High-Throughput Screening and Prediction Model Building for Novel Hemozoin Inhibitors Using Physicochemical Properties

Nguyen Tien Huy,^{a,b,c} Pham Lan Chi,^{b,d} Jun Nagai,^{b,e} Tran Ngoc Dang,^{f,g} Evaristus Chibunna Mbanefo,^{b,c,d} Ali Mahmoud Ahmed,^h Nguyen Phuoc Long,^{c,f} Le Thi Bich Thoa,^{c,f} Le Phi Hung,^{c,f} Afaf Titouna,^{b,d} Kaeko Kamei,ⁱ Hiroshi Ueda,^{b,e} Kenji Hirayama^{b,d}

Department of Clinical Product Development, Institute of Tropical Medicine (NEKKEN), Nagasaki University, Sakamoto, Nagasaki, Japan^a; Graduate School of Biomedical Sciences, Nagasaki University, Bunkyo-machi, Nagasaki, Japan^b; Online Research Club, Nagasaki University, Sakamoto, Nagasaki, Japan^c; Department of Immunogenetics, Institute of Tropical Medicine (NEKKEN), Nagasaki University, Sakamoto, Nagasaki, Japan^d; Department of Pharmacology and Therapeutic Innovation, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan^e; University of Medicine and Pharmacy at Ho Chi Minh City, Ho Chi Minh City, Vietnam^f; Department of Health Care Policy and Management, Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Japan^g; Faculty of Medicine, Al-Azhar University, Cairo, Egypt^h; Department of Biomolecular Engineering, Kyoto Institute of Technology, Sakyo-ku, Kyoto, Japanⁱ

ABSTRACT It is essential to continue the search for novel antimalarial drugs due to the current spread of resistance against artemisinin by *Plasmodium falciparum* parasites. In this study, we developed *in silico* models to predict hemozoin inhibitors as a potential first-step screening for novel antimalarials. An *in vitro* colorimetric high-throughput screening assay of hemozoin formation was used to identify hemozoin inhibitors from 9,600 structurally diverse compounds. The physicochemical properties of positive hits and randomly selected compounds were extracted from the ChemSpider database; they were used for developing prediction models to predict hemozoin inhibitors using two different approaches, i.e., traditional multivariate logistic regression and Bayesian model averaging. Our results showed that a total of 224 positive-hit compounds exhibited the ability to inhibit hemozoin formation, with 50% inhibitory concentrations (IC₅₀s) ranging from 3.1 μ M to 199.5 μ M. The best model according to traditional multivariate logistic regression included the three variables octanol-water partition coefficient, number of hydrogen bond donors, and number of atoms of hydrogen, while the best model according to Bayesian model averaging included the three variables octanol-water partition coefficient, number of hydrogen bond donors, and index of refraction. Both models had a good discriminatory power, with area under the curve values of 0.736 and 0.781 for the traditional multivariate model and Bayesian model averaging, respectively. In conclusion, the prediction models can be a new, useful, and cost-effective approach for the first screen of hemozoin inhibition-based antimalarial drug discovery.

KEYWORDS HTS, antimalarial, compounds, hemozoin, *in silico* model, physical properties, screening

Hemozoin is a crystalline pigment product that is synthesized by hemoparasites, including *Plasmodium* species, from the hemoglobin degradation process (1). Hemozoin formation is an adaptation of the parasite to be protected against toxic heme (2), which is released as a by-product of hemoglobin degradation in the *Plasmodium* food vacuole. Within the infected red blood cells, the parasites digest hemo-

Received 25 July 2016 Returned for modification 28 August 2016 Accepted 14 October 2016

Accepted manuscript posted online 5 December 2016

Citation Huy NT, Chi PL, Nagai J, Dang TN, Mbanefo EC, Ahmed AM, Long NP, Thoa LTB, Hung LP, Titouna A, Kamei K, Ueda H, Hirayama K. 2017. High-throughput screening and prediction model building for novel hemozoin inhibitors using physicochemical properties. *Antimicrob Agents Chemother* 61:e01607-16. <https://doi.org/10.1128/AAC.01607-16>.

Copyright © 2017 American Society for Microbiology. All Rights Reserved.

Address correspondence to Nguyen Tien Huy, tienhuy@nagasaki-u.ac.jp, or Kenji Hirayama, hiraken@nagasaki-u.ac.jp.

N.T.H. and P.L.C. contributed equally to this work.

globin as a main source of amino acids for their growth and development (3). Due to the toxic effect of the released heme (4), it is imperative for *Plasmodium* to evolve effective heme homeostasis mechanisms, one of which is hemozoin formation (5).

The rapid spread of resistance to artemisinin-based combination therapies among *Plasmodium falciparum* parasites has been identified as a major global challenge in the fight against malaria (6, 7). Although the development of an effective malaria vaccine is the most effective control measure, there is still no vaccine available for preventing this disease (8). To date, only one malaria vaccine candidate has reached phase III clinical trials (9). It is essential to continue the search for novel antimalarial drugs, especially for countries where malaria is endemic. An ideal target is the blocking of the heme detoxification pathway of the parasite (10–13). Indeed, this mechanism is also one of the main targets of current antimalarial drugs like quinine and has been the major target of several antimalarial screening projects. Unlike chloroquine, to which resistance resulting from mutation of the membrane transport protein that effluxes chloroquine out of the food vacuole is widespread (1), quinine still has strong antimalarial activity against chloroquine-resistant strains, although reduced efficacy has been noticed recently (14). This makes hemozoin inhibition a good target for novel antimalarial drug development.

Hemozoin formation is a physicochemical process that occurs in the presence of parasite proteins (15–18) and/or lipids (19, 20). Recently, commercial lipophilic detergents, including Tween 20 and Nonidet P-40 (NP-40), have been identified as surrogate substances to promote the crystallization of heme under relevant conditions (21, 22). This artificial system is amenable for use in high-throughput hemozoin inhibition assays for screening novel antimalarials (21). However, it is still time-consuming and requires expensive and specialized instruments and laborious preparation. Therefore, the execution of *in silico* models or other machine-learning models, such as Bayesian modeling, is ideal for screening millions of chemical compounds to prioritize them for high-throughput screening (HTS), leading to valuable hit rates with fewer test compounds. Recently, Wicht et al. showed that Bayesian models can be effective tools to predict hemozoin inhibitor compounds, with high enrichment rates in comparison to those of conventional random screening (23). Making *in silico* models is not only valuable for future HTS, it is also a good way to drive benefit from all available data, even data for inactive compounds, from preceding screens. In this study, we developed a model to predict hemozoin inhibitors using the physicochemical properties of chemical compounds.

RESULTS

High-throughput screening using the heme crystallization assay. Pyridine molecules formed coordinate bonds to free iron from noncrystallized heme molecules and produced a pyridine-heme complex with strong absorption at 405 nm (24). The robustness and reproducibility of the assay were improved by optimizing the concentrations and volumes of compounds, hemin, and detergent solutions. As a result, the Z factors of all plates were higher than 0.5, which is an essential minimum value for validation of HTS assays. In other words, a high degree of reproducibility and a large dynamic range were achieved for the assay (24).

A total of 9,600 diversely selected compounds (the core library), assigned randomly from more than 200,000 compounds in the chemical library of The Drug Discovery Initiative, Tokyo University (<http://www.ddi.u-tokyo.ac.jp/en/#5>), were used in the HTS assay. Active compounds were identified as compounds with absorbance three standard deviations above that of the dimethyl sulfoxide (DMSO) negative control. The absorbance values from 384-well plates were described in heat maps (see Fig. S1 at https://www.researchgate.net/publication/309208397_Supplemental_material_Hig). Evident red color in the heat maps represents correlative compounds, which were likely to strongly inhibit the crystallization of free heme. In total, 394 active compounds (4.1% of the 9,600 screened compounds) were identified by the high-throughput screening assay.

FIG 1 Scheme for building a prediction model of antihemozoin compounds. First, positive hits were identified from 9,600 structurally diverse compounds by high-throughput screening (HTS) and a dose-response assay of hemozoin formation. Second, the physical properties of 224 true-positive hits and 199 random negative compounds were extracted using the ChemSpider software. Third, prediction models were built by the traditional approach versus the Bayesian approach using these physical properties.

Dose-response assay for positive compounds. The 394 active compounds resulting from the HTS assay were subsequently tested in dose-response assays to exclude false positives and to determine the half-maximal (50%) inhibitory concentrations (IC_{50} s). A positive hit was identified as an active compound exhibiting a typical sigmoid dose-response curve (see Fig. S2 at https://www.researchgate.net/publication/309208397_Supplemental_material_Hig). False positives for which the absorbance values could not generate typical sigmoid dose-response curves were likely due to the compounds' colors or aggregation.

Finally, 224 of the 394 active compounds were shown to be positive hits (Fig. 1). Therefore, among 9,600 chemical compounds tested, the combination of high-throughput screening and dose-response assays identified 224 positive hits (resulting in a hit rate of 2.34%). The positive hits exhibited IC_{50} s from 3.1 μ M to 199.5 μ M; 9 of the 224 positive hits exhibited IC_{50} s of less than 5 μ M.

Development of prediction models. The physical and chemical properties were extracted for all 224 positive hits, as well as for 199 negative compounds, i.e., compounds without antihemozoin activity, which were randomly selected from the original 9,600 compounds of the core library (Fig. 2). Then, 70% of the data ($n = 285$, complete case analysis) was used to develop the best models using different approaches: a traditional method (i.e., univariate logistic regression followed by multivariate regression) and the Bayesian model averaging (BMA) method.

FIG 2 Results from antihemozoin HTS of 9,600 diverse compounds. Dots represents % hemozoin inhibition of 9,600 compounds, including 224 true-positive hits (closed dots above cutoff line), 170 false-positive compounds (open dots above cutoff line), and negatives (dots under cutoff line). The cutoff value was determined as the average absorbance value for the negative DMSO control plus 3 standard deviations. The true-positive hits were identified by the subsequent dose-response assay.

In the traditional approach, the octanol-water partition coefficient (Log P), KOC (at pH 5.5 and pH 7.4), distribution coefficient (Log D) (at pH 5.5 and pH 7.4), index of refraction, molar refractivity, number of hydrogen bond donors, number of freely rotating bond donors, number of rule-of-five violations, density, surface tension, and number of atoms of hydrogen, oxygen, and nitrogen yielded *P* values of <0.1 by univariate logistic regression analysis. The multivariate logistic regression of these properties showed that the ability of positive hits to inhibit hemozoin formation was significantly correlated with Log P, number of hydrogen bond donors, and number of atoms of hydrogen, with *P* values of <0.05 (Table 1). The equation for the best multivariate model is as follows: $\text{Logit}(\text{probability}) = -0.739 + 0.671 \times \text{Log P} + 0.484 \times N_1 - 0.099 \times N_2$, where “probability” represents the probability of antihemozoin activity of a particular compound, while N_1 and N_2 stand for the number of hydrogen bond donors and the number of atoms of hydrogen, respectively.

In the BMA approach, first, all variables that yielded *P* values of <0.1 by univariate logistic regression were submitted to BMA. Later, the BMA process suggested the five most parsimonious models on the basis of Bayesian information criterion (BIC) values (Table 2). Among them, the best model included the following variables, which resulted in the smallest BIC values: Log P, number of hydrogen bond donors, and index of refraction. The equation for the best BMA is as follows: $\text{Logit}(\text{probability}) = -23.62 + 0.592 \times \text{Log P} + 0.351 \times N_1 + 1.322 \times N_3$, where probability and N_1 are as described for the multivariate model equation and N_3 stands for the index of refraction.

Validation of prediction models. After we successfully developed the two best models (i.e., the multivariate model and BMA), we validated them using the remaining 30% of the data ($n = 121$). Figure 3 shows the areas under the curve (AUCs) (left) and the accuracies (right) of these best models, which indicate that the discriminatory power of BMA is better than that of the multivariate model. The AUCs, however, were 0.736 and 0.781 for the multivariate model and BMA, respectively, which implies that both models have good discriminatory power.

The multivariate model with a cutoff of 0.536 resulted in optimal sensitivity and specificity and maximum accuracy of 65.6%, 77.2%, and 71.1%, respectively, while BMA with a cutoff of 0.465 resulted in optimal sensitivity and specificity and maximum accuracy of 79.7%, 66.7%, and 73.5%, respectively.

DISCUSSION

The inhibition of hemozoin formation, proposed as the major mechanism of current antimalarials, such as quinine and chloroquine (25), was the foundation of the research

TABLE 1 Univariate and multivariate analyses of positive-hit versus negative-hit compounds

Predictor ^a	Univariate analysis ^b		Multivariate analysis ^c	
	OR (95% CI)	P value	Adjusted OR (95% CI)	P value
Log P	1.54 (1.29–1.83)	<0.0001	2.04 (1.27–3.27)	<u>0.0028</u>
BCF, pH 5.5	1.00 (1.00–1.00)	0.2181	—	—
BCF, pH 7.4	1.00 (1.00–1.00)	0.185	—	—
KOC, pH 5.5	1.00 (1.00–1.00)	0.052	1.00 (1.00–1.00)	0.9400
KOC, pH 7.4	1.00 (1.00–1.00)	0.041	1.00 (1.00–1.00)	0.4344
Log D, pH 5.5	1.34 (1.20–1.50)	<0.0001	1.05 (0.69–1.62)	0.7939
Log D, pH 7.4	1.34 (1.18–1.52)	<0.0001	0.88 (0.60–1.30)	0.5458
Avg mass (Da)	1.00 (1.00–1.00)	0.251	—	—
Density (g/cm ³)	17.10 (3.59–82.10)	0.0004	0.35 (0.01–18.5)	0.6088
Index of refraction ^d	3.72 (2.38–5.81)	<0.0001	1.55 (0.64–3.76)	0.3304
Molar refractivity (cm ³)	1.01 (1.00–1.02)	0.0686	1.0 (0.99–1.06)	0.0976
Molar vol (cm ³)	1.00 (0.99–1.00)	0.4041	—	—
Monoisotopic mass (Da)	1.00 (1.00–1.00)	0.3383	—	—
No freely rotating bonds	0.92 (0.82–1.01)	0.0933	0.95 (0.79–1.16)	0.6712
No H bond acceptors	0.98 (0.86–1.11)	0.7777	—	—
No H bond donors	1.40 (1.13–1.73)	0.0017	1.38 (1.03–1.87)	<u>0.0308</u>
No. of rule-of-5 violations	4.00 (1.45–10.98)	0.0071	3.42 (0.72–16.31)	0.1218
No. of Br	1.79e + 7 (0–Inf)	0.9855	—	—
No. of C	1.02 (0.96–1.07)	0.4763	—	—
No. of Cl	1.37 (0.79–2.37)	0.2517	—	—
No. of F	0.70 (0.46–1.06)	0.1000	—	—
No. of H	0.94 (0.91–0.98)	0.0105	0.87 (0.77–0.98)	<u>0.0293</u>
No. of N	1.25 (1.06–1.46)	0.0066	1.09 (0.82–1.46)	0.5150
No. of O	0.81 (0.70–0.93)	0.0049	0.81 (0.59–1.10)	0.1795
No. of S	1.13 (0.82–1.57)	0.4311	—	—
Polar surface area (Å ²)	1.00 (0.99–1.01)	0.2108	—	—
Polarizability (×10 ²⁴ cm ³)	1.02 (0.99–1.05)	0.1016	—	—
Surface tension (dyne/cm)	1.04 (1.02–1.06)	<0.0001	1.01 (0.96–1.06)	0.4724

^aLog P, octanol-water partition coefficient; Log D, distribution coefficient; BCF, bioconcentration factor; KOC, adsorption coefficient.

^bOR, odds ratio; 95% CI, 95% confidence interval.

^cSignificant P values (<0.05) from multivariate analysis are underlined. —, variable was not included in multivariate analysis because it had a P value of ≥0.1 in univariate analysis.

^dThe original scale has been multiplied by 10.

on novel antimalarials via high-throughput screening assay. A total of 224 positive hits out of 9,600 library compounds exhibited the ability to inhibit the hemozoin formation, with IC₅₀s ranging from 3.1 μM to 199.5 μM. Analysis of the physical and chemical properties of these positive hits showed positive correlation between Log P, index of refraction, number of hydrogen bond donors, and capability to inhibit the hemozoin formation (Table 2).

The 2.34% hit rate fulfilled in our study is considerably higher than the 0.42% in previous research by Sandlin et al. (21). The compound concentration used in our assays was 220 μM, which is higher than the concentrations used in the assays of Sandlin et al. Besides this difference, the 9,600 compounds used in this study, which were assigned as a core chemical library from more than 200,000 compounds with various structures, could be completely different from the 38,400 compounds used by Sandlin et al. or the 5,000 compounds used by Wicht et al. (23). Consequently, the difference in hit rates is likely due to the difference in the concentrations at which the compounds were tested. The main advantage of our study over that done by Wicht et al. (23) is that our study used two models, multivariate logistic regression and BMA, rather than BMA alone.

The octanol-water partition coefficient (Log P) of a compound expresses the tendency of a compound to partition between lipophilic phase and aqueous phase, known as lipophilicity (26). The capability of compounds for inhibiting the formation of hemozoin is probably related to their lipophilicity, as there is very strong evidence supporting the theory that hemozoin formation is lipid mediated (27–29). On examining the trophozoite stage of red blood cells (RBCs) infected with *Plasmodium falciparum*

TABLE 2 The five most parsimonious models selected by the Bayesian model averaging approach^a

Model	Explanatory variable	Coefficient	P value	BIC ^b	Posterior probability ^c
1	Log P	0.5924	3.05e−09	−1,278	0.318
	No H bond donors	0.3514	0.00495		
	Index of refraction ^d	1.322	0.00215		
	Intercept = −23.6		0.00192		
2	Log P	0.5328	1.26e−10	−1,276	0.118
	Index of refraction ^d	1.487	1.79e−05		
	Intercept = −25.5		0.00098		
3	No. of H	−0.0519	1.18e−09	−1,276	0.117
	Log P	0.6506	5.37e−05		
	No H bond donors	0.3600	0.00104		
	Index of refraction ^d	1.149	0.01276		
	Intercept = −20.05				
4	No. of O	−0.1374	9.2e−09	−1,275	0.066
	Log P	0.5774	0.008566		
	No H bond donors	0.3778	0.001341		
	Index of refraction ^d	1.247	0.000925		
	Intercept = −22.0				
5	Log P	0.6267	2.83e−10	−1,275	0.054
	No H bond donors	0.3862	1.16e−05		
	No freely rotating bonds	−0.0946	0.000514		
	Index of refraction ^d	1.206			
	Intercept = −21.43				

^aThrough the BMA process, 18 models were selected, and the 5 best models are presented. The cumulative posterior probability is equal to 0.6727.

^bBIC, Bayesian information criteria. The smallest BIC suggests the model with maximum parsimony (i.e., minimum explanatory variables and maximum discrimination power).

^cPosterior probability is the probability of a model being a correct model in the BMA process.

^dThe original scale has been multiplied by 10.

by electron microscopy, Pisciotta et al. found nanosphere lipid droplets containing hemozoin crystals (30). The crystallization of β -hemozoin usually occurs in a hydrophobic environment, which is preferred for the formation of the hydrogen bonds that create the linkage in the hydrophilic ferriprotoporphyrin IX's [Fe^{III}PPIX] propionate (31, 32). All

FIG 3 Discriminatory-power comparison of the best multivariate model and the best BMA on the basis of AUC (left) and accuracy (right). The best multivariate model consists of the three variables Log P value, number of hydrogen bond donors, and number of atoms of hydrogen, while the best BMA consists of the three variables Log P value, number of hydrogen bond donors, and index of refraction. The discriminatory power of BMA is better than that of the multivariate model in terms of AUC and accuracy.

these causes make the lipophilicity an important property of a compound that enables it to inhibit β -hematin or hemozoin formation. On the cellular level, lipophilic compounds can permeate through the lipid bilayer membrane of the food vacuole of *P. falciparum* and, therefore, can easily reach hemozoin crystals.

The index of refraction, also known as the refractive index, is an optical property defined as the difference between the velocities of light in vacuum and in a medium in which it propagates. In the Lorentz-Lorenz equation, the refractive index is estimated using molar refraction, which is the sum of the contributions of corresponding atoms and bonds (33). Hence, the index of refraction related to the polarizability, purity, and density of an organic compound is applied to evaluate the characteristics of the material (34). Moreover, the presence and quantity of some heavy atoms and functional groups with high refractive indices, such as sulfur (35), halogen elements (especially bromine and iodine) (36), and phosphorus (37), play an important role in increasing the molar refraction. Therefore, the presence of heavy atoms increases the index of refraction. However, our models could not detect an association of an antihemozoin effect with any specific heavy atom, probably due to the small sample size of each atom. Nevertheless, we were not able to find any research on the relationship between the refractive indices of compounds and their antihemozoin effects. Therefore, compounds with high refractive indices could be an interesting topic for antimalarial studies in the future.

The number of hydrogen bond donors plays an important role in β -hematin or hemozoin crystal inhibition through the intramolecular formation of hydrogen bonds between neighboring complexes in the crystal (38). For instance, hydrogen bond donors of known antimalarial drugs, such as halofantrine and quinoline, form a hydrogen bond bridge with β -hematin in parasite food vacuoles (38, 39). Thus, these antimalarials are likely to inhibit hemozoin formation via the hydrogen bond conformation (39). However, the number of hydrogen bond donors of compound candidates, which is positively related to the antihemozoin capability of compounds, should be considered to satisfy Lipinski's rule of five for an orally bioavailable drug development (40).

In this study, we proposed two different approaches in the development of prediction models for calculating the probability of inhibition of hemozoin formation by each compound. The AUCs of both the multivariate model and BMA indicate that both models can be applied in a real setting (41). Interestingly, when tested against five well-known antimalarial drugs, including chloroquine, quinine, amodiaquine, halofantrine, and artemisinin, BMA accurately predicted all four well-known antihemozoin drugs (with probabilities of 0.63 for chloroquine, 0.60 for quinine, 0.88 for amodiaquine, and 0.94 for halofantrine) and one non-hemozoin-inhibiting drug (artemisinin, with a probability of 0.12), while the multivariate model correctly predicted four drugs, chloroquine (probability of 0.58), amodiaquine (probability of 0.78), halofantrine (0.93) (probability of >0.536), and artemisinin (probability of 0.20), but wrongly predicted quinine (probability of 0.42) as a nonhemozoin inhibitor (probability of <0.536).

The prediction models also have some advantages for antimalarial design. First, while other approaches, such as the development of analogs of existing agents or natural products, mainly detect new antimalarials by chemical modifications of known compounds (42), new antimalarial compounds can be discovered by the prediction equation based on the well-known metabolic target. Thus, the models help researchers to find good chemical groups for synthetic compounds. Second, expensive equipment and specialized laboratory wares are not essential in these prediction models. Therefore, millions of library compounds can be screened *in silico* by using the models. Third, the relationship between the properties of compounds and antihemozoin activity is also interpreted from the models. It can be the first clue for understanding the mechanism of action of antimalarials. In addition, we proved that BMA is likely a good approach in the development of prediction models because it considers the "uncertainty" in model selections. Hence, the habit of building the only right model in the

traditional approach should be compared with other approaches that consider uncertainty in model selection (e.g., BMA) in studies similar to this one.

Despite the benefits of the high-throughput screening assay mentioned above, the study has several limitations. First, in the antihemozoin assay, we did not remove all of the soluble contents before dissolving the noncrystallized heme by adding pyridine solution (25). The lack of this step probably resulted in the false positives due to the compounds' color and/or aggregation. These were eliminated in the second step using the dose-response assay. Second, the interaction between the positive-hit compounds and intraparasitic conditions was not fully evaluated in this paper, although the assay was performed under conditions that closely mimic the physiological environment in the parasite food vacuole. It is known that only a small fraction of hemozoin inhibitors possess antimalarial activity *in vitro*. Our ongoing experiments revealed a total of 23 positive-hit compounds and two negative-hit compounds that exhibited antimalarial activity with IC_{50} s of less than 10 μ M. Among them, four positive-hit compounds showed IC_{50} s of less than 1 μ M. However, using the antihemozoin assay for HTS, we could lower the *in vitro* antimalarial assay workload by approximately 40 times. The last limitation is that the prediction models have not yet been validated in external samples.

In conclusion, the *in vitro* high-throughput hemozoin formation assay was performed with a high degree of reproducibility and robustness. A total of 224 true-positive hits were identified from the core library with a hit rate of 2.34%. The prediction models based on physicochemical parameters represent a new, useful, and cost-effective approach for antimalarial drug discovery in developing countries. Moreover, the physicochemical properties, namely, Log P, the index of refraction, and the number of hydrogen bond donors, should be investigated further in order to identify their effects on the antihemozoin activity of compounds.

MATERIALS AND METHODS

Materials. Hemin chloride (heme) and quinine were purchased from Sigma. The core library, which contains 9,600 structurally diverse compounds, was from the Open Innovation Center for Drug Discovery, University of Tokyo (Tokyo, Japan). The detergent NP-40 served as a mediator for β -hematin formation due to its stability, low cost, and low IC_{50} , beside its similarity to the natural lipid particles of the parasite's vacuole (21, 43). Dimethyl sulfoxide (DMSO), from Wako Pure Chemicals, Osaka, Japan, was chosen as the negative control because of its proven inability to inhibit the heme crystallization reaction (44).

HTS using an antihemozoin assay. High-throughput screening using the antihemozoin assay was performed in 384-well plates, using quinine and DMSO as the positive and negative controls, respectively (21, 24). Quinine and candidate compounds were dissolved in DMSO to achieve a final concentration of 220 μ M. Using an automated dispenser (Multi-Dispenser EDR 384, BioTec, Japan), 5 μ l of each compound was transferred to each well of the assay plates (see Fig. S1 at https://www.researchgate.net/publication/309208397_Supplemental_material_Hig). Following the transfer of compounds, a Multidrop combi dispenser (Thermo Fisher Scientific) was used to distribute 20 μ l of hemin solution (10 mM heme in DMSO and 100 mM acetate buffer, pH 4.8) and 10 μ l of detergent NP-40 into each well of the plates. The assay mixture was incubated at 37°C for 250 min (43). Afterwards, pyridine solution was added and the mixture shaken for 10 min. To dissolve bubbles, 10 μ l of acetone was added to each well, and the plate was finally transferred into a multiplate reader to detect noncrystallized heme at 405/705 nm using the colorimetric method (24, 45).

Antihemozoin dose-response assay. Active compounds identified by the previously described high-throughput screening using the antihemozoin assay were tested in dose-response assays. Quinine was also used as a positive control in each assay plate. Each compound was retested with the hemozoin inhibition assay in 384-well plates at concentrations ranging from 0 μ M to 208 μ M. The absorbance values of each compound measured at 405/750 nm were dependent on the difference in the concentration of the compound. Data were analyzed to determine the half-maximal inhibitory concentration (IC_{50}) for each compound, relying on a sigmoid dose-response curve fitted by GraphPad Prism software, version 5.00 (45).

Physicochemical properties of positive hits and a representative sample of negative compounds. The average mass, octanol-water partition coefficient (Log P), distribution coefficient (Log D), bioconcentration factor (BCF), adsorption coefficient (KOC), number of rule-of-five violations, number of hydrogen bond acceptors, number of hydrogen bond donors, freely rotating bonds, polar surface area, index of refraction, molar refractivity, molar volume, polarizability, flash point, boiling point, enthalpy of vaporization, and number of atoms of chemical elements (such as bromine, carbon, chlorine, fluorine, hydrogen, nitrogen, oxygen, and sulfur) of each of the positive hits and a sample of negative compounds were retrieved from ChemSpider (www.chemspider.com), as predicted by Advanced Chemistry Development (ACD/Laboratories) software (46).

Statistical analysis. (i) Missing data analysis. We used complete-case analysis, which eliminates compounds/cases with missing data (i.e., physicochemical properties of a compound) so only complete

compounds/cases are left. The rates of missing data varied from property to another. Therefore, the incomplete cases ranged from 0.2% to 3.5% of the compounds due to lack of one of the physicochemical properties.

(ii) Univariate and multivariate logistic regression. The outcome variable was the ability to inhibit the hemozoin formation of a compound, determined using two values: 1 if the compound could inhibit hemozoin formation (exhibiting a typical sigmoid dose-response curve with an IC_{50} of $<200 \mu M$) and 0 if the compound could not. The predictor variables were the physicochemical properties of a compound.

First, we performed univariate logistic regression to examine the association between physicochemical properties and the ability to inhibit hemozoin formation. Second, variables with P values of <0.1 were submitted to multivariate analysis to find the independent predictors of inhibition of hemozoin formation. Significance was set at a P value of <0.05 in the multivariate regression analysis.

(iii) Development and validation of the prediction models. The development and validation of the prediction models consisted of the following steps: (1) The original data were randomly divided into training and testing sets at a ratio of 70:30, respectively (2). The training data set was constructed to develop prediction models using two approaches: the traditional approach, in which univariate logistic regression was followed by multivariate regression as described above, and the Bayesian model averaging (BMA) approach to select the best prediction models (3). The discriminatory powers of the best prediction models obtained from different approaches were compared on the basis of the area under the curve (AUC) from the receiver operating characteristic (ROC) curve and accuracy (47).

Basically, the purpose of BMA is to search for the most parsimonious model (i.e., a model with the minimum number of explanatory variables and the maximum discriminatory power) (48). In brief, 2^k possible models (not including interaction models) can be constructed if there are k explanatory variables. Among the 2^k models, the best models are suggested based on the Bayesian information criterion (BIC), in which a smaller BIC value indicates a better model. Therefore, unlike the traditional approach mentioned above, BMA considers the uncertainty in the model selection process. Recently, BMA has been receiving more attention in prognostic-model studies (49–51).

All analyses were performed using R software version 3.2.2 (The R Foundation for Statistical Computing).

ACKNOWLEDGMENTS

This work was partly supported by Platform for Drug Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

The funder had no role in study design, data collection and analysis, preparation of the manuscript, or decision to publish.

REFERENCES

- Hänscheid T, Egan TJ, Grobusch MP. 2007. Haemozoin: from melatonin pigment to drug target, diagnostic tool, and immune modulator. *Lancet Infect Dis* 7:675–685. [https://doi.org/10.1016/S1473-3099\(07\)70238-4](https://doi.org/10.1016/S1473-3099(07)70238-4).
- Ihekwereme CP, Esimone CO, Nwanegbo EC. 2014. Hemozoin inhibition and control of clinical malaria. *Adv Pharmacol Sci* 2014:984150. <https://doi.org/10.1155/2014/984150>.
- Zhang J, Krugliak M, Ginsburg H. 1999. The fate of ferriprotoporphyrin IX in malaria infected erythrocytes in conjunction with the mode of action of antimalarial drugs. *Mol Biochem Parasitol* 99:129–141. [https://doi.org/10.1016/S0166-6851\(99\)00008-0](https://doi.org/10.1016/S0166-6851(99)00008-0).
- Fitch CD, Chevli R, Kanjanangkulpan P, Dutta P, Chevli K, Chou AC. 1983. Intracellular ferriprotoporphyrin IX is a lytic agent. *Blood* 62:1165–1168.
- Moore LR, Fujioka H, Williams PS, Chalmers JJ, Grimberg B, Zimmerman PA, Zborowski M. 2006. Hemoglobin degradation in malaria-infected erythrocytes determined from live cell magnetophoresis. *FASEB J* 20:747–749.
- Dondorp AM, Yeung S, White L, Nguon C, Day NP, Socheat D, von Seidlein L. 2010. Artemisinin resistance: current status and scenarios for containment. *Nat Rev Microbiol* 8:272–280. <https://doi.org/10.1038/nrmicro2331>.
- Dondorp AM, Ringwald P. 2013. Artemisinin resistance is a clear and present danger. *Trends Parasitol* 29:359–360. <https://doi.org/10.1016/j.pt.2013.05.005>.
- Targett GA, Moorthy VS, Brown GV. 2013. Malaria vaccine research and development: the role of the WHO MALVAC Committee. *Malar J* 12:362. <https://doi.org/10.1186/1475-2875-12-362>.
- Agnandji ST, Lell B, Fernandes JF, Abossolo BP, Methogo BG, Kabwende AL, Adegnika AA, Mordmüller B, Issifou S, Kremsner PG. 2012. A phase 3 trial of RTS, S/AS01 malaria vaccine in African infants. *N Engl J Med* 367:2284–2295. <https://doi.org/10.1056/NEJMoa1208394>.
- Egan TJ, Hunter R, Kaschula CH, Marques HM, Mispilon A, Walden J. 2000. Structure-function relationships in aminoquinolines: effect of amino and chloro groups on quinoline-hematin complex formation, inhibition of beta-hematin formation, and antiparasmodial activity. *J Med Chem* 43:283–291. <https://doi.org/10.1021/jm990437l>.
- Adams PA, Berman PA, Egan TJ, Marsh PJ, Silver J. 1996. The iron environment in heme and heme-antimalarial complexes of pharmacological interest. *J Inorg Biochem* 63:69–77. [https://doi.org/10.1016/0162-0134\(95\)00212-X](https://doi.org/10.1016/0162-0134(95)00212-X).
- Sullivan DJ. 2002. Theories on malarial pigment formation and quinoline action. *Int J Parasitol* 32:1645–1653. [https://doi.org/10.1016/S0020-7519\(02\)00193-5](https://doi.org/10.1016/S0020-7519(02)00193-5).
- Ridley RG, Dorn A, Vippagunta SR, Vennerstrom JL. 1997. Haematin (haem) polymerization and its inhibition by quinoline antimalarials. *Ann Trop Med Parasitol* 91:559–566. <https://doi.org/10.1080/00034983.1997.11813174>.
- Dinio T, Gorka AP, McGinniss A, Roepe PD, Morgan JB. 2012. Investigating the activity of quinine analogues versus chloroquine resistant *Plasmodium falciparum*. *Bioorg Med Chem* 20:3292–3297. <https://doi.org/10.1016/j.bmc.2012.03.042>.
- Sullivan DJ, Jr, Gluzman IY, Goldberg DE. 1996. *Plasmodium* hemozoin formation mediated by histidine-rich proteins. *Science* 271:219–222. <https://doi.org/10.1126/science.271.5246.219>.
- Huy NT, Serada S, Trang DT, Takano R, Kondo Y, Kanaori K, Tajima K, Hara S, Kamei K. 2003. Neutralization of toxic heme by *Plasmodium falciparum* histidine-rich protein 2. *J Biochem* 133:693–698. <https://doi.org/10.1093/jb/mvg089>.
- Jani D, Nagarkatti R, Beatty W, Angel R, Slebodnick C, Andersen J, Kumar S, Rathore D. 2008. HDP—a novel heme detoxification protein from the malaria parasite. *PLoS Pathog* 4:e1000053. <https://doi.org/10.1371/journal.ppat.1000053>.
- Nakatani K, Ishikawa H, Aono S, Mizutani Y. 2014. Identification of essential histidine residues involved in heme binding and hemozoin

- formation in heme detoxification protein from *Plasmodium falciparum*. *Sci Rep* 4:6137. <https://doi.org/10.1038/srep06137>.
19. Dorn A, Stoffel R, Matile H, Bubendorf A, Ridley RG. 1995. Malarial haemozoin/beta-haematin supports haem polymerization in the absence of protein. *Nature* 374:269–271. <https://doi.org/10.1038/374269a0>.
 20. Tripathi AK, Gupta A, Garg SK, Tekwani BL. 2001. In vitro beta-hematin formation assays with plasma of mice infected with *Plasmodium yoelii* and other parasite preparations: comparative inhibition with quinoline and endoperoxide antimalarials. *Life Sci* 69:2725–2733. [https://doi.org/10.1016/S0024-3205\(01\)01349-2](https://doi.org/10.1016/S0024-3205(01)01349-2).
 21. Sandlin RD, Carter MD, Lee PJ, Auschwitz JM, Leed SE, Johnson JD, Wright DW. 2011. Use of the NP-40 detergent-mediated assay in discovery of inhibitors of β -hematin crystallization. *Antimicrob Agents Chemother* 55:3363–3369. <https://doi.org/10.1128/AAC.00121-11>.
 22. Huy NT, Uyen DT, Maeda A, Oida T, Harada S, Kamei K. 2007. Simple colorimetric inhibition assay of heme crystallization for high-throughput screening of antimalarial compounds. *Antimicrob Agents Chemother* 51:350–353. <https://doi.org/10.1128/AAC.00985-06>.
 23. Wicht KJ, Combrinck JM, Smith PJ, Egan TJ. 2015. Bayesian models trained with HTS data for predicting β -haematin inhibition and in vitro antimalarial activity. *Bioorg Med Chem* 23:5210–5217. <https://doi.org/10.1016/j.bmc.2014.12.020>.
 24. Rush MA, Baniecki ML, Mazitschek R, Cortese JF, Wiegand R, Clardy J, Wirth DF. 2009. Colorimetric high-throughput screen for detection of heme crystallization inhibitors. *Antimicrob Agents Chemother* 53:2564–2568. <https://doi.org/10.1128/AAC.01466-08>.
 25. Tekwani BL, Walker LA. 2005. Targeting the hemozoin synthesis pathway for new antimalarial drug discovery: technologies for in vitro beta-hematin formation assay. *Comb Chem High Throughput Screen* 8:63–79. <https://doi.org/10.2174/1386207053328101>.
 26. Tetko IV, Tanchuk VY, Villa AE. 2001. Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices. *J Chem Inf Comput Sci* 41:1407–1421. <https://doi.org/10.1021/ci010368v>.
 27. Bendrat K, Berger BJ, Cerami A. 1995. Haem polymerization in malaria. *Nature* 378:138. <https://doi.org/10.1038/378138a0>.
 28. Dorn A, Vippagunta SR, Matile H, Bubendorf A, Vennerstrom JL, Ridley RG. 1998. A comparison and analysis of several ways to promote haematin (haem) polymerisation and an assessment of its initiation in vitro. *Biochem Pharmacol* 55:737–747. [https://doi.org/10.1016/S0006-2952\(97\)00509-1](https://doi.org/10.1016/S0006-2952(97)00509-1).
 29. Fitch CD, Cai G-z, Chen Y-F, Shoemaker JD. 1999. Involvement of lipids in ferriprotoporphyrin IX polymerization in malaria. *Biochim Biophys Acta* 1454:31–37. [https://doi.org/10.1016/S0925-4439\(99\)00017-4](https://doi.org/10.1016/S0925-4439(99)00017-4).
 30. Pisciotto JM, Coppens I, Tripathi AK, Scholl PF, Shuman J, Bajad S, Shulaev V, Sullivan DJ. 2007. The role of neutral lipid nanospheres in *Plasmodium falciparum* haem crystallization. *Biochem J* 402:197–204. <https://doi.org/10.1042/BJ20060986>.
 31. Egan TJ, Chen JYJ, de Villiers KA, Mabotha TE, Naidoo KJ, Ncokazi KK, Langford SJ, McNaughton D, Pandiancherri S, Wood BR. 2006. Haemozoin (β -haematin) biomineralization occurs by self-assembly near the lipid/water interface. *FEBS Lett* 580:5105–5110. <https://doi.org/10.1016/j.febslet.2006.08.043>.
 32. Huy NT, Maeda A, Uyen DT, Trang DTX, Sasai M, Shiono T, Oida T, Harada S, Kamei K. 2007. Alcohols induce beta-hematin formation via the dissociation of aggregated heme and reduction in interfacial tension of the solution. *Acta Trop* 101:130–138. <https://doi.org/10.1016/j.actatropica.2007.01.001>.
 33. Gharagheizi F, Ilani-Kashkouli P, Kamari A, Mohammadi AH, Ramjugernath D. 2014. Group contribution model for the prediction of refractive indices of organic compounds. *J Chem Eng Data* 59:1930–1943. <https://doi.org/10.1021/je5000633>.
 34. Katritzky AR, Sild S, Karelson M. 1998. General quantitative structure-property relationship treatment of the refractive index of organic compounds. *J Chem Inf Comput Sci* 38:840–844. <https://doi.org/10.1021/ci980028i>.
 35. Liu J-G, Nakamura Y, Shibasaki Y, Ando S, Ueda M. 2007. High refractive index polyimides derived from 2,7-bis (4-aminophenylsulfanyl) thi-anthrene and aromatic dianhydrides. *Macromolecules* 40:4614–4620. <https://doi.org/10.1021/ma070706e>.
 36. Gaudiana RA, Minns RA, Rogers HG. 1992. High refractive index polymers. US patent 5,132,430.
 37. Olshavsky M, Allcock HR. 1997. Polyphosphazenes with high refractive indices: optical dispersion and molar refractivity. *Macromolecules* 30:4179–4183. <https://doi.org/10.1021/ma961628q>.
 38. de Villiers KA, Marques HM, Egan TJ. 2008. The crystal structure of halofantrine-ferriprotoporphyrin IX and the mechanism of action of arylmethanol antimalarials. *J Inorg Biochem* 102:1660–1667. <https://doi.org/10.1016/j.jinorgbio.2008.04.001>.
 39. Buller R, Peterson ML, Almarsson Ö, Leiserowitz L. 2002. Quinoline binding site on malaria pigment crystal: a rational pathway for antimalaria drug design. *Crystal Growth Design* 2:553–562. <https://doi.org/10.1021/cg025550i>.
 40. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. 2012. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 64:4–17. <https://doi.org/10.1016/j.addr.2012.09.019>.
 41. Zou KH, O'Malley AJ, Mauri L. 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 115:654–657. <https://doi.org/10.1161/CIRCULATIONAHA.105.594929>.
 42. Rosenthal PJ. 2003. Antimalarial drug discovery: old and new approaches. *J Exp Biol* 206:3735–3744. <https://doi.org/10.1242/jeb.00589>.
 43. Carter MD, Phelan VV, Sandlin RD, Bachmann BO, Wright DW. 2010. Lipophilic mediated assays for β -hematin inhibitors. *Comb Chem High Throughput Screen* 13:285–292. <https://doi.org/10.2174/138620710790980496>.
 44. Kurosawa Y, Dorn A, Kitsuji-Shirane M, Shimada H, Satoh T, Matile H, Hofheinz W, Masciadri R, Kansy M, Ridley RG. 2000. Hematin polymerization assay as a high-throughput screen for identification of new antimalarial pharmacophores. *Antimicrob Agents Chemother* 44:2638–2644. <https://doi.org/10.1128/AAC.44.10.2638-2644.2000>.
 45. Vargas S, Ndjoko loset K, Hay A-E, loset J-R, Wittlin S, Hostettmann K. 2011. Screening medicinal plants for the detection of novel antimalarial products applying the inhibition of β -hematin formation. *J Pharm Biomed Anal* 56:880–886. <https://doi.org/10.1016/j.jpba.2011.06.026>.
 46. Huy NT, Shima Y, Maeda A, Men TT, Hirayama K, Hirase A, Miyazawa A, Kamei K. 2013. Phospholipid membrane-mediated hemozoin formation: the effects of physical properties and evidence of membrane surrounding hemozoin. *PLoS One* 8:e70025. <https://doi.org/10.1371/journal.pone.0070025>.
 47. Jin H, Ling CX. 2005. Using AUC and accuracy in evaluating learning algorithms. *IEEE Trans Know Data Eng* 17:299–310. <https://doi.org/10.1109/TKDE.2005.50>.
 48. Jennifer A, Hoeting DM, Adrian E, Raftery Chris T, Volinsky. 1999. Bayesian model averaging: a tutorial. *Stat Sci* 14:382–417. <https://doi.org/10.1214/ss/1009212519>.
 49. Wang D, Zhang W, Bakhai A. 2004. Comparison of Bayesian model averaging and stepwise methods for model selection in logistic regression. *Stat Med* 23:3451–3467. <https://doi.org/10.1002/sim.1930>.
 50. Binh TQ, Thu NTT, Phuong PT, Nhung BT, Nhung TTH. 2015. CDKN2A-rs10811661 polymorphism, waist-hip ratio, systolic blood pressure, and dyslipidemia are the independent risk factors for prediabetes in a Vietnamese population. *BMC Genet* 16:107. <https://doi.org/10.1186/s12863-015-0266-0>.
 51. Tran TS, Hirst JE, Do MAT, Morris JM, Jeffery HE. 2013. Early prediction of gestational diabetes mellitus in Vietnam: clinical impact of currently recommended diagnostic criteria. *Diabetes Care* 36:618–624. <https://doi.org/10.2337/dc12-1418>.