Journal of Medicinal Chemistry

pubs.acs.org/jmc

Article

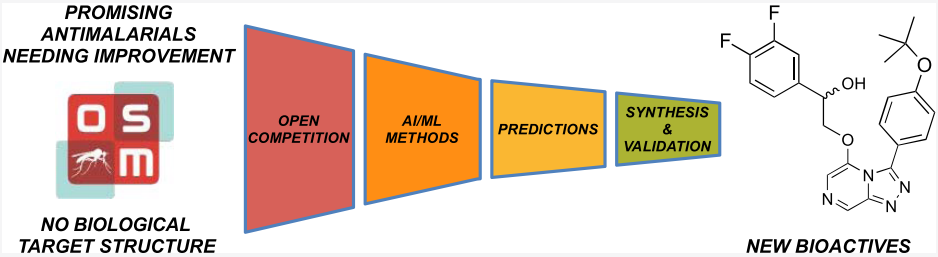
An Open Drug Discovery Competition: Experimental Validation of Predictive Models in a Series of Novel Antimalarials

Edwin G. Tse, Laksh Aithani, Mark Anderson, Jonathan Cardoso-Silva, Giovanni Cincilla, Gareth J. Conduit, Mykola Galushka, Davy Guan, Irene Hallyburton, Benedict W. J. Irwin, Kiaran Kirk, Adele M. Lehane, Julia C. R. Lindblom, Raymond Lui, Slade Matthews, James McCulloch, Alice Motion, Ho Leung Ng, Mario Öeren, Murray N. Robertson, Vito Spadavecchio, Vasileios A. Tatsis, Willem P. van Hoorn, Alexander D. Wade, Thomas M. Whitehead, Paul Willis, and Matthew H. Todd*

Cite This: https://doi.org/10.1021/acs.jmedchem.1c00313

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: The Open Source Malaria (OSM) consortium is developing compounds that kill the human malaria parasite, *Plasmodium falciparum*, by targeting *Pf*ATP4, an essential ion pump on the parasite surface. The structure of *Pf*ATP4 has not been determined. Here, we describe a public competition created to develop a predictive model for the identification of *Pf*ATP4 inhibitors, thereby reducing project costs associated with the synthesis of inactive compounds. Competition participants could see all entries as they were submitted. In the final round, featuring private sector entrants specializing in machine learning methods, the best-performing models were used to predict novel inhibitors, of which several were synthesized and evaluated against the parasite. Half possessed biological activity, with one featuring a motif that the human chemists familiar with this series would have dismissed as "ill-advised". Since all data and participant interactions remain in the public domain, this research project "lives" and may be improved by others.

■ INTRODUCTION

Efficiency in the early stages of the drug discovery pipeline, from hit identification to lead optimization, 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 optimize 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-optimization 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 synthesize 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 synthesizing 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).

Received: February 18, 2021

Downloaded via UNIV OF CAMBRIDGE on November 9, 2021 at 12:03:04 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ACS Publications

© XXXX The Authors. Published by American Chemical Society

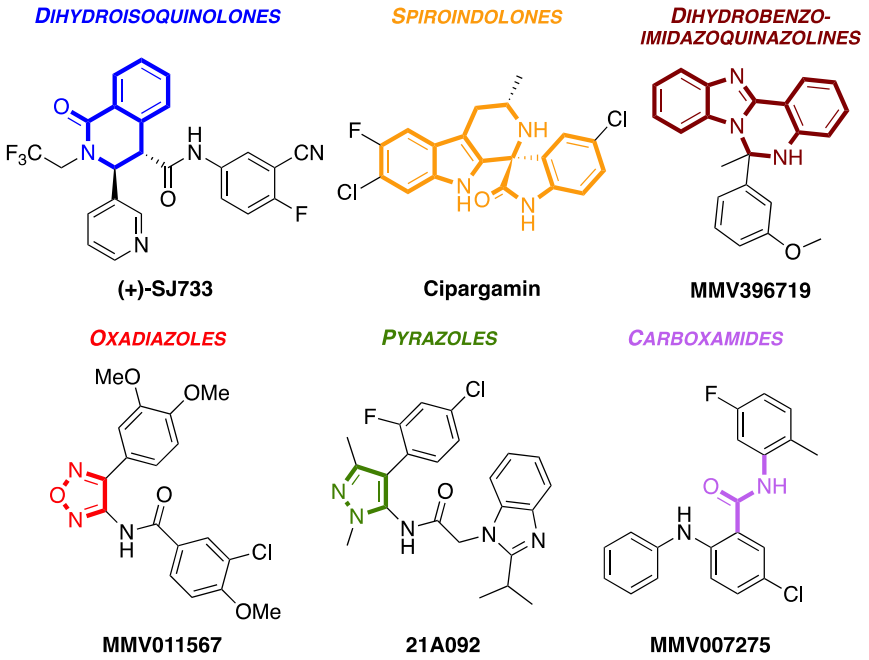


Figure 1. Examples of diverse chemotypes (colored) that have been linked to *Pf*ATP4. Each of the compounds gives rise to effects on the parasite's internal Na^+ concentration and pH that are consistent with *Pf*ATP4 inhibition.[20,21,24]

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-optimization process more effectively. Recent years have seen the increased use of computational methods such as these to aid the drug discovery process.[8−11] 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) utilize 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 the human malaria parasite, *Plasmodium falciparum*, is the essential P-type ATPase *Pf*ATP4, which localizes 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 of *Pf*ATP4 in the mechanism of action of a wide range of antiplasmodial compounds identified in phenotypic screens comes from several sources, including from analysis of mutations in resistant lines and from a range of physiological and biochemical assays (measurement of parasite cytosolic Na^+ concentration ($[Na^+]$) and pH, as well as parasite volume and Na^+-ATPase activity). *Pf*ATP4 has been implicated as the target for spiroindolone cipargamin[17,19] (currently in Phase III clinical development), 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 (Figure 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 such a diverse range of molecules might 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.

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

https://doi.org/10.1021/acs.jmedchem.1c00313
J. Med. Chem. XXXX, XXX, XXX−XXX

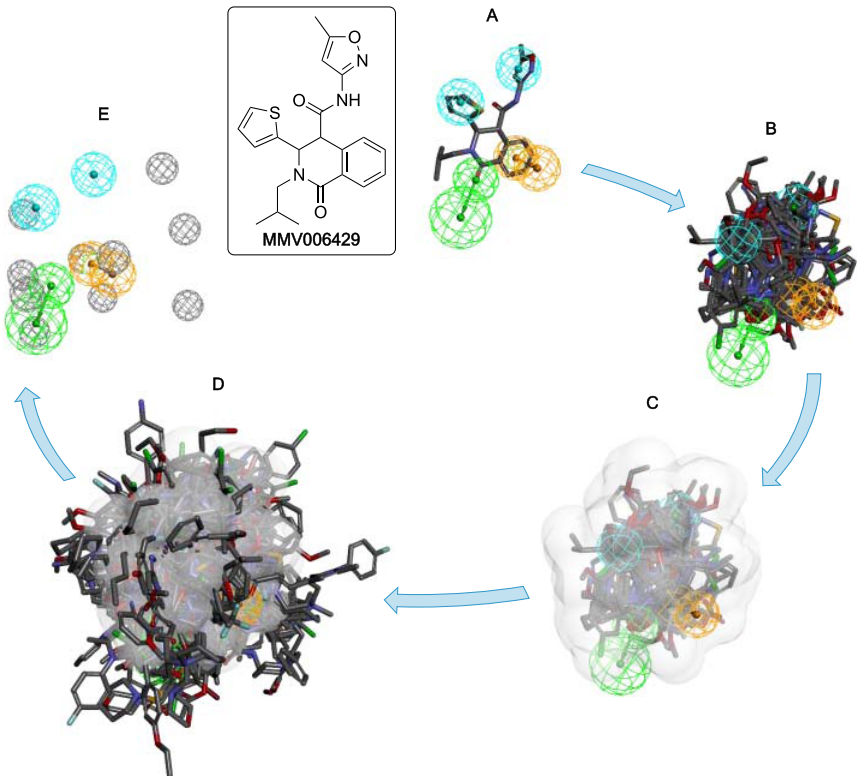


Figure 2. Model creation workflow. (A) Four-feature pharmacophore model chosen for further development with MMV006429 mapped. (B) All 28 active compounds used in Round 0 superimposed onto the four-feature model. (C) Shape feature added based on poses in (B). (D) Inactive molecules from the data set mapped.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

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 intraseries 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-optimization stage, with minor structural modifications being made in the search for improved solubility, potency, and metabolic clearance. As is typical in such a search, analogues 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 data set of analogues with their associated activities, whether against the parasite or derived from biochemical ([Na^+], pH, 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. There is obviously a rich history of QSAR-based approaches that might be called upon. 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 modeling in phenotypic drug discovery.

https://doi.org/10.1021/acs.jmedchem.1c00313
J. Med. Chem. XXXX, XXX, XXX−XXX

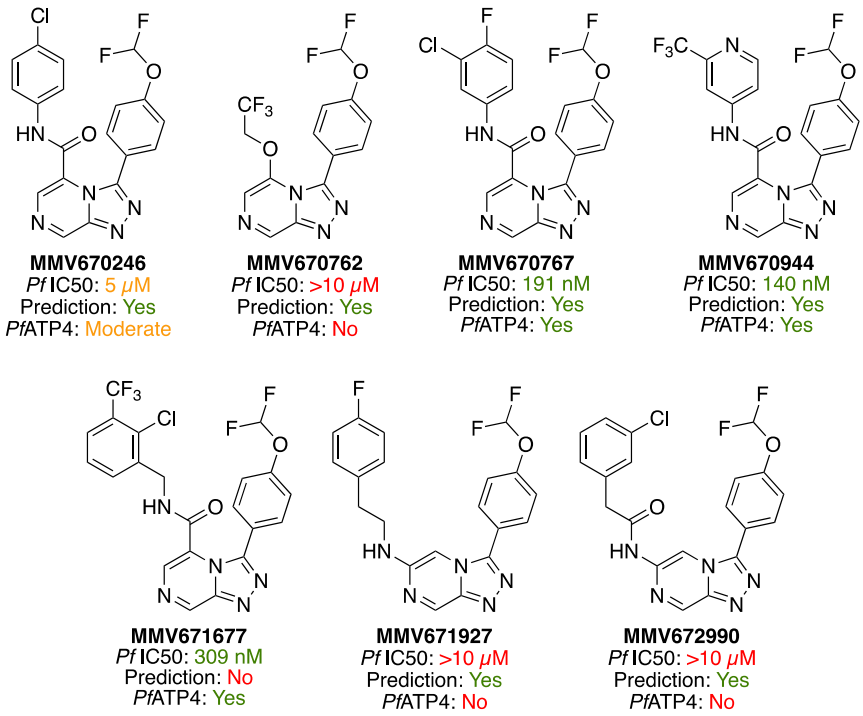


Figure 3. Poor correlation was seen between the first model's predictions and experimental data. While there is an excellent correlation between in vitro parasite killing potency and the ability to dysregulate parasite ion homeostasis, the majority of the model predictions did not correlate well with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

■ 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] Using Discovery Studio from Accelrys (now BIOVIA) to process 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 (Figure 2A).

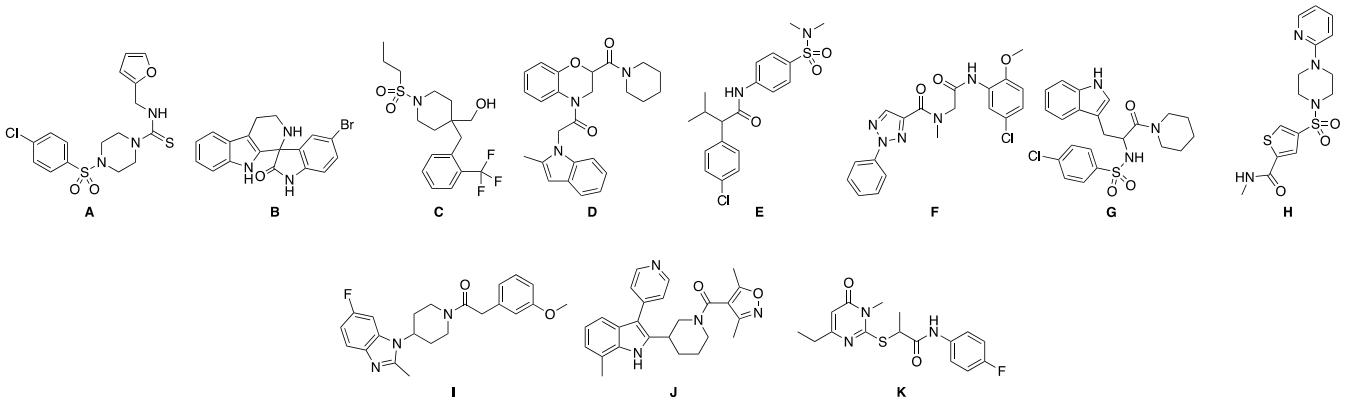
The 28 active compounds were mapped to the model and a shape feature was created (Figure 2B). It was thought that this could give a general idea of the shape of the compound binding site (Figure 2C). Exclusion features were next added in areas where high scoring, inactive ligands penetrated outside of 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 (Figure 3). The test set was selected to be structurally diverse, including features known to be associated with inactivity (e.g., transposition of the northwest pendant) but also features where minor variations were known to be important for activity (aromatic substituents in the pendant amide). It was suggested that the 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.

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 cytosolic Na^+ concentration (at 1 μM) and pH (at 5 μM). None of the compounds were found to increase the parasite's cytosolic [Na^+] or pH, which confirmed that the model required further optimization and led to the start of a crowdsourced attempt to solve this challenge.

Round 1. The first full round of the predictive modeling competition was run between 2016 and 2017 and was intended to elicit the participation of members of the wider scientific community with expertize 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 data set 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 compounds were at the same time screened for their effects on parasite cytosolic [Na^+] concentration and pH and the data

https://doi.org/10.1021/acs.jmedchem.1c00313
J. Med. Chem. XXXX, XXX, XXX−XXX

Table 1. Summary of the Results from Round 1 of the Predictive Modeling Competition[a]



entrant	description of model	correctly predictive 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	semisupervised machine learning used to construct QSAR models. Molecules were featurized by either graph convolutional techniques or with 1024 Bit ECFP4 descriptors.	**B**, **F**	runner-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 modeling 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

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

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] The 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, creator of the previous model in Round 0), and Prof. Alexander Tropsha (University of North Carolina, Chapel Hill)) who evaluated the top 20 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 20 rankings, with a further model a close third place.[39]

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. A possible reason for this was the dissimilarity of the structures in the OSM Series 4 data set and the contents of the MMV Pathogen Box. Of note was, again, the striking diversity of chemotypes (A−K, Table 1) sharing a target. Interestingly, opinions of the performance of the models in this round differed between laboratory chemists (who regarded the 2/20 hit rate as not being practically helpful) vs cheminformatics-based entrants and judges (who regarded the 2/20 hit rate from a structurally diverse set of 400 compounds that was not strongly correlated with the training set, as a respectable outcome).

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 analogues" were still arising in the experimental program. 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 specialized 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 data set, including all of 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 data set of activity from recently synthesized analogues was kept back to serve as the basis for judging model fitness.

By the conclusion of Round 2 (a period of ∼10 weeks), 10 entries had been submitted, five of which were from returning participants (Table 2). In a similar fashion, submissions were

https://doi.org/10.1021/acs.jmedchem.1c00313
J. Med. Chem. XXXX, XXX, XXX−XXX

Table 2. Summary of the Results from Round 2 of the Predictive Modeling Competition

entrant (affiliation)	description of model[a]	precision of accurate predictions (active and inactive)[b]	result
Jonathan Cardoso-Silva (King's College London)	Network-based piecewise linear regression for QSAR modeling.[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 regularization, and learning step size = 0.01).	**91%**[c]	**winner (company)**
Mykola Galushka (Auromind)	SMILES variational autoencoder to generate chemical compounds fingerprint and cascade models Naive Bayes classifier with multilayer 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 three corrections method[44] and JCLogP, optimized for mean absolute error.	**82%**	**winner (noncompany)**
Ben Irwin, Mario Öeren, Tom Whitehead (Optibrium/Intellegens)	Deep imputation[45−47] with quantum mechanical StarDrop6.6 AutoModeller and pK_a 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 optimized for mean absolute error.	58%	runner-up
Slade Matthews (The University of Sydney)	Random forest model using 200 mordred descriptors based on optimized 3D structures. Training RMSE = 0.805.	N.A.	runner-up
Ho-Leung Ng (Kansas State University)	QSAR model based on 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.	71%	runner-up
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 the Supporting Information (SI) for full experimental details. [b]Based on regression prediction. [c]Based on classification prediction.

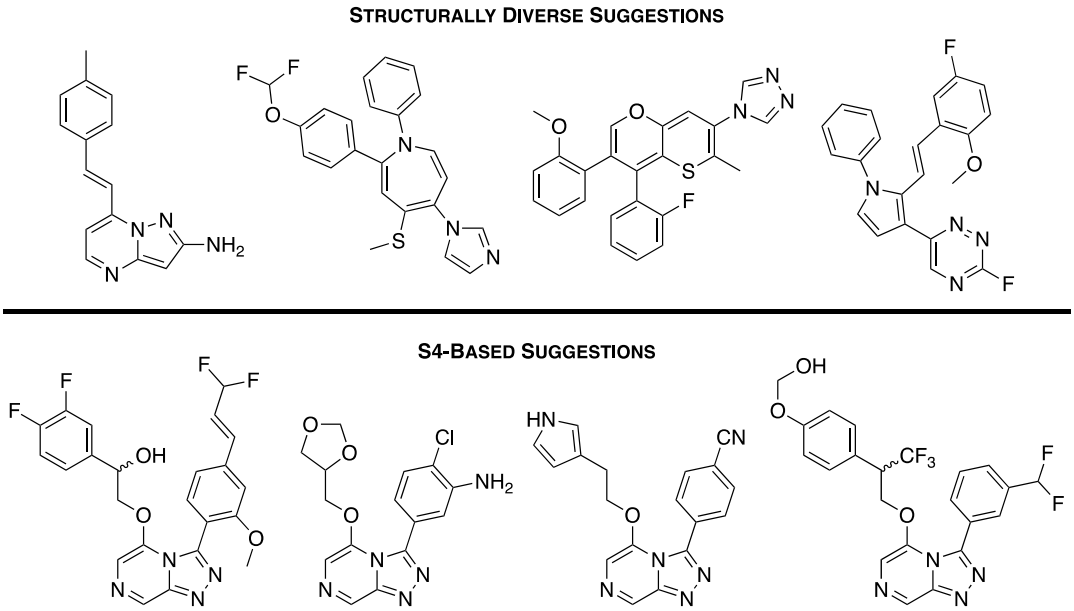


Figure 4. Examples of the suggested compounds predicted by the winning entrants from Round 2.

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 34 compounds.

The precision of each model was calculated according to: precision = $x/(x + y)$, where x is the number of correct predictions (active and inactive combined) and y is the number of false-positive predictions.[42]

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 synthesized and validated experimentally. In addition to optimizing 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, and these were triaged with some minor human inputs from the computational and synthetic teams; these inputs varied from team to team and typically involved selecting between the highest-scoring compounds. Synthetic tractability is often an issue when predictive models do not take into account known synthetic pathways, though there is significant

https://doi.org/10.1021/acs.jmedchem.1c00313
J. Med. Chem. XXXX, XXX, XXX−XXX

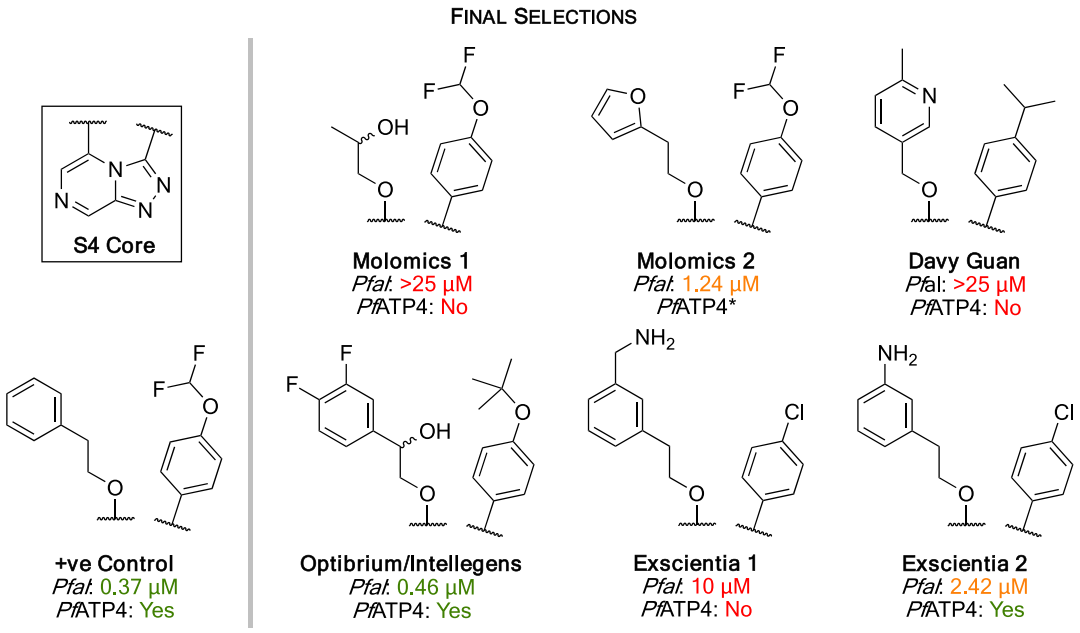


Figure 5. Six chosen suggested compounds for experimental validation. The predictions were synthesized (see the SI) and their potencies and MoAs (Figure S9) were experimentally validated. Three compounds were found to be active. *PfATP4 activity was not obtained for this compound.

activity at present to improve the incorporation of synthetic planning into library suggestion (Figure 4).[49,50]

The initial list was narrowed to focus on six predicted triazolopyrazine compounds (Figure 5). The six compounds were successfully synthesized 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 (cytosolic [Na^+]) assays to confirm that the MoA had not changed following these structural changes.

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 cutoff 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 cytosolic [Na^+] 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 (Figure 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, lies in the class of inactive compounds subject to human retrospective wisdom (i.e., the "Could Have Told You That" class). In contrast, 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). Finally, 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 (where related substituents such as -OMe have been observed to perform poorly, and much time had been spent in the production of inactive variants); yet, this compound displayed good potency and is a particularly useful outcome (i.e., the "Machine Overlords" class).

To gain more insight, and to improve these potential antimalarials, further iterations of these models are needed. The open nature of competitions and of the overarching consortium is that anyone may work on improvements since everyone has access to all of the data, making this a "living" research project. A potential explanation for the predicted hit rate not being higher is the relatively small data set (∼400 compounds) from which each model was developed, potentially compromising perfectly reasonably computational approaches yet representing a fairly typical situation for lead optimization. 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 this endeavor has already been used as an unvarnished 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 were less important than the extent to which

https://doi.org/10.1021/acs.jmedchem.1c00313
J. Med. Chem. XXXX, XXX, XXX−XXX

entrants worked together openly to improve the underlying research.[41]

CONCLUSIONS

With hit identification and lead optimization 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 synthesized 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 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.

EXPERIMENTAL SECTION

General Information. Reagents were purchased from either Sigma–Aldrich, Alfa Aesar, Acros, Merck, Fischer Scientific, Matrix Scientific, Ajax or Fluorochem. Unless otherwise specified, the reagents were used without further purification. Anhydrous solvents were obtained by drying over activated 3 Å molecular sieves. Argon gas was used as acquired. Reduced pressure means under rotary evaporation at 40 °C from 900 to 50 mbar. Flash chromatography was performed on a Biotage Selekt. Analytical thin-layer chromatography was performed on Merck Silica Gel 60 F_{254} precoated aluminum plates (0.2 mm) and visualized with UV irradiation (254 nm) and potassium permanganate. High-temperature reactions were carried out in silicone oil baths, controlled by a temperature probe in the oil bath.

Melting points (mp) were recorded on a Stuart SMP10 at 2 °C min^{-1} (capillaries ø = 1.8–1.9 mm, 100 mm). Nuclear magnetic resonance spectroscopy was carried out at 300 K on Bruker spectrometers: either AVANCE 200 (^{1}H at 200 MHz), AVANCE 300 (^{1}H at 300 MHz, ^{13}C at 75 MHz), AVANCE III 400 (^{1}H at 400 MHz, ^{13}C at 101 MHz), or AVANCE III 500 (^{1}H at 500 MHz, ^{13}C at 126 MHz). Spectra were processed using Mestrelab Research Mnova. Deuterated solvents ($CDCl_3$, DMSO-d_6, CD_3OD) were obtained from the Cambridge Isotope Laboratories. ^{1}H and ^{13}C chemical shifts are reported in parts per million (ppm) with respect to tetramethylsilane (TMS) at 0.00 ppm. The chemical shifts of the spectra were calibrated to residual solvent peaks (^{1}H: $CHCl_3$ 7.26 ppm, dimethyl sulfoxide (DMSO) 2.50 ppm, MeOH 3.31 ppm, TMS 0.00 ppm; ^{13}C: $CHCl_3$ 77.16 ppm, DMSO 39.52 ppm, MeOH 49.00 ppm, TMS 0.00 ppm). ^{1}H signal multiplicity is reported as: singlet (s), doublet (d), triplet (t), quartet (q), pentet (p) and combinations thereof, or multiplet (m). Broad signals are designated broad (br). Coupling constants (J) are reported in Hertz (Hz). Integrals are relative. $_{app}$ = apparent when the multiplicity was unexpected, e.g., coincidental or unresolved. Low-resolution mass spectrometry (m/z) was carried out on a Finnigan quadrupole ion trap mass spectrometer using electrospray ionization (ESI) or atmospheric-pressure chemical ionization (APCI). High-resolution mass spectrometry (HRMS) was performed on a Bruker 7T FT-ICR using ESI or APCI. Positive and negative detection is indicated by the charge of the ion, e.g., $[M + H]^+$ indicates positive ion detection. Analytical liquid chromatography-mass spectrometry (LCMS) was performed on an Agilent Infinity 1290 II system consisting of a quaternary pump (G7111A) and a diode array detector WR (G7115A) coupled to an InfinityLab LC/MSD (G6125B) using ESI. An Agilent Poroshell 120 EC-C18 column (2.7 μm, 3.0 mm × 50 mm) was eluted at a flow rate of 1.5 mL/min with a mobile phase of 0.05% formic acid in H_2O and 0.05% formic acid in MeCN.

The purity of all evaluated compounds was >95% as determined by NMR spectroscopy (provided for all compounds evaluated biologically).

Compounds 2-chloro-6-hydrazinylpyrazine (**1**), (*E*)-2-chloro-6-(2-(4-(difluoromethoxy)benzylidene)hydrazinyl)pyrazine (**5**), 5-chloro-3-(4-(difluoromethoxy)phenyl)-[1,2,4]triazolo[4,3-*a*]pyrazine (**9**), and 3-(4-(difluoromethoxy)phenyl)-5-phenethoxy-[1,2,4]triazolo-[4,3-*a*]pyrazine (**+ve Control** compound in Figure 4) were previously synthesized according to literature procedures.[51]

General Procedure 1: Condensation of Hydrazinylpyrazine with Aldehyde. Compound **1** (1 equiv) was dissolved in EtOH (112 mM). Aldehyde (1 equiv) was added and the reaction stirred at rt overnight. The suspension was filtered and washed with cold EtOH to give the corresponding hydrazone that was used without further purification.

General Procedure 2: Cyclization of Hydrazone to Triazolopyrazine Core. The product from General Procedure 1 (1 equiv) was dissolved in CH_2Cl_2 (112 mM). $PhI(OAc)_2$ (1 equiv) was added and the reaction stirred at rt overnight. The reaction was quenched with sat. $NaHCO_3$ solution, diluted with CH_2Cl_2, and the organic layer was separated. The aqueous layer was extracted with CH_2Cl_2 (2×) and the combined organic layers were washed with sat. $NaHCO_3$ solution, brine, dried ($MgSO_4$), filtered, and concentrated under reduced pressure to give the crude product, which was purified by automated flash chromatography on silica to give the corresponding triazolopyrazine core.

General Procedure 3: Reduction of Esters to Alcohols. Ester (1 equiv) was dissolved in anhydrous tetrahydrofuran (THF) (566 mM) and cooled to 0 °C. $LiAlH_4$ (1 M in THF, 2 equiv) was added dropwise, and the reaction mixture stirred for 10 min at 0 °C, then at rt. Upon completion, the reaction was diluted with THF and cooled to 0 °C. H_2O (1 mL/1 g of $LiAlH_4$) was added followed by 15% aq. NaOH (1 mL/1 g of $LiAlH_4$) and H_2O (3 mL/1 g of $LiAlH_4$). The mixture was allowed to warm to rt and stirred for 15 min. $MgSO_4$ was added, and the reaction mixture was filtered through a pad of celite and concentrated under reduced pressure to give the crude product, which was purified by automated flash chromatography on silica to give the corresponding alcohol.

General Procedure 4: Nucleophilic Displacement of Triazolopyrazine Core Chlorine with Alcohol. Alcohol (1.0 equiv) was added to PhMe (168 mM) along with triazolopyrazine core (1.0 equiv), KOH (3.0 equiv), and 18-crown-6 (0.1 equiv). The reaction was stirred at rt until completion as indicated by thin-layer chromatography (TLC) (100% EtOAc). The reaction was diluted with H_2O and then extracted with EtOAc (3×). The combined organic layers were washed with H_2O until the aqueous layer became neutral, followed by brine, dried ($MgSO_4$), filtered, and concentrated under reduced pressure to give the crude product, which was purified by automated flash chromatography on silica to give the corresponding ether-linked product.

(*E*)-2-(2-(4-(*tert*-Butoxy)benzylidene)hydrazinyl)-6-chloropyrazine (2). Prepared according to General Procedure 1 from: compound **1** (750 mg, 5.19 mmol, 1 equiv) and 4-(*tert*-butoxy)-benzaldehyde (904 μL, 5.19 mmol, 1 equiv) to give **2** as a pearlescent pale yellow powder (1.29 g, 82%). Mp 207–211 °C; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 11.47 (s, 1H), 8.53 (s, 1H), 8.03 (s, 1H), 8.02 (s, 1H), 7.65 (d, J = 8.6 Hz, 2H), 7.02 (d, J = 8.6 Hz, 2H), and 1.33 (s, 9H); ^{13}C NMR (126 MHz, DMSO-d_6) δ: 156.5, 152.4, 145.6, 142.5, 132.1, 129.1, 128.7, 127.6, 123.5, 78.6, and 28.6; m/z (ESI+) 305 ($[M + H]^+$, 100%).

(*E*)-2-Chloro-6-(2-(4-isopropylbenzylidene)hydrazinyl)-pyrazine (3). Prepared according to General Procedure 1 from: compound **1** (1.00 g, 6.92 mmol) and 4-isopropylbenzaldehyde (1.05 mL, 6.92 mmol) to give **3** as a yellow powder (1.49 g, 78%). Mp 231–233 °C; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 11.51 (s, 1H), 8.54

https://doi.org/10.1021/acs.jmedchem.1c00313
J. Med. Chem. XXXX, XXX, XXX–XXX

(s, 1H), 8.04 (s, 2H), 7.65 (d, J = 8.2 Hz, 2H), 7.30 (d, J = 8.2 Hz, 2H), 2.92 (p, J = 6.9 Hz, 1H), and 1.21 (d, J = 6.9 Hz, 6H); ^{13}C NMR (126 MHz, DMSO-d_6) δ: 152.4, 150.1, 145.6, 142.8, 132.2, 132.1, 128.7, 126.8, 126.7, 33.4, and 23.7; m/z (ESI+) 275 ([M + H]$^+$, 100%).

(*E*)-2-Chloro-6-(2-(4-chlorobenzylidene)hydrazinyl)pyrazine (4). Prepared according to General Procedure 1 from: compound **1** (1.00 g, 6.92 mmol) and 4-chlorobenzaldehyde (972 mg, 6.92 mmol) to give **4** as a yellow powder (1.53 g, 82%). Mp 230−233 °C; ^{1}H NMR (500 MHz, DMSO-d_6) δ: 11.64 (s, 1H), 8.58 (s, 1H), 8.07 (s, 1H), 8.05 (s, 1H), 7.77 (d, J = 8.5 Hz, 2H), and 7.48 (d, J = 8.5 Hz, 2H); ^{13}C NMR (126 MHz, DMSO-d_6) δ: 152.2, 145.5, 141.3, 133.8, 133.4, 132.6, 128.9, and 128.3 (one obscured signal); m/z (ESI+) 267 ([M + H]$^+$, 100%).

3-(4-(*tert*-Butoxy)phenyl)-5-chloro-[1,2,4]triazolo[4,3-*a*]pyrazine (6). Prepared according to General Procedure 2 from: compound **2** (750 mg, 2.46 mmol); purified by automated flash chromatography on silica (25−100% EtOAc in hexanes) to give **6** as a pale yellow powder (595 mg, 80%). Mp 143−147 °C; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 9.31 (s, 1H), 7.85 (s, 1H), 7.52 (d, J = 8.7 Hz, 2H), 7.12 (d, J = 8.7 Hz, 2H), and 1.43 (s, 9H); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 158.2, 148.4, 147.3, 143.1, 132.4, 129.8, 123.0, 122.2, 121.0, 79.7, and 29.0; m/z (ESI+) 303 ([M + H]$^+$, 100%).

5-Chloro-3-(4-isopropylphenyl)-[1,2,4]triazolo[4,3-*a*]pyrazine (7). Prepared according to General Procedure 2 from: compound **3** (750 mg, 2.73 mmol); purified by automated flash chromatography on silica (25−100% EtOAc in hexanes) to give **7** as a pale orange powder (662 mg, 89%). Mp 133−137 °C; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 9.32 (s, 1H), 7.85 (s, 1H), 7.54 (d, J = 8.3 Hz, 2H), 7.37 (d, J = 8.1 Hz, 2H), 3.02 (p, J = 6.9 Hz, 1H), and 1.32 (d, J = 6.9 Hz, 6H); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 152.0, 148.6, 147.3, 143.1, 131.5, 129.7, 126.2, 124.0, 122.2, 34.3, and 24.0; m/z (ESI+) 273 ([M + H]$^+$, 100%).

5-Chloro-3-(4-chlorophenyl)-[1,2,4]triazolo[4,3-*a*]pyrazine (8). Prepared according to General Procedure 2 from: compound **4** (750 mg, 2.81 mmol); purified by automated flash chromatography on silica (25−100% EtOAc in hexanes) to give **8** as an orange powder (634 mg, 85%). Mp 180−183 °C; ^{1}H NMR (300 MHz, $CDCl_3$) δ: 9.29 (s, 1H), 7.86 (s, 1H), 7.55 (d, J = 8.6 Hz, 2H), and 7.48 (d, J = 8.5 Hz, 2H); ^{13}C NMR (75 MHz, $CDCl_3$) δ: 147.2, 142.9, 137.2, 132.6, 129.8, 128.3, 125.1, and 121.8 (one obscured signal); m/z (ESI+) 265 ([M + H]$^+$, 100%).

Methyl 2-(3,4-Difluorophenyl)-2-hydroxyacetate (10). 2-(3,4-Difluorophenyl)-2-hydroxyacetic acid (850 mg, 4.52 mmol, 1 equiv) and *p*-TsOH monohydrate (17.2 mg, 0.09 mmol, 0.02 equiv) were dissolved in MeOH (3.12 mL, 1.45 M) and the reaction heated to reflux (80 °C). The reaction was cooled to rt and the solvent removed. EtOAc was added to the residue and the organic layer washed with H_2O, sat. $NaHCO_3$ solution, brine, dried ($MgSO_4$), filtered, and concentrated under reduced pressure to give **10** as clear colorless oil that solidified on standing (738 mg, 81%). No further purification required. Mp 44−49 °C; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.30−7.24 (m, 1H), 7.23−7.02 (m, 2H), 5.14 (d, J = 5.0 Hz, 1H), 3.78 (s, 3H), and 3.52 (d, J = 5.1 Hz, 1H); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 173.6, 151.5 (d, J = 12.6 Hz), 149.5 (d, J = 12.6 Hz), 135.1, 122.8, 117.5 (d, J = 17.5 Hz), 115.8 (d, J = 18.4 Hz), 71.8, and 53.5; m/z (ESI+) 225 ([M + Na]$^+$, 100%).

Methyl 2-(3,4-Difluorophenyl)-2-((tetrahydro-2*H*-pyran-2-yl)oxy)acetate (11). Compound **10** (600 mg, 2.97 mmol, 1.0 equiv) was dissolved in CH_2Cl_2 (10.7 mL, 277 mM), *p*-TsOH (102 mg, 0.59 mmol, 0.2 equiv), and 3,4-dihydro-2*H*-pyran (0.3 mL, 3.26 mmol, 1.1 equiv) were added and the reaction stirred at rt. The reaction was quenched with ice cold H_2O and the organic layer separated. The aqueous layer was extracted with CH_2Cl_2 (3×) and the combined organic layers washed with H_2O, brine, dried ($MgSO_4$), filtered, and concentrated under reduced pressure to give the crude product, which was purified by automated flash chromatography on silica (6−50% ethyl acetate in hexanes) to give **11** as a viscous dark orange oil (395 mg, 46%). ^{1}H NMR (500 MHz, $CDCl_3$, present as a mixture of diastereomers) δ: 7.38−7.28 (m, 2H), 7.24−7.10 (m, 4H), 5.27 (s, 1H), 5.18 (s, 1H), 4.86 (t, J = 3.0 Hz, 1H), 4.57 (t, J = 3.4 Hz, 1H), 3.72 (s, 6H), 3.55−3.44 (m, 4H), and 1.93−1.36 (m, 12H); ^{13}C NMR (126 MHz, $CDCl_3$, present as a mixture of diastereomers) δ: 171.3, 170.7, 150.7 (dd, J = 249.4, 12.0 Hz), 150.6 (dd, J = 249.3, 13.1 Hz), 150.5 (dd, J = 249.3, 13.4 Hz), 150.4 (dd, J = 248.5, 12.6 Hz), 133.8 (dd, J = 5.4, 4.2 Hz), 133.7−132.9 (m), 123.8 (dd, J = 6.6, 3.6 Hz), 123.4 (dd, J = 6.4, 3.7 Hz), 117.5 (t, J = 18.1 Hz, 2C), 116.6 (d, J = 18.0 Hz), 116.4 (d, J = 18.4 Hz), 97.4, 97.0, 75.8, 74.5, 62.5, 62.3, 52.62, 52.58, 30.3, 30.2, 25.3 (2C), 19.1, and 18.8; m/z (ESI+) 309 ([M + Na]$^+$, 100%).

2-(3,4-Difluorophenyl)-2-((tetrahydro-2*H*-pyran-2-yl)oxy)-ethan-1-ol (12). Prepared according to General Procedure 3 from: compound **11** (300 mg, 1.05 mmol); purified by automated flash chromatography on silica (25−100% EtOAc in hexanes) to give **12** as a viscous pale yellow oil (150 mg, 55%). ^{1}H NMR (500 MHz, $CDCl_3$, present as a mixture of diastereomers) δ: 7.28−6.92 (m, 6H), 4.92−4.81 (m, 1H), 4.81−4.72 (m, 1H), 4.68 (dd, J = 6.8, 4.6 Hz, 1H), 4.50 (dd, J = 5.6, 2.8 Hz, 1H), 4.00 (dt, J = 11.0, 5.2 Hz, 1H), 3.73−3.62 (m, 4H), 3.56 (tt, J = 10.2, 4.6 Hz, 2H), 3.32 (dt, J = 10.9, 4.7 Hz, 1H), 3.05−2.96 (m, 1H), 2.18−2.07 (m, 1H), and 1.91−1.36 (m, 12H); ^{13}C NMR (126 MHz, $CDCl_3$, present as a mixture of diastereomers) δ: 150.5 (dd, J = 248.8, 12.8 Hz), 150.4 (dd, J = 248.2, 12.7 Hz), 150.1 (dd, J = 248.3, 12.7 Hz), 149.9 (dd, J = 247.6, 12.6 Hz), 137.7−136.7 (m), 136.3−135.5 (m), 122.9 (dd, J = 6.3, 3.6 Hz), 122.7 (dd, J = 6.3, 3.6 Hz), 117.4 (d, J = 17.3 Hz), 117.2 (d, J = 17.2 Hz), 115.9 (d, J = 17.7 Hz), 115.8 (d, J = 17.8 Hz), 99.6, 98.1, 79.6, 78.8, 67.5, 66.5, 63.9, 62.9, 31.1, 30.7, 25.30, 25.26, 20.3, and 19.6; m/z (ESI+) 281 ([M + Na]$^+$, 100%).

(6-Methylpyridin-3-yl)methanol (13). Prepared according to General Procedure 3 from: 6-methylnicotinic acid (750 mg, 5.47 mmol); purified by automated flash chromatography on silica (1−10% MeOH in CH_2Cl_2) to give **13** as a yellow oil (103 mg, 15%). ^{1}H NMR (500 MHz, $CDCl_3$) δ: 8.42 (d, J = 2.0 Hz, 1H), 7.61 (dd, J = 7.9, 2.2 Hz, 1H), 7.14 (d, J = 7.9 Hz, 1H), 4.66 (s, 2H), and 2.53 (s, 3H) (alcohol OH signal not seen); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 157.6, 147.8, 135.8, 133.6, 123.3, 62.8, and 30.1; m/z (ESI+) 146 ([M + Na]$^+$, 100%). Spectroscopic data matched those in the literature.[53]

***tert*-Butyl (3-(2-Hydroxyethyl)benzyl)carbamate (14).** Prepared according to General Procedure 3 from: 2-(3-(((*tert*-butoxycarbonyl)amino)methyl)phenyl)acetic acid (500 mg, 1.88 mmol); purified by automated flash chromatography on silica (1−10% MeOH in CH_2Cl_2) to give **14** as a viscous clear colorless oil (332 mg, 70%). ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.27 (t, J = 7.8 Hz, 1H), 7.18−7.05 (m, 3H), 4.90 (br s, 1H), 4.29 (br d, J = 5.5 Hz, 2H), 3.84 (t, J = 6.6 Hz, 2H), 2.84 (t, J = 6.6 Hz, 2H), and 1.46 (s, 9H) (alcohol OH signal not seen); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 156.1, 139.3, 139.1, 129.0, 128.2, 128.1, 125.7, 79.7, 63.7, 44.7, 39.2, and 28.5; m/z (ESI+) 274 ([M + Na]$^+$, 100%).

Methyl 2-(3-((*tert*-Butoxycarbonyl)amino)phenyl)acetate (15). Boc_2O (363 mg, 1.66 mmol, 1.1 equiv) was added to a solution of 3-aminophenylacetic acid methyl ester (250 mg, 1.51 mmol, 1.0 equiv) in CH_2Cl_2 (4 mL) and the reaction was stirred at rt overnight. Sat. NH_4Cl solution (6 mL) was added and the organic layer separated. The aqueous layer was extracted with CH_2Cl_2 (2 × 8 mL) and the combined organic layers were dried (Na_2SO_4), filtered, and concentrated under reduced pressure to give the crude product, which was purified by automated flash chromatography on silica (12−50% ethyl acetate in hexanes) to give **15** as a clear colorless oil (174 mg, 43%). ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.33 (br s, 1H), 7.23 (d, J = 4.4 Hz, 2H), 6.99−6.90 (m, 1H), 6.51 (br s, 1H), 3.68 (s, 3H), 3.59 (s, 2H), and 1.51 (s, 9H); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 172.0, 152.8, 138.7, 135.0, 129.3, 124.0, 119.4, 117.4, 80.7, 52.2, 41.2, and 28.4; m/z (ESI+) 288 ([M + Na]$^+$, 100%).

***tert*-Butyl (3-(2-Hydroxyethyl)phenyl)carbamate (16).** Prepared according to General Procedure 3 from: compound **15** (130 mg, 0.49 mmol); purified by automated flash chromatography on silica (12−100% EtOAc in hexanes) to give **16** as a viscous clear colorless oil (71.3 mg, 61%). ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.30 (br s, 1H), 7.22 (t, J = 7.7 Hz, 1H), 7.16 (d, J = 8.1 Hz, 1H), 6.90 (d, J = 7.3 Hz, 1H), 6.55 (br s, 1H), 3.84 (t, J = 6.5 Hz, 2H), 2.83 (t, J = 6.5

https://doi.org/10.1021/acs.jmedchem.1c00313
J. Med. Chem. XXXX, XXX, XXX−XXX

Hz, 2H), and 1.51 (s, 9H) (alcohol OH signal not seen); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 152.9, 139.7, 138.7, 129.2, 123.9, 119.2, 116.8, 80.7, 63.6, 39.3, and 28.5; *m/z* (ESI+) 260 ([M + Na]$^+$, 100%). Spectroscopic data matched those in the literature.[54]

2-(Furan-2-yl)ethan-1-ol (19). Prepared according to General Procedure 3 from: 2-(furan-2-yl)acetic acid (500 mg, 3.96 mmol); purified by automated flash chromatography on silica (12−100% EtOAc in hexanes) to give **16** as a yellow oil (291 mg, 65%). ^{1}H NMR (500 MHz, $CDCl_3$) δ: 7.69−7.19 (m, 1H), 6.30 (dd, *J* = 2.9, 2.0 Hz, 1H), 6.09 (d, *J* = 3.1 Hz, 1H), 3.85 (t, *J* = 6.3 Hz, 2H), and 2.88 (t, *J* = 6.3 Hz, 2H) (alcohol OH signal not seen); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 153.0, 141.6, 110.4, 106.5, 61.1, and 31.6.

3-(4-(*tert*-Butoxy)phenyl)-5-(2-(3,4-difluorophenyl)-2-((tetrahydro-2*H*-pyran-2-yl)oxy)ethoxy)-[1,2,4]triazolo[4,3-*a*]pyrazine (20). Prepared according to General Procedure 4 from: compound 12 (100 mg, 0.39 mmol) and compound 6 (117 mg, 0.39 mmol); purified by automated flash chromatography on silica (25−100% EtOAc in hexanes) to give 20 as an orange powder (95.9 mg, 47%). Mp 72−75 °C; ^{1}H NMR (500 MHz, $CDCl_3$, present as a mixture of diastereomers) δ: 9.02 (s, 1H), 9.01 (s, 1H), 7.62 (dd, *J* = 8.6, 1.5 Hz, 4H), 7.38 (s, 1H), 7.29 (s, 1H), 7.14 (dd, *J* = 8.6, 1.5 Hz, 4H), 7.09−7.01 (m, 1H), 7.00 (dt, *J* = 9.9, 8.2 Hz, 1H), 6.95 (ddd, *J* = 10.6, 7.6, 2.1 Hz, 1H), 6.80 (dq, *J* = 6.4, 2.0 Hz, 1H), 6.73 (ddd, *J* = 10.2, 7.6, 2.1 Hz, 1H), 6.60 (dq, *J* = 6.1, 1.9 Hz, 1H), 4.89 (t, *J* = 5.5 Hz, 1H), 4.70 (dd, *J* = 6.7, 4.8 Hz, 1H), 4.44 (dd, *J* = 9.9, 5.3 Hz, 1H), 4.37 (q, *J* = 5.1, 4.3 Hz, 2H), 4.29 (dd, *J* = 9.8, 6.9 Hz, 1H), 4.22 (ddd, *J* = 19.1, 9.9, 5.3 Hz, 2H), 3.77 (ddd, *J* = 10.7, 7.1, 3.3 Hz, 1H), 3.47−3.40 (m, 1H), 3.37 (ddd, *J* = 11.7, 9.1, 3.0 Hz, 1H), 3.25 (dt, *J* = 11.0, 4.6 Hz, 1H), 1.87−1.45 (m, 12H), 1.44 (s, 9H), and 1.43 (s, 9H); ^{13}C NMR (126 MHz, $CDCl_3$, present as a mixture of diastereomers) δ: 157.8, 157.7, 150.5 (dd, *J* = 249.7, 12.6 Hz), 150.42 (dd, *J* = 250.0, 12.5 Hz), 150.36 (dd, *J* = 249.0, 13.0 Hz), 150.2 (dd, *J* = 248.7, 12.5 Hz), 147.90, 147.88, 147.3, 147.2, 144.04, 143.99, 137.1, 137.0, 136.2−135.5 (m), 134.6 (t, *J* = 4.2 Hz), 131.7, 131.6, 123.3 (dd, *J* = 6.3, 3.5 Hz), 123.1, 123.0, 122.6 (dd, *J* = 6.2, 3.6 Hz), 122.3, 122.1, 117.6 (d, *J* = 17.4 Hz), 117.4 (d, *J* = 17.4 Hz), 116.1 (d, *J* = 17.7 Hz), 115.8 (d, *J* = 17.9 Hz), 109.1, 108.7, 99.3, 96.6, 79.6, 74.0, 73.7, 73.4, 63.2, 62.4, 30.6, 30.5, 29.04, 29.03, 25.3, 25.2, 19.7, and 19.1 (two obscured signals); *m/z* (ESI+) 525 ([M + H]$^+$, 100%).

***tert*-Butyl (3-(2-((3-(4-Chlorophenyl)-[1,2,4]triazolo[4,3-*a*]pyrazin-5-yl)oxy)ethyl)benzyl)carbamate (21).** Prepared according to General Procedure 4 from: compound **14** (150 mg, 0.60 mmol) and compound **8** (158 mg, 0.60 mmol); purified by automated flash chromatography on silica (25−100% EtOAc in hexanes) to give **21** as a light brown powder (209 mg, 73%). ^{1}H NMR (500 MHz, $CDCl_3$) δ: 8.98 (s, 1H), 7.59 (d, *J* = 8.5 Hz, 2H), 7.39 (d, *J* = 8.5 Hz, 2H), 7.30 (s, 1H), 7.19 (t, *J* = 7.5 Hz, 1H), 7.14 (d, *J* = 7.6 Hz, 1H), 6.86 (br s, 1H), 6.75 (d, *J* = 7.4 Hz, 1H), 4.86 (br s, 1H), 4.43 (t, *J* = 6.6 Hz, 2H), 4.22 (d, *J* = 5.6 Hz, 2H), 2.94 (t, *J* = 6.5 Hz, 2H), and 1.43 (s, 9H); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 156.0, 147.9, 146.3, 143.9, 139.7, 136.6, 136.5, 136.4, 132.1, 129.1, 128.1, 127.7, 127.6, 126.33, 126.26, 108.5, 79.7, 71.2, 44.6, 34.5, and 28.5; *m/z* (ESI+) 480 ([M + H]$^+$, 100%), 502 ([M + Na]$^+$, 52%).

***tert*-Butyl (3-(2-((3-(4-Chlorophenyl)-[1,2,4]triazolo[4,3-*a*]pyrazin-5-yl)oxy)ethyl)phenyl)carbamate (22).** Prepared according to General Procedure 4 from: compound **16** (35.0 mg, 0.15 mmol) and compound **8** (39.1 mg, 0.15 equiv); purified by automated flash chromatography on silica (25−100% EtOAc in hexanes) to give **22** as a light brown powder (30.1 mg, 44%). ^{1}H NMR (500 MHz, $CDCl_3$) δ: 9.02 (s, 1H), 7.58 (d, *J* = 8.5 Hz, 2H), 7.42 (d, *J* = 8.5 Hz, 2H), 7.31 (s, 1H), 7.18−7.06 (m, 3H), 6.55 (d, *J* = 7.3 Hz, 1H), 6.43 (br s, 1H), 4.44 (t, *J* = 6.5 Hz, 2H), 2.93 (t, *J* = 6.5 Hz, 2H), and 1.51 (s, 9H); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 152.7, 148.0, 146.4, 144.0, 138.9, 137.2, 136.6, 136.4, 132.2, 129.4, 128.2, 126.4, 123.1, 118.5, 117.2, 108.5, 80.9, 71.2, 34.6, and 28.5; *m/z* (ESI+) 466 ([M + H]$^+$, 100%), 488 ([M + Na]$^+$, 45%).

2-((3-(4-(*tert*-Butoxy)phenyl)-[1,2,4]triazolo[4,3-*a*]pyrazin-5-yl)oxy)-1-(3,4-difluorophenyl)ethan-1-ol (Optibrium/Intellegens). Compound **20** (70.0 mg, 0.13 mmol, 1 equiv) was dissolved in EtOH (1.63 mL, 82 mM). $CuCl_2 \cdot 2H_2O$ (1.14 mg, 6.67 μmol, 5 mol %) was added and the reaction heated at reflux. The solvent was removed and EtOAc was added. The mixture was washed with H_2O (3×), then brine, and the organic layer was dried ($MgSO_4$), filtered, and concentrated under reduced pressure to give the crude product, which was purified by automated flash chromatography on silica (25−100% EtOAc in hexanes) to give Optibrium/Intellegens as a pale orange powder (49.2 mg, 84%). Mp 75−79 °C; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 9.05 (s, 1H), 7.66 (d, *J* = 8.6 Hz, 2H), 7.17 (d, *J* = 8.6 Hz, 2H), 7.17−7.06 (m, 1H), 7.03 (ddd, *J* = 10.2, 7.4, 1.6 Hz, 1H), 6.93−6.87 (m, 1H), 4.78 (dt, *J* = 7.4, 3.3 Hz, 1H), 4.26 (dd, *J* = 9.3, 3.3 Hz, 1H), 4.19−4.08 (m, 1H), 2.00 (d, *J* = 3.7 Hz, 1H), and 1.42 (s, 9H) (alcohol OH signal not seen); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 158.0, 150.6 (dd, *J* = 249.5, 12.7 Hz), 150.4 (dd, *J* = 249.7, 12.6 Hz), 147.8, 146.9, 143.8, 137.3, 135.2−135.0 (m), 131.7, 123.2, 122.6, 122.2 (dd, *J* = 6.4, 3.6 Hz), 117.8 (d, *J* = 17.4 Hz), 115.4 (d, *J* = 18.2 Hz), 108.5, 79.9, 75.0, 70.6, and 29.0; *m/z* (ESI+) 441 ([M + H]$^+$, 100%); HRMS (ESI+) found 441.1739 [M + H]$^+$; $C_{23}H_{22}F_2N_4O_3H^+$ requires 441.1738.

3-(4-Isopropylphenyl)-5-((6-methylpyridin-3-yl)methoxy)-[1,2,4]triazolo[4,3-*a*]pyrazine (Davy Guan). Prepared according to General Procedure 4 from: compound **13** (100 mg, 0.81 mmol) and compound **7** (221 mg, 0.81 mmol); purified by automated flash chromatography on silica (25−100% EtOAc in hexanes) to give a white powder (181 mg, 62%). Repurified by automated reversed-phase flash chromatography on silica (5−100% MeOH in H_2O) to give a white powder (40.6 mg). Repurified by automated reversed-phase flash chromatography on silica (5−75% MeOH in H_2O) to give **Davy Guan** as a white powder (16.3 mg, 6%). Mp decomposed >150 °C; ^{1}H NMR (400 MHz, CD_3OD) δ: 9.00 (s, 1H), 8.33 (d, *J* = 2.0 Hz, 1H), 7.65 (s, 1H), 7.54−7.48 (m, 3H), 7.23 (d, *J* = 8.0 Hz, 1H), 7.12 (d, *J* = 8.2 Hz, 2H), 5.32 (s, 2H), 2.87 (p, *J* = 6.9 Hz, 1H), 2.53 (s, 3H), and 1.20 (d, *J* = 6.9 Hz, 6H); ^{13}C NMR (101 MHz, CD_3OD) δ: 160.0, 153.0, 152.3, 149.9, 148.9, 146.0, 139.2, 136.6, 131.9, 128.7, 126.6, 125.8, 124.8, 110.2, 71.3, 35.2, 24.2, and 23.7; *m/z* (ESI+) 360 ([M + H]$^+$, 100%); HRMS (ESI+) found 360.1832 [M + H]$^+$; $C_{21}H_{21}N_5OH^+$ requires 360.1824.

(3-(2-((3-(4-Chlorophenyl)-[1,2,4]triazolo[4,3-*a*]pyrazin-5-yl)oxy)ethyl)phenyl)methanamine (Exscientia 1). Compound **21** (150 mg, 0.31 mmol, 1.00 equiv) was dissolved in CH_2Cl_2 (0.91 mL, 345 mM). Trifluoroacetic acid (TFA) (0.27 mL, 3.50 mmol, 11.2 equiv) was added and the reaction stirred at rt overnight. The solvent was removed and the residue directly purified by automated flash chromatography on silica (1−10% MeOH in CH_2Cl_2, then 100% MeOH) to give a sticky brown solid (135 mg). Repurified by automated reversed-phase flash chromatography on silica (5−100% MeOH in H_2O) to give **Exscientia 1** as a white powder (77.0 mg, 65%). Mp 74−77 °C; ^{1}H NMR (400 MHz, CD_3OD) δ: 8.97 (s, 1H), 7.68 (d, *J* = 8.5 Hz, 2H), 7.54 (s, 1H), 7.50 (d, *J* = 8.5 Hz, 2H), 7.32−7.25 (m, 2H), 7.07 (s, 1H), 6.98−6.89 (m, 1H), 4.60 (t, *J* = 6.2 Hz, 2H), 4.03 (s, 2H), and 3.02 (t, *J* = 6.2 Hz, 2H) (amine NH_2 signal not seen); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 149.0, 147.6, 145.9, 139.7, 137.5, 136.2, 134.7, 133.5, 130.6, 130.4, 130.1, 129.2, 128.1, 127.6, 110.0, 72.4, 44.2, and 35.2; *mj/z* (ESI+) 380 ([M + H]$^+$, 100%); HRMS (ESI+) found 380.1282 [M + H]$^+$; $C_{20}H_{18}ClN_5OH^+$ requires 380.1278.

3-(2-((3-(4-Chlorophenyl)-[1,2,4]triazolo[4,3-*a*]pyrazin-5-yl)oxy)ethyl)aniline (Exscientia 2). Compound **22** (21.0 mg, 0.05 mmol, 1.00 equiv) was dissolved in CH_2Cl_2 (0.13 mL, 345 mM). TFA (0.04 mL, 3.50 mmol, 11.2 equiv) was added and the reaction stirred at rt overnight. The solvent was removed and the residue directly purified by automated flash chromatography on silica (1−10% MeOH in CH_2Cl_2) to give a yellow film (23.5 mg). Repurified by automated reversed-phase flash chromatography on silica (5−100% MeOH in H_2O) to give **Exscientia 2** as a white powder (11.4 mg, 69%). Mp 80−84°C; ^{1}H NMR (500 MHz, CD_3OD) δ: 8.93 (s, 1H), 7.61 (d, *J* = 8.6 Hz, 2H), 7.52 (s, 1H), 7.46 (d, *J* = 8.6 Hz, 2H), 6.94 (t, *J* = 7.8 Hz, 1H), 6.55 (d, *J* = 9.3 Hz, 1H), 6.37 (s, 1H), 6.26 (d, *J* = 7.5 Hz, 1H), 4.51 (t, *J* = 6.4 Hz, 2H), and 2.84 (t, *J* = 6.4 Hz, 2H) (amine NH_2 signal not seen); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 149.0, 149.0,

https://doi.org/10.1021/acs.jmedchem.1c00313
J. Med. Chem. XXXX, XXX, XXX−XXX

147.8, 145.9, 139.2, 137.5, 136.0, 133.5, 130.2, 129.1, 127.4, 119.5, 116.4, 114.9, 110.0, 72.8, and 35.3; *m/z* (ESI+) 366 ([M + H]$^+$, 100%); HRMS (ESI+) found 366.1122 [M + H]$^+$; $C_{19}H_{16}ClN_5OH^+$ requires 366.1122.

1-((3-(4-(Difluoromethoxy)phenyl)-[1,2,4]triazolo[4,3-*a*]-pyrazin-5-yl)oxy)propan-2-ol (Molomics 1). To a mixture of lactic acid (0.83 mL, 11.1 mmol, 1.0 equiv), 3,4-dihydro-2*H*-pyran (3.04 mL, 33.3 mmol, 3.0 equiv) and CH_2Cl_2 (2.98 mL, 3.73 M) in an ice bath was added pyridinium *p*-toluenesulfonate (279 mg, 1.11 mmol, 0.1 equiv) and pyridine (1 drop). The reaction was stirred at rt overnight. CH_2Cl_2 was added and the solution washed with 5% $NaHCO_3$, H_2O (2×), dried ($MgSO_4$), filtered, and concentrated under reduced pressure to give a mixture of **17a** and **17b** as a clear colorless oil (1.09 g, 38%). Used without further purification. *m/z* (ESI+) 209 (**17b**, [M + Na]$^+$, 100%); *m/z* (ESI+) 281 (**17a**, [M + Na]$^+$, 100%). The mixture of **17a** and **17b** (700 mg, 2.71 mmol) was subjected to General Procedure 3 and purified by automated flash chromatography on silica (12−100% EtOAc to hexanes) to give a ∼1:2 mixture of **18a** and **18b** as a clear colorless oil (96.3 mg, 17%). ^{1}H NMR (500 MHz, $CDCl_3$) δ: 4.72 (dd, *J* = 5.1, 2.8 Hz, 1H), 4.56 (dd, *J* = 4.3, 2.8 Hz, 2H), 3.96−3.90 (m, 1H), 3.86 (dtd, *J* = 11.2, 7.6, 7.1, 3.5 Hz, 3H), 3.74 (dt, *J* = 9.6, 6.7 Hz, 2H), 3.64 (t, *J* = 6.5 Hz, 4H), 3.58 (dd, *J* = 11.6, 3.5 Hz, 1H), 3.54−3.43 (m, 4H), 3.39 (dt, *J* = 9.6, 6.5 Hz, 2H), 1.92−1.33 (m, 33H), and 1.21 (d, *J* = 6.4 Hz, 3H); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 99.2, 99.1, 75.1, 67.6, 66.3, 63.4, 63.0, 62.6, 32.7, 31.2, 30.9, 29.6, 25.6, 25.5, 22.6, 20.2, 19.8, and 17.8; *m/z* (ESI+) 183 (**18a**, [M + Na]$^+$, 100%); *m/z* (ESI+) 211 (**18b**, [M + Na]$^+$, 100%). The mixture of **18a** and **18b** (70.0 mg, 0.44 mmol) was subjected to General Procedure 4 with compound **9** (130 mg, 0.44 mmol) and purified by automated flash chromatography on silica (25−100% EtOAc in hexanes) to give a ∼1:1 mixture of **23a** and **23b** as a brown powder (87.0 mg, 47%). ^{1}H NMR (500 MHz, $CDCl_3$) δ: 9.05 (s, 1H), 9.03 (s, 1H), 7.73 (t, *J* = 9.0 Hz, 4H), 7.33 (s, 1H), 7.29 (s, 1H), 7.29−7.20 (m, 4H), 6.64 (t, *J* = 73.3 Hz, 1H), 6.63 (t, *J* = 73.3 Hz, 1H), 4.52 (t, *J* = 3.7 Hz, 1H), 4.24−4.14 (m, 3H), 4.16−4.09 (m, 2H), 3.97−3.88 (m, 1H), 3.82 (dtd, *J* = 11.7, 8.8, 8.2, 3.8 Hz, 2H), 3.66 (dt, *J* = 9.7, 6.6 Hz, 1H), 3.54−3.40 (m, 2H), 3.29 (dt, *J* = 9.5, 6.2 Hz, 1H), 1.85−1.30 (m, 18H), and 1.07 (d, *J* = 6.5 Hz, 3H); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 152.7, 152.5, 147.91, 147.88, 146.3, 144.3, 144.2, 136.6, 136.3, 132.6, 132.5, 125.0, 124.9, 118.9, 118.5, 115.9 (t, *J* = 260.7 Hz), 115.7 (t, *J* = 261.1 Hz), 108.5, 108.2, 99.2, 98.7, 74.0, 71.0, 70.8, 67.2, 62.7, 62.4, 53.6, 30.85, 30.82, 29.2, 28.4, 25.5, 25.4, 22.8, 19.9, 19.3, and 18.2; *m/z* (ESI+) 421 (**23a**, [M + H]$^+$, 100%); *m/z* (ESI+) 449 (**23b**, [M + H]$^+$, 100%). The mixture of **23a** and **23b** (56.0 mg, 0.13 mmol, 1 equiv) was dissolved in EtOH (1.62 mL, 82 mM). $CuCl_2{\cdot}2H_2O$ (1.14 mg, 6.66 μmol, 5 mol %) was added and the reaction heated at reflux. The solvent was removed and EtOAc was added. The mixture was washed with H_2O (3×), then brine, and the organic layer was dried ($MgSO_4$), filtered, and concentrated under reduced pressure to give the crude product, which was purified by automated flash chromatography on silica (1−15% MeOH in CH_2Cl_2) to give a brown oil (39.8 mg). Repurified by automated reversed-phase flash chromatography on silica (5−100% MeOH in H_2O) to give **Molomics 1** as a white powder (10.3 mg, 23%). Mp 137−140 °C; ^{1}H NMR (400 MHz, $CDCl_3$) δ: 9.07 (s, 1H), 7.75 (d, *J* = 8.7 Hz, 2H), 7.32 (s, 1H), 7.29 (d, *J* = 8.6 Hz, 2H), 6.62 (t, *J* = 73.0 Hz, 1H), 4.18−4.12 (m, 1H), 4.04−3.90 (m, 2H), and 1.06 (d, *J* = 6.3 Hz, 3H); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 152.5, 147.9, 146.1, 144.0, 136.9, 132.5, 125.3, 119.1, 115.6 (t, *J* = 262.1 Hz), 108.5, 75.6, 65.5, and 18.7; *m/z* (ESI+) 337 ([M + H]$^+$, 100%); HRMS (ESI+) found 337.1107 [M + H]$^+$; $C_{15}H_{14}F_2N_4O_3H^+$ requires 337.1112.

5-((3-(4-(Difluoromethoxy)phenyl)-[1,2,4]triazolo[4,3-*a*]-pyrazin-5-yl)oxy)pentan-1-ol (Molomics 1′). Isolated from the same reaction as for **Molomics 1** to give **Molomics 1′** as a white powder (16.1 mg, 36%). Mp 104−108 °C; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 9.01 (s, 1H), 7.70 (d, *J* = 8.7 Hz, 2H), 7.28 (s, 1H), 7.24 (d, *J* = 8.6 Hz, 2H), 6.65 (t, *J* = 73.2 Hz, 1H), 4.20 (t, *J* = 6.0 Hz, 2H), 3.54 (t, *J* = 6.3 Hz, 2H), 1.79−1.54 (m, 2H), 1.53−1.30 (m, 2H), and 1.14 (ddd, *J* = 11.7, 4.6, 2.5 Hz, 2H); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 152.5, 147.9, 146.3, 144.3, 136.3, 132.6, 125.1, 118.6, 115.7 (t, *J* = 261.3 Hz), 108.2, 71.0, 62.5, 32.0, 28.5, and 22.3; *m/z* (ESI+) 365 ([M + H]$^+$, 100%); HRMS (ESI+) found 365.1424 [M + H]$^+$; $C_{17}H_{18}F_2N_4O_3H^+$ requires 365.1425.

3-(4-(Difluoromethoxy)phenyl)-5-(2-(furan-2-yl)ethoxy)-[1,2,4]triazolo[4,3-*a*]pyrazine (Molomics 2). Prepared according to General Procedure 4 from: compound **19** (100 mg, 0.89 mmol) and compound **9** (265 mg, 0.89 mmol); purified by automated flash chromatography on silica (25−100% EtOAc in hexanes) to give **Molomics 2** as a brown powder (213 mg, 64%). Mp 120−123°C; ^{1}H NMR (500 MHz, $CDCl_3$) δ: 9.03 (s, 1H), 7.65 (d, *J* = 8.7 Hz, 2H), 7.33 (s, 1H), 7.31−7.27 (m, 1H), 7.17 (d, *J* = 8.7 Hz, 2H), 6.59 (t, *J* = 73.3 Hz, 1H), 6.26 (dd, *J* = 3.1, 1.9 Hz, 1H), 5.82 (d, *J* = 3.1 Hz, 1H), 4.47 (t, *J* = 6.3 Hz, 2H), and 2.98 (t, *J* = 6.3 Hz, 2H); ^{13}C NMR (126 MHz, $CDCl_3$) δ: 152.5 (t, *J* = 2.7 Hz), 150.1, 147.9, 146.5, 143.9, 142.0, 136.8, 132.5, 124.9, 118.7, 115.7 (t, *J* = 261.1 Hz), 110.6, 108.4, 107.1, 68.6, and 27.5; *m/z* (ESI+) 373 ([M + H]$^+$, 100%); HRMS (ESI+) found 373.1118 [M + H]$^+$; $C_{18}H_{14}F_2N_4O_3H^+$ requires 373.1112.

***In Vitro* Antiplasmodial Activity (Drug Discovery Unit, University of Dundee).**[55] Cultures of the widely used malaria reference strain of chloroquine-sensitive *P. falciparum* strain 3D7 were maintained in a 5% suspension of human red blood cells cultured in RPMI 1640 medium supplemented with 0.5% Albumax II (available from Gibco Life Technologies, San Diego, CA, cat. no. 11021-037), 12 mM sodium bicarbonate, 0.2 mM hypoxanthine, (pH 7.3), and 20 mg/L gentamicin at 37 °C, in a humified atmosphere of 1% O_2, 3% CO_2 with a gas balance of nitrogen. Growth inhibition of *P. falciparum* cultures was quantified in a 10-point dose−response curve with a one in three dilution series. Top concentrations of either 10 or 25 μM were used, indicating the upper limit of detection of the assay in each instance. This 384 well plate-based fluorescence assay utilizes the binding of SYBRgreen I (Thermo Fisher Scientific/Invitrogen cat. no. S7585) to double-stranded DNA, which greatly increases the fluorescent signal at 528 nm after excitation at 485 nm. Mefloquine was used as a drug control to monitor the quality of the assay (Z' = 0.6−0.8, where Z' is a measure of the discrimination between the positive and negative controls on a screen plate). Dose−response curves were determined from a minimum of three independent experiments. Compound bioactivity was expressed as IC_{50}, the concentration of compound causing 50% inhibition. IC_{50} values were determined from a minimum of three independent experiments. All data was processed using IDBS ActivityBase; raw data was converted into percent inhibition through linear regression by setting the high inhibition control as 100% and the no inhibition control as 0%. Quality control criteria for passing plates were as follows: Z' > 0.5, S:B > 3, %CV (no inhibition control) < 15. The formula used to calculate was $Z' = 1 - \frac{3 \times (\mathrm{StDev_{high}} + \mathrm{StDev_{low}})}{\mathrm{ABS}(\mathrm{Mean_{high}} - \mathrm{Mean_{low}})}$. All IC_{50} curve fitting was undertaken using XLFit version 4.2 using Model 205 with the following four parametric equation: $y = A + {}^{B-A}/_{1+(Cx)^D}$, where A = % inhibition at bottom, B = % inhibition at top, C = IC_{50}, D = slope, x = inhibitor concentration, and y = % inhibition. If the curve did not reach 100% of inhibition, B was fixed to 100 only when at least 50% of inhibition was reached.

***P. falciparum* ATP4 Assay (Lehane and Kirk Labs, Australian National University).** To measure ion concentrations inside the parasite, *P. falciparum* trophozoites were isolated from their host erythrocytes by brief exposure to saponin (0.05% w/v final concentration), then loaded with either the Na^+-sensitive dye SBFI (for measurements of cytosolic Na^+ concentration)[17] or the pH-sensitive dye BCECF (for measurements of cytosolic pH).[56] Fluorescence measurements and calibrations were performed at 37 °C essentially as described previously, either in 96 well plates using a Tecan Infinite M1000 PRO plate reader or in individual 1 mL suspensions using a PerkinElmer LS 50B fluorescence spectrometer.[17,21,22]

https://doi.org/10.1021/acs.jmedchem.1c00313
J. Med. Chem. XXXX, XXX, XXX−XXX

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jmedchem.1c00313.

Round 2 predictive models, and ^{1}H and ^{13}C NMR spectra for all synthesized compounds (PDF)

Table of molecular formula strings, in vitro potency values, and calculated properties for target compounds (CSV)

Summary of Round 1 Competition Entries (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Matthew H. Todd − *School of Pharmacy, University College London, London WC1N 1AX, U.K.;* orcid.org/0000-0001-7096-4751; Email: matthew.todd@ucl.ac.uk

Authors

Edwin G. Tse − *School of Pharmacy, University College London, London WC1N 1AX, U.K.;* orcid.org/0000-0002-9133-9274

Laksh Aithani − *Exscientia Ltd., The Schrödinger Building, Oxford OX4 4GE, U.K.*

Mark Anderson − *Drug Discovery Unit, Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, Dundee DD1 5EH, U.K.*

Jonathan Cardoso-Silva − *Department of Informatics, Faculty of Natural and Mathematical Sciences, King's College London, London WC2B 4BG, U.K.*

Giovanni Cincilla − *Molomics, Barcelona Science Park, Barcelona 08028, Spain;* orcid.org/0000-0002-5242-0707

Gareth J. Conduit − *Intellegens Ltd., Eagle Labs, Cambridge CB4 3AZ, U.K.; Theory of Condensed Matter Group, Cavendish Laboratories, University of Cambridge, Cambridge CB3 0HE, U.K.*

Mykola Galushka − *Auromind Ltd, Belfast BT9 6EU, U.K.*

Davy Guan − *School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia;* orcid.org/0000-0001-6290-3166

Irene Hallyburton − *Drug Discovery Unit, Division of Biological Chemistry and Drug Discovery, School of Life Sciences, University of Dundee, Dundee DD1 5EH, U.K.*

Benedict W. J. Irwin − *Theory of Condensed Matter Group, Cavendish Laboratories, University of Cambridge, Cambridge CB3 0HE, U.K.; Optibrium Ltd. Blenheim House, Cambridge CB25 9QE, U.K.*

Kiaran Kirk − *Research School of Biology, Australian National University, Canberra ACT 2601, Australia*

Adele M. Lehane − *Research School of Biology, Australian National University, Canberra ACT 2601, Australia;* orcid.org/0000-0002-0050-9101

Julia C. R. Lindblom − *Research School of Biology, Australian National University, Canberra ACT 2601, Australia*

Raymond Lui − *School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia*

Slade Matthews − *School of Medical Sciences, The University of Sydney, Sydney, NSW 2006, Australia;* orcid.org/0000-0002-1652-543X

James McCulloch − *Kellerberrin, Sydney, NSW 2041, Australia*

Alice Motion − *School of Chemistry, The University of Sydney, Sydney, NSW 2006, Australia;* orcid.org/0000-0002-5859-7888

Ho Leung Ng − *Department of Biochemistry and Molecular Biophysics, Kansas State University, Manhattan, Kansas 66506, United States;* orcid.org/0000-0002-6415-1938

Mario Öeren − *Optibrium Ltd. Blenheim House, Cambridge CB25 9QE, U.K.*

Murray N. Robertson − *Strathclyde Institute Of Pharmacy And Biomedical Sciences, University of Strathclyde, Glasgow G4 ORE, U.K.;* orcid.org/0000-0001-9543-7667

Vito Spadavecchio − *Interlinked Therapeutics LLC, Portland, Oregon 97214, United States*

Vasileios A. Tatsis − *Exscientia Ltd., The Schrödinger Building, Oxford OX4 4GE, U.K.*

Willem P. van Hoorn − *Exscientia Ltd., The Schrödinger Building, Oxford OX4 4GE, U.K.*

Alexander D. Wade − *Theory of Condensed Matter Group, Cavendish Laboratories, University of Cambridge, Cambridge CB3 0HE, U.K.;* orcid.org/0000-0003-1500-3733

Thomas M. Whitehead − *Intellegens Ltd., Eagle Labs, Cambridge CB4 3AZ, U.K.*

Paul Willis − *Medicines for Malaria Venture, 1215 Geneva 15, Switzerland*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jmedchem.1c00313

Author Contributions

E.G.T., A.M., and M.N.R. carried out the synthetic chemistry. A.M.L., J.C.R.L., and K.K. planned and/or carried out the Na^+ and pH assays and/or advised on the results. I.H. developed the *Pf* assay, manually prepared and analyzed compounds, collated all data, and contributed to the paper (Experimental section and data table). M.A. carried out routine *Pf* assay and analysis. M.N.R., D.G., H.L.N., J.C.-S, B.W.J.I., M.Ö., T.M.W., G.J.C., A.D.W., L.A., V.A.T., W.P.v.H., J.M., V.S., R.L., S.M., G.C., and M.G. contributed models to the competitions and collaborated on the outputs during the competitions. P.W. helped conceive the study. M.H.T. founded OSM, conceived the project and secured (AI3SD-FundingCall1_029) or co-secured (LP150101226, with P.W. and K.K.) the funding. All authors contributed to the manuscript.

Funding

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 M.H.T.), the US National Science Foundation (1833181) (to H.L.N.), the EPSRC Centre for Doctoral Training in Computational Methods for Materials Science (EP/L015552/2) (to A.D.W.), the EPSRC Future Continuous Manufacturing and Advanced Crystallization (CMAC) Research Hub (EP/L015552/1) (to M.N.R.), and the Royal Society (to G.J.C.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank those contributors to Open Source Malaria not named as authors for their help with this research. We

thank Exscientia for funding the open access charges for this paper.

ABBREVIATIONS USED

ACT, artemisinin-based combination therapy; AI, artificial intelligence; ELN, electronic laboratory notebook; HTS, high-throughput screen; ML, machine learning; MMV, Medicines for Malaria Venture; MoA, mechanism of action; OSM, Open Source Malaria; QSAR, quantitative structure−activity relationship; SAR, structure−activity relationship; SERCA, sarco/endoplasmic reticulum Ca^{2+}-ATPase

REFERENCES

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

(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 Discovery* **2017**, *16*, 531−543.

(3) Sellwood, M. A.; Ahmed, M.; Segler, M. H. S.; Brown, N. Artificial Intelligence in Drug Discovery. *Future Med. Chem.* **2018**, *10*, 2025−2028.

(4) Griffen, E. J.; Dossetter, A. G.; Leach, A. G. Chemists: AI is Here; Unite to Get the Benefits. *J. Med. Chem.* **2020**, *63*, 8695−8704.

(5) Tyrchan, C.; Evertsson, E. Matched Molecular Pair Analysis in Short: Algorithms, Applications and Limitations. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 86−90.

(6) Neves, B. J.; Braga, R. C.; Melo-Filho, C. C.; Moreira-Filho, J. T.; Muratov, E. N.; Andrade, C. H. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Front. Pharmacol.* **2018**, *9*, 1275.

(7) Muhammed, M. T.; Aki-Yalcin, E. Homology Modelling in Drug Discovery: Overview, Current Applications, and Future Perspectives. *Chem. Biol. Drug Des.* **2019**, *93*, 12−20.

(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.* **2019**, *59*, 947−961.

(9) Brown, N.; Fiscato, M.; Segler, M. H. S.; Vaucher, A. C. GuacaMol: Benchmarking Models for de Novo Molecular Design. *J. Chem. Inf. Model.* **2019**, *59*, 1096−1108.

(10) Walters, W. P.; Murcko, M. Assessing the Impact of Generative AI on Medicinal Chemistry. *Nat. Biotechnol.* **2020**, *38*, 143−145.

(11) Stokes, J. M.; Yang, K.; Swanson, K.; Jin, W.; Cubillos-Ruiz, A.; Donghia, N. M.; MacNair, C. R.; French, S.; Carfrae, L. A.; Bloom-Ackermann, Z.; Tran, V. M.; Chiappino-Pepe, A.; Badran, A. H.; Andrews, I. W.; Chory, E. J.; Church, G. M.; Brown, E. D.; Jaakkola, T. S.; Barzilay, R.; Collins, J. J. A Deep Learning Approach to Antibiotic Discovery. *Cell* **2020**, *180*, 688−702.

(12) Mugumbate, G.; Abrahams, K. A.; Cox, J. A. G.; Papadatos, G.; van Westen, G.; Lelièvre, J.; Calus, S. T.; Loman, N. J.; Ballell, L.; Barros, D.; Overington, J. P.; Besra, G. S. Mycobacterial Dihydrofolate Reductase Inhibitors Identified Using Chemogenomic Methods and *In Vitro* Validation. *PLoS One* **2015**, *10*, No. e0121492.

(13) Homeyer, N.; van Deursen, R.; Ochoa-Montaño, B.; Heikamp, K.; Ray, P.; Zuccotto, F.; Blundell, T. L.; Gilbert, I. H. A Platform for Target Prediction of Phenotypic Screening Hit Molecules. *J. Mol. Graphics Modell.* **2020**, *95*, No. 107485.

(14) Qinghaosu Antimalaria Coordinating Research Group. Antimalaria Studies on Qinghaosu. *Chin. Med. J.* **1979**, *92*, 811−816.

(15) Hamilton, W. L.; Amato, R.; van der Pluijm, R. W.; Jacob, C. G.; Quang, H. H.; Thuy-Nhien, N. T.; Hien, T. T.; Hongvanthong, B.; Chindavongsa, K.; Mayxay, M.; Huy, R.; Leang, R.; Huch, C.; Dysoley, L.; Amaratunga, C.; Suon, S.; Fairhurst, R. M.; Tripura, R.; Peto, T. J.; Sovann, Y.; Jittamala, P.; Hanboonkunupakarn, B.; Pukrittayakamee, S.; Chau, N. H.; Imwong, M.; Dhorda, M.; Vongpromek, R.; Chan, X. H. S.; Maude, R. J.; Pearson, R. D.; Nguyen, T.; Rockett, K.; Drury, E.; Gonçalves, S.; White, N. J.; Day, N. P.; Kwiatkowski, D. P.; Dondorp, A. M.; Miotto, O. Evolution and Expansion of Multidrug-Resistant Malaria in Southeast Asia: A Genomic Epidemiology Study. *Lancet Infect. Dis.* **2019**, *19*, 943−951.

(16) Tse, E. G.; Korsik, M.; Todd, M. H. The Past, Present and Future of Anti-Malarial Medicines. *Malar. J.* **2019**, *18*, No. 93.

(17) Spillman, N. J.; Allen, R. J. W.; McNamara, C. W.; Yeung, B. K. S.; Winzeler, E. A.; Diagana, T. T.; Kirk, K. Na^+ Regulation in the Malaria Parasite *Plasmodium falciparum* Involves the Cation ATPase PfATP4 and Is a Target of the Spiroindolone Antimalarials. *Cell Host Microbe* **2013**, *13*, 227−237.

(18) Kirk, K. Ion Regulation in the Malaria Parasite. *Annu. Rev. Microbiol.* **2015**, *69*, 341−359.

(19) Rottmann, M.; McNamara, C.; Yeung, B. K. S.; Lee, M. C. S.; Zou, B.; Russell, B.; Seitz, P.; Plouffe, D. M.; Dharia, N. V.; Tan, J.; Cohen, S. B.; Spencer, K. R.; González-Páez, G. E.; Lakshminarayana, S. B.; Goh, A.; Suwanarusk, R.; Jegla, T.; Schmitt, E. K.; Beck, H.-P.; Brun, R.; Nosten, F.; Renia, L.; Dartois, V.; Keller, T. H.; Fidock, D. A.; Winzeler, E. A.; Diagana, T. T. Spiroindolones, a Potent Compound Class for the Treatment of Malaria. *Science* **2010**, *329*, 1175−1180.

(20) Jiménez-Díaz, M. B.; Ebert, D.; Salinas, Y.; Pradhan, A.; Lehane, A. M.; Myrand-Lapierre, M.-E.; O'Loughlin, K. G.; Shackleford, D. M.; de Almeida, M. J.; Carrillo, A. K.; Clark, J. A.; Dennis, A. S. M.; Diep, J.; Deng, X.; Duffy, S.; Endsley, A. N.; Fedewa, G.; Guiguemde, W. A.; Gómez, M. G.; Holbrook, G.; Horst, J.; Kim, C. C.; Liu, J.; Lee, M. C. S.; Matheny, A.; Martínez, M. S.; Miller, G.; Rodríguez-Alejandre, A.; Sanz, L.; Sigal, M.; Spillman, N. J.; Stein, P. D.; Wang, Z.; Zhu, F.; Waterson, D.; Knapp, S.; Shelat, A.; Avery, V. M.; Fidock, D. A.; Gamo, F.-J.; Charman, S. A.; Mirsalis, J. C.; Ma, H.; Ferrer, S.; Kirk, K.; Angulo-Barturen, I.; Kyle, D. E.; DeRisi, J. L.; Floyd, D. M.; Guy, R. K. (+)-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.* **2014**, *111*, E5455−E5462.

(21) Lehane, A. M.; Ridgway, M. C.; Baker, E.; Kirk, K. Diverse Chemotypes Disrupt Ion Homeostasis in the Malaria Parasite. *Mol. Microbiol.* **2014**, *94*, 327−339.

(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.* **2018**, *8*, No. 8795.

(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.* **2015**, *5*, 149−162.

(24) Vaidya, A. B.; Morrisey, J. M.; Zhang, Z.; Das, S.; Daly, T. M.; Otto, T. D.; Spillman, N. J.; Wyvratt, M.; Siegl, P.; Marfurt, J.; Wirjanata, G.; Sebayang, B. F.; Price, R. N.; Chatterjee, A.; Nagle, A.; Stasiak, M.; Charman, S. A.; Angulo-Barturen, I.; Ferrer, S.; Jiménez-Díaz, M. B.; Martínez, M. S.; Gamo, F. J.; Avery, V. M.; Ruecker, A.; Delves, M.; Kirk, K.; Berriman, M.; Kortagere, S.; Burrows, J.; Fan, E.; Bergman, L. W. Pyrazoleamide Compounds are Potent Antimalarials that Target Na^+ Homeostasis in Intraerythrocytic *Plasmodium falciparum*. *Nat. Commun.* **2014**, *5*, No. 5521.

(25) Williamson, A. E.; Ylioja, P. M.; Robertson, M. N.; Antonova-Koch, Y.; Avery, V.; Baell, J. B.; Batchu, H.; Batra, S.; Burrows, J. N.; Bhattacharyya, S.; Calderon, F.; Charman, S. A.; Clark, J.; Crespo, B.; Dean, M.; Debbert, S. L.; Delves, M.; Dennis, A. S. M.; Deroose, F.; Duffy, S.; Fletcher, S.; Giaever, G.; Hallyburton, I.; Gamo, F.-J.; Gebbia, M.; Guy, R. K.; Hungerford, Z.; Kirk, K.; Lafuente-Monasterio, M. J.; Lee, A.; Meister, S.; Nislow, C.; Overington, J. P.; Papadatos, G.; Patiny, L.; Pham, J.; Ralph, S. A.; Ruecker, A.; Ryan, E.; Southan, C.; Srivastava, K.; Swain, C.; Tarnowski, M. J.; Thompson, P.; Turner, P.; Wallace, I. M.; Wells, T. N. C.; White, K.; White, L.; Willis, P.; Winzeler, E. A.; Wittlin, S.; Todd, M. H. Open Source Drug Discovery: Highly Potent Antimalarial Compounds Derived from the Tres Cantos Arylpyrroles. *ACS Cent. Sci.* **2016**, *2*, 687−701.

(26) In Vivo Efficacy. https://github.com/OpenSourceMalaria/Series4/wiki/In-Vivo-Efficacy (accessed Sept 26, 2019).

https://doi.org/10.1021/acs.jmedchem.1c00313
J. Med. Chem. XXXX, XXX, XXX−XXX

(27) Evaluation of Series 4 Compounds vs ATP4-Resistant Mutants. http://malaria.ourexperiment.org/biological_data/11448/post.html (accessed Oct 14, 2019).
(28) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Zpitzer, M.; Zhao, S. Applications of Machine Learning in Drug Discovery and Development. *Nat. Rev. Drug Discovery* **2019**, *18*, 463−477.
(29) Stephen Chan, H. C.; Shan, H.; Dahoun, T.; Vogel, H.; Yuan, S. Advancing Drug Discovery via Artificial Intelligence. *Trends Pharmacol. Sci.* **2019**, *40*, 592−604.
(30) Bentzien, J.; Muegge, I.; Hamner, B.; Thompson, D. C. Crowd Computing: Using Competitive Dynamics to Develop and Refine Highly Predictive Models. *Drug Discovery Today* **2013**, *18*, 472−478.
(31) Pharmacophore Modelling of the Malaria Box PfATP4 Active Compounds. http://malaria.ourexperiment.org/pharmacophore_modelling_/7971/post.html (accessed Sept 26, 2019).
(32) Using the Pharmacophore Model to search Commercial Compounds for new leads. http://malaria.ourexperiment.org/pharmacophore_modelling_/12498/post.html (accessed: Sept 26, 2019).
(33) Maybridge Screening Collection. https://www.maybridge.com/portal/alias__Rainbow/lang__en/tabID__146/DesktopDefault.aspx (accessed Sept 26, 2019).
(34) COMPETITION: A Predictive Model for Series Four. https://github.com/OpenSourceMalaria/OSM_To_Do_List/issues/421 (accessed Sept 26, 2019).
(35) Todd, M. H. Six Laws of Open Source Drug Discovery. *ChemMedChem* **2019**, *14*, 1804−1809.
(36) Ion Regulation Data for OSM Competition. https://docs.google.com/spreadsheets/d/1WWP8fE3X2BLzZ7jOm6bRWpnHZqVJBf8XgIYEb7hshXU/edit?usp=sharing (accessed Sept 26, 2019).
(37) Hars, A.; Shaosong, O. Working for Free? Motivations for Participating in Open-Source Projects. *Int. J. Electron. Commer.* **2002**, *6*, 25−39.
(38) Summary of Competition Entries. https://docs.google.com/spreadsheets/d/1pY6sYXIw66jnzUO3CoP8HceYdDjLRvwg5_pLkBY1Wek/edit?usp=sharing (accessed Sept 26, 2019).
(39) VS Results https://docs.google.com/spreadsheets/d/1uaQ_mSVY6vQSbnDHD5gf232ySH-KVjUg--dQCsilGWI/edit?usp=sharing (accessed Sept 26, 2019).
(40) Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 493−500.
(41) COMPETITION ROUND 2: A Predictive Model for Series 4. https://github.com/OpenSourceMalaria/Series4_PredictiveModel/issues/1 (accessed Sept 26, 2019).
(42) Round 2 Results. https://docs.google.com/spreadsheets/d/1ZPJlMM7znFa056yjOuIycLgzV7-ENt1mMSUJKgH9TxY/edit?usp=sharing (accessed Sept 12, 2019).
(43) Cardoso-Silva, J.; Papageorgiou, L. G.; Tsoka, S. Network-Based Piecewise Linear Regression for QSAR Modelling. *J. Comput. Aided Mol. Des.* **2019**, *33*, 831−844.
(44) Sure, R.; Grimme, S. Corrected Small Basis Set Hartree-Fock Method for Large Systems. *J. Comput. Chem.* **2013**, *34*, 1672−1685.
(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.* **2019**, *59*, 1197−1204.
(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.* **2020**, *60*, 2848−2857.
(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 Discovery* **2020**, *2*, FDD38.
(48) Hunt, P.; Hosseini-Gerami, L.; Chrien, T.; Plante, J.; Ponting, D. J.; Segall, M. Predicting pKa Using a Combination of Semi-Empirical Quantum Mechanics and Radial Basis Function Methods. *J. Chem. Inf. Model.* **2020**, *60*, 2989−2997.
(49) InfoChem IC*SYNTH*. https://www.infochem.de/synthesis/ic-synth (accessed Apr 21, 2020).
(50) Nicolaou, C. A.; Watson, I.; LeMasters, M. A.; Masquelin, T.; Wang, J. Context Aware Data-Driven Retrosynthetic Analysis. *J. Chem. Inf. Model.* **2020**, *60*, 2728−2738.
(51) Tse, E. G.; Houston, S. D.; Williams, C. M.; Savage, G. P.; Rendina, L. M.; Hallyburton, I.; Anderson, M.; Sharma, R.; Walker, G. S.; Obach, R. S.; Todd, M. H. Non-Classical Phenyl Bioisosteres as Effective Replacements in a Series of Novel Open Source Antimalarials. *J. Med. Chem.* **2020**, *63*, 11585−11601.
(52) Data Finders: How can artificial intelligence and deep-learning platforms be used to combat malaria? https://themedicinemaker.com/discovery-development/data-finders (accessed Jun 18, 2020).
(53) Kyba, E. P.; Liu, S. T.; Chockalingam, K.; Reddy, B. R. A General Synthesis of Substituted Fluorenones and Zazfluorenones. *J. Org. Chem.* **1988**, *53*, 3513−3521.
(54) Ohtera, A.; Miyamae, Y.; Yoshida, K.; Maejima, K.; Akita, T.; Kakizuka, A.; Irie, K.; Masuda, S.; Kambe, T.; Nagao, M. Identification of a New Type of Covalent PPARγ Agonist using a Ligand-Linking Strategy. *ACS Chem. Biol.* **2015**, *10*, 2794−2804.
(55) Baragaña, B.; Forte, B.; Choi, R.; Hewitt, S. N.; Bueren-Calabuig, J. A.; Pisco, J. P.; Peet, C.; Dranow, D. M.; Robinson, D. A.; Jansen, C.; Norcross, N. R.; Vinayak, S.; Anderson, M.; Brooks, C. F.; Cooper, C. A.; Damerow, S.; Delves, M.; Dowers, K.; Duffy, J.; Edwards, T. E.; Hallyburton, I.; Horst, B. G.; Hulverson, M. A.; Ferguson, L.; JiménezDíaz, M. B.; Jumani, R. S.; Lorimer, D. D.; Love, M. S.; Maher, S.; Matthews, H.; McNamara, C. W.; Miller, P.; O'Neill, S.; Ojo, K. K.; Osuna-Cabello, M.; Pinto, E.; Post, J.; Riley, J.; Rottmann, M.; Sanz, L. M.; Scullion, P.; Sharma, A.; Shepherd, S. M.; Shishikura, Y.; Simeons, F. R. C.; Stebbins, E. E.; Stojanovski, L.; Straschil, U.; Tamaki, F. K.; Tamjar, J.; Torrie, L. S.; Vantaux, A.; Witkowski, B.; Wittlin, S.; Yogavel, M.; Zuccotto, F.; Angulo-Barturen, I.; Sinden, R.; Baum, J.; Gamo, F.-J.; Mäser, P.; Kyle, D. E.; Winzeler, E. A.; Myler, P. J.; Wyatt, P. G.; Floyd, D.; Matthews, D.; Sharma, A.; Striepen, B.; Huston, C. D.; Gray, D. W.; Fairlamb, A. H.; Pisliakov, A. V.; Walpole, C.; Read, K. D.; Van Voorhis, W. C.; Gilbert, I. H. Lysyl-tRNA Synthetase as a Drug Target in Malaria and Cryptosporidiosis. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 7015−7020.
(56) Saliba, K. J.; Kirk, K. pH Regulation in the Intracellular Malaria Parasite, *Plasmodium falciparum*. *J. Biol. Chem.* **1999**, *274*, 33213−33219.

https://doi.org/10.1021/acs.jmedchem.1c00313
J. Med. Chem. XXXX, XXX, XXX−XXX