large-scale clustering of protein sequences based on the extraction of maximal cliques. Comput. Biol. Chem. *28*, 211–218.

Montague, M.G., and Hutchison, C.A., 3rd. (2000). Gene content phylogeny of herpesviruses. Proc. Natl. Acad. Sci. USA *97*, 5334–5339.

Mostafavi, S., Ray, D., Warde-Farley, D., Grouios, C., and Morris, Q. (2008). GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome Biol. *9* (*Suppl 1*), S4.

Painter, H.J., Altenhofen, L.M., Kafsack, B.F.C., and Llinás, M. (2013). Whole-genome analysis of Plasmodium spp. utilizing a new Agilent Technologies DNA microarray platform. In Malaria: Methods and Protocols, 2nd edition (New York: Humana Press).

Tsuda, K., Shin, H., and Schölkopf, B. (2005). Fast protein classification with multiple networks. Bioinformatics *21* (*Suppl 2*), ii59–ii65.

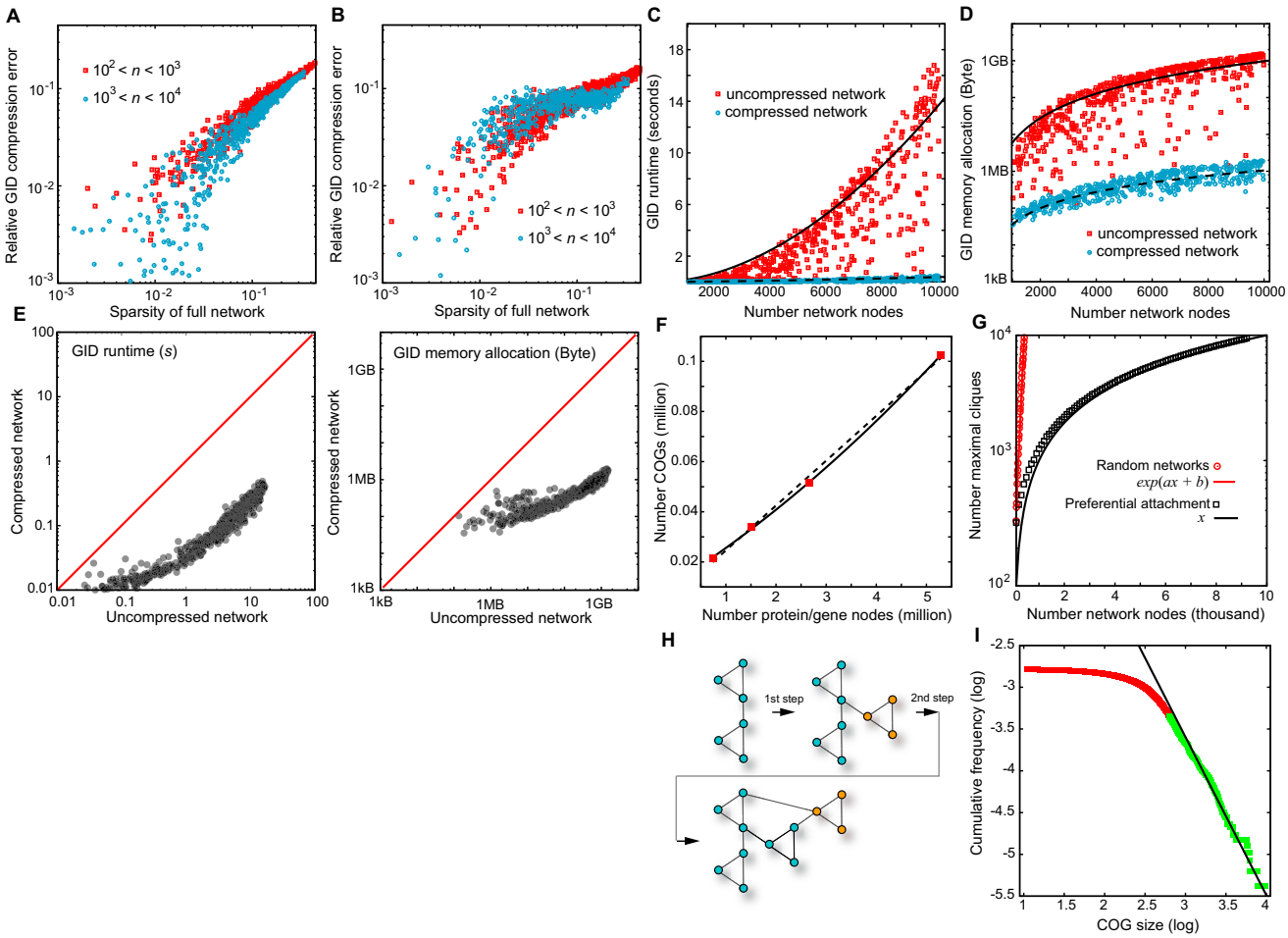

Figure S1. Network Compression and Compressible Network Model, Related to Figure 1

(A) Scatter plot of relative error in GID through compression after intra-COG network compression in relationship to initial edge density (*sparsity*) of randomly generated networks. Sparsity is defined as the ratio of given symmetric links over the maximum number $n\,(n-1)/2$ of links in a network with n nodes. In these random networks, COGs were modeled as fully connected network cliques with variable, random sizes. Given a number of random size cliques, the networks were produced by adding random links, defined as pairs of nodes chosen from a homogeneous distribution between nodes, until a desired sparsity level was reached.

(B) Scatter plot of relative error in GID after intra-COG and inter-COG compression. In both cases, sparser networks lead to smaller errors.

(C and D) Computational runtime (C) and memory allocation (D) for GID on random networks on full, uncompressed networks (red data points) against GID on compressed networks (intra- and inter-clique compression). Solid and dotted lines represent least square fits to data using the function $a\,n^b$, where n is the number of nodes in the network. For the uncompressed networks, the fit gives $a = 5 \times 10^{-7}$ and $b = 1.86$ (GID computer runtime), and $a = 1.40$ and $b = 2.21$ (memory allocation); for the compressed networks, it gives $a = 5 \times 10^{-7}$ and $b = 1.47$ (computer time), and $a = 1.40$ and $b = 1.47$ (memory allocation).

(E) Scatter log-plot of computational single CPU runtime (left panel) and memory allocation (right panel) of uncompressed (full) network on the x axis against compressed values on the y axis with the same data as in (C, D). Network compression always produced an advantage in both computational runtime and in memory demand.

(F) Number of identified COGs over total number of gene/protein nodes follows a linear trend; points (red boxes) correspond to intrinsic evolutionary network data from STRING database versions 6.2, 7.1, 8.2, and 9.0; least-square fit to a quadratic polynomial with $a + bx + cx^2$ returned $a = 1.1 \times 10^{-2}$, $b = 2.2 \times 10^{-2}$ and $c = 8.2 \times 10^{-4}$ (solid line); dashed line least-square fit to linear function.

(G) Networks evolving by preferential attachment have a linear number of maximal cliques in the number of network nodes (black open boxes, black solid graph); in contrast, for Erdős-Rényi random networks, this number increases exponentially (red open circles; red solid graph least square fit function $\exp(ax + b)$ with $a = 0.01$ and $b = 5.07$. These random networks had the same number of links as those networks generated through preferential attachment; links in random networks were generated as Erdős-Rényi random graphs by choosing pairs of random nodes (from a uniform distribution) until the target number of links was reached. Maximal cliques calculated with the Bron-Kerbosch algorithm in a *Matlab* (version 7.1) computer program implementation.

(H) Graphical representation of the first two steps of clique-based preferential attachment with two initial cliques of size three.

(I) Cumulative distribution of clique sizes of the supergenomic network follows a power law with exponent 1.94 ± 0.03; this means that the size distribution of cliques/COGs in the supergenomic network follows a power law with $\gamma = 2.94 \pm 0.03$; solid line depicts least square fit power law with exponent 1.94 over a range indicated by green data points.

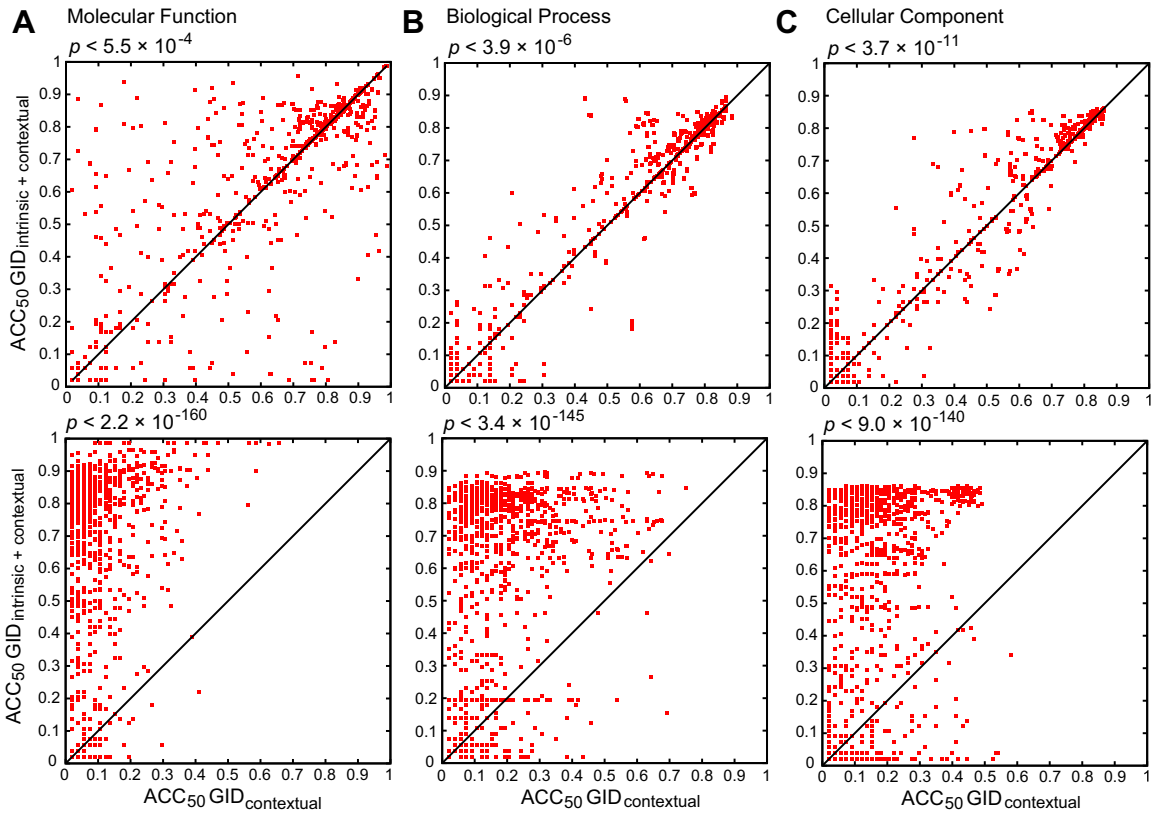

Figure S2. Computational Validation of Supergenomic Network Predictions, Related to Figure 2

(A) Accuracy (ACC_{50}) in a series of one thousand leave-one-out experiments evaluated over all three GO categories.

(B and C) In the Molecular Function category, GID on the combined network (compressed intrinsic and contextual on the *y*-axes in panels) was significantly more accurate (p values from Wilcoxon paired signed ranks test) than both the intrinsic and the contextual network alone (*x* axes in panels). As demonstrated by accuracy values on the *x* axes, both intrinsic and contextual networks were individually significant for predicting GO terms (Wilcoxon paired ranks test resulted in $p < 4.2 \times 10^{-121}$ when compared against random predictions). Test sets were the SwissProt 1,000 protein sequences set annotated by Molecular Function terms with full EC codes, and a new set of equal size annotated by the GOSlim subset of GO in terms of (B) Biological Process and (C) Cellular Component category.

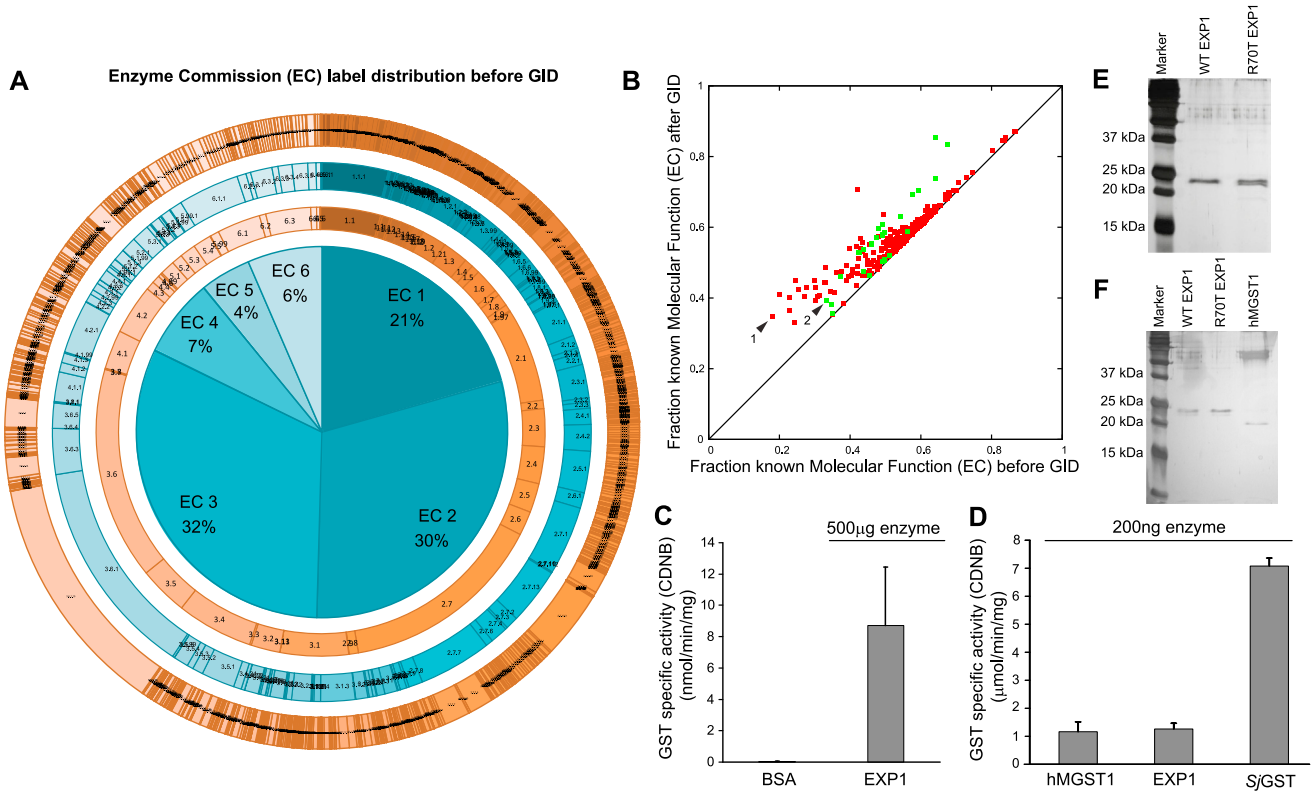

Figure S3. Supergenomic Network Prediction and Experimental Validation of Function, Related to Figure 3

(A) Functional distribution of initial input 315,692 GID labels for given EC numbers in the Molecular Function category before competitive GID. A detailed list of the output functional predictions after GID is given in Table S1.

(B) Scatter plot of fractions of all gene products, before GID and after GID, with preannotated input and additionally predicted output Molecular Functions (EC), respectively. Each point represents the genome/proteome of one species with coordinates that are the fraction of the whole genome that have EC Molecular Function labels (x axis), and the fraction of the whole genome that have EC Molecular Function labels but including GID predictions (y axis); GID predictions only with *Z* scores above 2 were counted; red boxes indicate archaea and prokaryotes, green boxes eukaryotes. The two extreme cases (arrows) which before GID had the smallest fraction of genes annotated with Molecular Function EC codes are (1) *Bacillus anthracis Sterne* with 18% of 5,237 annotated genes (among archaea and bacteria), and (2) *Plasmodium falciparum* 3D7 with 33% of 5,135 (among eukaryota).

(C) GST specific activity for BSA (negative control with 2,152 µg total protein) and *P. falciparum* EXP1 resulting from spectrometry data in Figure 3E in the main text with average EXP1 protein yield of 500 µg.

(D) Comparison of specific activities of EXP1 (1.22 ± 0.43 µmol/min/mg) to *E. coli* expressed and purified human microsomal GST hMGST1 (1.09 ± 0.39 µmol/min/mg) and to *S. japonicum* cytosolic GST (7.08 ± 0.37 µmol/min/mg). Mean values and standard errors from at least three experimental replications.

(E) SDS-PAGE silver stain of purified WT and mutant (R70T) EXP1 protein.

(F) SDS-PAGE silver stain of purified WT EXP1, R70T EXP1, and human MGST1 protein. MGST1 has a predicted molecular weight of 17.6 kDa.

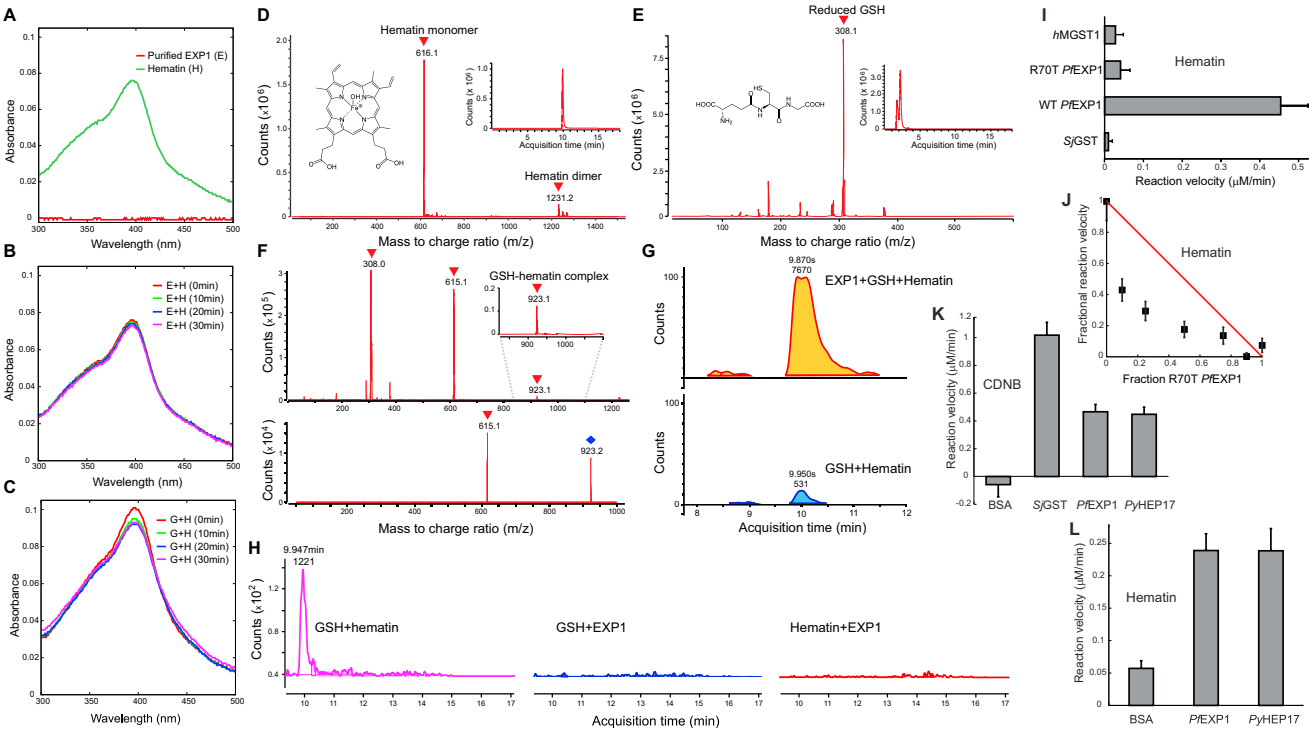

Figure S4. Hematin Is a Substrate of EXP1, Related to Figure 4

(A) A flat absorbance spectrum around 400 nm indicates that no endogenous heme or hematin from the *E. coli* expression host was purified along with EXP1 protein (*E*); addition of exogenous hematin (*H*) leads to a characteristic Soret peak at 395 nm.

(B) EXP1 together with hematin does not lead to a measurable decrease of absorbance at the Soret peak over time; see also panel (H) for corresponding mass spectrometry data.

(C) Spontaneous reaction between reduced glutathione (*G*) and hematin is monitored through a reduction of peak absorbance.

(D) Mass spectrometry detection of hematin at mass-to-charge ratio of *m/z* 616.1 for the ion $[Hematin+H]^+$; inset shows the corresponding peak at 9.9 min acquisition (retention) time which indicates chromatographic separation from the analytical column in liquid chromatography mass spectrometry (LC-MS, see Experimental Procedures).

(E) Mass spectrometry detection of reduced glutathione (GSH) at *m/z* 308.0 $[GSH+H]^+$ and corresponding LC-MS retention time profile (inset).

(F) Detection of the hematin-GSH product complex at *m/z* 923.1, which was consistent with the sum of the individual components; electrospray ionization control of the selected product peak (diamond) confirmed that its composition included hematin as signalized by the secondary hematin peak at *m/z* 615.1 $[Hematin]^+$.

(G) Acquisition time profiles at ~9.9 min were used to measure the relative abundance of the reaction product: in comparison to the reaction without the GST enzyme, the addition of EXP1 resulted in an increase of hematin-GSH product by a factor of (7670/531 − 1) = 14.4.

(H) Comparison of product accumulation detected through LC-MS at 9.9 min retention time in the presence of GSH and hematin (left panel) to two negative controls with GSH and EXP1 (middle panel) as well as with hematin and EXP1 (right panel) showed no detectable product in both negative cases.

(I) Initial reaction velocity of in vitro GST activity toward hematin for four different enzymes: three integral membrane GSTs (*h*MGST1, R70T *Pf*EXP1, WT *Pf*EXP1) and one cytosolic GST (*Sj*GST).

(J) Cooperative inhibition of WT PfEXP1 GST activity toward hematin through R70T indicates that multiple EXP1 units form a functional GST complex in vitro. In contrast, a noncooperative interaction between R70T and WT EXP1 would result in fractional reaction velocities along the solid line. Residual R70T EXP1 activity subtracted as blank.

(K and L) Reaction velocity comparison between the EXP1 *Plasmodium yoelii* ortholog *Py*HEP17 against other membrane (*Pf*EXP1) and cytosolic (*Sj*GST) GSTs, measured for both CDNB (K) and hematin (L). Negative controls with BSA, where in (L) the residual activity was due to the spontaneous conjugation of GSH and hematin. Protein concentration was 10 nM in 1 ml volumes in all cases. Mean values with standard errors from at least three replications.

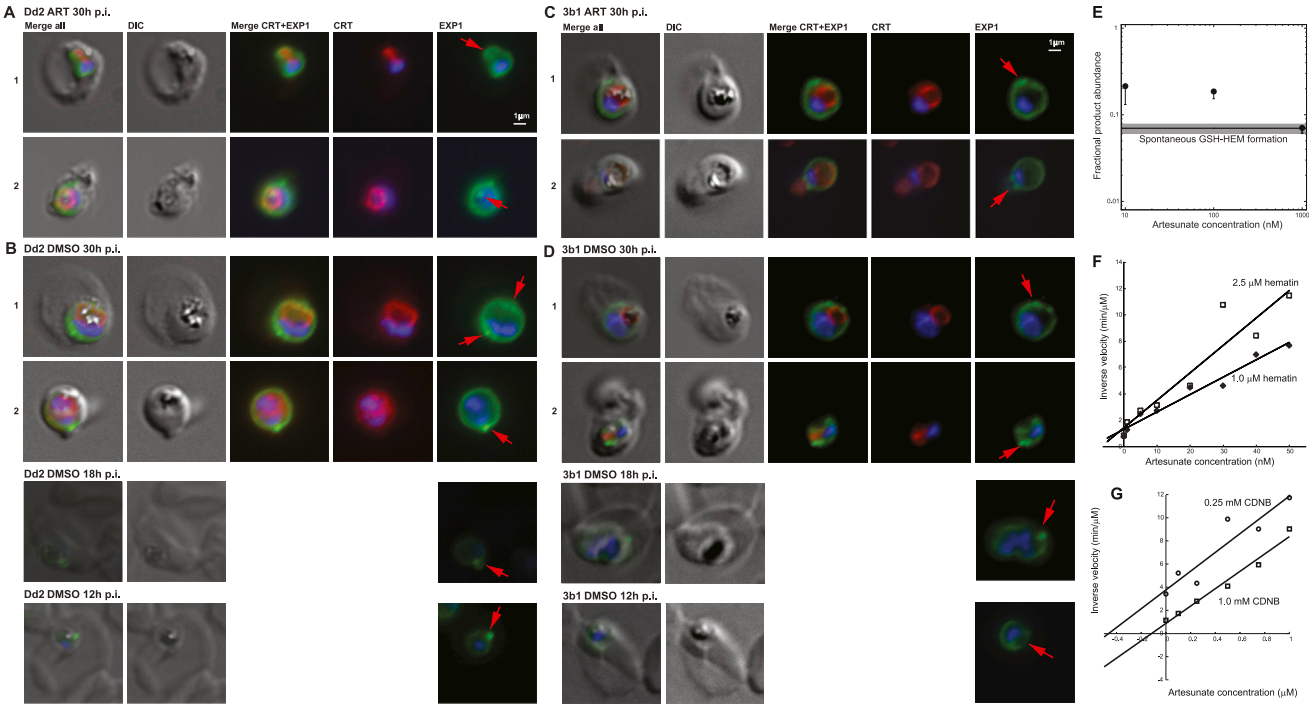

Figure S5. Cellular Localization and Inhibition of EXP1, Related to Figure 5

(A) Trophozoite stage Dd2 parasites (starting ~24 hr post-invasion, p.i.) were exposed to 3 × IC_{50} concentration of ART for 6 hr, revealing a spherical and peripheral EXP1 expression signal that appeared to correspond to a cytostomal invagination (A1) and to an internalized compartment/vesicle with localized EXP1 expression (A2).

(B) Unchallenged parasites (6 hr exposure to solvent DMSO starting 24 hr post-RBC invasion) display similar morphology and EXP1 expression patterns (peripheral, partly internalized foci, largely independent of the CRT signal) to the drug-exposed parasites at 30, 18, and 12 hr post-invasion.

(C) Trophozoite stage 3b1 parasites (starting ~24 hr post-invasion) were exposed to 3 × IC_{50} concentration of ART for 6 hr, revealing a spherical and peripheral EXP1 expression signal that appeared to correspond to a cytostomal invagination (A1) and to an internalized compartment/vesicle with sharply localized EXP1 expression (A2).

(D) Unchallenged parasites (6 hr exposure to solvent DMSO starting 24 hr post RBC invasion) display similar morphology and EXP1 expression patterns (peripheral, partly internalized foci, largely independent of the CRT signal) to the drug-exposed parasites at 30, 18, and 12 hr post-invasion.

(E) ART-mediated inhibition of glutathione(GSH)-hematin(HEM) product formation measured through LC-MS.

(F) Dixon plot for ART inhibition of EXP1 GST activity toward hematin. Intersection of fitted lines at −1.5 nM.

(G) In the absence of hematin, Dixon plot for EXP1 GST activity toward CDNB inhibited by ART indicates a weaker and uncompetitive inhibition with IC_{50} = 184 nM.

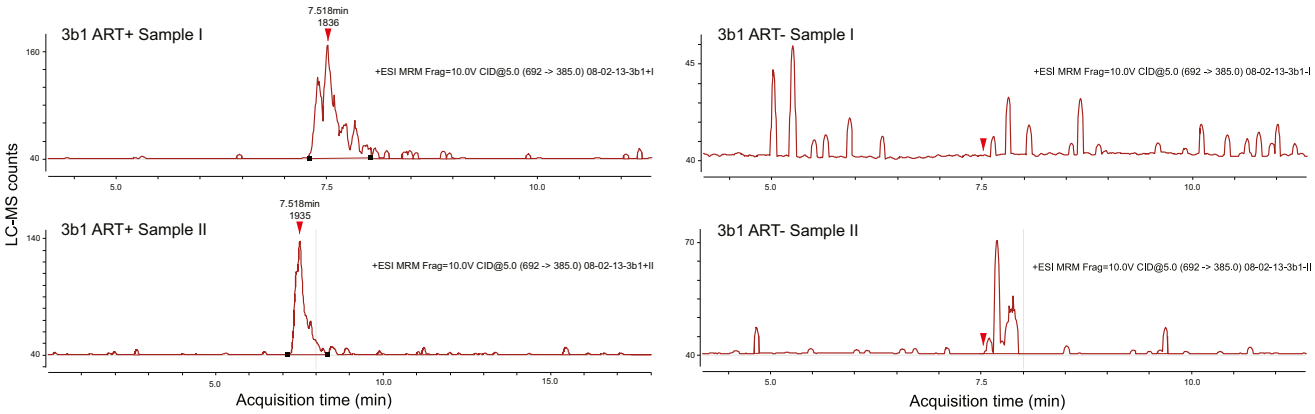

Figure S6. EXP1 Is Associated with ART Metabolism In Vivo, Related to Figure 6

LC-MS retention time peaks (7.518 min, red arrows) as reference points in abundance measurements from the analytical column in 3b1 parasite samples with (ART^+) and without (ART^-) drug exposure (see Experimental Procedures). Two independent samples were prepared for each of the two drug conditions (ART^+, ART^-).