In silico activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen

David Plouffe*, Achim Brinker*, Case McNamara*, Kerstin Henson*, Nobutaka Kato*, Kelli Kuhen*, Advait Nagle*, Francisco Adrián*, Jason T. Matzen*, Paul Anderson*, Tae-gyu Nam†, Nathanael S. Gray*‡, Arnab Chatterjee*, Jeff Janes*, S. Frank Yan*, Richard Trager*, Jeremy S. Caldwell*, Peter G. Schultz*†§, Yingyao Zhou*, and Elizabeth A. Winzeler*¶∥

*Genomics Institute of the Novartis Research Foundation, San Diego, CA 92121; and Departments of †Chemistry and ¶Cell Biology, The Scripps Research Institute, La Jolla, CA 92037

The growing resistance to current first-line antimalarial drugs represents a major health challenge. To facilitate the discovery of new antimalarials, we have implemented an efficient and robust high-throughput cell-based screen (1,536-well format) based on proliferation of *Plasmodium falciparum* (Pf) in erythrocytes. From a screen of ≈1.7 million compounds, we identified a diverse collection of ≈6,000 small molecules comprised of >530 distinct scaffolds, all of which show potent antimalarial activity (<1.25 μM). Most known antimalarials were identified in this screen, thus validating our approach. In addition, we identified many novel chemical scaffolds, which likely act through both known and novel pathways. We further show that in some cases the mechanism of action of these antimalarials can be determined by *in silico* compound activity profiling. This method uses large datasets from unrelated cellular and biochemical screens and the guilt-by-association principle to predict which cellular pathway and/or protein target is being inhibited by select compounds. In addition, the screening method has the potential to provide the malaria community with many new starting points for the development of biological probes and drugs with novel antiparasitic activities.

antifolates | cheminformatics | high-throughput screening | *Plasmodium falciparum*

Parasite resistance has rendered some of the least expensive, traditional antimalarial drugs ineffective. Moreover, because the likelihood is high that resistance will emerge to the current first-line drugs, artemisinin-based combination therapies, there is currently great interest in finding the next generation of antimalarial drugs. Insofar as malaria affects many countries with poor public health resources, attributes of an ideal treatment for malaria are different from those for diseases of industrialized countries. An ideal antimalarial should be inexpensive to synthesize, have good oral bioavailability, have short treatment regimens, be well tolerated by the patient, and be stable at room temperature. One approach to the discovery of such antimalarial agents involves the identification of new therapeutic targets that then form the basis for chemical screens to identify small molecules that modulate the target's activity *in vivo*. Although such an approach has been highly productive in general, it has not worked well for many infectious agents. In many cases, these target-based screens reveal small molecules with potent activity against an enzyme but that are still unable to clear an infection, either because the target is not really essential to the microbe's viability in the host or because the compound is unable to inhibit the target in the *in vivo* environment (1).

An alternative and more traditional approach is to perform cell-based screens directly against living organisms in which a small molecule is tested in an unbiased fashion against all targets required for viability simultaneously. The disadvantage is that once a compound with potent cellular activity is discovered, lead optimization is hindered without knowing which protein target the compound inhibits. Various strategies for target deconvolution have been developed, including selection of resistant mutants, biochemical affinity-based methods, and cDNA complementation. Nonetheless, this remains a challenging and time-consuming task. However, with the automation and miniaturization of cellular screening systems, we can now obtain unprecedented amounts of data for a single small molecule across a diverse collection of cellular screens. Because compounds with similar activities against a pathway or a target are likely to have similar profiles across screens, we hypothesized that a comprehensive evaluation of these large-scale datasets might provide insights into a compound's possible mechanism of action (MOA) through an *in silico* guilt-by-association approach. Here, we report the application of such an approach to a large cell-based screen for compounds with antimalarial activity. From a fluorescence-based screen (2) of 1.7 million compounds, we identified a subset of ≈17,000 compounds with potent antimalarial activity in a cellular assay (<1.25 μM) that were then evaluated across 131 unrelated cellular and enzymatic screens. *In silico* compound activity profiling has revealed the cellular pathway and/or protein target for a number of selected compounds.

Results

Development of a Malaria High-Throughput Screening Method. Because most published assays for antimalarial activity include an unacceptable use of radioactivity (3–5), expensive reagents, excessive liquid transfer steps, or time-dependent steps (2, 6–10), all of which are not compatible with automated 1,536-well high-throughput screens (HTS), most large chemical libraries have not yet been screened for antimalarial activity. We thus set out to adapt a method for use with the small assay volumes (≈8 μl) of 1,536-well HTS. Our assay is based on a published method (2, 7) in which an increase in parasite nuclei in red blood cell culture is measured after staining with SYBR green I, a dye that fluoresces when bound to
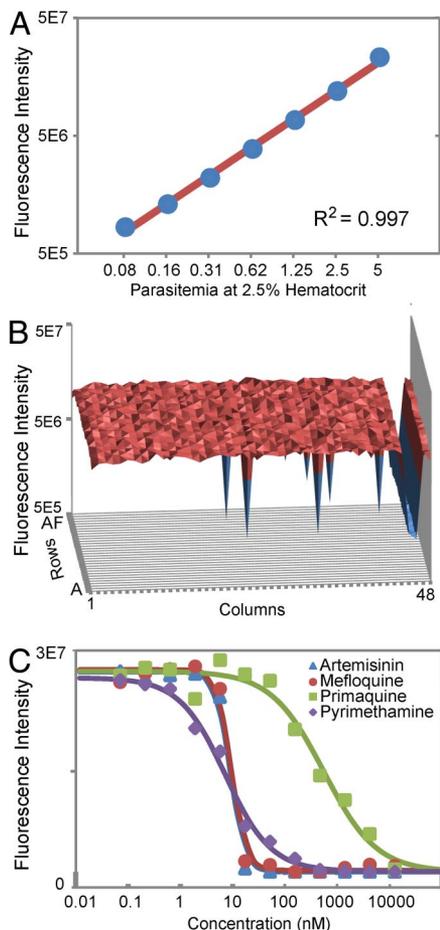

CHEMISTRY

MICROBIOLOGY

Fig. 1. Validation of the antimalarial HTS assay. (*A*) Test results from a titration experiment. A 1:2 dilution of a parasite culture of Pf strain 3D7 starting out at a 5% parasitemia and 2.5% hematocrit was performed with screening media. The trend line was fitted with an R^2 of 0.997 from a 0.078% to 5% parasitemia (logarithmically scaled in both dimensions). (*B*) Fluorescence intensity across an entire assay plate. Columns 45 and 46 were treated with 12.5 μM mefloquine, and columns 47 and 48 were treated with 0.125% DMS0. (*C*) 3D7 dose–response plot with antimalarial control compounds. The EC_{50} values of artemisinin, mefloquine, pyrimethamine, and primaquine are 9.0 nM, 9.5 nM, 7.2 nM, and 548 nM, respectively.
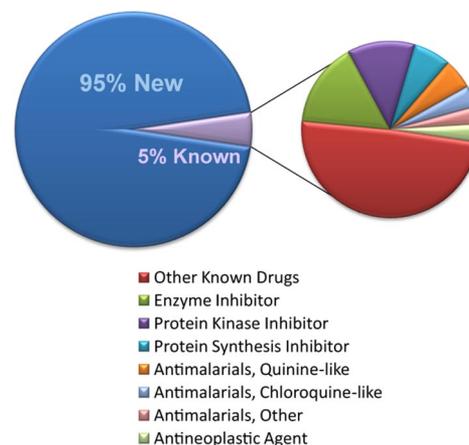

Fig. 2. Chemotype and MOA distributions of validated compounds. Chemotypes distinct from any known drugs were found in 95% of the validated hits. Known chemotypes are further categorized based on their MOA assignments from MeSH.

nucleic acids. Because erythrocytes do not contain nuclei, increases in DNA and RNA are directly attributable to parasite proliferation. Optimization of various parameters, including top and bottom microtiter plate reading, drug-parasite incubation times, lysis methods, and staining time, enabled signal-to-noise ratios of up to 10 and produced excellent correlations between parasitemia (Pf strain 3D7) and fluorescence signal (Fig. 1*A*). The mean *Z* factor for plates in the validation screens was 0.63 (SD = 0.37, median = 0.78). Even though albumax was used in the screening media and resulted in small protein shifts, we were able to readily determine the EC_{50} values of mefloquine (9.5 ± 0.8 nM), pyrimethamine (7.2 ± 1.4 nM), artemisinin (9.0 ± 0.7 nM), and primaquine (548 ± 200 nM) (Fig. 1*B*), which were in good agreement with literature values.

To conduct the full library screen (1.7 million compounds) the fully integrated and automated screening system (www.gnfsystems.com; described in ref. 11) was used. Because of the sheer numbers of plates involved and the limited number of incubators, libraries were divided into sets of ≈300,000 compounds and analyzed over a 12-week period. A typical microtiter plate read from the screen is shown in Fig. 1*B*. Using a criterion of a 50% inhibition in parasite growth at 1.25 μM relative to control plates, we identified ≈17,000

primary hits. Most metrics indicated that the data were of high quality. First, the library fortuitously contained redundant compounds, and in most cases they were rediscovered (e.g., five of the six wells containing pyrimethamine were considered "hits," six of the eight wells containing chloroquine were considered hits, and five of the five containing aminodiaquine were considered hits), suggesting a low false negative rate for the screen. Second, we searched each scaffold against PubChem (http://pubchem.ncbi.nlm.nih.gov) and the World Drug Index (http://scientific.thomsonreuters.com/products/wdi) and retrieved annotations for 394 compounds (12). This set included most of the known antimalarial drugs along with closely related molecules. As previously observed (13), we were able to confirm that the antihistamine astemizole shows antimalarial activity. Hydroxyfenone, which is closely related to the arrhythmia drug propafenone, which had previously been identified as having antimalarial activity, was found (14), as was perhexiline (14). A comprehensive search of the whole chemical library confirmed that our hit list captured most compounds with known antimalarial activities with a *P* value of 10^{-10}. Based on automated evaluation of the MeSH heading (Medical Subject Headings; see www.nlm.nih.gov/mesh), there were also 23 misses, including those with activities of >1 μM [e.g., phenylnorstatine, which is a substructure of a complete peptidomimetic protease inhibitor, may not have the required potency at 1.25 μM (15); pepstatin has an EC_{50} of between 3 and 30 μM (16); betulinic acid has an EC_{50} of >4 μM (17); bredinin has an EC_{50} of 50 μM (18); 5′-methylthioadenosine has an EC_{50} of 80 μM (19)] and those that need to be metabolized (e.g., proguanil) or compounds that were not in the library (e.g., atovaquone). Impressively, only five known antimalarials remained either inconclusive or false negatives. Of the 17,000 primary hits, 6,549 compounds were available that were interesting and potent enough to function as possible drug discovery leads, and these were tested in dose–response format. Of these, 5,973 showed EC_{50} values of <1.25 μM and 648 showed EC_{50} values of <100 nM. When chemical powders were obtained and retested, >80% were reconfirmed.

Clustering by Chemical Scaffold. To assess the utility of the screen in terms of novel structural motifs, confirmed HTS hits were first clustered based on the chemical similarity (20), resulting in the identification of ≈530 different classes (cluster size ≥3) with a similarity metric of 0.85 or higher. The distribution of different scaffolds is shown in Fig. 2. Although 5% of the structures are similar to known drugs, the majority of the remaining scaffolds were not previously classified as having antimalarial activity. Therefore,

Plouffe *et al.*

our HTS approach significantly expanded the chemical space of potential new-generation antimalarials. Among some of the classes we discovered were a variety of drug-like scaffolds containing piperzines, pyrazoles, imidazoles, pyazines, and other heterocyclic scaffolds all obeying the Lipinski's rule of five (21). None of these compounds contained known metal-binding motifs (such as quinoliones) or chemically reactive functionalities (such as Michael acceptors, chloromethyl ketones, or hydrazones), making them highly attractive leads.

Clustering by Historical Activities. To identify compounds that might be broadly toxic, we collected data for each of the 8,457 primary hits across 131 historical, mostly cell-based screens, which had been carried out with the same basic collection of compounds (comprised of different assay formats, readouts, therapeutic areas, or signal qualities). Compounds that are active in a large range of cellular assays are likely to be toxic or nonselective for parasites, and many such compounds were readily identified by our profile analysis. Among the 72 compounds that were hits in at least 40 cellular screens were tonzonium bromide (a surfactant), valinomycin (a pore-forming antibiotic), acetarsol emetine (inhibits polypeptide chain elongation), sangivamycin (an apoptosis-inducing nucleoside analog), anisomycin (an antibiotic), malachite-green-oxalate, and fascaplysin (a tyrosine kinase inhibitor). However, the majority (7,375 of those primary hits with data) of compounds were found in <15 other assays and in some cases compounds were only hits in the malaria assay. The group included chloroquine, amindiaquine, quinine, quinacrine, artemisinin, benflumetol (lumefantrine), and some more obscure compounds with known antimalarial activity, including ciproquinate, bialamicol, and several used to treat parasitic infections (e.g., trichomonacid, acranil, and propamidine isethionate).

Using ''Guilt by Association'' to Determine the Mechanism of Action of Small Molecules. The advantage of a cell-based screen is that it represents an unbiased picture of the known and unknown cellular pathways that can be modulated by small molecules. The chief disadvantage of cellular screening is that the protein target of a small molecule discovered in the process remains unknown. Lead optimization steps (including optimization of affinity, selectivity, and pharmacological profile) are likely to be more efficient if a protein target is known. Thus, the determination of a compound's exact protein target or MOA is desirable even though the process can be a time-consuming one. Therefore, the availability of rapid, low cost ways to predict what pathways or targets an uncharacterized cellular hit might affect in the cell would be quite attractive. Because this process is analogous to predicting how uncharacterized proteins might function in the cell by using global gene expression profiling and "guilt-by-association" analysis, we sought to test whether we could use the same analytical approach, which has been applied successfully to gene expression analysis, to our chemical screening data. We thus applied a semi-supervised clustering algorithm (22) to the historical screening data. We first began with all compounds annotated under the same MOA (62 MeSH groups in total) and created a "metaprofile" for this group using data from the 131 historical screens. We then ranked the profiles of all of the uncharacterized compounds according to their similarity to the metaprofile and descended through this list of compounds until the cluster with the best enrichment score was identified (see *Methods*). Permutation testing was also performed to determine whether the enrichment would be obtained by chance. Thus, we were able to confirm that in 31 cases (of the 62 different MeSH groups associated with hits from the screen), compounds with similar annotations showed much more similar activity profiles than what one would expect in a collection of 8,457 small molecules (Table 1).

One of the more striking of the 31 MeSH groups contained 30 compounds, 14 of which had MeSH annotations. Of these 14, 11 had MeSH annotations indicating that they prevent protein syn-

Table 1. Results of clustering analysis

Annotation category	$\log_{10}P$	$P_{permute}$	In current MeSH	Cluster size
MeSH82005493:Folic Acid Antagonists	−38.78	0.00	15	26
MeSH82000563:Amebicides	−29.46	0.00	11	16
MeSH82000969:Antinematodal Agents	−28.84	0.00	12	29
MeSH82000970:Antineoplastic Agents	−28.71	0.00	18	122
MeSH82018712:Analgesics, Non-Narcotic	−27.92	0.00	12	30
MeSH82000871:Anthelmintics MeSH82009125:Muscle Relaxants, Central MeSH82018727:Muscarinic Antagonists	−26.10	0.00	12	45
MeSH82011500:Protein Synthesis Inhibitors	−25.36	0.00	11	30
MeSH82000891:Anti-Infective Agents, Local	−25.20	0.00	8	8
MeSH82000317:Adrenergic alpha-Antagonists	−23.72	0.00	12	60
MeSH82000981:Antiprotozoal Agents	−23.38	0.00	8	8
MeSH82000962:Antimalarials	−20.80	0.00	21	213
MeSH82000923:Anticestodal Agents MeSH82005369:Filaricides	−20.74	0.00	6	6
MeSH82000889:Anti-Arrhythmia Agents	−20.41	0.00	20	457
MeSH82000894:Anti-Inflammatory Agents, Non-Steroidal MeSH82018501:Antirheumatic Agents	−18.82	0.00	6	6
MeSH82005456:Fluorescent Dyes	−18.82	0.00	6	6
MeSH82000935:Antifungal Agents	−16.81	0.00	5	5
MeSH82000900:Anti-Bacterial Agents	−15.09	0.00	7	63
MeSH82000893:Anti-Inflammatory Agents	−14.78	0.00	6	32
MeSH82007476:Ionophores MeSH82003049:Coccidiostats	−14.35	0.00	4	4
MeSH82002614:Chelating Agents	−14.35	0.00	4	4
MeSH82000972:Antineoplastic Agents, Phytogenic	−13.95	0.00	5	16
MeSH82000903:Antibiotics, Antineoplastic	−13.65	0.00	5	18
MeSH82050257:Tubulin Modulators	−13.18	0.00	4	6
MeSH82004701:Enzyme Inhibitors	−12.21	0.00	5	5
MeSH82014344:Trypanocidal Agents	−11.55	0.03	4	9
MeSH82000959:Antihypertensive Agents	−10.96	0.03	5	57

Compounds with 62 different MeSH headings were identified in the primary screen. For each MeSH heading group, we tested to see whether activity profiles across 131 screens were more similar than would be expected by chance. The $\log_{10}P$ is an estimate of the probability that compounds with a given MeSH heading (number shown in ''In current MeSH'') would be found together in a group given by ''Cluster size'' by chance in an analysis of 5,973 compounds. $P_{permute}$ gives the P value obtained by permutation testing. Only statistically significant clusters are shown.

thesis. The probability of this degree of segregation occurring by chance is 10^{-25} (Fig. 3A). The group included emetine, muconomycin A, anisomycin (inhibits peptidyl transferase of the 80S ribosome system), cephaelin, echinomycin A (binds DNA and inhibits RNA synthesis), cycloheximide (inhibits protein synthesis by binding to RNA), and puromycin, all of which are active in various assays measuring cellular proliferation. One of the three false positives in the group was verapamil, a compound that is thought to block calcium channels but that also may block protein synthesis by inhibiting the uptake of thymidine, uridine, and leucine (23). The other false positives (muconomycin A and sangivamycin) probably do inhibit protein synthesis (24, 25) but do not have MeSH headings indicating as much. Most of these compounds would not be useful as antimalarials but could be useful as chemical probes to query protein function, for finding new protein targets for antimalarials that might be more selective for parasites, or as leads that could be engineered to be specific for the parasite. Interestingly, although these compounds shared similar activity profiles, their chemical structures were very different from one another (similarity score of 0.33, $P = 0.65$; see *Methods*). This is somewhat expected, however, because small molecules that target the same cellular process may bind different protein targets and have different structures, but produce similar chemical phenotypes.
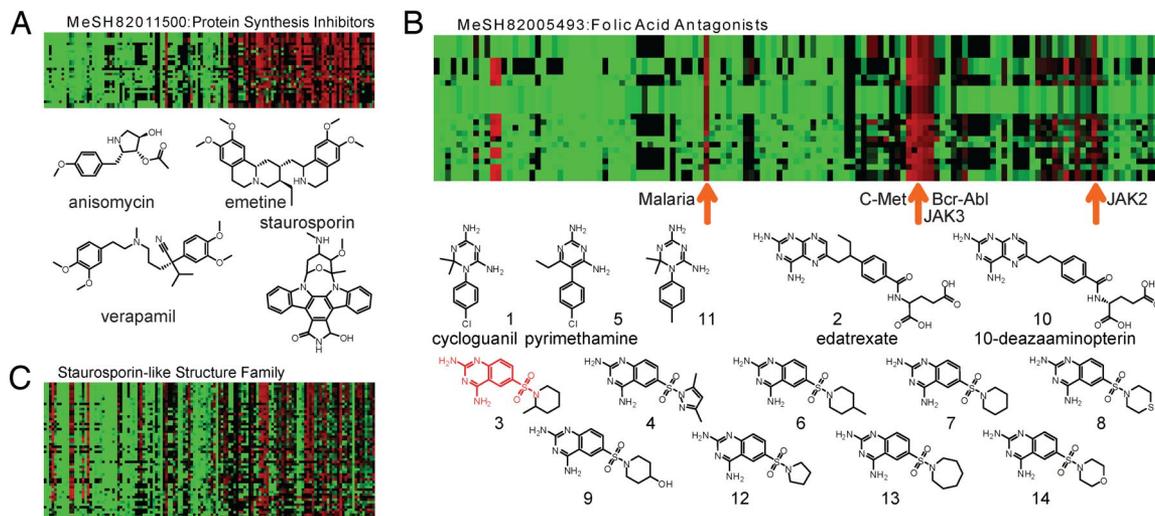
CHEMISTRY

MICROBIOLOGY

Fig. 3. Several statistically significant MOA and SAR clusters. (*A*) Cluster of 30 compounds containing 11 annotated protein synthesis inhibitors. (*B*) Cluster containing 26 compounds, among which 15 are antimalarials known to be folic acid antagonists. (*C*) SAR-driven cluster containing 44 staurosporine-like structures sharing the same profile.

One other notable group identified by the above MOA enrichment analysis consisted of 26 compounds, including several copies of the known antifolate malarial drug pyrimethamine (Fig. 3*C* and Table 2). This group also included the other antifolate drug cycloguanil, the cancer drug edatrexate, and 11 uncharacterized compounds. All of the characterized compounds in this group are known to act by directly inhibiting dihydrofolate reductase (DHFR), thus blocking folate biosynthesis, an essential intermediate in DNA metabolism (reviewed in ref. 26). The probability of finding 15 of the 16 known antifolates in the same cluster of 26 can be estimated at 1 in 10^{-39}. In addition to being active in the parasite proliferation assay, all 26 also had activity in several other assays that involved a cell line whose proliferation was made dependent on the activation of various tyrosine kinases such as the janus-related kinase 2, the janus-related kinase 3, Brc-Abl, or mesenchymal epithelial transition factor (c-met). Although it is feasible that the compounds could be hitting different targets within the folate biosynthesis pathway, most-common-substructure analysis (27) revealed that all of the compounds shared a distinct common pharmacophore, a diaminopyrimidine moiety (Fig. 3*C*), suggesting that all likely interact with the target of pyrimethamine and edatrexate, DHFR. Interestingly, many compounds with a diaminopyrimidine group have unrelated profiles or have no activity against parasites, indicating that this group is necessary but not sufficient for forming a complex with DHFR.

Testing Mechanism-of-Action Predictions by Using Docking Studies and Drug-Resistant Parasites. To test whether the uncharacterized compounds in the folate cluster, predicted by the clustering algorithm, might be interacting with *P. falciparum* DHFR, we tested them against Pf strain W2. This parasite isolate carries a triple point mutation (S108N/N51I/C59R) in DHFR that renders parasites 225 times more resistant to pyrimethamine (reviewed in ref. 28). With one exception, all compounds showed substantially less activity against W2 (Table 2). Because strain W2 is not isogenic with 3D7, is also resistant to quinoline-type antibiotics, and carries an amplification in the multidrug resistant transporter, these differences in sensitivity could be due to other mutations elsewhere in the genome. However, 59.4% of the antimalarial hits in this study were equipotent against strain W2, and thus these differences are unlikely to be due to other differences between the W2 and 3D7 genomes or other artifacts of screening. As further confirmation, we docked them to the active site of the *P. falciparum* DHFR structure

(29) using AutoDock (30). As can be seen in Fig. 4, all compounds fit well into the active site of DHFR. Interestingly, the 10-deazaaminopterin that was relatively active against W2 was not able to achieve this activity as a consequence of increased interactions with the cysteine in the W2 active site at position 10. Rather, it is the innate homology of the inhibitor to the native ligand, dihydrofolate, and the inherent flexibility derived from this design. Notably, this is also crucial to the success of the potent antimalarial WR99210, which also overcomes W2 active-site mutations (31). As further proof, searches of patent databases indicated that the compounds had previously been reported as potentially acting against DHFR (32). We thus conclude that in some cases compounds with similar activities across cellular screens have similar MOA and that this cheminformatic approach allowed the target of the 11 uncharacterized compounds to be determined by using guilt by association. These data also highlight the portion of the molecule that likely has critical interactions with the protein target and can thus accelerate structure–activity relationship studies.

Discussion

One of the challenges of modern drug discovery lies in determining the mechanism of action of active compounds identified in cell-based assays. Most biochemical approaches such as affinity chromatography, a method in which derivatized versions of the active compound are synthesized and used to purify proteins that bind the compound (33), typically require a high affinity between the target and lead. Furthermore, for some microbes, such as malaria parasites, large quantities of protein for biochemistry may be difficult to acquire. Genetic approaches have also been used to find the target of small molecules, but these are also relatively serial in nature, sometimes requiring long periods to select for resistance in the laboratory or the field (10). Unless one has candidate genes in mind, finding the gene bearing the mutation can be more even challenging, and one must remember that a gene involved in resistance to a compound may not be a compound's target. However, the availability of genome sequences and large collections of expression data have altered the way that we determine what genes are doing for the cell, and new methods have arisen that can be used to discover patterns in immense biological datasets. It seems likely that large datasets will prove just as useful in chemical biology. Here, we show that when enough high-throughput screening data are available, compounds may segregate based on their mechanism of action. This *in silico* compound activity profiling approach can also

Table 2. Primary hits that share a similar activity profile to known antifolates across 131 screens

ID	Structure (common name)	3D7, μM	W2, μM	Ref.
1 (four times)	Cycloguanil	0.011	>1.25	
2 (three times)	Edatrexate	0.027	0.111	
3		0.205	>1.25	32
4		0.071	>1.25	32
5 (five times)	Pyrimethamine	0.066	>1.25	
6		0.106	0.715	32
7		0.020	>1.25	32
8		0.116	>1.25	32
9		0.012	>1.25	32

Table 2. (continued)

ID	Structure (common name)	3D7, μM	W2, μM	Ref.
10	10-Deazaaminopterin	0.066	0.032	
11		0.011	0.742	36
12		0.060	>1.25	32
13		0.098	>1.25	32
14		0.047	>1.25	32

All, except one (not shown), were reconfirmed in subsequent testing. Compounds **1**, **2**, and **5** were present and were registered as hits multiple times.

National Institutes of Health Roadmap initiatives should improve the percentage over the time. In contrast, the large amount of information in data sources such as patent collections has yet to be

Fig. 4. Predicted mode of binding of the uncharacterized inhibitors to wild-type PfDHFR. Blind docking studies between the diaminopyrimidine class of inhibitors to the crystal structure of PfDHFR revealed a clustering of hits to the active site that shared the same general mode of binding. (*A*) For comparison, top hits for each of the 11 uncharacterized diaminopyrimidine class of inhibitors superimposed. (*B*) Top docking result for compound 5 (yellow) shown superimposed to the crystallized complex between NADPH (gray), the antifolate inhibitor WR99210 (red), and PfDHFR (green).

substitute for initial structure–activity relationship studies because both potency and selectivity are revealed simultaneously, further accelerating the process of drug development.

MOA prediction of uncharacterized compounds has traditionally depended on their structural similarity to known drugs; i.e., SAR. Our clustering analysis can also work if it starts with compounds sharing the same scaffold instead of MOA. For example, staurosporine is a compound that is active against a wide variety of protein kinases. Of the 8,457 hits, structure-clustering analysis showed 48 compounds similar to staurosporine (similarity ≥0.85, $P = 0.001$), and 44 of these were contained in an activity profile group of 55 compounds (Fig. 3*B*). The probability of this occurring by chance is $<10^{-100}$. However, in our MOA-driven antifolate prediction, the diaminopyrimidine group only shares a low similarity score of 0.30 ($P = 0.75$) and 0.48 ($P = 0.11$) with pyrimethamine and edatrexate, respectively. Their MOA cannot be predicted if the scaffold-similarity approach is taken.

One chief limitation of MOA-driven clustering analysis is the lack of chemical annotation for compounds that can be used in guilt-by-association methods; PubChem or MeSH annotations were only available for ≈5% of compounds in the screen. The fast developing

CHEMISTRY

MICROBIOLOGY

more effectively exploited. Another limitation in using historical screening data and guilt-by-association approaches is that many of the compounds that were discovered as active in the antimalarial screen have little activity in the other screens that we have run with our compound set. From a drug discovery point of view, these may be the most interesting because they are likely to be more selective. For malaria the goal is thus to discover more high-throughput assays that will eventually allow different compounds that currently have similar profiles to segregate away from one another. Such assays may include tests for the ability to kill gametocytes (artemisinin, primaquine), tests for the ability to inhibit liver stage development (primaquine, atovaquone), pathway-based assays, or enzymatic assays such as those that may be performed against targets involved in fatty acid biosynthesis or mitochondrial function (atovaquone). Similar approaches have been described in cancer, where a set of compounds can be run against a panel of 60 different cell line (34). The inclusion of screening data from a variety of different drug-resistant strains with different phenotypic sensitivities to compounds may also be revealing (such as the 3D7 and W2 screens performed here). Alternatively, testing each compound against a panel of isogenic parasite strains each of which overexpresses a particular protein could be informative in much the same way it has been informative in yeast (35). Automated high-content imaging of parasites that have been treated with different inhibitors may also provide clues about how the compounds are acting in the cell. Finally, the establishment of an infrastructure to share this large set of active compounds with the greater malaria research community may provide a method for further annotating the MOA of these compounds, as well as for accelerating antimalarial drug development.

Materials and Methods

Experimental Protocols. Approximately 1.7 million compounds were screened in 1,536-well format for antimalarial activity [see supporting information (SI) Materials and Methods]. Using Genomics Institute of the Novartis Research Foundation (GNF) on-line screening equipment, 3 μl of screening medium [RPMI (without phenol red, with L-glutamine), 4.16 mg/ml albumax, 0.013 mg/ml hypoxanthine, 1.73 mg/ml glucose, 0.18% NaHCO$_3$, 0.031M Hepes, 2.60 mM NaOH, 0.043 mg/ml gentamicin] was dispensed into 1,536-well, black, clear-bottom plates (Greiner). With a PinTool (GNF Systems), 10 nl of compound in DMSO was transferred into the assay plates along with control compounds (1.25 μM in 0.125% DMSO). Next, 5 μl of the parasite suspension in screening medium (see above) was dispensed into the assay plates such that the final parasitemia was 0.3% and the final hematocrit was 2.5%. Mefloquine (12.5 μM) and DMSO (0.125%) were used within the assay plates to serve as background and baseline controls, respectively. The assay plates were transferred to off-line incubators that contained airtight incubation units. The units were gassed daily with 93% nitrogen, 4% carbon dioxide, and 3% oxygen during the 72-h incubation at 37°C. Two microliters of detection reagent consisting of 10$\times$ SYBR Green I (Invitrogen; supplied in 10,000$\times$ concentration) in lysis buffer (20 mM Tris·HCl, 5 mM EDTA, 0.16% Saponin wt/vol, 1.6% Triton X vol/vol) was dispensed into the assay plates. The assay plates were left at room temperature for 24 h and read off-line by using several Acquest GT multimode readers (Molecular Devices).

Data Analysis. An activity matrix of 8,457 compounds across 131 biological assays was constructed by using the GNF in-house HTS database (complete details are available in SI Materials and Methods). Compounds were clustered into scaffold families based on their structural similarities, and their mode-of-action annotations were retrieved from the MeSH database whenever possible through an in-house compound annotation pipeline. Each scaffold family and each MeSH term served as a piece of independent knowledge that guided the ontology-based pattern identification (OPI) algorithm to group the compounds into statistically significant clusters that share unusually similar activity profiles. A total of 530 scaffold families and 62 MeSH categories were studied. A cluster of 26 compounds enriched in antifolates was identified; docking simulations were carried out to study the binding conformations of predicted compound members of that MOA.

1. Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL (2007) Drugs for bad bugs: Confronting the challenges of antibacterial discovery. *Nat Rev Drug Discovery* 6:29–40.
2. Smilkstein M, *et al.* (2004) Simple and inexpensive fluorescence-based technique for high-throughput antimalarial drug screening. *Antimicrob Agents Chemother* 48:1803–1806.
3. Desjardins RE, Canfield CJ, Haynes JD, Chulay JD (1979) Quantitative assessment of antimalarial activity in vitro by a semiautomated microdilution technique. *Antimicrob Agents Chemother* 16:710–718.
4. Elabbadi N, Ancelin ML, Vial HJ (1992) Use of radioactive ethanolamine incorporation into phospholipids to assess in vitro antimalarial activity by the semiautomated microdilution technique. *Antimicrob Agents Chemother* 36:50–55.
5. Iber PK, Pavanand K, Wilks NE, Colwell EJ (1975) Evaluation of in vitro drug sensitivity of human Plasmodium falciparum by incorporation of radioactive isoleucine. *J Med Assoc Thailand* 58:559–566.
6. Baniecki ML, Wirth DF, Clardy J (2007) High-throughput Plasmodium falciparum growth assay for malaria drug discovery. *Antimicrob Agents Chemother* 51:716–723.
7. Bennett TN, *et al.* (2004) Novel, rapid, and inexpensive cell-based quantification of antimalarial drug efficacy. *Antimicrob Agents Chemother* 48:1807–1810.
8. Druilhe P, *et al.* (2001) A colorimetric in vitro drug sensitivity assay for Plasmodium falciparum based on a highly sensitive double-site lactate dehydrogenase antigen-capture enzyme-linked immunosorbent assay. *Am J Trop Med Hyg* 64:233–241.
9. Corbett Y, *et al.* (2004) A novel DNA-based microfluorimetric method to evaluate antimalarial drug activity. *Am J Trop Med Hyg* 70:119–124.
10. Johnson JD, *et al.* (2007) Assessment and continued validation of the malaria SYBR green I-based fluorescence assay for use in malaria drug screening. *Antimicrob Agents Chemother* 51:1926–1933.
11. Chanda SK, Caldwell JS (2003) Fulfilling the promise: Drug discovery in the postgenomic era. *Drug Discovery Today* 8:168–174.
12. Zhou Y, *et al.* (2007) Large-scale annotation of small-molecule libraries using public databases. *J Chem Inf Model* 47:1386–1394.
13. Chong CR, *et al.* (2006) A clinical drug library screen identifies astemizole as an antimalarial agent. *Nat Chem Biol* 2:415–416.
14. Weisman JL, *et al.* (2006) Searching for new antimalarial therapeutics amongst known drugs. *Chem Biol Drug Design* 67:409–416.
15. Boss C, *et al.* (2003) Inhibitors of the Plasmodium falciparum parasite aspartic protease plasmepsin II as potential antimalarial agents. *Curr Med Chem* 10:883–907.
16. Liu J, *et al.* (2006) Plasmodium falciparum ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proc Natl Acad Sci USA* 103:8840–8845.
17. Steele JC, Warhurst DC, Kirby GC, Simmonds MS (1999) In vitro and in vivo evaluation of betulinic acid as an antimalarial. *Phytother Res* 13:115–119.
18. Webster HK, Whaun JM (1982) Antimalarial properties of bredinin. Prediction based on identification of differences in human host-parasite purine metabolism. *J Clin Invest* 70:461–469.
19. Sufrin JR, *et al.* (1995) Methionine recycling pathways and antimalarial drug design. *Antimicrob Agents Chemother* 39:2511–2515.
20. Yan SF, Asatryan H, Li J, Zhou Y (2005) Novel statistical approach for primary high-throughput screening hit selection. *J Chem Inf Model* 45:1784–1790.
21. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Delivery Rev* 46:3–26.
22. Zhou Y, *et al.* (2008) Evidence-based annotation of the malaria parasite's genome using comparative expression profiling. *PLoS ONE* 3:e1570.
23. Weir MR, Peppler R, Gomolka D, Handwerger BS (1993) Calcium channel blockers inhibit cellular uptake of thymidine, uridine and leucine: The incorporation of these molecules into DNA, RNA and protein in the presence of calcium channel blockers is not a valid measure of lymphocyte activation. *Immunopharmacology* 25:75–82.
24. Brinckerhoff CE, Lubin M (1977) Prolonged inhibition of protein and glycoprotein synthesis in tumor cells treated with muconomycin A. *J Natl Cancer Inst* 58:605–609.
25. Cohen MB, Glazer RI (1985) Cytotoxicity and the inhibition of ribosomal RNA processing in human colon carcinoma cells. *Mol Pharmacol* 27:308–313.
26. Yuthavong Y, Kamchonwongpaisan S, Leartsakulpanich U, Chitnumsub P (2006) Folate metabolism as a source of molecular targets for antimalarials. *Future Microbiol* 1:113–125.
27. Ghose AK, Herbertz T, Salvino JM, Mallamo JP (2006) Knowledge-based chemoinformatic approaches to drug discovery. *Drug Discovery Today* 11:1107–1114.
28. Gregson A, Plowe CV (2005) Mechanisms of resistance of malaria parasites to antifolates. *Pharmacol Rev* 57:117–145.
29. Yuvaniyama J, *et al.* (2003) Insights into antifolate resistance from malarial DHFR-TS structures. *Nat Struct Biol* 10:357–365.
30. Morris GM, *et al.* (1998) Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. *J Comput Chem* 19:1639–1662.
31. Yuthavong Y, *et al.* (2005) Malarial (Plasmodium falciparum) dihydrofolate reductase-thymidylate synthase: Structural basis for antifolate resistance and development of effective inhibitors. *Parasitology* 130:249–259.
32. Elslager EF, *et al.* (1984) Folate antagonists. 22. Antimalarial and antibacterial effects of 2,4-diamino-6-quinazolinesulfonamides. *J Med Chem* 27:1740–1743.
33. Knockaert M, *et al.* (2000) Intracellular targets of cyclin-dependent kinase inhibitors: Identification by affinity chromatography using immobilised inhibitors. *Chem Biol* 7:411–422.
34. Weinstein JN, *et al.* (1997) An information-intensive approach to the molecular pharmacology of cancer. *Science* 275:343–349.
35. Giaever G, *et al.* (1999) Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat Genet* 21:278–283.
36. Hansch C, Dietrich SW, Fukunaga JY (1981) Inhibition of bovine and rat liver dihydrofolate reductase by 4,6-diamino-1,2-dihydro-2,2-dimethyl-1-(4-substituted-phenyl)-s-triazines. *J Med Chem* 24:544–549.

Plouffe *et al.*