

## In Silico Models for Drug Resistance

Segun Fatumo, Marion Adebiyi, and Ezekiel Adebiyi

### Abstract

Resistance to drugs that treat infectious disease is a major problem worldwide. The rapid emergence of drug resistance is not well understood. We present two in silico models for the discovery of drug resistance mechanisms and for combating the evolution of resistance, respectively. In the first model, we computationally investigated subgraphs of a biological interaction network that show substantial adaptations when cells transcriptionally respond to a changing environment or treatment. As a case study, we investigated the response of the malaria parasite *Plasmodium falciparum* to chloroquine and tetracycline treatments. The second model involves a machine learning technique that combines clustering, common distance similarity measurements, and hierarchical clustering to propose new combinations of drug targets.

**Key words** In silico, Drug, Resistance, Model, Mechanism

---

## 1 Introduction

Controlling infectious diseases is becoming more difficult as a result of the emergence of resistance to available drugs on the market. Drug resistance has emerged in the most dangerous diseases affecting humans, including malaria, tuberculosis, and HIV infection. These diseases have increased the disease burden particularly in developing countries, especially in Africa.

In this report, we present two in silico models, one for the discovery of drug resistance mechanisms and another for combating the evolution of drug resistance. Although we have adapted and developed these models for malaria research, they can be employed in the study of other infectious diseases. The first model has not been previously published. A model similar to our second model has been developed for the treatment of gastrointestinal stromal tumor (GIST) (1). With tumors as heterogeneous as GIST, up to five different types of secondary mutations can occur in the same patient. The aim is not only to wait for mutations to emerge before selecting the right compound but also to predict and group mutations according to likelihood, enabling clinicians to prescribe an appropriate drug as soon as a patient displays a particular mutation.

This dynamic multidrug-targeted prevention technique has been proposed in the treatment of chronic myeloid leukemia and the positive results obtained with the newly introduced drugs nilotinib and dasatinib suggested that a combination of two or three kinase inhibitors, when carefully selected to cover all known resistant mutations, could shut off all mechanisms of escape.

### **1.1 *In Silico* Modeling**

An improved knowledge of genomics and of the structure of individual proteins has helped us increase our understanding of biological systems. However, insight into functional interactions between the key components of cells, organs, and systems helps us in understanding their physiology. Perturbations in these interactions lead to various diseases. We therefore must compute these interactions to determine the characteristics of the system when it changes from the healthy to the diseased state. With the development of powerful computing hardware and algorithms and an increasing number of pathway databases and models of cells, tissues, and organs, we can now explore functionality in a mathematical manner from the level of genes to the physiological function of whole organs and regulatory systems (2). The simplified mathematical representation of the dynamics of a system is called modeling (3). Modeling has become an important research area in biology and bioinformatics. We use models to explain experimental observations. Hence, we can also use them to test a hypothesis about biological function. We also use models for storing experimental data on biological molecules and processes in databases so as to analyze them (4). While modeling of individual reactions has been under way for a long time, we have only recently begun to appreciate the importance of modeling complex reactions, biochemical pathways, and networks (5). Because experimental data on biochemical reactions are insufficient and difficult, expensive, and time consuming to obtain, computational models of biological networks help in filling this data gap. We use computational models both for simulation and for metabolic engineering (2). Using computational simulation of complex biological networks, we can not only validate the conclusions drawn by experimental studies but also propound fresh hypotheses for further experimental validation. This iterative process of experimental studies and computational simulation has helped us develop highly sophisticated and realistic models, e.g., models of heart cells (6).

---

## **2 Materials and Concepts**

### **2.1 *DNA Microarray***

The advent of DNA microarray high-throughput profiling experiments has allowed us to explore a major subset or all genes of an organism under a variety of conditions such as alternative treatments (drug-influenced condition vs. condition influenced by

factors considered normal), mutants, developmental stages, and time points. For example, the technique enables us to classify tumor samples (7), to define small sets of potential marker genes to distinguish leukemia (8), and to discover regulatory mechanisms (9, 10). For example, without prior information, the structure and function of the network that regulates the SOS pathway in *Escherichia coli* was elucidated via transcription profiles (11).

## **2.2 Biochemical Metabolic Network**

Biochemical investigations especially in the past 40 years have revealed an increasingly consistent image of cellular metabolism; see, for example, Berg et al. (12). This is especially true for less complex organisms such as *E. coli* (13). However, this approach used alone provides a rather static image of the cell and thus investigations have been performed to discover cellular adaptation programs in response to changing environments such as nutrient excess, starvation, and other stresses (14). These observations originally followed linear interaction and reaction cascades; studies investigated single knockouts and tediously tracked transcripts for single genes, compounds, and proteins that might be influenced; see, for example, Neidhardt (15). By combining metabolic network data and microarray data, data on the physical and chemical interactions of proteins can be integrated. For example, knowledge of protein–protein interaction gained from the use of high-throughput techniques (16) applied to the analysis of gene expression data revealed novel regulatory circuits (17). Moreover, knowledge of biochemical network interactions has been used to support the clustering procedure for gene expression profiles of yeast (18, 19).

## **2.3 BioCyc: A Collection of Biochemical Pathway Databases**

BioCyc (20) is a collection of more than 200 pathway/genome databases, containing whole databases dedicated to certain organisms. For example, EcoCyc, which falls under the giant umbrella of BioCyc, is a highly detailed bioinformatics database on the genome and metabolic reconstruction of *E. coli*, including thorough descriptions of various signaling pathways. The EcoCyc database can serve as a paradigm and model for any reconstruction. Additionally, MetaCyc, an encyclopedia of metabolic pathways, contains a wealth of information on metabolic reactions derived from more than 600 different organisms, including *Plasmodium* and *Homo sapiens*.

## **2.4 Pathway Tools**

Pathway Tools is a bioinformatics package that assists in the construction of pathway/genome databases such as EcoCyc (21). Developed by Peter Karp and his associates at the SRI International Bioinformatics Group, Pathway Tools comprises several separate units that work together to generate new pathway/genome databases (20). First, PathoLogic takes an annotated genome of an organism and infers probable metabolic pathways, allowing the creation of a pathway/genome database for the organism.

Pathway Hole Filler can then be applied to predict likely genes to fill “holes” (missing steps) in predicted pathways. Thereafter, the Pathway Tools Navigator and Editor functions let users visualize, analyze, access, and update the database. Thus, by using PathoLogic and encyclopedias such as MetaCyc, an initial fast reconstruction can be developed automatically, and then, using the other units of Pathway Tools, a detailed manual update, curation, and verification step is possible.

---

### 3 First Model: In Silico Model for Deducing Drug Resistance Mechanisms

In the first model, we sought to reveal subgraphs of a biological interaction network that show substantial adaptations when cells transcriptionally respond to a changing environment or treatment. As a case study, we investigated the response of the malaria parasite *Plasmodium falciparum* to chloroquine and tetracycline treatments.

This work was designed to unveil the mechanisms that culminated in the widespread resistance of this parasite to these drugs. We hope that our results will be useful in developing combinations of antiresistance drugs for malaria patients. Simple clustering of gene expression on the metabolic network of *P. falciparum* can yield subgraphs (clusters or features) that are either stimulated or repressed when the organism attempts to resist a particular treatment given to a malaria patient. König and Eils (22) and König et al. (23) demonstrated a similar mechanism with tryptophan-treated cells and in the heterofermentative bacterium *E. coli* in response to oxygen deprivation (24).

Following this line of work, we made the discoveries reported here. Before we indicate these, we note that the microarray datasets that we analyzed for tetracycline and chloroquine do not contain many differentially regulated reactions. The possibility remains that studying drug resistance mechanisms of the malaria parasites at the transcriptional level of their proteins is not reliable (Karine Le Roch, personal communication).

Using the tetracycline microarray data, Dahl et al. (25) indicated that tetracyclines specifically block expression of the apicoplast genome and concluded that the loss of apicoplast function in the progeny of treated parasites leads to a slow but potent antimalarial effect. From the clusters we extracted, we show that this slow antimalarial effect is due in particular to excess glucose that is being made available. The fatty acid production is upregulated (beta oxidation, starting at acetyl-coA) together with the farnesyl pathway, which is needed for cholesterol and also leads to fatty acids and membrane components. We also discovered important genes and reactions that participated in the resistance mechanism of *P. falciparum* to tetracycline.

From the chloroquine microarray data, we found that tryptophanyl-tRNA synthetase production in the apicoplast is upregulated. Others have hypothesized that resistant *P. falciparum* parasites have a mechanism for releasing chloroquine via an efflux process (26, 27). We prove in this work that the upregulated tryptophanyl-tRNA synthetase production in the apicoplast suggests that this efflux process may have been made possible (caused) by the apicoplast, the mini bacterium living inside the malaria parasite. We hypothesize that when our results are experimentally proved, in particular for the case of chloroquine, our findings may lead to better and more cost-effective agents for eradication of the parasite from the human blood stream.

### 3.1 Gene Expression Data Used

Serial Analysis of Gene Expression (SAGE) tags of chloroquine-treated cells were obtained from the work of Gunasekera et al. (24) and the microarray data were obtained from Gunasekera et al. (28). Data from the microarray response to tetracycline treatment were taken from Dahl et al. (25). Chloroquine is designed to inhibit the parasitic enzyme heme polymerase and tetracycline is designed to inhibit the cytosolic ribosomes. Additionally, Dahl et al. (25) showed the antimalarial effect of chloroquine against the apicoplast genome of *P. falciparum*. In the following discussion, “chloroquine drug influence” refers to the microarray data on chloroquine treatment vs. control (in cases when this is not so, we will explicitly state this).

In Gunasekera et al. (28), the parasite culture preparation and RNA preparation/hybridization were done as follows. Blood-stage *P. falciparum* parasites were maintained in vitro at 37°C in RPMI 1640 (Roswell Park Memorial Institute) medium (Invitrogen, Carlsbad, CA) containing 25 mM HEPES, 0.2% sodium bicarbonate, 50 µg/mL hypoxanthine, 25 µg/mL gentamicin, 5% heat-inactivated human O<sup>+</sup> serum, 5% bovine serum albumin (Albumax II, Invitrogen), and 5% human O<sup>+</sup> blood, following standard protocols (29). 3D7 strain parasites were used for all experiments. Mixed-stage 3D7 parasites were treated with 120 and 400 nM chloroquine for 30 min and 6 h, alongside matched controls, yielding six samples (0 nM—30 min, 120 nM—30 min, 400 nM—30 min, 0 nM—6 h, 120 nM—6 h, and 400 nM—6 h). Two separate starting cultures at 8% parasitemia but with different stage profiles were subjected to each of the six treatments. The first consisted of approximately 1.7% rings, 2.5% early trophozoites, 3.4% late trophozoites, and 0.15% schizonts, and the second contained 1.7% rings, 5.4% early trophozoites, 0.6% late trophozoites, and 0.15% schizonts. Hence a total of 12 different cell states, representing parasites under varying drug concentrations (three), drug exposures (two), and staging profiles (two), were assayed. Total RNA was harvested at the end of each time point using the Tri-Reagent BD protocol (Molecular Research Center, Cincinnati, OH), labeled by a strand-specific protocol and hybridized to a custom-made

high-density oligonucleotide array containing 260,596 25-mer probes from a predicted coding sequence of the parasite genome and 106,630 probes from a noncoding sequence (30). Probes mapping to coding sequences were used to compute gene expression levels by means of the match-only integral distribution algorithm (MOID) (31). We normalized the expression data using an established variance normalization method (32).

The malaria parasites preparation, culture, and microarray analysis by Dahl et al. (25) were performed using the following setup. *P. falciparum* parasites were cultured in human erythrocytes maintained at 2% hematocrit in RPMI 1640 medium with 0.5% (wt/vol) bovine serum albumin in 92% N<sub>2</sub>, 5% CO<sub>2</sub>, and 3% O<sub>2</sub>. Synchrony was maintained by serial sorbitol treatments. Strain 3D7 was used here. Parasites stably expressing green fluorescent protein fused to an acyl carrier protein apicoplast-targeting sequence (ACP<sub>1</sub>-GFP), kindly provided by Geoff McFadden (33), were maintained in medium containing 100 nM pyrimethamine. Dually transfected parasites stably expressing a red fluorescent protein fused to an acyl carrier protein apicoplast-targeting signal and a yellow fluorescent protein fused to a citrate synthetase mitochondrial targeting signal (ACP<sub>1</sub>-DsRed and CS<sub>1</sub>-YFP), also kindly provided by Geoff McFadden (34), were maintained in medium containing 5 nM WR99210.

Synchronized parasites were treated at the late ring/early trophozoite stage (approximately 20 h postinvasion) with 1 μM doxycycline or an equivalent volume of dimethyl sulfoxide for 24 h, until they reached the late schizont stage. The parasites were then subcultured and maintained in drug-free medium for an additional 35 h. Infected erythrocytes were collected every 5 h, lysed with 0.1% saponin for 5 min, centrifuged at 12,000 × *g* at 4°C, flash-frozen in an ethanol–dry ice bath, and stored at –80°C. Total parasite RNA was harvested using TRIzol reagent (Invitrogen). For each sample, 12 μg of total parasite RNA was reverse transcribed into cDNA containing amino-allyl-dUTP (Ambion, Invitrogen) using SuperScript II RNase H-Reverse Transcriptase (Invitrogen) and then coupled to succidimyl ester Cy5 dye (Amersham, GE Healthcare, Chalfont St. Giles, UK), as described previously (35). Cy5-labeled sample cDNA and a reference pool of Cy3-labeled cDNA representing all life cycle stages were competitively hybridized to a *P. falciparum* 70-mer microarray as described by Bozdech et al. (36). The microarrays were scanned using a GenePix 4000B scanner, and images were analyzed using GenePix3 Software (Molecular Devices, Sunnyvale, CA), stored, and normalized using the NOMAD database (<http://ucsf-nomad.sourceforge.net>). Expression data were log transformed and mean centered.

### **3.2 Mapping of SAGE Tags to Genes**

We mapped all SAGE tags to the genes they represented as follows. We used the standalone Blast from NCBI (<ftp://ftp.ncbi.nlm.nih>).

[gov/blast/executables](http://blast.ncbi.nlm.nih.gov/blast/executables)) and the databases of coded regions of *P. falciparum* and blasted all SAGE tags against all open reading frames selecting only the perfect matches.

### **3.3 Model for Analyzing Gene Expression Data on Metabolic Networks**

To analyze the gene expression data above on the metabolic network of *P. falciparum*, we used the following computational pipelines:

1. Construction of the metabolic network from Plasmocyc obtained from BioCyc.
2. Network clustering using a simulated annealing and a Kernighan–Li clustering procedure.
3. Mapping gene expression data onto the reactions.
4. Feature extraction using a combinatorial approach.
5. Analysis of stimulated and repressed pathways.

We elaborate on these in the following sections. The pipelines explored here have been used by König et al. (23), but these investigators used another feature extraction technique, the Haar wavelet transform. We explored a novel feature extraction technique based on a combinatorial approach. We confirm further the results obtained via the pipelines above using the Haar wavelet transform. This transform was done using the clusters due to the consecutive-ones clustering technique (23).

### **3.4 Construction of the Metabolic Network from Plasmocyc (BioCyc)**

We constructed our network from the metabolic reaction database Plasmocyc. The metabolites were taken as nodes. Two metabolites were connected by an edge if an enzymatic reaction existed that had them as an educt or product, respectively (23). We discarded highly connected metabolites such as water, CO<sub>2</sub>, and adenosine triphosphate. These metabolites are needed in many reactions and are therefore unspecific in the metabolic network.

### **3.5 Network Clustering Using Kernighan–Li and Simulated Annealing Algorithms**

Here we describe how the network given above will be clustered to group enzymes into parts of the network with their major connections. Formally, given the metabolic network as graph  $G(V,E)$  with node set  $V$  (metabolites) and edge set  $E$  (reactions), the goal of our clustering here is to identify clusters of  $G$  where each cluster was given by the node set of a highly connected subgraph. Note that the clusters are not required to be mutually disjoint.

For the network clustering problem, we used both the simulated annealing and Kernighan–Li (37–39) algorithms. We then applied our feature extraction technique, the combinatorial approach on both clusters obtained via these algorithms. The idea is that if the clusters are similar (similarly ranked) in both results, this will help confirm our findings.

In the following, we explain each clustering technique briefly. We adapted these algorithms to cluster the metabolic network described above. For more details, readers should see Kernighan

and Li (38), Dutt (39), and Brown and Huntley (37). The Kernighan–Li algorithm was designed to solve the following combinatorial problem: given a graph  $G$  with costs on its edges, partition the nodes of  $G$  into subsets no larger than a given maximum size, so as to minimize the total cost of the edges cut. We explain the two-way uniform partition of  $G$  using Kernighan and Li; its application in performing multiple-way partitions (as we did in this work) is achieved using the two-way procedure that allows us to partition into unequal-sized sets. Formally, let  $G(V, E)$  be a graph with node set  $V(G)$  and edge set  $E(G)$ , where there is a positive cost  $c(\{v_i, v_j\})$  associated with every edge  $\{v_i, v_j\} \in E(G)$  that may, for example, represent the width of the corresponding link. The problem is to partition  $V(G)$  into partitions  $P_1$  and  $P_2$  so that  $-1 \leq |P_1| - |P_2| \leq 1$ , and the cost of the cut-set  $\sum c(\{v_i, v_j\})$  is minimized, where  $v_i$  and  $v_j$  belong to different partitions. The resulting effect of this partitioning is that nodes that are densely connected to each other are placed near each other.

Let us take the following notations. Given a partition  $P_1$  and  $P_2$  of  $V(G)$ , for each  $u \in V(G)$ , let us define the external cost  $E_u$  and the internal cost  $I_u$  of  $u$  as follows:

$$E_u = \sum_{v \in P_i} c(\{u, v\}), \text{ where } i = 1, 2, u \text{ does not belong to } P_i, \text{ and } \{u, v\} \in E(G)$$

$$I_u = \sum_{v \in P_i} c(\{u, v\}), \text{ where } i = 1, 2, u \text{ does belong to } P_i, \text{ and } \{u, v\} \in E(G)$$

We define the  $D$  value of node  $u$  as  $D_u = E_u - I_u$ . This is the gain (reduction in the cost of the cut-set) obtained by moving  $u$  from its current partition. Thus if  $u \in P_1$  and  $v \in P_2$ , then it is easy to see that the gain  $G_{u,v}$  associated with swapping the pair of nodes  $(u, v)$  is  $D_u + D_v - 2c(\{u, v\})$  if  $\{u, v\} \in E(G)$  and  $D_u + D_v$  otherwise.

Assume that there are  $n = 2m$  nodes in  $G$ , and the initial partitions are  $P_1$  and  $P_2$ , with  $|P_1| = |P_2| = m$ . Let  $P_1 = \{u_1, u_2, \dots, u_m\}$  and  $P_2 = \{v_1, v_2, \dots, v_m\}$ . A main data structure used in the Kernighan and Li algorithm is the symmetric cost matrix  $C$ , where  $C_{u,v} = c(\{u, v\})$  if  $\{u, v\} \in E(G)$  and  $C_{u,v} = 0$  otherwise. First the  $D$  value of each node  $u$  is computed using  $C$ . Then, that pair of nodes  $(u_{i1}, v_{j1})$  is chosen for swapping that has the maximum value of  $G_{u_i, v_j}$ . Node  $u_{i1}$  is removed from  $P_1$ ,  $v_{j1}$  is removed from  $P_2$ ,  $(u_{i1}, v_{j1})$ , and inserted in an ordered set  $S$  of node pairs, and the  $D$  value of each node  $u$  is updated to reflect the fact that the pair  $(u_{i1}, v_{j1})$  has been swapped between the partitions. This procedure is iterated  $m$  times until  $P_1$  and  $P_2$  become empty, with the node pair  $(u_{ik}, v_{jk})$  inserted in  $S$  in the  $k$ th iteration,  $1 \leq k \leq m$  to give  $S = [(u_{i1}, v_{j1}), (u_{i2}, v_{j2}), \dots, (u_{im}, v_{jm})]$ . All partial sums  $S_k = \sum_{t=1, \dots, k} G_{u_{it}, v_{jt}}$  are computed, and  $p$  is chosen such that the partial sum  $S_p$  is the maximum. The sets of node pairs that are actually swapped are then  $\{(u_{i1}, v_{j1}), \dots, (u_{ip}, v_{jp})\}$ , such that the maximum gain  $G = S_p$  is obtained. This whole process is called a pass. A number of passes are made until the maximum gain  $G$

obtained is 0. This is a local maxima with respect to the initial partitions  $P_1$  and  $P_2$ . Empirical evidence shows that the number of passes required to achieve a local maxima is 2–4.

The simulated annealing algorithm (37) for the partitional clustering as we required in this work was designed based on the following problem formation.

Let

$Q$  be the set of all objects to be clustered (here, metabolites),

$n=|Q|$  be the number of objects in  $Q$ ,

$k \leq n$  be the maximum number of clusters,

$P = \{p: \text{for every } i \in \{1, \dots, n\}, p_i \in \{1, \dots, k\}\}$  be the set of all partitionings,

$J: P \rightarrow R$  be the internal clustering criterion;

Then

Minimize  $J(p)$  (here, based on minimizing the total cost of the edges cut)

Subject to

$p \in P$

The algorithm requires the perturbation operator  $\delta$  and the annealing schedule  $(\text{MaxIt}, T_o, \alpha, T_f)$ . The perturbation operator for partitional clustering switches a randomly chosen object  $i$  in  $Q$  from one cluster to another randomly chosen cluster. A set  $L$  contains the cluster labels used in  $\mathbf{p}$ . Similarly,  $L^c$  contains the labels not used in  $\mathbf{p}$ . The switching procedure first selects an integer  $m$  in the range  $[0, |L|]$ . If  $m$  equals 0 and there exists an unused cluster label (i.e.,  $|L| < k$ ), then object  $i$  is placed in its own singleton cluster. Otherwise,  $i$  switches to another, existing cluster. The computational effort is made fair by allowing each run a fixed number of trial perturbations. The total number of perturbations tried in any run is  $\text{MaxIt} \cdot \text{NumTemp}$ , where  $\text{MaxIt}$  is a fixed multiple of the number of objects to be clustered and  $\text{NumTemp}$  is a user-defined constant. The solution is made accurate using a very conservative annealing schedule (40, 41).

### **3.6 Mapping Gene Expression Data onto Reactions and Feature Extraction**

To do the mapping, each reaction documented in the metabolic network created above is linked to the gene(s) that produced the enzymes that catalyze it. This way, the gene expression values obtained from the microarray experiment (for each time point) replace the corresponding reaction of the metabolic network. We take the average reading for the case where a reaction is catalyzed by more than one gene.

The features extraction, i.e., the discovery of clusters whose genes are differentially expressed, in particular within different time points of the microarray under the control condition (no drug)

and the drug-influenced condition, is carried out using a combinatorial approach. We explain this below.

In the combinatorial approach we developed here, all possible combinations of sums and differences of expression values in each cluster are calculated. Note that we do not need all combinations, only half, because the other half can be obtained by the multiplication of  $-1$  with one-half. We explain the process further using a small example: If we have a cluster with three reactions and we have already mapped the expression values to the corresponding genes of the reactions, let these expression values be 1, 2, and 3. Then we will have four possible combinations, namely  $+1 +2 +3$ ,  $+1 +2 -3$ ,  $+1 -2 +3$ , and  $+1 -2 -3$ . Next, if we compared (subtracted) all combinations, the largest difference would be taken as  $P$  value and the clusters are ranked according to their  $P$  values. In the actual sense, the rationale behind choosing the largest  $P$  value is that it indicates the best probability that exists for the group of genes in the cluster in question not to be differentially expressed.

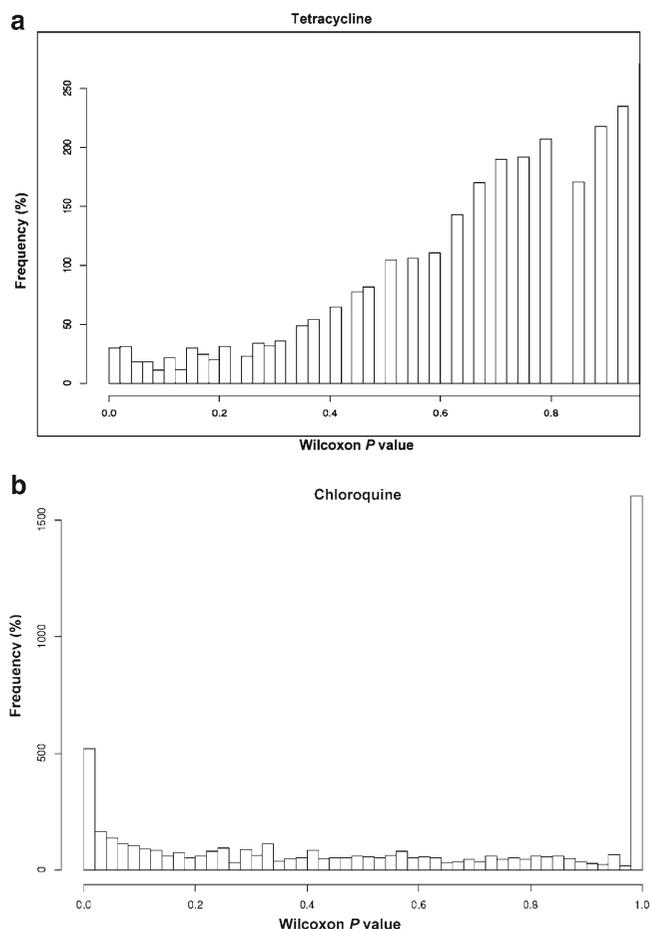
For each cluster, all combinations are calculated as described above. However, this is done for each experiment (time point) separately. Once all combinations are calculated, a Wilcoxon test is done to distinguish differences between the two different states (in our case control vs. drug). For each cluster, this is done for every calculated combination. Once all Wilcoxon tests for all clusters and all combinations are done, the  $P$  values are corrected for multiple testing. The clusters are then ranked according to the lowest  $P$  value that was achieved for the respective clusters.

### **3.7 Analysis of Stimulated or Repressed Pathways**

The analysis of stimulated or repressed pathways was done manually and included an in-depth literature search. First, per cluster, we identified the product/function of each gene and the metabolic pathways in which each is functionally active. We did this using plasmoDB. For each drug, we identified the genes functionally active in the pathways it targeted, expecting our pattern extraction tool to capture the distinct differential expression of these genes between the drug-induced and the control samples. We looked for cases that did not show this format (such cases have been found) to give us hints on what collection of genes differentially coexpressed might deactivate the effectiveness of the drug on the targeted pathway.

### **3.8 Results and Discussion**

Figure 1 shows the histograms of the Wilcoxon  $P$  value of each gene expression under tetracycline (A) and chloroquine (B) treatment conditions compared with its gene expression under no drug influence. There are many more discriminative coexpression patterns in the chloroquine data than in the tetracycline data. Using these data, we list in Tables 1 and 2 the genes that are at least 95% significantly differentially expressed under tetracycline influence and genes that are at least 99.999% significantly differentially



**Fig. 1** Histograms of the Wilcoxon  $P$  values of each gene's expression under tetracycline treatment (**a**) and chloroquine treatment (**b**) compared with its expression under no drug influence. The  $x$ -axis lists ranges of all the  $P$  values estimated and the  $y$ -axis shows the frequency of each

expressed under chloroquine influence, respectively, with their corresponding Wilcoxon  $P$  values.

Currently from PlasmoCyc, 691 reactions of the malaria parasite have been curated and documented. We consider here top-ranking reactions of the parasite whose enzymes were significantly differentially expressed under drug treatment conditions compared with their expression under no drug influence for tetracycline and chloroquine (Tables 3 and 4, respectively). The first and second columns give the reaction's common name and unique ID in PlasmoCyc. The second column gives their Wilcoxon test  $P$  values. The Wilcoxon test is applicable here because we do not have the requirement for normally distributed data. The lower the  $P$  values of a reaction under the drug influence vs. control, the more highly significant the possibility that the reaction may have contributed to

**Table 1**  
**Sixty-five genes that are significantly differentially expressed ( $P$  value  $\leq 0.05$ ) under tetracycline treatment**

Gene ID	Wilcoxon $P$ values
pla_ORF78	0.003636625
MAL13P1.271	0.04490200
MAL13P1.312	0.02048920
MAL13P1.304	0.001829776
PFI0495w	0.01004454
PF14_0114	0.0006560272
PF10_0319	0.0284207
PF10_0026	0.02048920
PF14_0582	0.02048920
PF14_0294	0.0004955335
PF14_0695	0.02048920
PFB0425c	0.01004454
PF10_0313	0.01449251
PFE0230w	0.01209398
PFC1065w	0.0284207
PFD1090c	0.01209398
pla_tufA1	0.00555959
PF14_0278	0.0008579387
PFE0755c	0.00829316
PFA0430c	0.03872114
PFI0990c	0.03872114
MAL6P1.93	0.01727119
wPF10_0061	0.01727119
pla_ORF470	0.03872114
pla_rps11	0.02418426
PFL0635c	0.02418426
pla_rps17	0.03872114
MAL8P1.71	0.0284207
PF11_0086	0.03324143
PF14_0175	0.02418426
MAL6P1.104	0.04490200

(continued)

**Table 1**  
**(continued)**

<b>Gene ID</b>	<b>Wilcoxon <i>P</i> values</b>
PF14_0409	0.01449251
PFD0260c	0.003636625
PFL2335w	0.01727119
PFL0835w	0.003636625
PFL1125w	0.03872114
PFD0400w	0.04490200
PF13_0332	0.01209398
PFE1455w	0.004513053
MAL13P1.33	0.0284207
PFC0260w	0.01727119
PFI1500w	0.03324143
PFC0750w	0.002316434
PF10_0213	0.00555959
PF14_0529	0.004513053
PFD0845w	0.001829776
PF13_0332	0.00829316
pla_rps7	0.02418426
PFD0885c	0.03324143
pla_tRNA-Gln	0.02418426
pla_tRNA-Gly	0.03872114
PFE1375c	0.0284207
pla_tRNA-Trp	0.03872114
PFL2325c	0.03872114
PFD0970c	0.03324143
PF14_0093	0.0284207
MAL6P1.105	0.0284207
MAL13P1.261	0.003636625
PF11_0433	0.01727119
PF10_0336	0.03872114
PF13_0210	0.04490200
PF11_0289	0.03872114
PFL0290w	0.02418426
PFA0430c	0.00555959

**Table 2**  
**Ninety genes that are significantly differentially expressed**  
**(*P* value  $\leq 1.0e-5$ ) under chloroquine**

Gene ID	Wilcoxon <i>P</i> value
MAL13P1.245	7.396023e-07
MAL13P1.25	7.396023e-07
MAL6P1.181	7.396023e-07
MAL6P1.4	7.396023e-07
MAL6P1.60	7.396023e-07
MAL6P1.79	8.875228e-06
MAL7P1.104	5.177216e-06
MAL7P1.50	8.875228e-06
MAL8P1.22	7.396023e-07
MAL8P1.24	1.479205e-06
MAL8P1.97	7.396023e-07
PF07_0050	2.958409e-06
PF07_0055	7.396023e-07
PF07_0056	2.958409e-06
PF07_0111	7.396023e-07
PF07_0115	1.479205e-06
PF08_0008	7.396023e-07
PF08_0018	8.875228e-06
PF08_0021	7.396023e-07
PF08_0073	8.875228e-06
PF10_0002	7.396023e-07
PF10_0082	7.396023e-07
PF10_0132	8.875228e-06
PF10_0167	5.177216e-06
PF10_0177	2.958409e-06
PF10_0198	1.479205e-06
PF11_0021	2.958409e-06
PF11_0098	7.396023e-07
PF11_0127	7.396023e-07
PF11_0164	2.958409e-06

(continued)

**Table 2**  
**(continued)**

<b>Gene ID</b>	<b>Wilcoxon <i>P</i> value</b>
PF11_0236	2.958409e-06
PF11_0289	7.396023e-07
PF13_0295	7.396023e-07
PF13_0317	7.396023e-07
PF14_0061	8.875228e-06
PF14_0161	5.177216e-06
PF14_0212	1.479205e-06
PF14_0217	2.958409e-06
PF14_0231	7.396023e-07
PF14_0303	1.479205e-06
PF14_0336	7.396023e-07
PF14_0481	7.396023e-07
PF14_0497	2.958409e-06
PF14_0512	5.177216e-06
PF14_0611	2.958409e-06
PF14_0701	5.177216e-06
PF14_0715	8.875228e-06
PFA0290w	8.875228e-06
PFA0460c	9.61483e-06
PFB0470w	1.479205e-06
PFB0820c	7.396023e-07
PFB0845w	1.479205e-06
PFC0195w	2.958409e-06
PFC0370w	5.177216e-06
PFC0470w	5.177216e-06
PFC0495w	7.396023e-07
PFC0575w	7.396023e-07
PFC0785c	1.479205e-06
PFD0035c	7.396023e-07
PFD0215c	8.875228e-06
PFD0430c	7.396023e-07

(continued)

**Table 2**  
**(continued)**

<b>Gene ID</b>	<b>Wilcoxon <i>P</i> value</b>
PFD0490c	7.396023e-07
PFD0520c	7.396023e-07
PFD0820w	7.396023e-07
PFE0820c	7.396023e-07
PFE0890c	7.396023e-07
PFE0950c	1.479205e-06
PFE1300w	7.396023e-07
PFE1595c	1.479205e-06
PFE1605w	5.177216e-06
PFI0300w	8.875228e-06
PFI0315c	8.875228e-06
PFI0860c	8.875228e-06
PFI1080w	1.479205e-06
PFI1225w	5.916818e-06
PFI1420w	7.396023e-07
PFI1485c	7.396023e-07
PFL0370w	5.177216e-06
PFL0410w	1.479205e-06
PFL0920c	7.396023e-07
PFL1045w	7.396023e-07
PFL1150c	1.479205e-06
PFL1195w	2.958409e-06
PFL1270w	1.479205e-06
PFL1970w	2.958409e-06
PFL1980c	7.396023e-07
PFL2190c	8.875228e-06
PFL2390c	7.396023e-07
PFL2415w	7.396023e-07
PFL2555w	5.177216e-06

**Table 3**

**Twenty-two top-ranking reactions of the parasite whose enzymes were significantly differentially expressed under drug (tetracycline) influence vs. their expression under no drug influence**

Common reaction name	Unique ID in PlasmoCyc	Wilcoxon <i>P</i> value
Threonine-tRNA ligase	THREONINE-TRNA-LIGASE-RXN	0.2189208
Phenylalanine-tRNA ligase	ALANINE-TRNA-LIGASE-RXN	0.1781820
Ferrochelatase	PROTOHEMEFERROCHELAT-RXN	0.2591973
Adenylosuccinate lyase	AMPSYN-RXN	0.2320216
Adenylosuccinate lyase	AICARSYN-RXN	0.2320216
Fructose-bisphosphate aldolase	F16ALDOLASE-RXN	0.2415238
Lysine decarboxylase	LYSDECARBOX-RXN	0.2591973
Copper-exporting ATPase	3.6.3.4-RXN	0.1472773
Inositol-1,4,5-trisphosphate 5-phosphatase	3.1.3.56-RXN	0.2092226
Thiosulfate sulfurtransferase	THIOSULFATE-SULFURTRANSFERASE-RXN	0.2581537
UDP- <i>N</i> -acetylglucosamine- dolichyl-phosphate <i>N</i> -acetylglucosamine phosphotransferase	2.7.8.15-RXN	0.1718688
Adenylate kinase	ADENYL-KIN-RXN	0.0786467
Aromatic amino acid transferase	TYRAMINOTRANS-RXN	0.2415238
Aromatic amino acid transferase	PHEAMINOTRANS-RXN	0.2415238
Aspartate aminotransferase	ASPAMINOTRANS-RXN	0.2415238
Phenylalanine(histidine) aminotransferase	3-SULFINOALANINE- AMINOTRANSFERASE-RXN	0.2415238
Dihydrolipoamide <i>S</i> - acetyltransferase	RXN0-1133	0.1848905
Acetyl-coA C-acyltransferase	METHYLACETOACETYLCOATHIOL-RXN	0.1118256
Acetyl-coA C-acyltransferase	KETOACYLCOATHIOL-RXN	0.1118256
Histone acetyltransferase	HISTONE-ACETYLTRANSFERASE-RXN	0.07230222
Acetyl-coA C-acetyltransferase	ACETYL-COA-ACETYLTRANSFER-RXN	0.1118256
Pyruvate dehydrogenase (lipoamide)	RXN0-1134	0.2179721

**Table 4**

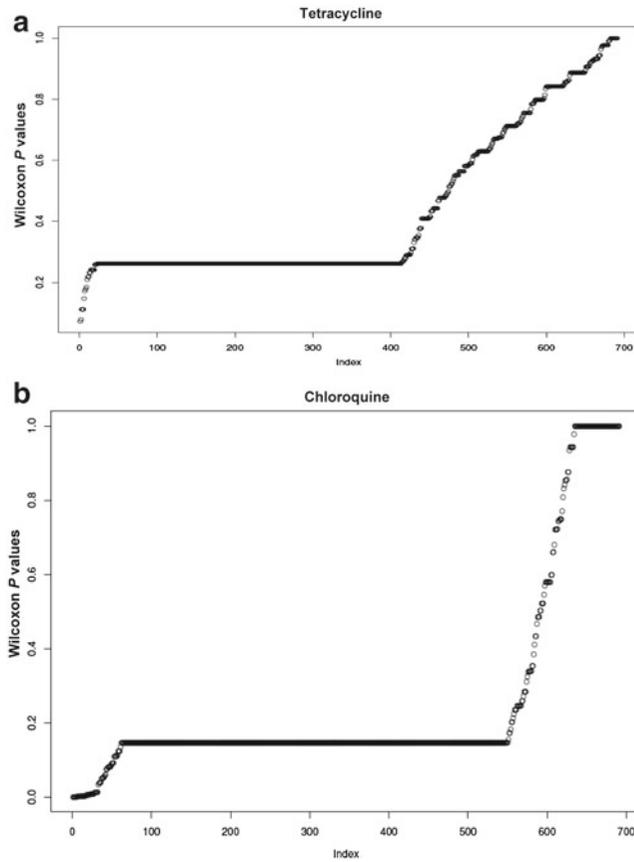
**Fifty-two top-ranking reactions of the parasite whose enzymes were significantly differentially expressed under drug (chloroquine) influence vs. their expression under no drug influence**

Reaction common name	Unique ID in Plasmocyc	Wilcoxon <i>P</i> value
Acetyl-coA carboxylase	RXN0-5055	0.01262057
Acetyl-coA carboxylase	ACETYL-COA-CARBOXYLTRANSFER-RXN	0.01262057
Biotin carboxylase	BIOTIN-CARBOXYL-RXN	0.01262057
Phosphopantothenate–cysteine ligase	P-PANTOCYSLIG-RXN	0.05755659
Long-chain-fatty-acid–coA ligase	RXN-7904	0.008890324
Long-chain-fatty-acid–coA ligase	R223-RXN	0.008890324
Long-chain-fatty-acid–coA ligase	ACYLCOASYN-RXN	0.008890324
Tyrosine–tRNA ligase	TYROSINE–TRNA-LIGASE-RXN	0.000641938
Methionine–tRNA ligase	METHIONINE–TRNA-LIGASE-RXN	0.03842444
Lysine–tRNA ligase	LYSINE–TRNA-LIGASE-RXN	0.0015584469
Leucine–tRNA ligase	LEUCINE–TRNA-LIGASE-RXN	0.01209398
Isoleucine–tRNA ligase	ISOLEUCINE–TRNA-LIGASE-RXN	0.0341075
Histidine–tRNA ligase	HISTIDINE–TRNA-LIGASE-RXN	0.003636625
Phosphoacetylglucosamine mutase	PHOSACETYLGLUCOSAMINEMUT-RXN	0.007259989
Mannose-6-phosphate isomerase	MANNPISOM-RXN	0.05335639
Ferrochelatae	PROTOHEMEFERROCHELAT-RXN	0.07540837
Guanylate cyclase	GUANYLCYC-RXN	0.04010272
Pseudouridylylate synthase	PSEUDOURIDYLATE-SYNTHASE-RXN	0.01381947
GDP-mannose 4,6-dehydratase	GDPMANDEHYDRA-RXN	0.09140153
1-Phosphatidylinositol-4,5-bisphosphate phosphodiesterase	3.1.4.11-RXN	8.875228e–06
Inositol-1,4,5-trisphosphate 5-phosphatase	3.1.3.56-RXN	0.006815353
Pyruvate, water dikinase	RXN0-308	0.09261214
Pantetheine-phosphate adenylyltransferase	PANTEPADENYLYLTRAN-RXN	0.0004955335
Adenylyltransferase	FADSYN-RXN	0.03872114
Mannose-1-phosphate guanylyltransferase	2.7.7.13-RXN	0.00829316

(continued)

**Table 4**  
**(continued)**

Reaction common name	Unique ID in Plasmocyc	Wilcoxon <i>P</i> value
Ribose-phosphate diphosphokinase	PRPPSYN-RXN	0.00250444
Pyruvate kinase	PEPDEPHOS-RXN	0.06123981
Ethanolamine kinase	ETHANOLAMINE-KINASE-RXN	0.08042361
Choline kinase	CHOLINE-KINASE-RXN	0.007112016
6-Phosphofructokinase	6PFRUCTPHOS-RXN	0.005852473
Diphosphate-fructose-6-phosphate 1-phosphotransferase	2.7.1.90-RXN	0.005852473
Glutathione transferase	GST-RXN	0.002898356
Glutathione transferase	GSHTRAN-RXN	0.002898356
Farnesyltranstransferase	FARNESYLTRANSTRANSFERASE-RXN	0.0002514648
Protein farnesyltranstransferase	2.5.1.58-RXN	0.0002514648
Formate C-acetyltransferase	RXN-1381	0.09165743
Histone acetyltransferase	HISTONE-ACETYLTRANSFERASE-RXN	0.08575171
Glycylpeptide <i>N</i> -tetradecanoyltransferase	2.3.1.97-RXN	0.002914033
Aminomethyltransferase	GCVT-RXN	0.002914033
Site-specific DNA-methyltransferase (cytosine-specific)	2.1.1.73-RXN	0.007795191
Cytochrome-b5 reductase	CYTOCHROME-B5-REDUCTASE-RXN	0.001432610
Sarcosine dehydrogenase	SARCOSINE-DEHYDROGENASE-RXN	0.002914033
Dimethylglycine dehydrogenase	DIMETHYLGLYCINE-DEHYDROGENASE-RXN	0.002914033
Pyridoxamine-phosphate oxidase	PMPOXI-RXN	0.003350211
Protoporphyrinogen oxidase	PROTOPORGENOXI-RXN	0.07529077
Pyruvate dehydrogenase (lipoamide)	RXN0-1134	0.05377831
Ferredoxin-NADP(+) reductase	FLAVONADPREDUCT-RXN	0.05004103
Ferredoxin-NADP(+) reductase	1.18.1.2-RXN	0.05004103
None	GDREDUCT-RXN	0.08158513
None	CDREDUCT-RXN	0.08158513
None	ADREDUCT-RXN	0.08158513
L-Lactate dehydrogenase	L-LACTATE-DEHYDROGENASE-RXN	0.00250444



**Fig. 2** Distribution of the sorted version of the  $P$  values for all reactions of the parasite for the tetracycline treatment (a) and the chloroquine treatment (b) compared with control. Each gene indexed is plotted on the  $x$ -axis and its corresponding Wilcoxon  $P$  value is plotted on the  $y$ -axis

the ability of the malaria parasite to resist these drugs. Figure 2 shows the distributions of the sorted versions of the  $P$  values for all reactions of the parasite for the tetracycline vs. control condition (A) and the chloroquine vs. control condition (B). Based on these findings, we listed all reactions whose  $P$  values are  $\leq 0.25$  for tetracycline and  $\leq 0.1$  for chloroquine.

Data in Table 1 suggested the following. Although a number of the genes in Table 1 are conserved 0.01004454 protein of unknown function, we were able to get important interpretation of the kind of results deducible from Table 1 via PFI0990. The gene PFI0990 is said to interact with the following genes: PF08\_0026 (conserved *Plasmodium* protein of unknown function), PFL1385C (a merozoite surface protein 9), and PFL1315W (a potassium channel protein). It was found that these genes are inhibited PFI0990 ([www.plasmodb.org](http://www.plasmodb.org)), which is heavily expressed (by our results in Table 1) under tetracycline treatment compared with its

normal expression in the absence of tetracycline treatment. This means that these genes must have been silenced for PFI0990 to be heavily expressed. First, PFL1385C (coding for a merozoite surface protein 9) confirm this statement by Dahl et al. (25): “Our results demonstrate that tetracyclines specifically block expression of the apicoplast genome, resulting in the distribution of nonfunctional apicoplasts into daughter merozoites.” And second, it is known potassium channels are found in most cell types and control a wide variety of cell functions. Therefore, the inhibition of PFL1315W looks to have contributed to the negative effect of tetracycline on the parasite. We also found PF10\_0061 (an apical membrane antigen 1) to be heavily expressed under tetracycline treatment compared with its normal expression in the absence of tetracycline treatment. Knowing the genes it interacts with can give us more insight into the biological mode of action of tetracycline.

In Table 2, little is known of the genes therein, interacted with. Information on gene PF07\_0056 obtained from plasmDB also gives us further information that can be deduced from Table 2, if more information about the genes therein are available, PF07\_0056, which is heavily expressed under chloroquine treatment, activates MAL8P1.23 which in turn activates PFF1300w (a pyruvate kinase). It is known that the enzyme pyruvate kinase affects the survival of red blood cells. In our prediction, via the chloroquine treatment, positively.

We did not find any significant differential expression between any clusters in the chloroquine SAGE and control samples. Le Roch also reached this conclusion (personal communication).

From the chloroquine microarray data (obtained using the combinatorial technique based on the Kernighan–Li clustering technique; second extracted subgraph), confirmed using the wavelets technique based on the consecutive-ones clustering technique (eighth extracted subgraph), we observed that tryptophanyl-tRNA synthetase production in the apicoplast is upregulated. Wellems and Plowe (42) state that “chloroquine’s efficacy is thought to lie in its ability to interrupt hemozoin detoxification in malaria parasites as they grow within their host’s red blood cells. Hemozoin is released in large amounts as the parasite consumes and digests hemoglobin in its digestive food vacuole. Hemozoin normally is detoxified by polymerization into innocuous crystals of hemozoin pigment and perhaps also by a glutathione-mediated process of destruction. Chloroquine binds with hemozoin in its  $\mu$ -oxodimer form and also adsorbs to the growing faces of the hemozoin crystal, disrupting detoxification and poisoning the parasite. Chloroquine-resistance *P. falciparum* survives by reducing accumulation of the drug in the digestive vacuole; however, the mechanism by which this happens has not been determined. Leading proposals include mechanisms that involve alterations of digestive vacuole pH or changes in the flux of chloroquine across the parasite’s cytoplasmic or digestive

vacuole membrane.” The second mechanism of flux of chloroquine was summarized by Krogstad et al. (43), who write that “... chloroquine-resistance *P. falciparum* accumulates less chloroquine than susceptible parasites. This observation suggests that chloroquine resistance in *P. falciparum* results from either decreased uptake or increased excretion of the drug by the resistant parasite ... resistance *P. falciparum* parasites have a mechanism for releasing chloroquine (an efflux process) (44). This efflux is either absent or greatly reduced in the susceptible parasite.”

Therefore, that tryptophanyl-tRNA synthetase production in the apicoplast is upregulated (in the chloroquine-induced microarray data) may suggest that this efflux process was made possible (caused) by the apicoplast, the mini bacterium living inside the malaria parasite. Ralph et al. (45) state that “it is not yet clear what the key function of the apicoplast is but the organelle is clearly indispensable. Curiously though, parasites cured of their apicoplasts do not die immediately. Rather, they fail to invade new host cells successfully. This suggests that apicoplasts provide some component essential to invasion and or [sic] establishment of the parasitophorous vacuole in the host cell” (46, 47).

Thus a combination of chloroquine with the agents that cured *P. falciparum* of its apicoplast may be helpful in preventing the parasite from invading new host cells, and this combination may also kill the parasite, because it could not then flux out accumulated chloroquine in its digestive food vacuole.

Analyzing the two sets of microarray data together here provides the opportunity to identify reactions that may be upregulated via treatment with both drugs. In this line, we found the following reactions: FARNESYLTRANSFERASE-RXN, TRYPTOPHAN-TRNA-LIGASE, THREONINE-TRNA-LIGASE-RXN, and ALANINE-TRNA-LIGASE-RXN. These reactions appear very important in the parasite quest to resist the two antimalaria drugs we have considered in this paper (tetracycline and chloroquine). Our study represents the first attempt to unveil this.

We also observed that many (19 of 22) of the enzymes encoded by the genes active in the pathway targeted by chloroquine have not been identified. We are following these leads and we believe that further findings will be possible when such information is available.

---

## 4 Second Model: In Silico Model to Combat Resistance

In the second model, we extended our algorithm (48, 49) using a machine learning approach. The resulting algorithm is able to identify novel combinable drug targets from the metabolic network of *P. falciparum*. Using this approach we identified, among others, 19 drug targets confirmed from the literature. The machine learning approach combines clustering, common distance similarity measurements, and hierarchical clustering to propose new

combinations of drug targets, see details in Fatumo et al. (50). Our result suggests that two or more enzymatic reactions from our list of drug targets that span across pathways could be combined to form an efficient malaria drug target, targeting distinct time points in the parasite's intraerythrocytic developmental cycle.

The metabolic network of *P. falciparum* was set up using the BioCyc database (<http://biocyc.org>) as described recently for *E. coli* (23). The metabolites were the nodes and the enzymatic reactions were the edges of the network. Our network yielded 554 metabolites and 575 reactions. Each compound can be substrate and product.

We set up a graph-based algorithm analyzing the structure of biochemical networks to infer differences (such as different paths) when exposed to changing nutrients and environmental conditions. Raymond and Segrè (51) showed that the access for metabolites changes drastically when oxygen is available. Following this strategy, we chose several sets of metabolites as sets of products. Then the investigated reaction was deleted from the network. The mutated network (the network with the deleted reaction) was investigated to determine whether the chosen products in each set could still be produced. We compared the number of products that could be produced in the wild-type network and the mutated network. The difference in the numbers gave an insight into whether the investigated reaction is essential or not.

#### **4.1 Verifying the Essentiality of a Knockout Reaction**

The algorithm investigates a reaction by deleting the reaction from the metabolic network and checking whether a chosen product can be produced without the deleted reaction.

#### **4.2 Creating the Variety of Products**

We assigned a list of all reactions in the neighborhood of compounds of the reaction under investigation. Thirty percent of all compounds of these reactions were set as a product to be produced by the remaining compounds. A total of 1,000 different combinations of the chosen product were assembled.

#### **4.3 Minimizing the Number of Reactants and Reactions to Produce the Products**

The algorithm checked every investigated reaction for a minimum number of needed reactions and reactants needed to produce the products. A “greedy” approach was employed for minimizing the number of reactants and reactions needed to produce the products.

#### **4.4 Comparing the Results of Wild-Type and the Mutated Network to Obtain the Essentiality of the Investigated Reaction**

We computed the average minimum sets of substrates for a knockout reaction in the mutated network vs. wild-type. Similarly, we computed the average minimum sets of reactions. We then compared the number of successful productions for the wild-type and knockout reactions. A total of 1,000 different sets of products were used.

#### **4.5 Gene Expression Analysis**

We identified 46 essential enzymatic reactions as reported by our algorithm. We used GENESIS (52), a sophisticated tool for analyzing gene expression data including clustering techniques, motif search, and visualization utilities, to analyze the essential reactions. Our DNA microarray data, which were obtained from Bozdech et al. (53) with 48 individual 1 h time points from the intraerythrocytic developmental cycle of *P. falciparum*, were organized by hierarchical clustering.

We clustered the 46 expressed genes into 6 groups according to their expression levels. Groups 1 and 2 had 4 enzymatic reactions, group 3 had 14 reactions, group 4 had 8 reactions, group 5 had 10 reactions, and group 6 had 6 reactions. We noticed that all the reactions in group 3 are responsible for transport and all coded for one gene; two reactions in group 6 also coded for the same gene. This left us with only 30 essential reactions as possible targets.

#### **4.6 Comparative Screening Analysis of Possible Drug Targets**

We found from the drug banks SIGMA and TDR targets for inhibitors or drugs for most of the possible drug targets we identified. We have at least one inhibitor/drug for 19 possible drug targets. We further did gene expression analysis of the 19 possible drug targets to determine whether two or more enzymatic reactions in the initial groups overlap with the new groups. It seems reasonable to combine the inhibitors for such possible drug targets because the resulting drug might attack the parasite at the same time point in its life cycle during its stay in the human red blood cells.

We clustered the gene expression data analysis using GENESIS (50). This analysis resulted in two new groups, with the initial groups 1, 5, and 6 now belonging to new group A and the initial groups 2 and 4 now belonging to group B. We hypothesize that it is beneficial to combine inhibitors/drugs for targets within each group.

---

## **5 Conclusion**

With the first in silico model, we were able to use the biochemical network of *P. falciparum* to deduce its drug resistance mechanism(s) using two sets of gene expression data obtained from treatment of the parasite with chloroquine and tetracycline. Our work is the first to develop and apply computational means toward the elucidation of these mechanisms in *P. falciparum*. Our work suggests viable mechanisms for the resistance of the malaria parasite to chloroquine and tetracycline. When these results are experimentally tested they may provide useful weapons to efficiently cleanse malaria parasites from the blood stream.

With the second in silico model, we established a machine learning tool that identified drug targets confirmed from the literature, which we then further analyzed using a sophisticated gene expression analysis tool. Our data were organized using common

distance similarity measurements and hierarchical clustering. Our results suggest that two or more enzymatic reactions from the list of our drug targets, which span about ten pathways, could be combinable if targeted at distinct pathways to produce an efficient malaria drug.

---

## Acknowledgments

Many thanks go to Karine Le Roch, Svetlana Bulashesva, Benedikt Brors, Gunnar Schramm, Anna-Lena Kranz, Roland Eils, and Rainer Koenig for many useful discussions and contributions.

## References

- Pierotti MA, Tamborini E, Negri T et al (2011) Targeted therapy in GIST: in silico modeling for prediction of resistance. *Nat Rev Clin Oncol* 8:161–170. doi:10.1038/nrclinonc.2011.3
- Noble D (2002) Modelling the heart – from genes to cells to the whole organ. *Science* 295(5560):1678–1682
- Hammer GL, Sinclair TR, Chapman SC, Oosterom EV (2004) Scientific correspondence on systems thinking, systems biology and the in silico plant. *Plant Physiol* 134:909–911
- Deville Y, Gilbert D, Helden JV, Wodak SJ (2003) An overview of data models for the analysis of biochemical pathways. *Brief Bioinform* 4(3):246–259
- Crampin EJ, Schnell S (2004) New approaches to modelling and analysis of biochemical reactions, pathways and networks. *Prog Biophys Mol Biol* 86(1):1–4
- Noble D, Rudy Y (2001) Models of cardiac ventricular action potentials: iterative interaction between experiment and simulation. *Phil Trans R Soc Lond A* 359:1127–1142
- Van't Veer LJ, Dai H, van de Vijver MJ et al (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415:530–536
- Stephanopoulos G, Hwang D, Schmitt WA, Mistra J (2002) Mapping physiological states from microarray expression measurements. *Bioinformatics* 18:1054–1063
- Gasch AP, Spellman PT, Kao CM et al (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* 11:4241–4257
- Spellman PT, Sherlock G, Zhang MQ et al (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridisation. *Mol Biol Cell* 9:3273–3297
- Gardner TS, di Bernardo D, Lorenz D, Collins JJ (2003) Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 301:102–105
- Berg JM, Tymoczko JL, Stryer L (2002) *Biochemistry*, 5th edn. W.H. Freeman, New York, p 1050
- Karp PD, Riley M, Pellegrini-Toole A (2002) The MetaCyc database. *Nucleic Acids Res* 30:59–61
- Khodursky AB, Peter BJ, Cozzarelli NR et al (2000) DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*. *Proc Natl Acad Sci USA* 97:12170–12175
- Neidhardt FC (1996) *Escherichia coli* and Salmonella: cellular and molecular biology. American Society for Microbiology, Washington, DC
- Uetz P, Giot L, Cagney G et al (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403:623–627
- Ideker T, Ozier O, Schwikowski B, Siegel AF (2002) Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* 18(Suppl 1):S233–S240
- Hanisch D, Zien A, Zimmer R, Lengauer T (2002) Co-clustering of biological networks and gene expression data. *Bioinformatics* 8(Suppl 1):S145–S154
- Zien A, Küffner R, Zimmer R, Lengauer T (2000) Analysis of gene expression data with pathway scores. *Proc Int Conf Intell Syst Mol Biol* 8:407–417
- Karp PD, Ouzounis CA, Moore-Kochlacs C et al (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res* 33:6083–6089

21. Francke C, Siezen RJ, Teusink B (2005) Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol* 13(11):550–558
22. König R, Eils R (2004) Gene expression analysis on biochemical networks using the Potts spin model. *Bioinformatics* 20:1500–1505
23. König R, Schramm G, Oswald M et al (2006) Discovering functional gene expression pattern in the metabolic network of *Escherichia coli* with wavelets transforms. *BMC Bioinformatics* 7:119
24. Gunasekera AM, Patankar S, Schug J et al (2003) Drug-induced alterations in gene expression of the asexual blood forms of *Plasmodium falciparum*. *Mol Microbiol* 50(4):1229–1239
25. Dahl EL, Shock JL, Shenai BR et al (2006) Tetracyclines specifically target the apicoplast of the malaria parasite *Plasmodium falciparum*. *Antimicrob Agents Chemother* 50(9):3124–3131
26. Booth KS, Lueker GS (1976) Testing for the consecutive ones property, interval graphs, and graph planarity using *PQ*-Tree algorithms. *J Comput Syst Sci* 13:335–379
27. Cohen J (1988) *Statistical power analysis for the behavioral sciences*, 2nd edn. Lawrence Erlbaum Associates, Hillsdale, NJ
28. Gunasekera AM, Myrick A, Le Roch K et al (2007) *Plasmodium falciparum*: genome wide perturbations in transcript profiles among mixed stage cultures after chloroquine treatment. *Exp Parasitol* 117:87–92
29. Trager W, Jensen JB (1976) Human malaria parasites in continuous culture. *Science* 193:673–675
30. Le Roch KG, Zhou Y, Blair PL et al (2003) Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* 301:1503–1508
31. Zhou Y, Abagyan R (2002) Match-only integral distribution (MOID) algorithm for high-density oligonucleotide array analysis. *BMC Bioinformatics* 3:3
32. Huber W, von Heydebreck A, Sultmann H et al (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 18(Suppl 1):S96–S104
33. Waller RF, Reed MB, Cowman AF, McFadden GI (2000) Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. *EMBO J* 19:1794–1802
34. van Dooren GG, Marti M, Tonkin CJ et al (2005) Development of the endoplasmic reticulum, mitochondrion and apicoplast during the asexual life cycle of *Plasmodium falciparum*. *Mol Microbiol* 57:405–419
35. Bozdech Z, Zhu J, Joachimiak MP et al (2003) Expression of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biol* 4:R9
36. Bozdech Z, Llinas M, Pulliam BL et al (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol* 1:E5
37. Brown DE, Huntley CL (1992) A practical application of simulated annealing to clustering. *Pattern Recog* 25:401–412
38. Kernighan BW, Lin S (1970) An efficient heuristic procedure for partitioning graphs. *Bell Syst Tech J* 49:291–307
39. Dutt S (1993) New faster Kernighan-Lin type graph-partitioning algorithms. In: *Proceedings of the 1993 IEEE/ACM international conference on computer-aided design*, Santa Clara, CA, pp 370–377
40. Aarts EHL, van Laarhoven PJM (1985) A new polynomial time cooling schedule. In: *Proceedings of the IEEE international conference on computer-aided design*, Santa Clara, CA, pp 206–208
41. Whites SR (1984) Concepts of scale in simulated annealing. In: *Proceedings of the IEEE international conference on computer-aided design*, Port Chester, NY, pp 646–651
42. Wellems T, Plowe CW (2001) Chloroquine-resistant malaria. *J Infect Dis* 184:770–776
43. Krogstad DJ, Schlesinger PH, Herwaldt BL (1988) Antimalarial agents: mechanism of chloroquine resistance. *Antimicrob Agents Chemother* 32:799–801
44. Krogstad DJ, Gluzman IY, Kyle DE et al (1987) Efflux of chloroquine from *Plasmodium falciparum*: mechanism of chloroquine resistance. *Science* 238:1283–1285
45. Ralph SA, D’Ombrain MC, McFadden GI (2001) The apicoplast as an antimalarial drug target. *Drug Resist Updat* 4:145–151
46. Fichera ME, Roos DS (1997) A plastic organelle as a drug target in apicomplexan parasites. *Nature* 390:407–409
47. He CY, Shaw MK, Pletcher CH et al (2001) A plastic segregation defect in the protozoan parasite *Toxoplasma gondii*. *EMBO J* 20:330–339
48. Fatumo S, Plaimas K, Mallm JP et al (2009) Estimating novel potential drug targets of *Plasmodium falciparum* by analysing the metabolic network of knock-out strains in silico. *Infect Genet Evol* 9(3):351–358
49. Fatumo S, Plaimas K, Adebiyi E, König R (2011) Comparing metabolic network models based on genomic and automatically inferred enzyme information from *Plasmodium* and its human host to define drug targets in silico. *Infect Genet Evol* 11(4):708–715
50. Fatumo S, Adebiyi E, Schramm G et al (2009) An *in-silico* approach to design efficient malaria drug targets to combat the malaria resistance

- problem. Presented at the Computer Science and Information Technology Spring Conference, Singapore, 17–20 Apr 2009. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=5169419&tag=1](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5169419&tag=1)
51. Raymond J, Segrè D (2006) The effect of oxygen on biochemical network and their evolution of complex life. *Science* 311:1764–1767
  52. Sturn A, Quackenbush J, Trajanoski Z (2003) Client-server environment for high-performance gene expression data analysis. *Bioinformatics* 19:772–773
  53. Bozdech Z, Llinas M, Pulliam BL et al (2003) The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol* 1:85–100