An Open Drug Discovery Competition: Experimental Validation of Predictive Models in a Series of Novel Antimalarials

Edwin G. Tse[1], Laksh Aithani[2], Mark Anderson[3], Jonathan Cardoso-Silva[4], Giovanni Cincilla[5], Gareth J. Conduit[6,7], Mykola Galushka[8], Davy Guan[9], Irene Hallyburton[3], Benedict W. J. Irwin[7,10], Kiaran Kirk[11], Adele M. Lehane[11], Julia C. R. Lindblom[11], Raymond Lui[9], Slade Matthews[9], James McCulloch[12], Alice Motion[13], Ho Leung Ng[14], Mario Öeren[10], Murray N. Robertson[15], Vito Spadavecchio[16], Vasileios A. Tatsis[2], Willem P. van Hoorn[2], Alexander D. Wade[7], Thomas M. Whitehead[6], Paul Willis[17] and Matthew H. Todd*[1]

Address: [1]School of Pharmacy, University College London, London WC1N 1AX, United Kingdom; [2]Exscientia Ltd., The Schrödinger Building, Oxford Science Park, Oxford OX4 4GE, United Kingdom; [3]Drug Discovery Unit, Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, Dundee DD1 5EH, United Kingdom; [4]Department of Informatics, Faculty of Natural and Mathematical Sciences, King's College London, London WC2B 4BG, United Kingdom; [5]Molomics, Barcelona Science Park, Barcelona, Spain; [6]Intellegens Ltd., Eagle Labs, Chesterton Road, Cambridge CB4 3AZ, United Kingdom; [7]Theory of Condensed Matter Group, Cavendish Laboratories, University of Cambridge, United Kingdom; [8]Auromind Ltd, 126 Eglantine Avenue, Belfast BT9 6EU, United Kingdom; [9]School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia; [10]Optibrium Ltd. Blenheim House, Denny End Road, Cambridge, United Kingdom; [11]Research School of Biology, Australian National University, Canberra, ACT 2601, Australia; [12]Kellerberrin, 6 Wharf Rd, Balmain, Sydney, NSW 2041, Australia; [13]School of Chemistry, The University of Sydney, Sydney, NSW 2006, Australia; [14]Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan KS 66506, USA; [15]Strathclyde Institute Of Pharmacy And Biomedical Sciences, University of Strathclyde, Glasgow G4 ORE, United Kingdom; [16]Interlinked Therapeutics LLC, Portland, OR 97214, USA; [17]Medicines for Malaria Venture, PO Box 1826, 20 rte de Pre-Bois, 1215 Geneva 15, Switzerland.

Email: Matthew H. Todd – matthew.todd@ucl.ac.uk
* Corresponding author

Abstract

The discovery of new antimalarial medicines with novel mechanisms of action is key to combating the problem of increasing resistance to our frontline treatments. The Open Source Malaria (OSM) consortium has been developing compounds ("Series 4") that have potent activity against *Plasmodium falciparum in vitro* and *in vivo* and that have been suggested to act through the inhibition of *Pf*ATP4, an essential membrane ion pump that regulates the parasite's intracellular Na^+ concentration. The structure of *Pf*ATP4 is yet to be determined. In the absence of structural information about this target, a public competition was created to develop a model that would allow the prediction of anti-*Pf*ATP4 activity among Series 4 compounds, thereby reducing project costs associated with the unnecessary synthesis of inactive compounds.

In the first round, in 2016, six participants used the open data collated by OSM to develop moderately predictive models using diverse methods. Notably, all submitted models were available to all other participants in real time. Since then further bioactivity data have been acquired and machine learning methods have rapidly developed, so a second round of the competition was undertaken, in 2019, again with freely-donated models that other participants could see. The best-performing models from this second round were used to predict novel inhibitory molecules, of which several were synthesised and evaluated against the parasite. One such compound, containing a motif that the human chemists familiar with this series would have dismissed as ill-advised, was active. The project demonstrated the abilities of new machine learning methods in the prediction of active compounds where there is no biological target structure, frequently the central problem in phenotypic drug discovery.

Since all data and participant interactions remain in the public domain, this research project “lives” and may be improved by others.

Keywords
*Pf*ATP4; predictive modelling; Open Source Malaria; drug discovery; machine learning

Introduction
Efficiency in the early stages of the drug discovery pipeline, from hit identification to lead optimisation, is key to the development of new drugs. The initial identification of a hit compound is typically carried out using one of two approaches. In *target-based drug discovery* the molecular target of interest is known [1]. With this knowledge, libraries containing many compounds are screened (experimentally or computationally) against the known target to identify promising candidates or chemical scaffolds for further development. Through testing these chemicals, the key binding interactions may be identified and more directed structure activity relationship (SAR) studies can be conducted to optimise activity.

Alternatively, if the biological target is not known, *phenotypic drug discovery* may be undertaken [2]. This process involves the initial identification of potent compounds that give rise to the desired effect (e.g. inhibition of cell growth), with target determination performed thereafter. The lead optimisation phase in this type of drug discovery is less streamlined than that in the former method as it is conducted without guidance from target binding interactions and often relies upon the intuition of the medicinal chemist to design and synthesise compounds to explore the SAR. There are a number of obvious limitations to this approach, including the personal bias/imagination of the scientist or the availability/cost of resources. As a result, good hypotheses or key insights may be overlooked, which can lengthen the time taken to identify a lead candidate and increase costs associated with synthesising complex molecules that are later revealed to be inactive. Nevertheless, the advantage of phenotypic drug discovery, which underpins its popularity, is that hit or lead compounds are already known to be effective in their overall role (*e.g.,* the killing of a pathogen).

To aid this latter approach and overcome the absence of knowledge of the target or its structure, computational models may be developed using artificial intelligence (AI) and machine learning (ML) [3,4]. Such approaches allow the activities of new compounds in a phenotypic-screening program to be predicted. For instance, matched molecular pair analysis [5] and quantitative structure activity relationship (QSAR) [6] models are commonly used in medicinal chemistry campaigns to determine the relationships between the physical and biological properties of a series of compounds. This information can then be used to guide the design of new active compounds. In those cases in which a target has been identified but its structure is not yet determined, a structural model may be developed based on a known close homolog of the target [7]. This method allows for docking studies to be conducted to examine potential binding interactions that may occur in the actual target, thus guiding the lead optimisation process more effectively. Recent years have seen the increased use of computational methods such as these to aid the drug discovery process [8,9,10,11]. For instance, there have been successes in the *in silico* target prediction of small molecules with activity against *Mycobacterium tuberculosis* [12,13].

In the case of the malaria parasite, the development of resistance to frontline treatments is an ever-present problem. Since the isolation of artemisinin from the plant *Artemisia annua* in 1971 by Tu Youyou and colleagues [14], this natural product and its derivatives have been used in some of the most effective treatments for malaria. The artemisinin-based combination therapies (ACTs) utilise a short-acting artemisinin derivative in combination with one or more complementary antimalarials that are long-acting and possess a different mechanism of action (MoA). The use of these combinations has, in part, been responsible for the slow development of resistance to ACTs, yet in recent years increasing numbers of

cases have emerged of reduced efficacy [15]. There is an urgent need for new medicines that possess novel MoAs [16].

One promising biological target in *Plasmodium falciparum* is the essential P-type ATPase *Pf*ATP4, which localises to the plasma membrane of the intraerythrocytic parasite and exports Na^+ while importing H^+ equivalents [17,18]. The structure of this membrane-bound protein remains unsolved. Evidence for the involvement if *Pf*ATP4 in the mechanism of action of antiplasmodial compounds comes from several sources, including parasite Na^+ and pH assays that implicated *Pf*ATP4 as the target for the spiroindolone cipargamin [17,19] (currently in Phase III clinical development), the dihydroisoquinolone (+)-SJ733 [20], and 28 compounds from the Medicines for Malaria Venture (MMV) Malaria Box [21] as well as 11 compounds from the MMV Pathogen Box [22]. These compounds represent a strikingly diverse range of chemotypes (Fig. 1) [23]. A homology model of *Pf*ATP4 was developed using crystal structures from the closest mammalian homolog, a sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) [20]. However, in the absence of a solved structure of *Pf*ATP4, ideally bound to small molecule inhibitors, it remains unclear how it is possible for such diverse molecules to share the same target. Indeed, a challenge to understanding such data is that structurally different molecules generating the same phenotype may be interacting with the biological target differently.

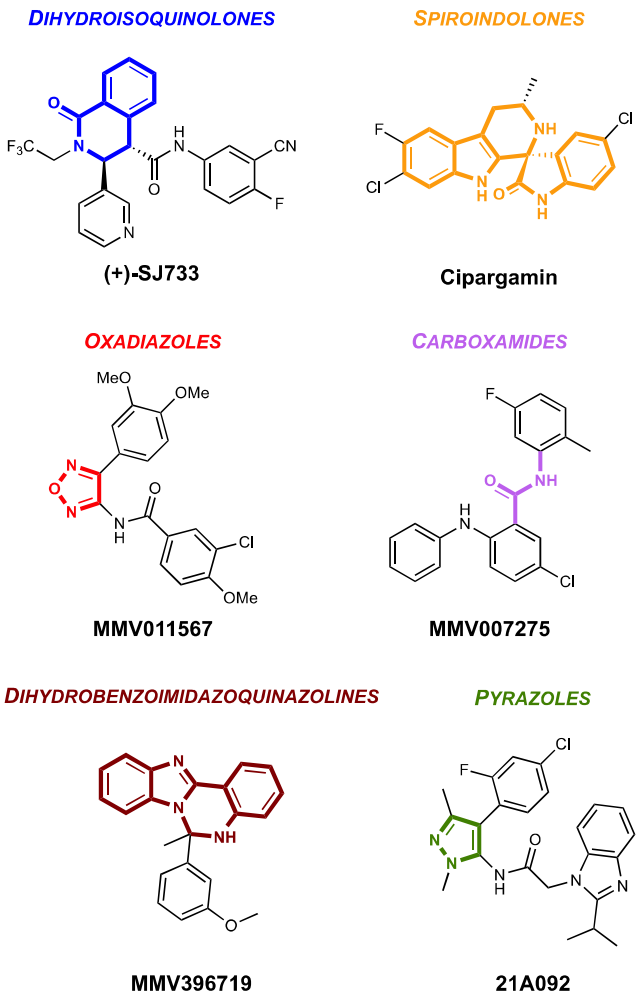


Fig. 1 Examples of the diverse chemotypes that have been linked to *Pf*ATP4. Each of the compounds give rise to effects on the parasite's internal Na^+ concentration and pH that are consistent with PfATP4 inhibition [20,21,24].

Since 2011, contributors to Open Source Malaria (OSM) have been evaluating several series of compounds originating from high-throughput screens (HTS) performed by pharmaceutical companies [25]. The recent focus of OSM has been on a class of triazolopyrazine-based compounds ("Series 4") that emerged from a screen carried out at Pfizer. There are currently more than 200 compounds in Series 4, with *in vitro* potencies against *P. falciparum* ranging from single-digit nanomolar to inactive. The highly promising nature of this series derives from several members having been found to be effective in the *in vivo* mouse model of the disease [26]. Based on preliminary investigations against *Pf*ATP4-resistant mutant strains (generated from the parent Dd2 strain by exposure to hits from the Malaria Box against *Pf*ATP4 [21]), Series 4 compounds are thought to target *Pf*ATP4 [27]. The intra-series similarity of their structures ought to imply a similarity in the way that the compounds interact with the target, but the interaction may differ from other compounds with the same phenotype.

The OSM Series 4 project is at the lead optimisation stage, with minor structural modifications being made in the search for improved solubility, potency and metabolic clearance. As is typical in such a search, analogs are being made that possess low potency, and these represent expensive "failures" (ca. $2K per compound for one postdoc-week per analogue). Better predictions of compound potency would save valuable resources and accelerate the science, so a predictive model was high on the list of priorities for the OSM consortium.

For the best means to develop such a model, we maintained an open mind. Available to us was a dataset of analogues with their associated activities, whether against the parasite or derived from biochemical (ion-regulation and/or ATPase) assays. Many of these compounds were from OSM Series 4, and there were also candidate antimalarials from other, structurally unrelated, series. It was possible to include "presumed inactives": randomly-selected molecules from commercial catalogues that were unlikely to display activity. A homology model (*vide supra*) was available that might permit a more target-based approach. Acknowledging these varied resources, we opted not to prescribe the approach to be taken and instead, in 2014, approached the scientific community simply with the need for a model that would allow us to predict the activity of hypothetical compounds. All data from OSM research projects are freely available to anyone online, representing an ideal starting point for such an open competition.

Between then and now there has been an explosion of interest in machine learning and AI methods in drug discovery [28,29]. While these new methods had the potential to be game-changing, there is the ever-present challenge in this sector of hype, in the sense that the actual capabilities of some of the newer technologies, outside of marketing statements, are sometimes not clear. In OSM the openness extends to the research process itself, allowing contributors to share what they are doing, rather than what they have done. The use of competitions to progress scientific research is not novel in itself, with previous examples of this in data analysis for drug discovery [30], but it is uncommon for competitions to be accompanied by the next crucial step: benchmarking by chemical synthesis and biological evaluation of predicted molecules. It is rarer still for science competitions to run completely openly, where everyone can see, and potentially incorporate, other entrants' solutions as they are submitted. We felt we could achieve two things by running this competition with OSM's open source ethos, in which those submitting entries would reveal their predictions in real time and, ideally, provide full methods (within the boundaries of commercial sensitivities). We would be able to approach the scientific problem along multiple paths, but we would also be able to provide a clear case study of the current effectiveness of predictive modelling in phenotypic drug discovery.

Results and Discussion

Round 0

An initial attempt by a single OSM contributor to develop a pharmacophore model was based around the known *Pf*ATP4 active compounds from the MMV Malaria Box [31,32]. By using Discovery Studio from Accelrys (now BIOVIA) to screen 28 active compounds with the Common Feature Pharmacophore Generation protocol, 10 four-feature models were produced. These were then narrowed down based on poses and score to one model that was developed further (Fig. 2A).

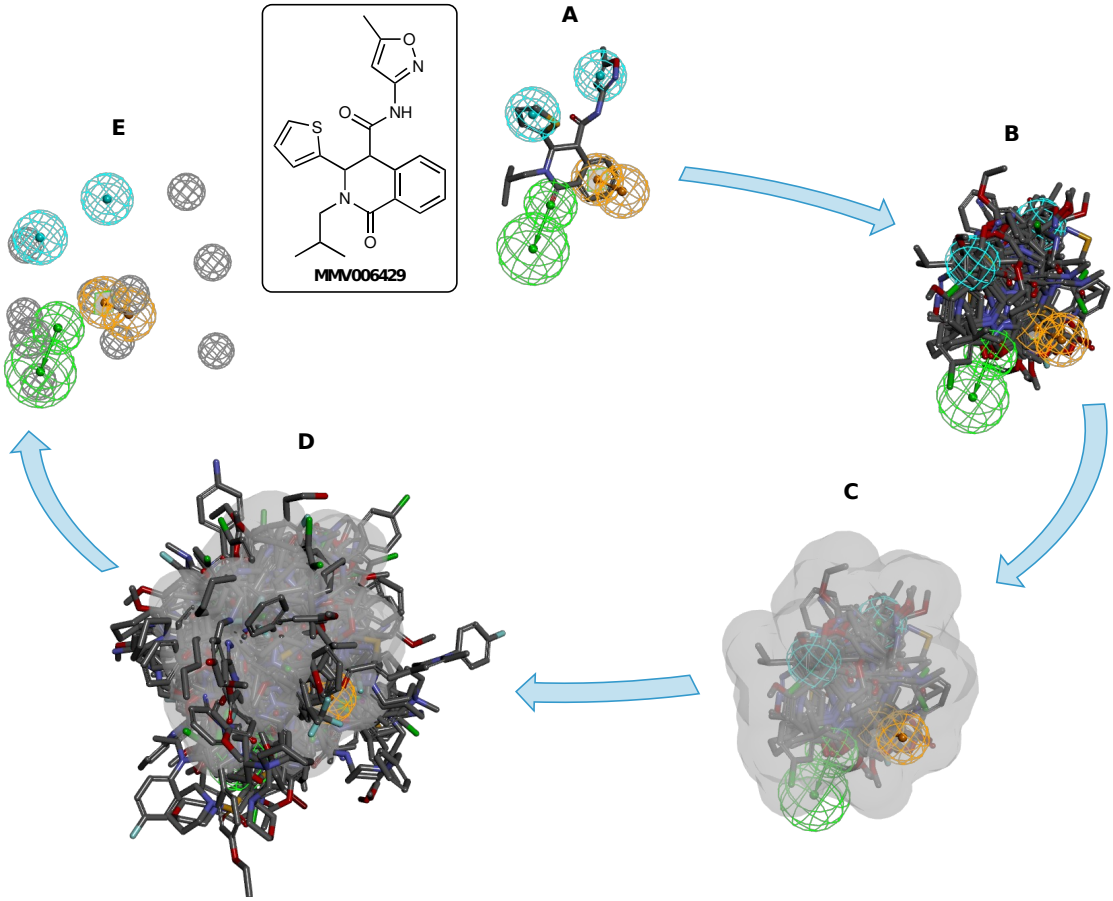


Fig. 2 Model creation workflow. A) The four-feature pharmacophore model chosen for further development with MMV006429 mapped. B) All 28 active compounds used in Round 0 superimposed onto the four-feature model. C) Shape feature added based on poses in B. D) Inactive molecules from the dataset mapped. E) Exclusion spheres added.

The 28 active compounds were mapped to the model and a shape feature was created (Fig. 2B). It was thought that this could give a general idea of the shape of the active site (Fig. 2C). Exclusion features were next added in areas where high scoring, inactive ligands penetrated outside of the shape figure. Unfortunately, when this model was applied in 2014 to a set of compounds that were evaluated for their ability to dysregulate ion homeostasis, the predictions were found to correlate poorly with the experimental potency results (Fig. 3). It was suggested that this lack of correlation could be due to factors not being taken into account by this first model (overlapping binding sites and compound chirality); a pharmacophore model explains aspects of the geometry of the interaction but not the details of the thermodynamics of the protein-small molecule contacts.

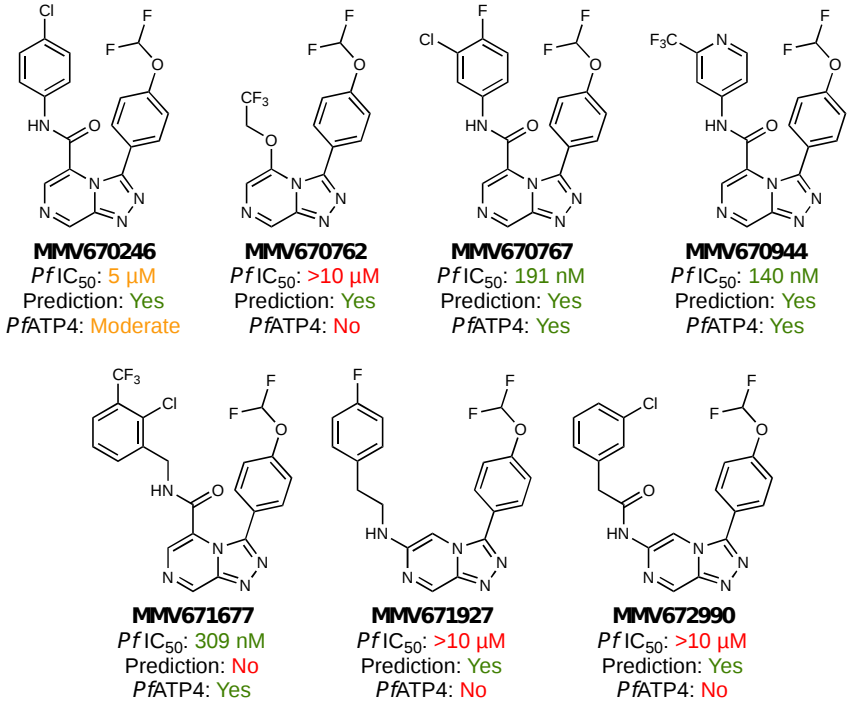


Fig. 3 Poor correlation was seen between the first model's predictions and experimental data. While there is excellent correlation between *in vitro* parasite killing potency and the ability to dysregulate parasite ion homeostasis ("*Pf*ATP4") activity, the majority of the model predictions did not correlate well with the experimental data. The compounds were tested for their effects on Na^+ regulation in saponin-isolated parasites (Dd2 strain) at 1 µM and for their effects on parasite pH at 5 µM; 'Yes': indicates that the compound gave rise to an increase in Na^+ concentration similar in magnitude to that of 50 nM cipargamin and a cytosolic alkalinisation, 'No': indicates that the compound did not affect the resting Na^+ concentration or pH, 'Moderate': indicates that the compound gave rise to an increase in pH, as well as an increase Na^+ concentration that was less than that observed on addition of 50 nM cipargamin.

This model was also used to screen [32] the Maybridge library of compounds [33] to identify a small and diverse selection of molecules to evaluate in biochemical assays. The results were filtered manually to give a final selection of 18 compounds that were subsequently evaluated for their effects on the parasite's internal Na^+ concentration (at 1 µM) and pH (at 5 µM). None of the compounds were found to increase the parasite's Na^+ concentration or pH, which confirmed that the model required further optimisation and led to the start of a crowdsourced attempt to solve this challenge.

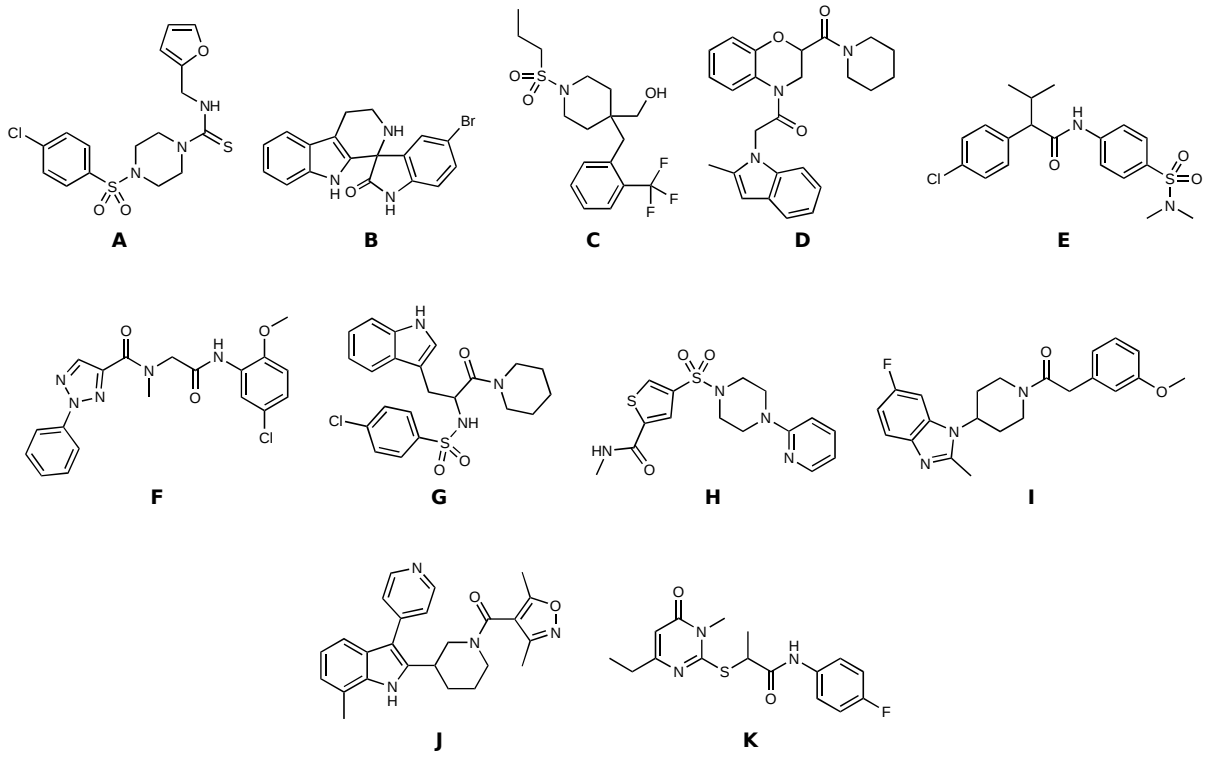
Round 1

The first full round of the predictive modelling competition was run between 2016 and 2017, and was intended to elicit the participation of members of the wider scientific community with expertise in computational chemistry [34]. The competition adhered to the open science principles underpinning the OSM consortium. Specifically, all participants were required to work openly for the duration of the competition, with working and data posted on open Electronic Laboratory Notebooks (ELN) that were made publicly available [35]. The participants were tasked with developing a predictive model using data provided by OSM that included a list of compounds with activity data for both *in vitro* whole cell potency and *Pf*ATP4 ion assays [36], along with the entire dataset of OSM compounds from previous series ((mostly presumed) inactives). Once the models were developed and deposited, the participants were provided with the molecular identifiers (*e.g.*, SMILES strings) for the 400 compounds contained within the MMV Pathogen Box and were required to rank them in

order of predicted activity in the ion assays. The Pathogen Box compounds were at the same time screened for their effects on parasite Na^+ concentration and pH and the data held back until the models had been submitted. A small cash prize inducement was employed to stimulate interest, despite the risk this brings of making the intrinsic reward for participation more extrinsic.[37]

Six diverse, fully-fledged entries were submitted from individuals working in both public and private sectors, with all working shared online (Table 1) [38]. These submissions were reviewed by a panel of four judges (Prof. Matthew Todd, A/Prof. Alice Motion (University of Sydney), Dr. Murray Robertson (University of Strathclyde and creator of the previous model in Round 0) and Prof. Alexander Tropsha (University of North Carolina, Chapel Hill)) that evaluated the top twenty ranked compounds from each model against the undisclosed Pathogen Box data. Two entrants developed models that were able to predict correctly two active compounds within their top twenty rankings, with a further model a close third place [39].

Table 1: Summary of the results from Round 1 of the predictive modelling competition.



Entrant	Description of Model	Correctly Predicted Actives	Result
Jonathan Cardoso-Silva	Gradient boosting model (using XGBoost) to predict actives and nonactives.	**B** just outside top 20	Runner-up
Giovanni Cincilla	*Pf*ATP4 Ion Regulation Activity classification model using: CDK descriptors [40], ECFC4 fingerprints and Random Forest.	**B**, **D**	Runner-up
Davy Guan	Semi-supervised machine learning, used to	**B**, **F**	Runner-

	construct QSAR models. Molecules were featurised by either Graph convolutional techniques or with 1024 Bit ECFP4 descriptors.		up
James McCulloch	Deep Neural Network ML using a vector of the chemo-physical properties of the target molecules.	**B**, **D**, **I** F just outside top 20	**Winner**
Ho-Leung Ng	QSAR model based on homology modelling of *Pf*ATP4 -Cresset Forge.	**K**, **D** **J** just outside top 20	**Winner**
Vito Spadavecchio	Library of 'common' transformations' as seen in CHEMBL.	**B**	Runner-up

Compounds A-K shown to be active from the MMV Pathogen Box screen against *Pf*ATP4 [22].

While this first round of the competition was successful in demonstrating the capabilities of the community to work openly and provide quality data, the models, though obtained with diverse methods, were not yet highly predictive. Of note was, again, the striking diversity of chemotypes (A–K, Table 1) sharing a target.

Round 2

Given the diverse, spontaneous inputs from the initial round of the open competition, and the high quality of the associated dialogue that had taken place on the relevant project website, GitHub, it was decided that a second round would be run in 2019 since “expensive failure analogs” were still arising in the experimental programme. The aim for this round was not only to allow for the entrants from Round 1 to improve upon the original models, but for new participants to get involved with inputs from larger companies that specialised in artificial intelligence and machine learning (AI/ML) approaches. Since the series had moved on in the interim (with further compounds being evaluated), the community had access to an expanded dataset, including all the data used as the test set for the previous round [22].

The competition’s second round was launched in July 2019 [41]. In this new phase of the competition it was the intention to use the best-performing models to perform the most important task of all: to predict new chemical matter that would be active (rather than merely look at the fit of retrospective data). Synthesis and evaluation of these predictions would then serve as model validation in a “real” case. A small, new dataset of activity from recently-synthesised analogs was kept back to serve as the basis for judging model fitness.

By the conclusion of Round 2 (a period of ~10 weeks), ten entries had been submitted, five of which were from returning participants (Table 2). In a similar fashion, the submissions were reviewed by a panel of four judges (Prof. Matthew Todd, Dr. Edwin Tse (UCL), Dr. Murray Robertson (Strathclyde) and Prof. Robert Glen (Cambridge)) who compared the predicted potencies against the experimentally-derived blood stage potency values for thirty-four compounds. The precision of each model was calculated according to: $precision = \frac{x}{x+y}$, where x is the number of correct predictions (active and inactive combined) and y is the number of false positive predictions [42].

Table 2: Summary of the results from Round 2 of the predictive modelling competition.

Entrant (Affiliation)	Description of Model[a]	Precision of	Result

		Accurate Predictions (Active and Inactive)[b]	
Jonathan Cardoso-Silva (King's College London)	Network-based piecewise linear regression for QSAR modelling [43].	36%	Runner-up
Giovanni Cincilla (Molomics)	*P. falciparum* inhibition classification model using: CDK descriptors [40], ECFC4 fingerprints and logistic regression (with: stochastic average gradient as solver, uniform regularisation and learning step size = 0.01).	91%[c]	**Winner (company)**
Mykola Galushka (Auromind)	SMILES variational auto-encoder to generate chemical compounds fingerprint and cascade models Naive Bayes classifier with Multi-layer perceptron regressor for filtering active components and identifying a specific potency value.	58%	Runner-up
Davy Guan (The University of Sydney)	Automated machine learning method with 21 quantum mechanical descriptors using the Hartree Fock with 3 corrections method [44] and JCLogP, optimised for Mean Absolute Error.	82%	**Winner (non-company)**
Ben Irwin/Mario Öeren/Tom Whitehead (Optibrium/Intellegens)	Deep imputation [45,46,47] with quantum mechanical StarDrop6.6 Automodeller and pKa descriptors [48].	81%	**Second place**
Raymond Lui (The University of Sydney)	Automated machine learning method using 59 permutation feature importance selected Mordred and quantum mechanical descriptors optimised for Mean Absolute Error.	58%	Runner-up
Slade Matthews (The University of Sydney)	Random forest model using 200 Mordred descriptors based on optimised 3D structures. Training RMSE = 0.805.	N.A.	Runner-up
Ho-Leung Ng (Kansas	QSAR model based on	71%	Runner-up

State University)	detailed homology modeling of PfATP4 and docking. 3D features are combined with 1D/2D QSAR features using XGBoost (gradient boosted trees) to make a regression model.		
Vito Spadavecchio (Interlinked TX)	Ensemble classification (logistic regression) and regression (MLP) using ECFP4 (Morgan radius 2).	79%[c]	Runner-up
Laksh Aithani/Willem van Hoorn (Exscientia)	Ridge regression model with alpha = 1. ECFP4 fingerprints with (Morgan radius 2) were the input to the model.	81%	**Second place**

[a]See SI for full experimental details. [b]Based on regression prediction. [c]Based on classification prediction.

It was originally intended for each of the four winning entrants (first and second place winners) to generate two new structures that were predicted to be active using their models: one possessing the Series 4 triazolopyrazine core and the other being structurally distinct. This would give a total of eight molecules to be synthesised and validated experimentally. In addition to optimising potency, model generators were tasked with keeping good solubility in mind as a design criterion. It became evident that certain suggested compounds were synthetically inaccessible, or would take major resources to pursue. The former is often an issue when predictive models do not take into account known synthetic pathways, though there is significant activity at present to improve the incorporation of synthetic planning into library suggestion [49,50]. The initial list was narrowed to focus on five predicted triazolopyrazine compounds (Fig. 4). The five compounds were successfully synthesised and subsequently evaluated for *in vitro* (growth-inhibition) activity against *P. falciparum* along with the previously reported positive control for the series [51]. In addition to the standard potency (*in vitro* growth) assay, these compounds were evaluated for their ability to inhibit *Pf*ATP4 in biochemical (Na^+ regulation) assays to confirm that the MoA had not changed following these structural changes.

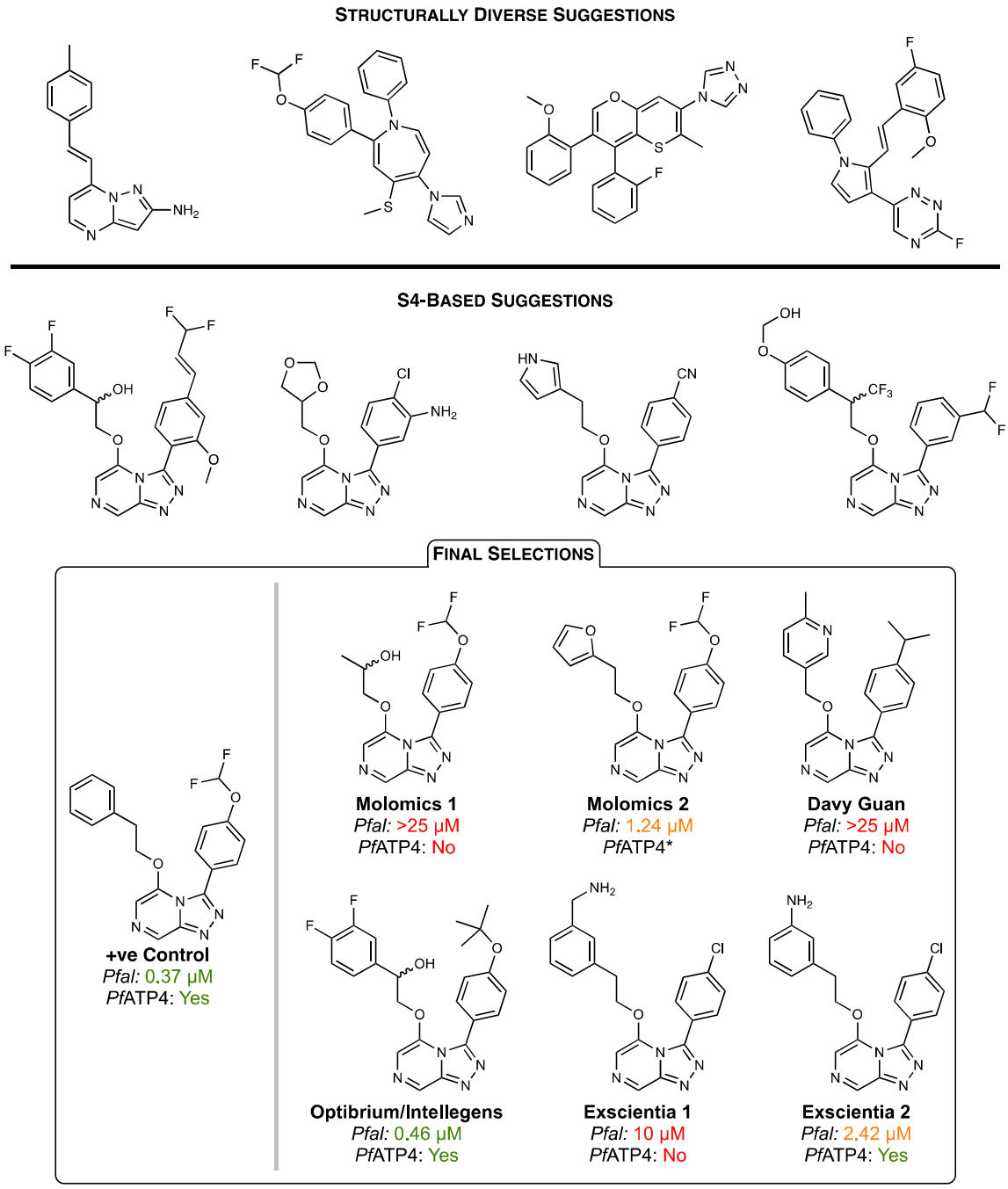


Fig. 4 Examples of the suggested compounds predicted by the winning entrants from Round 2 and the five chosen for experimental validation. The predictions were synthesised (see SI) and their potencies and MoAs (Fig. S9) experimentally validated. Three compounds were found to be active. **Pf*ATP4 activity was not obtained for this compound.

Three of the six compounds were found to be active (<1 µM) or moderately active (1–2.5 µM) in *in vitro* growth assays with asexual blood-stage *P. falciparum* (3D7) parasites, representing a hit rate of 50% on a small sample size. Up to this point a total of 398 compounds had been made and evaluated for *in vitro* activity in OSM Series 4, with the design of these compounds driven entirely by the intuition of medicinal chemists. By setting a potency cut-off of 2.5 µM (the upper limit of reasonable activity), the tally of active compounds discovered in this series stands at 165, representing a comparable human intuition-derived hit rate of 41% on a larger sample size. Most of the compounds were tested (blind) for their ability to disrupt Na^+ regulation in isolated asexual blood-stage parasites,

which confirmed an unchanged mechanism of action: two of the compounds found to be active in *in vitro* growth assays disrupted Na^+ regulation whereas the three compounds inactive in growth assays did not, at the concentrations tested (Fig. S9).

It is interesting to compare these results with the intuition of the chemists who have deep experience of this series and who are familiar with the SAR. A recurring observation was the sensitivity of the length of the ether linker between triazolopyrazine core and northwest phenyl group, with a spacer of two methylene units (between phenyl ring and oxygen) leading to far higher potencies than other lengths. The Davy Guan prediction involving the shorter linker, and the Molomics 1 prediction without the pendant phenyl ring, lie in the class of inactive compounds subject to human retrospective wisdom (i.e. the "I could have told you that" class). In contrast, the Exscientia compounds were thought by the human team to be likely to be potent, but only one performed well (i.e. the "that's odd" class). Lastly the Optibrium/Intellegens suggestion that included the *tert*-butyl pendant was thought by the human team to be a certain inactive, given what was known of variation in that part of the molecule, yet this compound displayed good potency and is a particularly useful outcome (i.e. the "I welcome our machine overlords" class).

To gain more insight, and to improve these potential antimalarials, further iterations of these models are needed. The open nature of the competitions and of the over-arching consortium is that anyone may work on improvements since everyone has access to all the data, making this a "living" research project. A potential explanation for the predicted hit rate not being higher is the relatively small dataset (~400 compounds) from which each model was developed, potentially compromising perfectly reasonably computational approaches yet representing a fairly typical situation for lead optimisation. Two further points are of particular note: 1) It was possible to involve leading experts from the private sector in an open competition to solve a public health challenge without those participants needing to compromise their competitive business advantage; indeed success in such an endeavour has been used as an open demonstration of capabilities [52]. 2) The private sector participants displayed high and sustained levels of collaborative working and commitment to a public good, in what is counter to the public's perception of the secretive nature of the modern pharmaceutical industry; indeed the "winning" and "losing" of the competition was less important than the extent to which entrants worked together openly to improve the underlying research [41].

Conclusion

With hit identification and lead optimisation being key steps in the development of any new drug, the continued advancements in machine learning and artificial intelligence approaches possess significant promise to streamline this process, which would result in more efficient medicinal chemistry campaigns. In the absence of target structural information, a crowdsourced approach was used to develop predictive models for a promising antimalarial series. Importantly, the winning models of the most recent competition round were used to generate novel compounds, which were then synthesised and evaluated for experimental validation of each model leading to a new counterintuitive "active". The simple open science and crowdsourcing principles used throughout this campaign are applicable to many medicinal chemistry projects, whereby the community's combined efforts can be used to accelerate the early stages of drug discovery and involve participants from public and private sectors. The work conducted here has been designed to be "living", in that all methods and results are publicly available and contributions can continue to be made by anyone because everyone has access to all data and ideas.

References

[1] Croston, G. E. The Utility of Target-Based Discovery. *Expert Opin. Drug Discov.* **12**, 427–429 (2017).

[2] Moffat, J. G., Vincent, F., Lee, J. A., Eder, J. & Prunotto, M. Opportunities and Challenges in Phenotypic Drug Discovery: An Industry Perspective. *Nat. Rev. Drug Discov.* **16**, 531–543 (2017).
[3] Sellwood, M. A., Ahmed, M., Segler, M. H. S. & Brown, N. Artifical Intelligence in Drug Discovery. *Future Med. Chem.* **10**, 2025–2028 (2018).
[4] Griffen, E. J., Dossetter, A. G., Leach, A. G. Chemists: AI is Here; Unite to Get the Benefits. *J. Med. Chem. In press* (2020).
[5] Tyrchan, C. & Evertsson, E. Matched Molecular Pair Analysis in Short: Algorithms, Applications and Limitations. *Comput. Struct. Biotechnol. J.* **15**, 86–90 (2017).
[6] Neves, B. J. *et al.* QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Front. Pharmacol.* **9**, No. 1275 (2018).
[7] Muhammed, M. T. & Aki-Yalcin, E. Homology Modelling in Drug Discovery: Overview, Current Applications, and Future Perspectives. *Chem. Biol. Drug Des.* **93**, 12–20 (2018).
[8] Sieg, J., Flachsenberg, F. & Rarey, M. In Need of Bias Control: Evaluating Chemica Data for Machine Learning in Structure-Based Virtual Screening. *J. Chem. Inf. Model* **59**, 947–961 (2019).
[9] Brown, N., Fiscato, M., Segler, M. H. S. & Vaucher, A. C. GuacaMol: Benchmarking Models for de Novo Molecular Design. *J. Chem. Inf. Model.* **59**, 1096–1108 (2019).
[10] Walters, W. P. & Murcko, M. Assessing the Impact of Generative AI on Medicinal Chemistry. *Nat. Biotechnol.* **38**, 143–145 (2020).
[11] Stokes, J. M. *et al.* A Deep Learning Approach to Antibiotic Discovery. *Cell* **180**, 688–702 (2020).
[12] Mugumbate, G. *et al.* Mycobacterial Dihydrofolate Reductase Inhibitors Identified Using Chemogenomic Methods and *In Vitro* Validation. *PLoS ONE* **10**, e0121492 (2015).
[13] Homeyer, N. *et al.* A Platform for Target Prediction of Phenotypic Screening Hit Molecules. *J. Mol. Graph. Model.* **95**, 107485 (2020).
[14] Qinghaosu Antimalaria Coordinating Research Group. Antimalaria Studies on Qinghaosu. *Chin. Med. J. (Engl.)* **92**, 811–816 (1979).
[15] Hamilton, W. L. *et al.* Evolution and Expansion of Multidrug-Resistant Malaria in Southeast Asia: A Genomic Epidemiology Study. *Lancet Infect. Dis.* **19**, 943–951 (2019).
[16] Tse, E. G., Korsik, M. & Todd, M. H. The Past, Present and Future of Anti-Malarial Medicines. *Malar. J.* **18**, No. 93 (2019).
[17] Spillman, N. J. *et al.* Na^+ Regulation in the Malaria Parasite *Plasmodium falciparum* Involves the Cation ATPase PfATP4 and Is a Target of the Spiroindolone Antimalarials. *Cell Host Microbe* **13**, 227–237 (2013).
[18] Kirk, K. Ion Regulation in the Malaria Parasite. *Annu. Rev. Microbiol.* **69**, 341–359 (2015).
[19] Rottmann, M. *et al.* Spiroindolones, a Potent Compound Class for the Treatment of Malaria. *Science* **329**, 1175–1180 (2010).
[20] Jiménez-Díaz, M. B. *et al.* (+)-SJ733, a Clinical Candidate for Malaria that acts through ATP4 to Induce Rapid Host-Mediated Clearance of *Plasmodium*. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E5455–E5462 (2014).
[21] Lehane, A. M., Ridgway, M. C., Baker, E. & Kirk, K. Diverse Chemotypes Disrupt Ion Homeostasis in the Malaria Parasite. *Mol. Microbiol.* **94**, 327–339 (2014).
[22] Dennis, A. S. M., Rosling, J. E. O., Lehane, A. M. & Kirk, K. Diverse Antimalarials from Whole-Cell Phenotypic Screens Disrupt Malaria Parasite Ion and Volume Homeostasis. *Sci. Rep.* **8**, No. 8795 (2018).
[23] Spillman, N. J. & Kirk, K. The Malaria Parasite Cation ATPase PfATP4 and its Role in the Mechanism of Action of a New Arsenal of Antimalarial Drugs. *Int. J. Parasitol. Drugs Drug Resist.* **5**, 149–162 (2015).
[24] Vaidya, A. B. *et al.* Pyrazoleamide Compounds are Potent Antimalarials that Target Na^+ Homeostasis in Intraerythrocytic *Plasmodium falciparum*. *Nat. Commun.* **5**, No. 5521 (2014).
[25] Williamson, A. E. *et al.* Open Source Drug Discovery: Highly Potent Antimalarial Compounds Derived from the Tres Cantos Arylpyrroles. *ACS Cent. Sci.* **2**, 687-701 (2016).

[26] In Vivo Efficacy (2018). Available at: https://github.com/OpenSourceMalaria/Series4/wiki/In-Vivo-Efficacy. (Accessed: 26th September 2019).
[27] Evaluation of Series 4 Compounds vs ATP4-Resistant Mutants (2015). Available at: http://malaria.ourexperiment.org/biological_data/11448/post.html. (Accessed: 14th October 2019).
[28] Vamathevan, J. *et al.* Applications of Machine Learning in Drug Discovery and Development. *Nat. Rev. Drug Discov.* **18**, 463–477 (2019).
[29] Chan, H. C. S., Shan, H., Dahoun, T., Vogel, H. & Yuan, S. Advancing Drug Discovery via Artificial Intelligence. *Trends Pharmacol. Sci.* **40**, 592–604 (2019).
[30] Bentzien, J., Muegge, I., Hamner, B., Thompson, D. C. Crowd Computing: Using Competitive Dynamics to Develop and Refine Highly Predictive Models. *Drug Discov. Today* **18**, 472–478 (2013).
[31] Pharmacophore Modelling of the Malaria Box PfATP4 Active Compounds (2015). Available at: http://malaria.ourexperiment.org/pharmacophore_modelling_/7971/post.html. (Accessed: 26th September 2019).
[32] Using the Pharmacophore Model to search Commercial Compounds for new leads (2015). Available at: http://malaria.ourexperiment.org/pharmacophore_modelling_/12498/post.html. (Accessed: 26th September 2019).
[33] Maybridge Screening Collection. Available at: https://www.maybridge.com/portal/alias__Rainbow/lang__en/tabID__146/DesktopDefault.aspx. (Accessed: 26th September 2019).
[34] COMPETITION: A Predictive Model for Series Four (2016). Available at: https://github.com/OpenSourceMalaria/OSM_To_Do_List/issues/421. (Accessed: 26th September 2019).
[35] Todd, M. H. Six Laws of Open Source Drug Discovery. *ChemMedChem* **14**, 1804–1809 (2019).
[36] Ion Regulation Data for OSM Competition. Available at: https://docs.google.com/spreadsheets/d/1WWP8fE3X2BLzZ7jOm6bRWpnHZqVJBf8XgIYEb7hshXU/edit?usp=sharing. (Accessed: 26th September 2019).
[37] Hars, Alexander, and Shaosong Ou. “Working for Free? Motivations for Participating in Open-Source Projects.” International Journal of Electronic Commerce, vol. 6, no. 3, 2002, pp. 25–39. JSTOR, www.jstor.org/stable/27751021. Accessed 18 June 2020.
[38] Summary of Competition Entries (2017). Available at: https://docs.google.com/spreadsheets/d/1pY6sYXIw66jnzUO3CoP8HceYdDjLRvwg5_pLkBY1Wek/edit?usp=sharing. (Accessed: 26th September 2019).
[39] VS Results. Available at: https://docs.google.com/spreadsheets/d/1uaQ_mSVY6vQSbnDHD5gf232ySH-KVjUg--dQCsilGWI/edit?usp=sharing. (Accessed: 26th September 2019).
[40] Steinbeck, C. *et al.* The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics. *J. Chem. Inf. Comput. Sci.* **43**, 493–500 (2003).
[41] COMPETITION ROUND 2: A Predictive Model for Series 4 (2019). Available at: https://github.com/OpenSourceMalaria/Series4_PredictiveModel/issues/1. (Accessed: 26th September 2019).
[42] Round 2 Results. Available at: https://docs.google.com/spreadsheets/d/1ZPJlMM7znFa056yjOuIycLgzV7-ENt1mMSUJKgH9TxY/edit?usp=sharing. (Accessed 12th September 2019).
[43] Cardoso-Silva, J., Papageorgiou, L. G. & Tsoka, S. Network-Based Piecewise Linear Regression for QSAR Modelling. *J. Comput. Aided Mol. Des.* **33**, 831–844 (2019).
[44] Sure, R. & Grimme, S. Corrected Small Basis Set Hartree-Fock Method for Large Systems. *J. Comput. Chem.* **34**, 1672–1685 (2013).
[45] Whitehead, T. M., Irwin, B. W. J., Hunt, P., Segall, M. D. & Conduit, G. J. Imputation of Assay Bioactivity Data Using Deep Learning. *J. Chem. Inf. Model*. **59**, 1197–1204 (2019).

[46] Irwin, B. W. J., Levell, J. R., Whitehead, T. M., Segall, M. D. & Conduit, G. J. Practical Applications of Deep Learning to Impute Heterogeneous Drug Discovery Data. *J. Chem. Inf. Model.* **60**, 2848-2857 (2020).
[47] Irwin, B. W. J., Mahmoud, S., Whitehead, T. M., Conduit, G. J. & Segall, M. D. Imputation versus Prediction: Applications in Machine Learning for Drug Discovery. *Future Drug. Discov.* **2** (2020).
[48] Hunt, P. A. *et al.* Predicting pKa Using a Combination of Semi-Empirical Quantum Mechanics and Radial Basis Function Methods. *J. Chem. Inf. Model.* Just Accepted (2020).
[49] InfoChem IC*SYNTH*. Available at: https://www.infochem.de/synthesis/ic-synth. (Accessed: 21st April 2020).
[50] Nicolaou, C. A., Watson, I., LeMasters, M. A., Masquelin, T. & Wang, J. Context Aware Data-Driven Retrosynthetic Analysis. *J. Chem. Inf. Model.* Just Accepted (2020).
[51] Tse, E. G. *et al.* Non-Classical Phenyl Bioisosteres as Effective Replacements in a Series of Novel Open Source Antimalarials. *J. Med. Chem. In press* (2020).
[52] Data Finders: How can artificial intelligence and deep-learning platforms be used to combat malaria? Medicine Maker, May 15th 2020 https://themedicinemaker.com/discovery-development/data-finders. (Accessed June 18th 2020).

Acknowledgements
This work has been supported by the Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery Network+, which has been funded by EPSRC (EP/S000356/1), the Australian Research Council (LP150101226), the Medicines for Malaria Venture (RD/11/0040) and University College London for funding (to MHT), the US National Science Foundation (1833181) (to HLN), the EPSRC Centre for Doctoral Training in Computational Methods for Materials Science (EP/L015552/2) (to ADW), the EPSRC Future Continuous Manufacturing and Advanced Crystallisation (CMAC) Research Hib (EP/L015552/1) (to MNR), the Royal Society (to GJC).

Author contributions

EGT, AM, MNR carried out the synthetic chemistry. AML, JCRL and KK planned and/or carried out the Na^+ and pH assays and/or advised on the results. IH developed the *Pf* assay, manually prepared and analysed compounds, collated all data and contributed to the paper (methods section and data table). MA carried out routine *Pf* assay and analysis. MNR, DG, HLN, JC-S, BWJI, MÖ, TMW, GJC, ADW, LA, VAT, WPvH, JM, VS, RL, SM, GC and MG contributed models to the competitions and collaborated on the outputs during the competitions. PW helped conceive the study. MHT founded OSM, conceived the project and secured (AI3SD-FundingCall1_029) or co-secured (LP150101226, with MMV and KK) the funding. All authors contributed to the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at XXX.