ACS Medicinal Chemistry Letters

LETTER

pubs.acs.org/acsmedchemlett

An Invitation to Open Innovation in Malaria Drug Discovery: 47 Quality Starting Points from the TCAMS

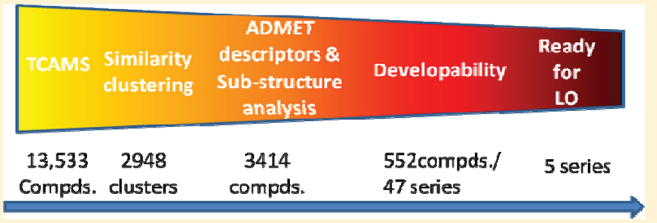
Félix Calderón,[*,†] David Barros,[†] José María Bueno,[†] José Miguel Coterón,[†] Esther Fernández,[†] Francisco Javier Gamo,[†] José Luís Lavandera,[†] María Luisa León,[†] Simon J. F. Macdonald,[‡,§] Araceli Mallo,[†] Pilar Manzano,[†] Esther Porras,[†] José María Fiandor,[*,†] and Julia Castro[*,†]

[†]Tres Cantos Medicines Development Campus, DDW, GlaxoSmithKline, Severo Ochoa 2, 28760 Tres Cantos, Spain

[‡]Medicines for Malaria Venture (MMV), 20, route de Pré-Bois-PO Box 1826, 1215 Geneva 15, Switzerland

Supporting Information

ABSTRACT: In 2010, GlaxoSmithKline published the structures of 13533 chemical starting points for antimalarial lead identification. By using an agglomerative structural clustering technique followed by computational filters such as antimalarial activity, physicochemical properties, and dissimilarity to known antimalarial structures, we have identified 47 starting points for lead optimization. Their structures are provided. We invite potential collaborators to work with us to discover new clinical candidates.



KEYWORDS: Malaria, TCAMS, lead optimization, open innovation

Malaria is a major global disease caused by parasites of the genus *Plasmodium*, which are transmitted to people when female anopheles mosquitoes feed on human blood.[1] In 2009, more than 200 million of cases of malaria were reported, causing nearly 1 million deaths, mostly among pregnant women and young children,[2] with *Plasmodium falciparum* and *Plasmodium vivax* being primarily responsible for the mortality and morbidity, respectively.[1−3] In regions where malaria is endemic, the humanitarian and economic burdens are considerable, and in 2007, the Bill and Melinda Gates Foundation, supported by other global health agencies, initiated an agenda, the ultimate aim of which is the eradication of malaria.

The reasons for the mortality and morbidity are varied and include access to medicines. However, *P. falciparum* resistant strains to standard antimalarial drugs have also developed.[4] As a consequence, there is an urgent requirement for new antimalarial drug, and this has triggered a great number of drug discovery and development programs from public institutions, private institutions, and public−private partnerships.[5,6] Many new antimalarial compounds and mechanisms have been investigated.[7,8]

The criteria for new antimalarial drugs are demanding; first, the drug must be safe and efficacious. Then, the profile of a new molecule should be better than that of existing drugs, it should be affordable (less than $1/treatment for an adult), and it should be active against resistant strains. Moreover, with the eradication strategy in mind, one component of the profile should ideally include activity against the hepatic or mosquito stage of the parasite lifecycle.

Most current antimalarial therapies generally only operate against four metabolic pathways of the parasite.[9] In contrast, sequencing of the *P. falciparum* genome has revealed more than 5000 genes, and a significant number of these genes are expected to encode for proteins that are essential for the intraerythrocytic stages of the parasite.[9]

Where should the search for leads for new antimalarial drugs start? In visionary initiatives, in 2010, St Jude's Children's Research Hospital[10] and Novartis[11] both published the structures of thousands of compounds that inhibit parasite growth, which represents a step change in the number of leads available for drug discovery programs. Also, in 2010, we at GlaxoSmithKline (GSK) published the Tres Cantos Antimalarial Set (TCAMS), 13533 compounds that are the result of screening nearly 2 million compounds from the GSK corporate collection.[12] The three sets of compounds are available for download from the Chembl-NTD database (http://www.ebi.ac.uk/chemblntd).

Being able to select a high-quality series for lead optimization from over 13000 potential starting points presents both an unprecedented opportunity and also a challenge for the medicinal chemist community. A clear strategy is required to rapidly identify those molecules that have both the best chance of being converted into differentiated antimalarial drugs and that are also likely to have the lowest risk of attrition in development. This letter describes our strategy in mining TCAMS to identify potential starting points for lead optimization programs. An elegant structure−activity relationship (SAR) analysis of TCAMS has already been described by Wawer and Bajorath;[13] however, our priority in this work is different in that we use

Received: June 8, 2011
Accepted: August 3, 2011
Published: August 03, 2011

ACS Publications © 2011 American Chemical Society

criteria designed to identify high-quality starting points suitable for oral drug discovery. These criteria have dictated the mining and filtering processes that we have used. This is just one way of mining TCAMS and one that may deprioritize compounds that represent perfectly good starting points if different mining or filtering processes were used or if exploratory SAR studies were carried out. Our aim was to select no more than five potential starting points for lead optimization. The main characteristics that the selected starting points should ideally possess are (a) the scaffold (chemotype) should be structurally different to known antimalarial scaffolds as various parasite strains are resistance to many of these drugs; (b) high tractability to facilitate rapid lead optimizations programs; (c) physicochemical profiles that are compatible with good oral absorption[14] and reasonable aqueous solubility; (d) no known toxicity issues; (e) druglike functionality; (f) no known intellectual property issues; and last but not least, (g) moderate to good antiplasmodial activity. We note in particular that being able to *control* and if necessary *reduce* lipophilicity is an important criterion in c and d,[15] and while low lipophilicity is desirable in the starting point, it is not essential.

The first step was to carry out a clustering exercise with all of the compounds in the set. From the plethora of methods available, we chose a standard agglomerative clustering technique to facilitate analysis of the data set using structural similarity tools. The "structural similarity" between clusters is defined by the Tanimoto similarity index between two compounds using Daylight fingerprints.[16] Following this algorithm, we obtained 2948 clusters. Most of the clusters contained less than 20 compounds, and 1120 clusters contain only one compound (in other words, they are singletons).[17]

Having clustered the TCAMS, we started the filtering process by assigning a score of up to 30 points to each compound in the clusters. We chose to prioritize compounds initially by their *P. falciparum* (3D7 strain) antimalarial activity by progressively assigning more points to the more potent compounds up to a maximum of 15 points for the most potent compounds. Compounds having an XC_{50}[12] less than 0.01 μM were assigned 15 points, those with an XC_{50} between 0.010 and 0.1 μM were assigned 10 points, those between 0.1 and 1 μM were assigned 4 points, and those between 1 and 2 μM were assigned 1 point. The second priority was the lipophilicity (clogP) of the compound as a preliminary measure of drug-likeness—compounds with a clogP < 4 scored 6 points; those with a clogP between 4 and 6 scored 3 points; and those with clogP > 6 scored 1 point. The third priority was given to the number of compounds in each cluster to prioritize clusters containing numerous compounds over those containing very few or only one (singletons). Thus, where a cluster contained more than 20 compounds, 6 points were added to the score of each compound in that cluster; where the cluster contained between 5 and 20 compounds, 3 points were added to each compound; and those containing fewer than 5 compounds, only 1 point was added. We judged that the more compounds a series contained, the more evidence there was that that cluster was likely to be amenable to lead optimization and that the overall profile of a lead compound could be improved by chemical modification. In contrast, with singletons, there is less *immediate* evidence that the overall profile of the compound can be improved. Note, however, that very potent singletons were still able to survive the filtering process. Finally, we also scored compounds by their molecular weight (<400, 3 points; 400−600, 2 points; and >600, 1 point).

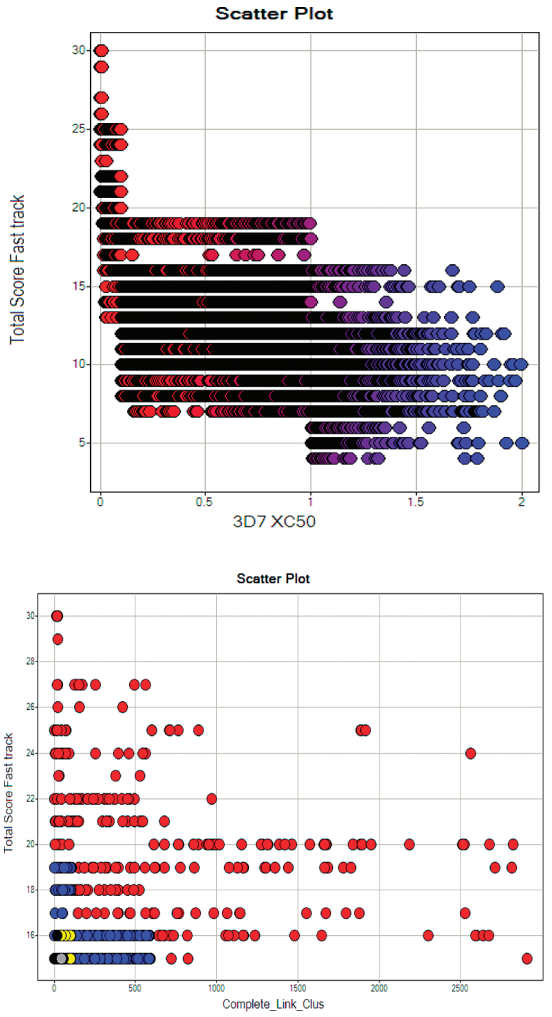


Figure 1. Top panel shows *P. falciparum* 3D7 XC_{50} vs total score (Total Score Fast-track) (see the text). Each spot represents a compound with the color of each compound/spot reflecting its potency with a gradient from most potent in red (≤0.5 μM) to least potent in blue (0.51−1 μM). Note that many compounds overlap resulting in black coloring. As might be expected, the compounds with the highest scores tend to be the most potent. The bottom panel shows those compounds scoring more than 15 with cluster number (Complete_Link_Clus) plotted vs score (Total Score Fast-track). The colors of the compounds reflect their potency with XC_{50} ≤0.5 (red), 0.51−1 (blue), 1−1.5 (yellow), and 1.51−5 μM (black). Compounds appearing in the same column are members of the same cluster. Note that some clusters only contain one compound (singletons).

Having scored all of the compounds, we discarded all of those scoring 15 points or less; only 3414 compounds in 467 clusters passed these filters (Figure 1). At this point in the filtering process, all of the surviving compounds should have a combination of good druglike properties, namely, their potency, lipophilicity, and molecular weight. A poor score against one property will be offset by an excellent score against another property. For example, any remaining singletons are likely to be highly potent.

In an attempt to improve the quality of the remaining compounds, the set was filtered further to remove less "druggable" molecules using filters, which are based on a selection of simple calculated molecular properties. These filters include assessments of heavy atom count (>60), the ratio of the total

number of nitrogen, oxygen, and sulfur atoms to carbon atoms (>1.5), the number of halogen atoms (>4),[18] lipophilicity (clogP > 6 without ionizable groups), and substructure filtering of well-known nondesiderable groups for stability or toxicity reasons were removed, including aldehydes, carboxylic acids, aromatic nitro groups, alkyl chains ≥3, primary alkyl amines, methoxy or ethoxy aromatics groups, phenols, and guanidine derivatives.

Next, using simple substructure searches, we removed those structures that were related to known antimalarials[17] to reduce the risk of developing a series with high potential for resistance. This removed approximately 3000 further compounds.

At this point, there were 322 compounds that remained grouped into 47 clusters comprising both series and singletons. Having identified the 47 top clusters, we added back into these clusters all of the compounds that had scored <15, taking the total number back up to 552 or so. The most druglike (often the most potent) representatives are shown (Figure 2). Inspection of a plot of the potency of these cluster representatives against

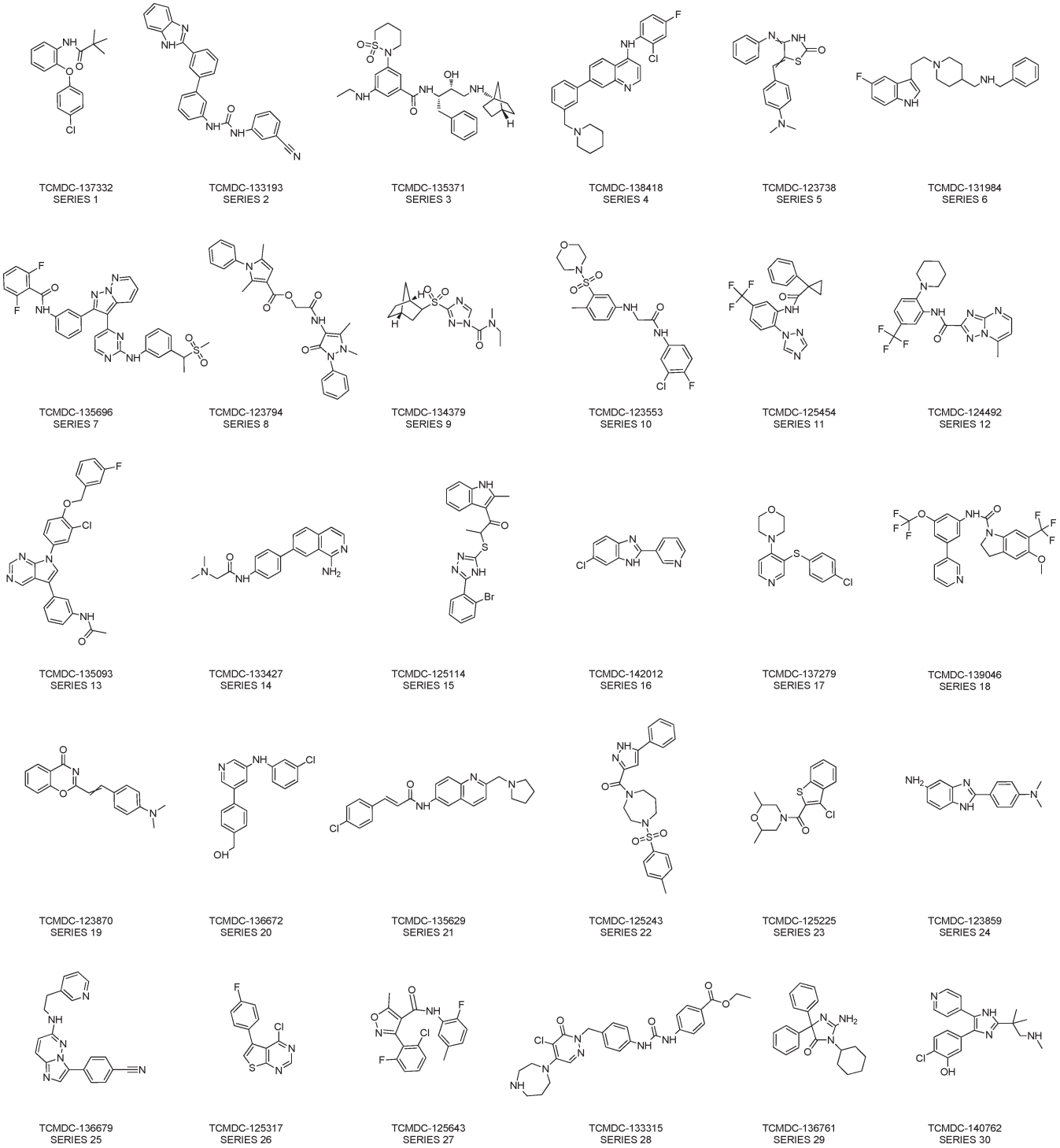


Figure 2. Continued

dx.doi.org/10.1021/ml200135p |*ACS Med. Chem. Lett.* 2011, 2, 741–746

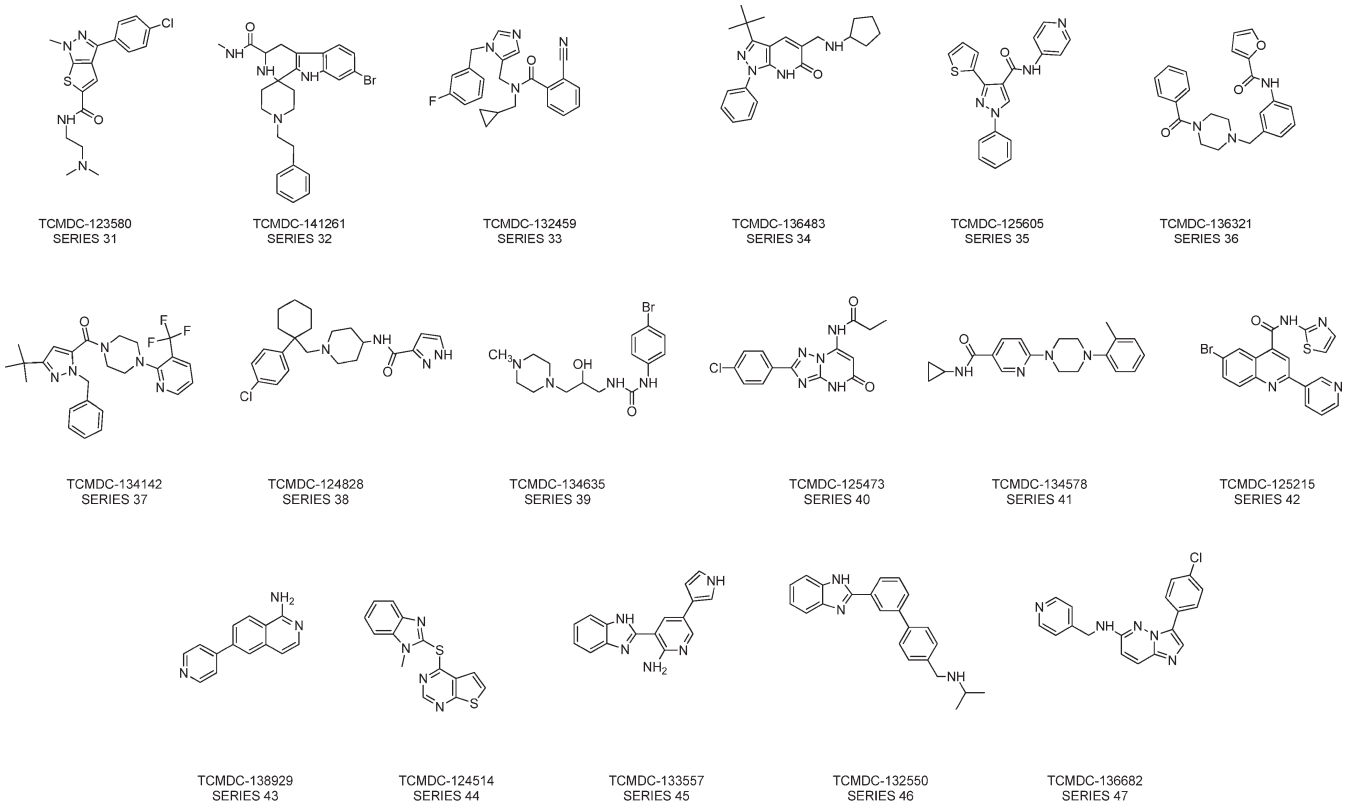


Figure 2. Most druggable/interesting structures from the 47 clusters that remain after the filtering process. Note that some of the scaffolds in these different series (clusters) are structurally very similar (e.g., series 16 and 24)—here, the algorithm has focused on the dissimilarity of the peripheral functionality.

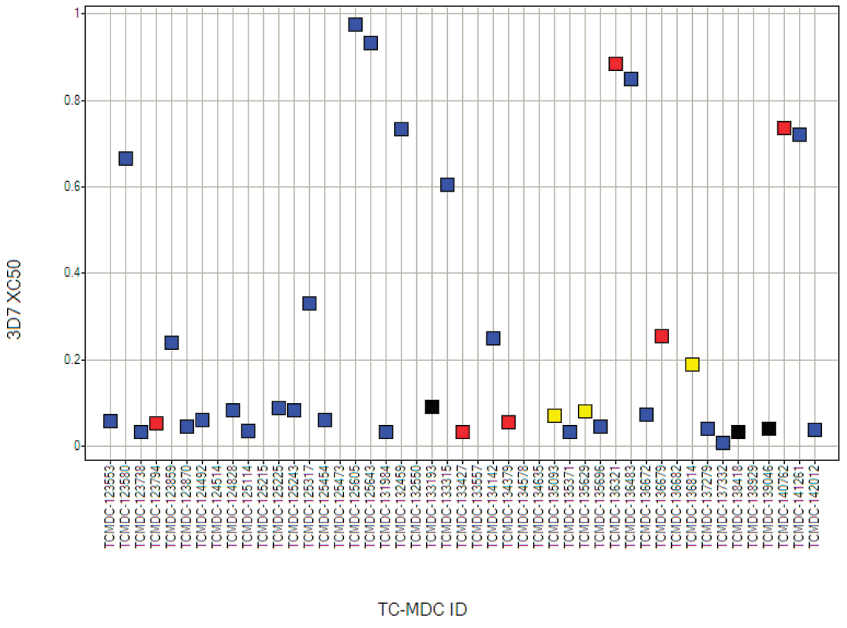


Figure 3. Plot showing the representatives of the 47 selected series comparing potency/clogP (red, cLogP ≤ 3; blue, clogP 3.1−5; yellow, clogP 5.1−6; and black, clogP > 6).

lipophilicity (Figure 3) shows most with excellent potency (XC_{50} < 100nM) and moderate to good lipophilicity. A list of the SMILES structures of these ∼500 compounds together with their cluster number, TCAMS reference number, pXC_{50}, and other physiochemical/structural descriptors is provided in the Supporting Information.

dx.doi.org/10.1021/ml200135p |*ACS Med. Chem. Lett.* 2011, 2, 741–746

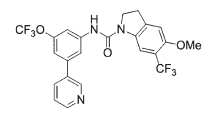
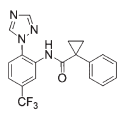
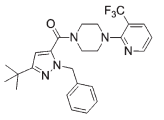
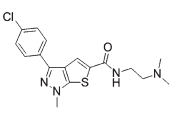
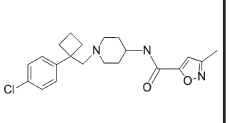
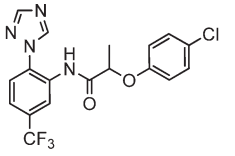
					
	TCMDC-139046	**TCMDC-125454**	**TCMDC-134142**	**TCMDC-123580**	**TCMDC-124833**
MW	497.4	372.4	471.5	362.9	387.9
clogP	6.18	3.49	4.61	3.58	3.6
Pf XC60 μM	0.04	0.06	0.25	0.66	0.09
Pf IC50 μM	0.08	0.03	0.10	0.95	0.08
hERG μM	>15	>50	>50	6.3	-
Cl mL/min.g (mouse/ human)	<0.3 / <0.3	<0.8 / <0.3	>30 / 12.7	4.0 / 0.3	5.6 / 18.8

Figure 4. Exemplar compounds and their profiles for the top five selected series. For comparison, chloroquine has *Pf* IC_{50} = 0.024 μM.

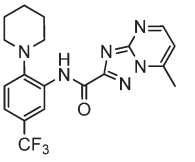
Forty-seven clusters of compounds were regarded as a more manageable number for a nonautomated inspection and selection process. So, in the last part of the filtering process, we manually inspected each cluster according to the following criteria: (a) the antiplasmodial activity based on its XC_{50}, (b) the "druggability" or appeal of the structure to the experienced medicinal chemist, (c) an assessment of the synthetic tractability of the compound, (d) the availability of SAR around the compound from within TCAMS, and (e) searching the literature for information about the compound and related substructures (e.g., had they previously been described as having antimalarial activity, were they in a crowded intellectual property space, or did they have activity against known biological or other targets?). This ultimately resulted in the selection of five series featuring indolines (series 18), aryl carboxamides (series 11), alkylpyrazoles (series 37), thienopyrazoles (series 31), and 4-aminopiperidines (series 38); they complied with our three main requirements, namely, potency, druggability, and tractability (Figure 4). All five series have undergone preliminary exploration at GSK.

In the indoline series, the exemplar TCMDC-139046, is highly potent compound with a *Pf* IC_{50} = 80 nM. While the molecular weight and lipophilicity are high, it has little cytotoxicity or hERG activity and is metabolically stable in mouse and human microsomes. This series is also known to have activity against the 5-HT2c receptor.[19] Remining TCAMS for analogues of TCMDC-139046, 62 compounds are found (the majority of which failed to survive the 15 point cutoff described earlier) having *Pf* XC_{50} values of 0.04−2.0 μM. Within the series are urea-linked indolines (*Pf* XC_{50} = 0.04−1.15 μM) and the analogous amide-linked indolines (*Pf* XC_{50} = 0.28−1.18 μM).

The exemplar compound from the second series termed the "aryl carboxamides" is TCMDC-125454, which is also very potent with *Pf* IC_{50} = 60 nM and is in fact a singleton. The cytotoxicity, hERG, and microsomal stability profiles of TCMDC-125454 are all druglike. We found other related carboxamide analogues in TCAMS (TCMDC-125023 and TCMDC-124492 (series 12) (Figure 5). All are attractive from a medicinal chemistry viewpoint for their low molecular weights (339−429), lipophilicity (clogP 2.48−4.72), and hydrogen bond acceptor and donor counts. All of the analogues feature a trisubstituted left-hand side aromatic ring linked through an amide in various ways to a right-hand aromatic ring.



TCMDC-125023
4 related aryloxy-lactic acid amides in TCAMS



TCMDC-124492
4 related aryloxy-lactic acid amides in TCAMS

Figure 5. Analogues of TCMDC-125454 in the aryl carboxamide series found in TCAMS.

The third series—the alkylpyrazoles—is exemplified by TCMDC-134142 with a *Pf* IC_{50} = 0.1 μM. It represents a novel antimalarial structural scaffold with an excellent cytotoxicity and hERG profile, with moderate lipophilicity and molecular weight, and poor microsomal stability. We found six analogues of TCMDC-134142 in TCAMS with a potency range of *Pf* XC_{50} of 0.18−2 μM.

The last two series are the thienopyrazoles exemplified by TCMDC-123580 *Pf* IC_{50} = 0.95 μM and 4-aminopiperidines series exemplified by TCMDC-124833 *Pf* XC_{50} = 0.09 μM. The comparatively lower molecular weights and lipophilicities of these exemplars are very attractive with the aminopiperidine having cytotoxicity, hERG, and microsomal stability profiles, which are druglike. Other analogues of both the thienopyrazoles (six analogues) and the aminopiperidine series (two analogues) were also found in TCAMS.

Having completed this selection process, we briefly investigated whether using other selection criteria such as ligand efficiency (LE) would have selected different chemical classes. We therefore selected those compounds from the TCAMS with a LE ≥ 0.35 (pXC_{50} × 1.37/number of heavy atoms) and belonging to a cluster with five or more compounds. We then applied the same structural/developability filters described above to these compounds, which generated 27 clusters (shown in the Supporting Information). Interestingly, 20 of these 27 chemical classes (75%) were also selected, by the "fast-track" approach described above. This perhaps suggests that while mining the TCAMS in different ways will lead to identification of different chemical classes, the overlap in finding the same series is expected to be high.

dx.doi.org/10.1021/ml200135p |*ACS Med. Chem. Lett.* 2011, 2, 741–746

The next step would be to assess the in vivo efficacy of the different selected series. Currently, we are following this strategy using a *P. berghei* mouse model adapted for screening to validate this theoretical analysis. The description of the methodology and results of the screen will be object of another publication.

In conclusion, we have described a fast-track filtering process of the TCAMS with a view to identifying high-quality starting points for lead optimization efforts. We have detailed here 552 such starting points, which are available as a resource for malaria drug discovery efforts. Furthermore, fuller profiles of exemplars from the five prioritized series are detailed. As with the publication of TCAMS, the publication of these chemical series is to contribute toward open-innovation in malaria drug discovery. We are unable to pursue lead optimization of 47 series simultaneously, and we invite and encourage other groups to collaborate with us in developing these series.

■ ASSOCIATED CONTENT

Supporting Information. Details of the ∼500 compounds identified from the filtering process. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: felix.r.calderon-romo@gsk.com.

Present Addresses

§GlaxoSmithKline, Gunnels Wood Road, Stevenage SG1 2NY.

Funding Sources

GlaxoSmithKline acknowledges financial support from Medicines for Malaria Venture (MMV).

■ ACKNOWLEDGMENT

GlaxoSmithKline acknowledges the perceptive and helpful advice from Jeremy Burrows and Mike Witty.

■ REFERENCES

(1) Greenwood, B. M.; Fidock, D. A.; Kyle, D. E.; Kappe, S. H. I.; Alonso, P. L.; Collins, F. H.; Duffy, P. E. Malaria: Progress, Perils and Prospects for Eradication. *J. Clin. Invest.* **2008**, *118*, 1266–1276.

(2) World Health Organization. World malaria report; http://www.who.int/malaria/publications/atoz/9789241563901/en/index.htmlæ, 2009.

(3) Anstey, N. M.; Russell, B.; Yeo, T. W.; Price, R. N. The Pathophysiology of vivax Malaria. *Trends Parasitol.* **2009**, *25*, 220–227.

(4) WHO. *Global Report on Antimalarial Drug Efficacy and Drug Resistance: 2000−2010*; WHO Press: Geneva, Switzerland, 2010.

(5) Nishtar, S. Public−Private "Partnerships" in Health—A Global Call to Action. *Health Res. Policy Syst.* **2004**, *2*, 5.

(6) Leroy, D.; Doerig, C. Drugging the Plasmodium Kinome: The Benefits of Academia−Industry Synergy. *Trends Pharmacol. Sci.* **2008**, *29*, 241–249.

(7) Lucumi, E.; Darling, C.; Jo, H.; Napper, A. D.; Chandramohanadas, R.; Fisher, N.; Shone, A. E.; Jing, H.; Ward, S. A.; Biagini, G. A.; DeGrado, W. F.; Dimond, S. L.; Greenbaum, D. C. Discovery of potent small molecule inhibitors of multi-drug resistant *Plasmodium falciparum* using a novel miniaturized high-throughput luciferase-based assay. *Antimicrob. Agents Chemother.* **2010**, *276*, 128–134.

(8) Baragana, B.; McCarthy, O.; Sanchez, P.; Bosch-Navarrete, C.; Kaiser, M.; Brun, R.; Whittingham, J. L.; Roberts, S. M.; Zhou, X.-X.; Wilson, K. S.; Johansson, N. G.; Gonzalez-Pacanowska, D.; Gilbert, I. H. Beta-Branched acyclic nucleoside analogues as inhibitors of *Plasmodium falciparum* dUTPase. *Bioorg. Med. Chem. Lett.* **2011**, *19*, 2378–2391.

(9) Gardner, M. J.; Hall, N.; Fung, E.; White, O.; Berriman, M.; Hyman, R. W.; Carlton, J. M.; Pain, A.; Nelson, k. E.; Bowman, S.; Paulsen, I. T.; James, K.; Eisen, J. A.; Rutherford, K.; Salzberg, S. L.; Craig, A.; Kyes, S.; Chan, M -S.; Nene, V.; Shallom, S. J.; Suh, B.; Peterson, J.; Angiuoli, S.; Pertea, M.; Allen, J.; Selengut, J.; Haft, D.; Mather, M. W.; Vaidya, A. B.; Martin, D. M. A.; Fairlamb, A. H.; Fraunholz, M. J.; Roos, D. S.; Ralph, S. A.; McFadden, G. I.; Cummings, L. M.; Subramanian, G. M.; Mungall, C.; Venter, J. C.; Carucci, D. J.; Hoffman, S. L.; Newbold, C.; Davis, R. W.; Fraser, C. M.; Barrell, B. Genome Sequence of the Human Malaria Parasite *Plasmodium falciparum*. *Nature* **2002**, *419*, 498–511.

(10) Guiguemde, W. A; Shelat, A. A.; Bouck, D.; Duffy, S.; Crowther, G. J.; Davis, P. H.; Smithson, D. C.; Connelly, M.; Clark, J.; Zhu, F.; Jimenez-Diaz, M. B.; Martinez, M. S.; Wilson, E. B.; Tripathi, A. K.; Gut, J.; Sharlow, E. R.; Bathurst, I.; Mazouni, F. E.; Fowble, J. W.; Forquer, I.; McGinley, P. L.; Castro, S.; Angulo-Barturen, I.; Ferrer, S.; Rosenthal, P. J.; DeRisi, J. L.; Sullivan, D. J.; Lazo, J. S.; Roos, D. S.; Riscoe, M. K.; Phillips, M. A.; Rathod, P. K.; Van Voorhis, W. C.; Avery, V. M.; Guy, R. K. Chemical genetics of *Plasmodium falciparum*. *Nature* **2010**, *465*, 311–315.

(11) Plouffe, D.; Brinker, A.; McNamara, C.; Henson, K.; Kato, N.; Kuhen, K.; Nagle, A.; Adrian, F.; Matzen, J. T.; Anderson, P.; Nam, T. G.; Gray, N. S.; Chatterjee, A.; Janes, J.; Yan, S. F.; Trager, R.; Caldwell, J. S.; Schultz, P. G.; Zhou, Y.; Winzeler, E. A. In silico activity profiling reveals the mechanism of action of antimalarials discovered in a high-throughput screen. *Proc. Natl. Acad. Sci.* **2008**, *105*, 9059–9064.

(12) Gamo, F. J.; Sanz, L. M.; Vidal, J.; de Cozar, C.; Alvarez, E.; Lavandera, J. L.; Vanderwall, D. E.; Green, D. V. S.; Kumar, V.; Hasan, S.; Brown, J. R.; Peishoff, C. E.; Cardon, L. R.; Garcia-Bustos, J. F. Thousands of Chemical Starting Points for Antimalarial Lead Identification. *Nature* **2010**, *465*, 305–312.

(13) Wawer, M.; Bajorath, J. Extracting SAR information from a large collection of anti-malarial screening hits Large Collection of Anti-Malarial Screening Hits by NSG-SPT analysis Analysis. *ACS Med. Chem. Lett.* **2011**, *2*, 201–206.

(14) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3–25.

(15) Leeson, P. D.; Springthorpe, B. The Influence of Drug-Like Concepts on Decision-Making in Medicinal Chemistry. *Nature Rev. Drug Discovery* **2007**, *6*, 881–890.

(16) Daylight Chemical Information Systems, Inc. *Daylight Theory Manual*; www.daylight.com/dayhtml/doc/theory/index.html, 2008.

(17) See the Supporting Information for more details.

(18) This count attempts to reflect the importance of different halogens: A fluorine (not part of a trifluromethyl group) counts one-half. A chlorine (not part of a trichloromethyl group), bromine, and iodine count one. A trifluro- or trichloromethyl group counts one.

(19) Bromidge, S. M.; Dabbs, S.; Davies, D. T.; Duckworth, D. M.; Forbes, I. T.; Ham, P.; Jones, G. E.; King, F. D.; Saunders, D. V.; Starr, S.; Thewlis, K. M.; Wyman, P. A.; Blaney, F. E.; Naylor, C. B.; Bailey, F.; Blackburn, T. P.; Holland, V.; Kennett, G. A.; Riley, G. J.; Wood, M. D. Novel and Selective 5-HT2C/2B Receptor Antagonists as Potential Anxiolytic Agents: Synthesis, Quantitative Structure-Activity Relationships, and Molecular Modeling of Substituted 1-(3-Pyridylcarbamoyl)indulines. *J. Med. Chem.* **1998**, *41*, 1598–1612.