JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim Article

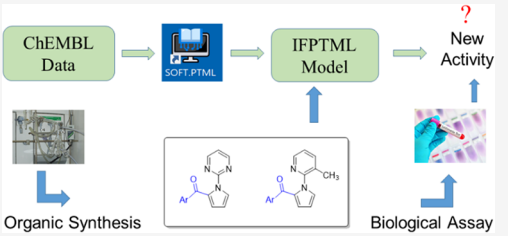
Prediction of Antileishmanial Compounds: General Model, Preparation, and Evaluation of 2-Acylpyrrole Derivatives

Carlos Santiago, Bernabé Ortega-Tenezaca, Iratxe Barbolla, Brenda Fundora-Ortiz, Sonia Arrasate, María Auxiliadora Dea-Ayuela, Humberto González-Díaz,* Nuria Sotomayor,* and Esther Lete*

Cite This: J. Chem. Inf. Model. 2022, 62, 3928−3940 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: In this work, the SOFT.PTML tool has been used to pre-process a ChEMBL dataset of pre-clinical assays of antileishmanial compound candidates. A comparative study of different ML algorithms, such as logistic regression (LOGR), support vector machine (SVM), and random forests (RF), has shown that the IFPTML-LOGR model presents excellent values of specificity and sensitivity (81−98%) in training and validation series. The use of this software has been illustrated with a practical case study focused on a series of 28 derivatives of 2-acylpyrroles **5a**,**b**, obtained through a Pd(II)-catalyzed C−H radical acylation of pyrroles. Their *in vitro* leishmanicidal activity against visceral (*L. donovani*) and cutaneous (*L. amazonensis*) leishmaniasis was evaluated finding that compounds **5bc** (IC_{50} = 30.87 μM, SI > 10.17) and **5bd** (IC_{50} = 16.87 μM, SI > 10.67) were approximately 6-fold more selective than the drug of reference (miltefosine) in *in vitro* assays against *L. amazonensis* promastigotes. In addition, most of the compounds showed low cytotoxicity, CC_{50} > 100 μg/mL in J774 cells. Interestingly, the IFPMTL-LOGR model predicts correctly the relative biological activity of these series of acylpyrroles. A computational high-throughput screening (cHTS) study of 2-acylpyrroles **5a**,**b** has been performed calculating >20,700 activity scores *vs* a large space of 647 assays involving multiple *Leishmania* species, cell lines, and potential target proteins. Overall, the study demonstrates that the SOFT.PTML all-in-one strategy is useful to obtain IFPTML models in a friendly interface making the work easier and faster than before. The present work also points to 2-acylpyrroles as new lead compounds worthy of further optimization as antileishmanial hits.



1. INTRODUCTION

Leishmaniasis is a parasitic disease, caused by *Leishmania* genus protozoan pathogens, that may present different clinical manifestations including cutaneous (CL), visceral or kala-azar (VL), post-kala-azar dermal leishmaniasis (PKDL), and mucocutaneous (MCL) leishmaniasis. As all neglected diseases, leishmaniasis remains a major global health problem as it is endemic in around 100 countries with more than 350 million people at risk.[1] Treatment of leishmaniasis relies mainly in a few drugs: pentavalent antimonials (ampB), paromomycin, pentamidine, liposomal amphotericin B, fluconazole, and miltefosine, depending on the etiological species, the infection type, and also the geographical region because of the increasing number of resistant strains. Additionally, the use of these drugs is associated with a number of severe side effects related to their toxicity.[2−5] Therefore, it is necessary to identify new effective antileishmanial compounds with chemotypes other than the ones in clinical use. In this context, nitrogen heterocycles can be considered privileged scaffolds because approximately 60% of U.S. FDA approved small-molecule drugs contain a nitrogen heterocycle.[6] In particular, the pyrrole core has attracted our attention because this motif is embedded in a variety of natural products (e.g., prodiginines,[7] bromopyrrole,[8] and spiroindimicin alkaloids[9]) with antiparasitic activity.[10] Regarding synthetic derivatives, pyridinyl aryl pyrroles **1** and **2** have proven to be inhibitors of casein kinase 1 that block the growth of *Leishmania major* promastigotes *in vitro*.[11] 1,2-Diarylpyrroles **3** have been identified as a new class of compounds active against the amastigote stay of *Leishmania infantum* by inhibiting the trypanothione reductase.[12] On the other hand, 2-acylpyrrole derivatives **4** also exhibited promising antileishmanial profiles (Figure 1).[13]

Additionally, it has been reported that pyrrole-indolinone SU11652, a sunitinib analog, targets the nucleoside diphosphate kinase from *Leishmania* parasites.[14] We recently synthesized 2-acylpyrroles through Pd(II)-catalyzed radical C−H acylation of pyrrole derivatives.[15] This efficient and flexible protocol allowed us to collect a small library of 2-acylpyrroles **5**, variably substituted on the aryl ring, and with a pyrimidine (series **5a**) or pyridine (series **5b**) ring linked to the nitrogen atom of the pyrrole nucleus (Figure 1). These structural features make our

Received: June 8, 2022
Published: August 10, 2022

https://doi.org/10.1021/acs.jcim.2c00731
J. Chem. Inf. Model. 2022, 62, 3928−3940

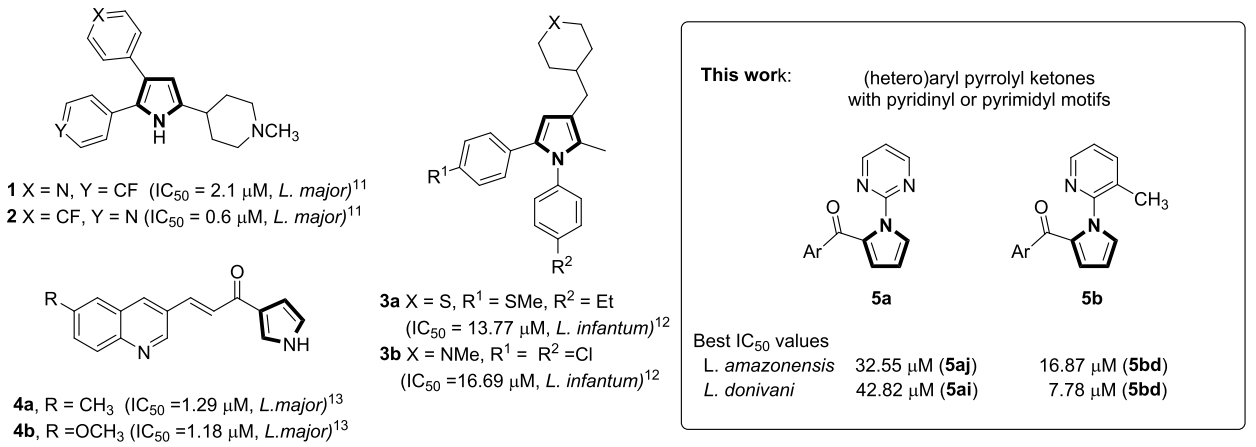


Figure 1. Antileishmanial activity of some synthetic compounds with the pyrrole motif (IC_{50} values are relative to in vitro assays on promastigotes for all compounds, except for **3**, where the values are relative to amastigotes).

pyrrole derivatives interesting candidates to be tested as potential antileishmanial compounds.

In this context, cheminformatic modeling can be a good option to reduce the development cost and increase the probability of finding new antileishmanial hits. Classic cheminformatic models focus on accelerating the antiparasitic drug discovery process by reducing the number of compounds to be assayed by trial-and-error tests. However, in addition to the large number of compounds to be tested, other factors may play a role slowing down this process. For example, the large number of combinations of biological parameters (MIC, IC_{50}, pK_i, *etc.*), parasite species, parasite stages, or target proteins greatly increases the time and cost *per* compound to be tested. Unfortunately, classic cheminformatic models fail to perform multiobjective optimization of antiparasitic compounds due to the difficulty of encoding multiple boundary conditions (parameter, protein, cell line, species, parasite stage, *etc.*) of assay and the need to obtain this information from many different data sources. We have recently reported the first PTML [(perturbation theory (PT) + machine learning (ML)] model that is capable of both explaining a very large dataset of preclinical assays of antileishmanial compounds and predicting the activity of new heterocycles (e.g., two series of pyrroloisoquinolines synthesized by our group) against different species of *Leishmania*.[16] Nevertheless, the development process of this first PTML model and its subsequent use for the prediction of new antileishmanial hits were laborious.

On the other hand, we have coined the term IFPTM [information fusion (IF) + perturbation theory (PT) + machine learning (ML)] for a new algorithm designed for multiobjective optimization of compounds. When the IF stage is missing, only the term PTML is used.[17,18] These IFPTML models have been used in medicinal chemistry, proteomics, metabolomics, and nanotechnology.[19,20]

The first phase (IF + PT) of these IFPTML models consists of merging information from different sources and/or transforming the original variables into PT operators (PTOs). These PTOs are new input variables useful for encoding information about multiple assay conditions from different sources. For example, PTOs can be used to encode information about protein targets, cell lines, microbial metabolic networks of target organisms, nanoparticle carriers of the drug, *etc.*[19,20] Next, the IFPTML workflow enters into the ML phase using classic ML algorithms. Until recently, training PTML models required running different software for each stage of the algorithm (IF, PT, and ML),[19,20] as was the case for our previous IFPTML model for antileishmanial compounds.[16] A calculation sheet was needed to run the first phase, ML software to seek the model, and a new calculation sheet to run predictions. This problem drew the attention of cheminformatics software developers to the need for new platforms to unify the different steps of IFPMTL analysis. To this end, we have introduced the QSAR-Co tool that jointly runs the PT and ML stages of the algorithm.[21] However, QSAR-Co cannot run IF procedures to calculate multilabel PTOs or reference functions that encode multiple assay conditions at the same time. Furthermore, QSAR-Co only calculates one class of PT operators, called single-condition moving averages. Consequently, it needs as many PTOs as boundary conditions are present in the problem, which implies a significantly higher number of variables to explore with respect to the multilabel PTOs used in IFPTML algorithms.[19,20] These PTOs have proven to be very useful in reducing the problem dimensionality, as in the case of ChEMBL antileishmanial pre-clinical assays dataset.[16]

Therefore, we introduced the SOFT.PTML studio tool, which has the possibility of calculating multilabel PTOs, including multilabel/multicondition reference functions, moving averages, co-variances, *etc.* SOFT.PTML has been used successfully in nanotechnology and medicinal chemistry.[22,23] In the present work, we report for the first time the use of SOFT.PTML to seek IFPTML models for antileishmanial compounds, performing a comparative study of different ML algorithms. We have also carried out a predictive study of a series of 2-acylpyrroles, previously synthesized by our group,[15] together with experimental preparation of new samples for assay and their *in vitro* leishmanicidal testing. Some of the 2-acylpyrroles tested compare favorably with respect to miltefosine (reference compound) in terms of activity and toxicity. This work opens a new experimental line of research focused on the synthesis and optimization of antileishmanial compounds as 2-acylpyrrole derivatives. It also lays the ground for the development of faster and user-friendlier IFPTML models for other neglected tropical diseases.

The general flowchart showing the interconnections between the different parts of this work: (1) chemoinformatics study, (2) organic synthesis, and (3) biological assays, is depicted in Figure 2.

https://doi.org/10.1021/acs.jcim.2c00731
J. Chem. Inf. Model. 2022, 62, 3928−3940

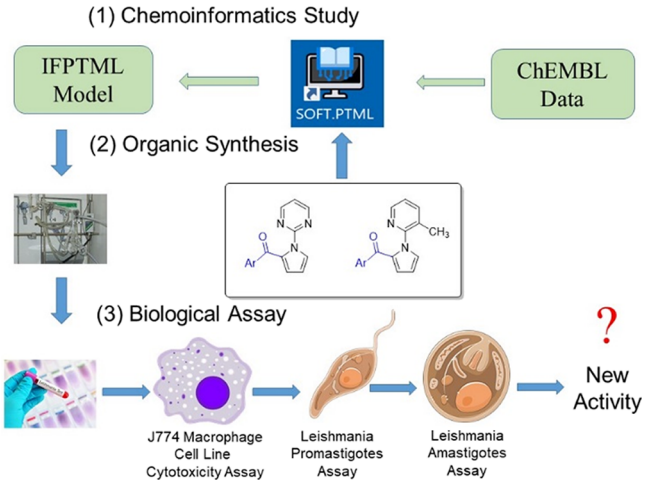


Figure 2. General workflow.

2. MATERIALS AND METHODS

2.1. Computational Methods. *2.1.1. IFPTML Model Basis.* An IFPTML model is proposed to calculate the values of the antileishmanial biological activity scoring function $f(v_{ij})_{calc}$ of the *i*th query compound in the *j*th assay with multiple boundary conditions $\mathbf{c}_j = [c_0, c_1, c_2, c_{jmax}]$. In these classification models, the $f(v_{ij})_{calc}$ function, which gets dimensionless values, is used to score the propensity of the *i*th compound to reach a certain level of the biological activity values v_{ij} (see next section).[24] Consequently, the values of $f(v_{ij})_{calc}$ can be used directly to compare the relative propensity of two different compounds to reach a certain level of biological activity in the *j*th assay compared to a threshold value $cutoff_j$. They can also be used to compare the behavior of the same compound in two different assays. The IFPTML model uses two types of input variables [functions of reference $f(v_{ij})_{ref}$ and perturbation theory operators $PTO_k(\mathbf{c}_j)$] to calculate the $f(v_{ij})_{calc}$ output values. Thus, the IFPTML model starts with the values of a function of reference $f(v_{ij})_{ref}$, which are used to characterize/identify the kind of biological activity to be modeled. Next, the values of the $PTO_k(\mathbf{c}_j)$ functions are used to measure the effect of perturbations over the biological activity outcome. $PTO_k(\mathbf{c}_j)$ functions quantify perturbations/deviations in the structure of the *i*th compound and/or in the conditions assay $\mathbf{c}_j$ compared to a set of reference compounds.[24] In the next section, the pre-processing of the raw data to construct the $f(v_{ij})_{ref}$ and $PTOk(\mathbf{c}_j)$ functions is explained. IFPTML linear models have the following form (eq 1):

$$f(v_{ij})_{calc} = a_0 + a_1 \cdot f(v_{ij})_{ref} + \sum_{k=1,j=0}^{k_{max},j_{max}} a_{kj} \cdot PTO_k(\mathbf{c}_j) \quad (1)$$

Figure 3 shows a workflow illustrating the integration of the different phases (IF + PT + ML) of the IFPTML analysis. First, the IF phase, which includes data collection, data rearrangement, and data fusion (horizontal and vertical), is run. The vertical IF involves aligning the output values v_{ij} for different output parameters (IC_{50}, K_i, *etc.*) in the same column. These parameters are then transformed into a Boolean variable $f(v_{ij})_{obs}$ using different cutoff values, (see next section). Horizontal IF involves merging multiple labels from pre-clinical assays to form multicondition label variables $\mathbf{c}_j = [c_1, c_2, \ldots c_j]$. Next, the PT phase is run by calculating the $PTO(D_k, \mathbf{c}_j)$ operators that can

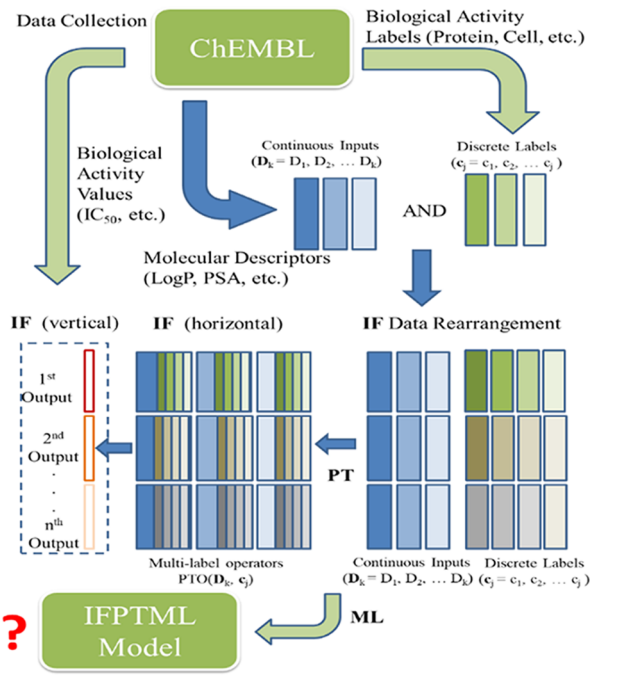


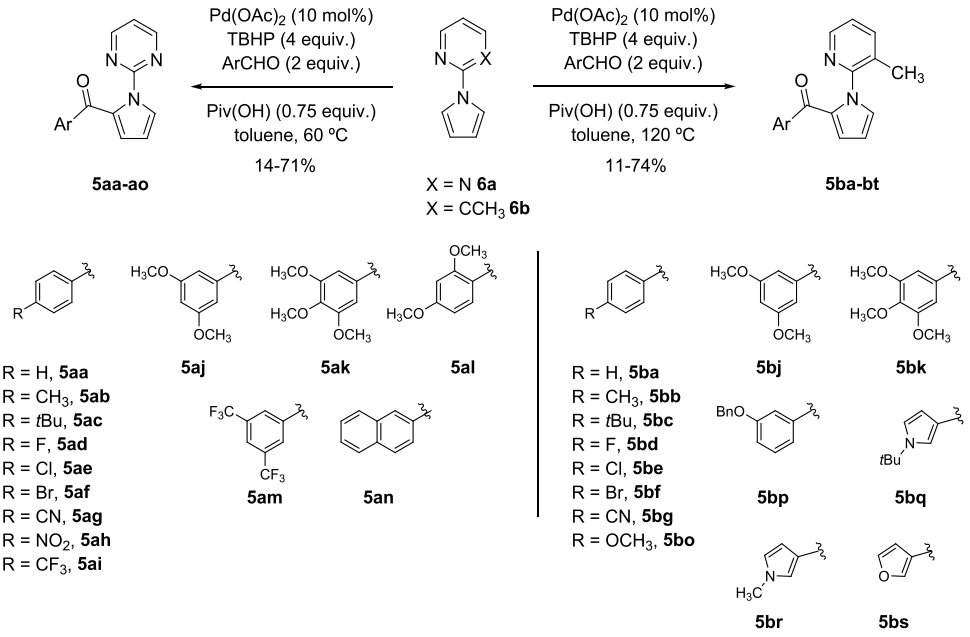
Figure 3. IFPTML method workflow illustrating the different phases (IF + PT + ML).

encode structural information (D_k) and multiple assay conditions ($\mathbf{c}_j$) at the same time. In the last phase, the different ML algorithms are run using the $PTO(D_k, \mathbf{c}_j)$ values as input. The following sections provide more details about the IFPTML steps.

2.1.2. Data Pre-processing and IF Output (Horizontal IF). The SOFT.PTML tool was used to pre-process a ChEMBL dataset of pre-clinical assays of antileishmanial compound candidates.[16] This tool was also used to perform training/validation of alternative IFPTML models. These data included 145,851 biological activity values (v_{ij}) for the *i*th compounds tested on the *j*th preclinical assays with boundary experimental conditions (labels) $\mathbf{c}_j = [c_0, c_1, c_2, \ldots c_n]$. The values v_{ij} are expressed as different antileishmanial activity parameters with label c_0, such as IC_{50}, K_i, EC_{50}, *etc.* The SOFT.PTML discretization procedure was used in order to convert all the values of biological activity v_{ij} into a Boolean objective function $f(v_{ij})_{obs} = 1$ or 0. The function was implemented in the software is $f(v_{ij})_{obs} = 1$ IF ($v_{ij} \geq cutoff_j$ AND $d(c_0) = 1$) OR ($v_{ij} \leq cutoff_j$ AND $d(c_0) = -1$) ELSE $f(v_{ij})_{obs} = 0$. In this equation, $\geq$ means higher than (>) or equal to (=) the cutoff. By analogy, $\leq$ means lower than (<) or equal to (=) the cutoff. The different $cutoff_j$ values are the cutoff values used for the different c_0-labaled biological activity parameters (IC_{50}, K_i, EC_{50}, *etc.*) in different *j*th assays. The desirability parameter $d(c_0) = 1$ or -1 had to be maximized or minimized to obtain an optimal biological effect. This IFPTML model is multioutput, i.e., it can predict multiple output probability of activity $p(f(v_{ij}) = 1)$ for the same *i*th drug entry. Consequently, to train the model, different values of the biological activity parameter v_{ij} (K_i, IC_{50}, etc.) have been transformed into Boolean variables $f(v_{ij}) = 1$ OR 0. Since each biological activity parameter v_{ij} (K_i, IC_{50}, *etc.*) has a different scale and optimal region, different values of cutoff were used for each parameter v_{ij}. These values of $cutoff_j$ were determined by trying to balance both the optimal activity region and the number of cases in the two classes $f(v_{ij}) = 1$ OR 0 for each

https://doi.org/10.1021/acs.jcim.2c00731
J. Chem. Inf. Model. 2022, 62, 3928−3940

Scheme 1. Synthesis of 2-Acylpyrroles 5a and 5b Screened against *L. amazonensis* and *L. donovani*



parameter. The more representative values of cutoff$_j$, desirability d(c_0), and number of cases n(c_0) in the dataset are listed in the Supporting Information 1 (CUTOFF sheet). The number of positive cases n(f(v_{ij}) = 1) and resulting values of the reference function $f(v_{ij})_{ref}$ = p(f(v_{ij}) = 1) priori probability for different biological activity parameters c_0 = property (units) were also listed.

2.1.3. Input IF and Codification (Horizontal IF). All the information on the boundary experimental conditions of the biological activity assays is codified by the vectors $\mathbf{c}_j$ = [c_0, c_1, c_2, ...c_{maxj}]. Members of this vector were regrouped, creating two partitions $P(c_j)_I = \mathbf{c}_I$ and $P(c_j)_{II} = \mathbf{c}_{II}$, which are also discrete variables. The value of the partition can be calculated by means of a horizontal IF process (see Figure 3) that consists on concatenating the text values of each label variable c_j included in the subset of variables I or II in the partition. That is, the first partition is calculated as $P(c_j)_I$ = concatenate $(c_j)_I$ = concatenate ([c_0, c_1, ... c_{max}])$_I$ = $\mathbf{c}_I$ and the second as $P(cj)_{II}$ = concatenate $(cj)_{II}$ = concatenate ([c_0, c_1, ... c_{max}])$_{II}$ = $\mathbf{c}_{II}$. These partitions independently encode information on the experimental conditions of preclinical assays $P(c_j)_I = \mathbf{c}_I$ = Concatenate([c_0, c_1, c_2, c_3, c_4, c_5]) = $c_0 - c_1 - c_2 - c_3 - c_4 - c_5$ and on the nature and quality of data $P(c_j)_I = \mathbf{c}_I$ = concatenate ([c_6, c_7, c_8, c_9, c_{10}]).

On the other hand, the structural information is also expressed as a vector $\mathbf{D}_{ki}$ = [D_{1i}, D_{2i}, D_{3i}, ...D_{maxi}]. The elements of this vector are: D_{1i} = ALOGP$_i$, the *n*-octanol/water partition coefficient; D_{2i} = PSA$_i$, the topological polar surface area, and D_{3i} = NVLR, the number of violations to Lipinski's rule of the structure of *i*th compound. The values of the molecular descriptors D_k were downloaded from ChEMBL and/or calculated with the software DRAGON[25] for new compounds. Next, the IF process was carried out to calculate the multicondition PT operators $PTO_k(\mathbf{D}_{ki}, \mathbf{c}_j)$. Each $PTO_k(\mathbf{D}_{ki}, \mathbf{c}_j)$ variable is the expression of the fusion of structural information from one or multiple elements of $\mathbf{D}_i$ and one or multiple elements of $\mathbf{c}_j$. Therefore, a $PTO_k(\mathbf{D}_{ki}, \mathbf{c}_j)$ is a function or operator (PTO_k) calculated to merge structural information $\mathbf{D}_i$ of the *i*th compound with the conditions $\mathbf{c}_i$ of the *j*th preclinical assay. The PTO_k can have many different forms and/or can be calculated for different subsets (domains) of the vectors $\mathbf{D}_{ki}$ and $\mathbf{c}_j$. First, the operators of type $PTO(D_{ki}, \mathbf{c}_I) = PTO(D_{1i}, \mathbf{c}_I)$, $PTO(D_{2i}, \mathbf{c}_I)$, *etc.* or type $PTO_k(\mathbf{D}_{ki}, \mathbf{c}_{II}) = PTO(D_{1i}, \mathbf{c}_{II})$, $PTO(D_{2i}, \mathbf{c}_{II})$, *etc.* are calculated. These calculated PTOs can take the form of multicondition moving average (MA) operators: $\Delta D_k(\mathbf{c}_I) = D_{ki} - \langle D_k(\mathbf{c}_I)\rangle$ or $\Delta D_k(\mathbf{c}_I) = D_{ki} - \langle D_k(\mathbf{c}_{II})\rangle$. They measure the deviation (Δ) of the structure of the *i*th compound expressed by D_{ki} with respect to the average/expected values $\langle D_k(\mathbf{c}_j)\rangle$ for all the compounds assayed under the same conditions $\mathbf{c}_I$ or $\mathbf{c}_{II}$.

2.1.4. SOFT.PTML In Silico Screening of New Compounds. The new model was used to study a series of 28 di(hetero)aryl ketone derivatives (2-acylpyrroles) synthesized in our group.[16] First, the SMILE codes of the 28 compounds were generated using ChemDraw Professional 20.1.[26] Next, the values of the input variables D_1 = LOGP, D_2 = PSA, and D_3 = NVLR were calculated. Then, these values were substituted on the SOFT.PTML model in order to obtain the probabilities of activity for each compound in different biological assays. A simulation of the biological response of 28 compounds + 1 control (miltefosine) in many different preclinical assays was carried out. These assays included a total of >50 different biological activity parameters [K_i(nM), IC_{50}(nM), Inhibiton (%), *etc.*)], 35 target proteins (P00374 dihydrofolate reductase, Q0GKD7 farnesyl pyrophosphate synthase, *etc.*), 28 cell lines (J774, HL-60, Jurkat, *etc.*), 40 assay organisms (*L. donovani*, *L. major*, *L. amazonensis*, *etc.*), and 2 microorganism development stages (amastigotes and promastigotes). In total, the outcome of the 29 compounds in 249 different pre-clinical assays was predicted.

2.1.5. Data Sampling. In order to select the training and validation sets, we carry out the following steps. First, we ordered all assays according to the labels of output property c_0 (IC_{50}, K_i, *etc.*) and the discrete values of the two main partitions, $\mathbf{c}_I$ and $\mathbf{c}_{II}$. After that, we assigned a value of set = train (t) or validation (v) from the beginning to the end following the pattern tttv. This allows us to use 75% (3/4) of the data for training the model and the remaining 25% (1/4) of the data for validation. The sorting of the cases by c_0, $\mathbf{c}_I$, and $\mathbf{c}_{II}$ ensures a

https://doi.org/10.1021/acs.jcim.2c00731
J. Chem. Inf. Model. 2022, 62, 3928−3940

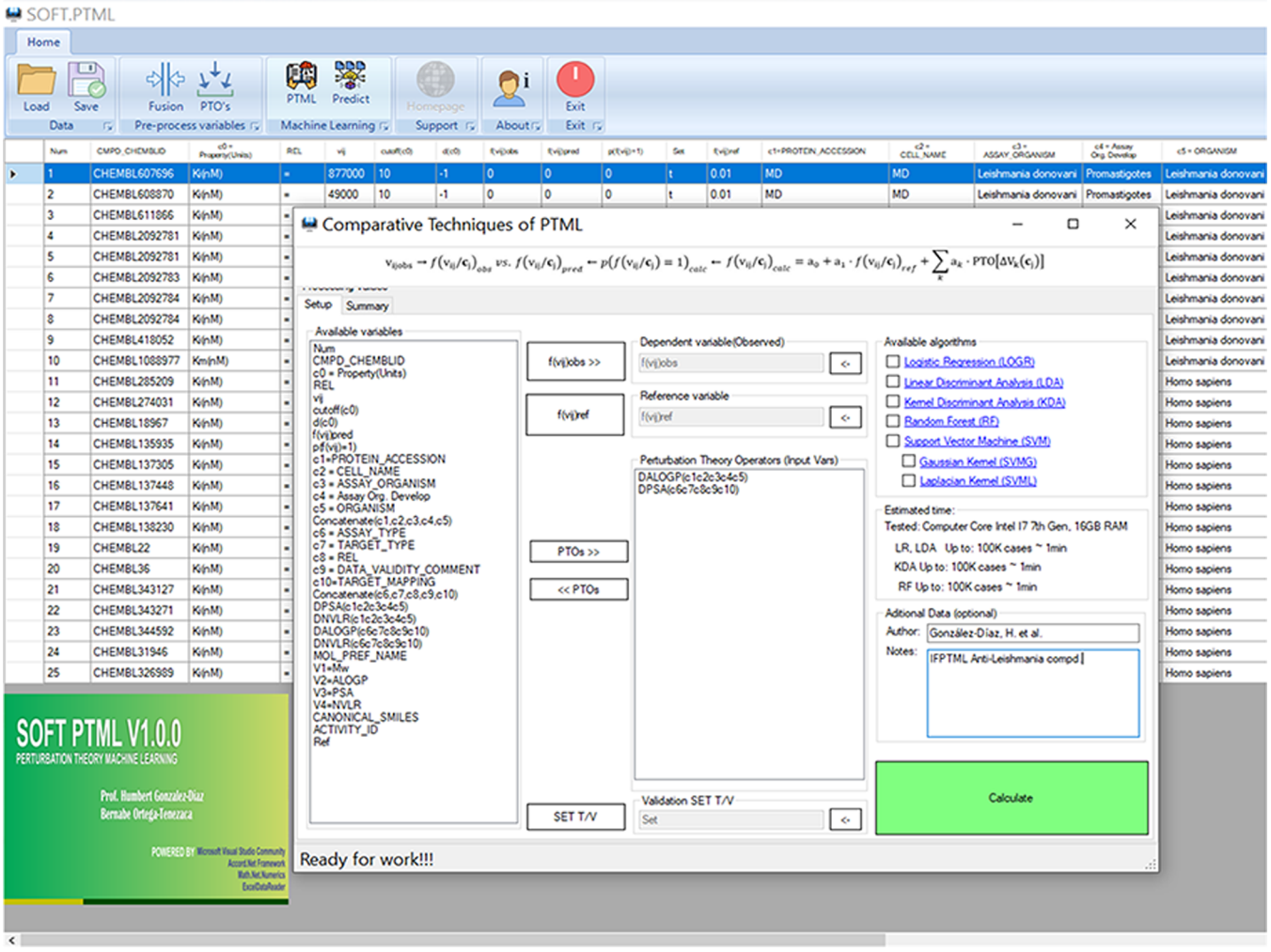


Figure 4. SOFT.PTML interface for analysis of ChEMBL antileishmanial compounds data.

more representative distribution of all data strata or data subsets (properties, target proteins, pathogen species, etc.) in both training and validation series. The random sorting of all cases within each subset (same property label c_0, or assay conditions c_I and c_{II}) ensured a higher randomness of the sampling. In such a way, we have carried out one stratified, random, and representative data sampling. As result, we included a total of 6711 positive assays ($f(v_{ij}) = 1$) and 102,678 negative or control assays ($f(v_{ij}) = 0$) in training series. In addition, we included a total of 2050 positive assays ($f(v_{ij}) = 1$) and 34,212 negative or control assays ($f(v_{ij}) = 0$) in validation series.

2.2. Experimental Methods. *2.2.1. Synthesis of 2-(Hetero)aroylpyrroles* ***5****.* Acylpyrroles **5aa**−**an** and **5ba**−**bs** were synthesized following a procedure previously developed by us (Scheme 1).[15] Acylation of 2-(1*H*-pyrrol-1-yl)pyrimidine **6a** and 3-methyl-2-(1*H*-pyrrol-1-yl)pyridine **6b** with the corresponding (hetero)aromatic aldehydes was carried out using $Pd(OAc)_2$ as the catalyst, TBHP as the oxidant, and pivalic acid as the additive in dry toluene as the solvent. The reactions were carried out in sealed tubes at 60 °C (for **6a**) or 120 °C (for **6b**) for 1.5−7 h.

2.3. Details for Pre-clinical Assays. *2.3.1. Parasites and Culture Procedure.* The following species of *Leishmania* were used: *L. donovani* (MHOM/IN/80/DD8) was purchased (ATCC, USA) and *L. amazonensis* (MHOM/Br/79/Maria) was kindly provided by Prof. Alfredo Toraño (Instituto de Salud Carlos III, Madrid). Promastigotes were cultured in Schneider's insect medium supplemented with 10% heat-inactivated fetal bovine serum (FBS) and 1000 U/L of penicillin plus 100 mg/L of streptomycin in 25 mL culture flasks at 26 °C.

2.3.2. In Vitro Promastigote Susceptibility Assay. The biological assay was carried out following previously published protocols.[27,28] Concisely, promastigotes (2.5×10^5 parasites/well) from the log phase have been cultured in 96-well plastic plates. New samples of the compounds have been prepared according to the protocol described above, and subsequently solutions of the chemical compound to be assayed have been dissolved in DMSO at 50 mg/mL. We performed serial dilutions 1:2 of the compounds in fresh culture medium (100, 50, 25, 12.5, 6.25, 3.12, 1.56, and 0.78 μg/mL) up to a 200 μL final volume. Growth control and signal-to-noise control were also included. The final solvent (DMSO) concentrations never exceeded 0.5% (v/v) warranting no effect on parasite proliferation or morphology. After 48 h at 26 °C, 20 μL of a 2.5 mM resazurin solution was added to each well and the plates were returned to the incubator for another 3 h. The relative fluorescence units (RFU) (535 nm−590 nm excitation−emission wavelength) was determined in a fluorimeter (Infinite 200Tecan i-Control). Growth inhibition (%) was calculated by 100 − [(RFU treated wells − RFU signal-to-noise)/(RFU untreated − RFU signal-to-noise) × 100]. All tests were carried out in triplicate. Miltefosine (Sigma-Merck, Madrid, Spain) was

https://doi.org/10.1021/acs.jcim.2c00731
J. Chem. Inf. Model. 2022, 62, 3928−3940

Table 1. SOFT.PTML Analysis Results

PTML model	ML	dataset	stat param	param value	set $f(v_{ij})_{pred}$	$f(v_{ij})_{obs}$ 0	$f(v_{ij})_{obs}$ 1
1	LOGR	training	Sp	0.9885	0	101,493	1233
			Sn	0.9885	1	1185	5478
			Ac	0.9779	total		
		validation	Sp	0.9888	0	33,828	407
			Sn	0.8191	1	384	1843
			Ac	0.9783	total		
2	RF	training	Sp	0.9985	0	102,519	503
			Sn	0.9985	1	159	6208
			Ac	0.9939	total		
		validation	Sp	0.9934	0	33,985	315
			Sn	0.8600	1	227	1935
			Ac	0.9851	total		
3	SVM	training	Sp	0.9972	0	102,388	988
			Sn	0.9972	1	290	5723
			Ac	0.9883	total		
		validation	Sp	0.9944	0	34,021	757
			Sn	0.6636	1	191	1493
			Ac	0.974	total		

used as the reference drug and was evaluated under the same conditions. The efficacy of each compound was estimated by calculating the IC_{50} (concentration of the compound that produced a 50% reduction in parasites) using a multinomial probit analysis incorporated in SPSS software v21.0. The selectivity index (SI) was calculated as the ratio between cytotoxicity (CC_{50}) and activity against parasites (IC_{50}).

2.3.3. In Vitro Intracellular Amastigote Susceptibility Assay. The assay was carried out as previously described.[29] Briefly, 5×10^4 J774 macrophages and stationary promastigotes in a 1:5 ratio were seeded in each well of a microtiter plate, suspended in 200 μL of culture medium, and incubated for 24 h at 33 °C in a 5% CO_2 chamber. After this first incubation, the temperature was increased up to 37 °C for another 24 h. Thereafter, cells were washed several times in culture medium by centrifugation at 1.500*g* for 5 min in order to remove free non-internalized promastigotes. Finally, the supernatant was replaced by 200 μL/well of culture medium containing 2-fold serial dilutions of the test compounds as in promastigotes assay. Growth control and signal-to-noise were also included. Following incubation for 48 h at 37 °C and 5% CO_2, the culture medium was replaced by 200 μL/well of the lysis solution (RPMI-1640 with 0.048% HEPES and 0.01% SDS) and incubated at room temperature for 20 min. Thereafter, the plates were centrifuged at 3.500*g* for 5 min and the lysis solution was replaced by 200 μL/well of Schneider's insect medium. The culture plates were then incubated at 26 °C for another 4 days to allow transformation of viable amastigotes into promastigotes and proliferation. Afterward, 20 μL/well of 2.5 mM resazurin was added and incubated for another 3 h. Finally, fluorescence emission was measured and IC_{50} was estimated as described above. All tests were carried out in triplicate. Miltefosine (Sigma-Merck, Madrid, Spain) was used as the reference drug and was evaluated under the same conditions. The IC_{50} and SI were calculated as in the previous section.

2.3.4. Cytotoxicity Assay on Macrophages. The assay was carried out as previously described.[30] J774 macrophages cell lines were seeded (5×10^4 cells/well) in 96-well flat-bottom microplates with 100 μL of RPMI 1640 medium. The cells were allowed to attach for 24 h at 37 °C and 5% CO_2, and the medium was replaced by different concentrations of the compounds in 200 μL of medium and exposed for another 24 h. Growth controls and signal-to-noise were also included. Afterward, a volume of 20 μL of 2.5 mM resazurin solution was added, and plates were returned to the incubator for another 3 h to evaluate cell viability. The reduction of resazurin was determined by fluorometry as in the promastigote assay. Each concentration was assayed three times. The cytotoxicity effect of compounds was defined as the 50% reduction of cell viability of treated culture cells with respect to untreated culture (CC_{50}) and was calculated using a multinomial probit analysis incorporated in SPSS software v21.0.

3. RESULTS AND DISCUSSION

3.1. SOFT.PTML Model. As mentioned in the introduction, the IFPTML algorithm is useful for finding predictive models for multiobjective optimization of compounds. In fact, we have already used IFPTML models for the study of new pyrroloisoquinolines *vs* different *Leishmania* species.[16] However, all steps of IFPTML analysis needed to be performed on different software and/or using different manual operations. Therefore, we decided to use our SOFT.PTML software for the development of IFPTML models for prediction of antileishmanial compounds.

The same dataset[16] containing n = 109,389 preclinical assays was selected and re-processed with the SOFT.PTML software. Figure 4 shows the user-friendly interface of the software with the IF, PT, and ML stages integrated in a single application, which allowed exploring different ML techniques in a more automatic way. Specifically, logistic regression (LOGR), support vector machine (SVM), and random forests (RF) algorithms were studied.[31−33] Table 1 summarizes the results obtained with the different algorithms (see the Supporting Information 1 for details of the dataset used and detailed results of the model for each case).

In this study, a total of 145,851 cases (pre-clinical assays outcomes) distributed in 109,389 cases in training series and 36,462 cases in validation series have been analyzed. The sensitivity (Sn) and specificity (Sp) values obtained from these algorithms were studied using the IFPTML strategy both in the

https://doi.org/10.1021/acs.jcim.2c00731
J. Chem. Inf. Model. 2022, 62, 3928−3940

Table 2. IC_{50} Leishmanicidal and Cytotoxic Effects from 2-Acylpyrrole Series 5a and 5b (Expressed as μM) on *In Vitro* Promastigote Assays

		L. amazonensis		*L. donovani*		macrophages J774
entry	comp.	IC_{50} ± SD (μM)[a]	SI[b]	IC_{50} ± SD (μM)[a]	SI[b]	CC_{50} ± SD (μM)[c]
1	**5aa**	259.58 ± 40.72	>1.55	N/A[d]		401.17[e]
2	**5ab**	N/A[d]		N/A[d]		381.24[e]
3	**5ac**	32.88 ± 0.74	>2.77	58.54 ± 7.04	>1.55	91.02 ± 6.55
4	**5ad**	N/A[d]		N/A[d]		270.19 ± 26.30
5	**5ae**	117.09 ± 10.92	>3.01	189.77 ± 6.76	>1.86	352,46[e]
6	**5af**	67.18 ± 4.94	>4.54	118.18 ± 25.81	>2.58	304.72[e]
7	**5ag**	N/A[d]		N/A[d]		364.59[e]
8	**5ah**	N/A[d]		N/A[d]		339.82[e]
9	**5ai**	48.02 ± 1.61	>3.13	42.82 ± 0.61	>3.51	150.36 ± 49.40[e]
10	**5aj**	32.55 ± 0.64	>2.71	210.90 ± 32.77	>0.42	88.29 ± 3.36[e]
11	**5ak**	N/A[d]		N/A[d]		87.27 ± 7.37[e]
12	**5al**	N/A[d]		N/A[d]		224.28 ± 46.89[e]
13	**5am**	57.34 ± 0.25	>4.53	198.54 ± 25.36	>1.31	259.56[e]
14	**5an**	36.11 ± 1.23	>3.12	58.50 ± 3.34	>1.93	112.82 ± 39.19[e]
15	**5ba**	152.77 ± 21.65	>2.50	N/A[d]		381.23[e]
16	**5bb**	149.20 ± 3.84	>2.43	119.64 ± 14.98	>3.02	361.87[e]
17	**5bc**	30.87 ± 3.11	>10.17	221.31 ± 36.90	>1.42	314.05[e]
18	**5bd**	16.87 ± 0.73	>10.67	7.78 ± 0.27	>23.15	136.96 ± 36.42
19	**5be**	51.46 ± 4.72	>2.89	20.69 ± 0.98	>7.19	148.74 ± 19.33
20	**5bf**	38.10 ± 0.85	>3.14	19.87 ± 1.47	>6.01	119.51 ± 37.36[e]
21	**5bg**	191.77 ± 12.04	>1.81	315.43 ± 60.80	>1.10	348.04[e]
22	**5bj**	71.06 ± 7.81	>1.53	43.51 ± 0.59	>0.80	108.45 ± 5.76[e]
23	**5bk**	N/A[d]		N/A[d]		283.78
24	**5bo**	224.26 ± 14.71	>1.53	172.92 ± 60.34	>1.98	342.07[e]
25	**5bp**	207.96 ± 23.62	>1.31	184.22 ± 8.67	>1.47	356.76[e]
26	**5bq**	209.16 ± 4.69	>1.56	92.81 ± 7.68	>1.08	325.31[e]
27	**5br**	226.54 ± 13.42	>1.66	N/A[d]		376.90[e]
28	**5bs**	N/A[d]		N/A[d]		396.40[e]
29	miltefosine	30.67 ± 8.80	1.80	0.24 ± 0.02	230.83	55.40 ± 4.19

[a]IC_{50}: concentration of the compound that produced a 50% reduction in parasites; SD: standard deviation. [b]SI: selectivity index, SI = CC_{50}/IC_{50}. [c]CC_{50}: concentration of the compound that produced a 50% reduction of cell viability in treated culture cells with respect to untreated ones. [d]N/A: not active at the maximum dose tested (100 μg/mL). [e]CC_{50} values, expressed as μM, correspond to 100 μg/mL, which was the higher doses tested.

training and external validation series of pre-clinical assays. It should be noted that all algorithms gave interesting results. However, the IFPTML-SVM model was discarded because it had a very low value of Sn = 0.6636 in validation series. The IFPTML-RF model was also discarded because although the results were very promising (Sp ≈ Sn = 0.8−0.98 range), the model itself is markedly more complex than the linear models. Therefore, the linear model IFPTML-LOGR was selected as the most appropriate based on the Ocam's razor or parsimony principle.[34,35] The equation of the IFPTML-LOGR model is the following (eq 2):

$$f(\nu_{ij})_{calc} = 8.350971 \cdot f(\nu_{ij})_{ref} + 0.885564 \cdot \Delta D_1(\mathbf{c}_I) + 0.024487 \cdot \Delta D_2(\mathbf{c}_I) - 1.388516 \cdot \Delta D_3(\mathbf{c}_I) - 0.853978 \cdot \Delta D_1(\mathbf{c}_{II}) - 0.025694 \cdot \Delta D_2(\mathbf{c}_{II}) + 1.517539 \cdot \Delta D_3(\mathbf{c}_{II}) + 5.090163 \quad (2)$$

$n = 109389$, $\chi^2 = 135169.7$, and $p < 0.05$

3.1.1. All-in-One vs Multisoftware Strategy. SOFT.PTML uses an all-in-one strategy implementing all stages (IF, PT, and ML) of the IFPTML algorithm in the same platform. To validate the all-in-one strategy, it was necessary to demonstrate that this software was capable of reproducing the results obtained with the multisoftware strategy. Therefore, we decided to compare the IFPTML-LDA model with the multisoftware strategy (eq 3) with the IFPTML-LOGR model because both models are linear equations with the same form, despite using different ML techniques (LDA and LOGR).

$$f(\nu_{ij})_{calc} = 59.918599 \cdot f(\nu_{ij})_{ref} + 1.631537 \cdot \Delta D_1(\mathbf{c}_I) + 0.041494 \cdot \Delta D_2(\mathbf{c}_I) - 2.675709 \cdot \Delta D_3(\mathbf{c}_I) - 1.562187 \cdot \Delta D_1(\mathbf{c}_{II}) - 0.041886 \cdot \Delta D_2(\mathbf{c}_{II}) + 2.649511 \cdot \Delta D_3(\mathbf{c}_{II}) - 25.182671 \quad (3)$$

$n = 109389$, $\chi^2 = 135169.7$, and $p < 0.05$

Both IFPTML models begin with a reference function $f(\nu_{ij})_{ref}$, whose value represents the prior probability with which a compound selected at random can give a positive outcome $f(v_{ij}) = 1$ of the specific parameter c_0 under conditions c_j. This function is used to specify within the equation the biological parameter (IC_{50}, K_i, MIC, etc.) to be studied. Next, the $PTO_{ki}(D_{ki}, c_j) = \Delta D_{ki}(c_j)$ values are added to measure variations/perturbations. It should be highlighted that both IFPTML-LOGR and IFPTML-LDA models reached similar values of Sn and Sp in training and validation series. It was also found that the ratio between the coefficients (a_k) of the

https://doi.org/10.1021/acs.jcim.2c00731
J. Chem. Inf. Model. 2022, 62, 3928−3940

Table 3. IC_{50} Leishmanicidal and Cytotoxic Effects from 2-Acylpyrroles 5af and 5bc (Expressed as μM) on *in Vitro* Amastigote Assay

		L. amazonensis		*L. donovani*		macrophages J774
entry	compound	IC_{50} ± SD (μM)[a]	SI[b]	IC_{50} ± SD (μM)[a]	SI[b]	CC_{50} ± SD (μM)[c]
1	**5af**	153.27 ± 9.11	>1.99	210.87 ± 30.26	>1.45	304.72[e]
2	**5bc**	60.55 ± 7.88	>5.19	N/A[d]		314.05[e]
3	miltefosine	47.55 ± 7.04	2.85	0.44 ± 0.05	307.70	135.93 ± 10.19

[a]IC_{50}: concentration of the compound that produced a 50% reduction in parasites; SD: standard deviation. [b]SI: selectivity index, SI = CC_{50}/IC_{50}. [c]CC_{50}: concentration of the compound that produced a 50% reduction of cell viability in treated culture cells with respect to untreated ones. [d]N/A: not active at the maximum dose tested (100 μg/mL). [e]CC_{50} values, expressed as μM, correspond to 100 μg/mL, which was the higher doses tested.

$PTO_{ki}(D_{ki}, c_j)$ variables in both models is exactly constant = 2.0. This indicates that, except for a scale factor of 2, both equations give equal weight to the different variables and should give similar results. In fact, we found a correlation coefficient of R = 0.98 for the $f(v_{ij})_{calc}$ values obtained with eq 2 *vs* eq 3. This result demonstrates that the all-in-one strategy implemented in SOFT.PTML is capable of reproducing the results obtained with the multisoftware strategy, using a single program with a user-friendly interface, which makes the work notably easier and faster.

3.1.2. Computational and Experimental Study of 2-Acylpyrroles. A case study is presented to illustrate the use of SOFT.PTML models for discovery of antileishmanial compounds in practice. As stated before, we focused on 2-acylpyrroles **5a** and **5b**, whose synthesis has been previously reported by us,[15] because they combined structural features of related pyrrole derivatives[11−14] with promising antileishmanial activity. To our knowledge, no previous studies on their antileishmanial activity have been reported. We describe herein the *in vitro* assays and in-depth computational screening of these compounds. First, the synthesis of new samples of 2-acylpyrroles **5a** and **5b** with pre-clinical assay quality was carried out. Next, these compounds were tested against two species of *Leishmania* in different development stages. These experimental studies included two biological activity parameters (IC_{50} and CC_{50}) for two *Leishmania* species (*L. amazonensis* and *L. donovani*). Finally, the study was closed with a wide computational screening of these compounds *vs* many *Leishmania* species in different stages and multiple target proteins.

3.1.3. Preparative Organic Synthesis. Our group has recently reported the synthesis of a variety of 2-(hetero)-aroylpyrroles through a Pd(II)-catalyzed acylation of pyrrole with aldehydes[36] in the presence of an oxidant, using 2-methylpyridinyl and 2-pyrimidyl as directing groups.[15] This radical C−H activation reaction[37−41] is a good catalytic alternative to classical acylation methods (Friedel−Crafts, Vilsmeier−Haack, or Houben−Hoesch type acylation reactions), which minimizes the production of waste as it does not require the use of stoichiometric amounts of Lewis or protic acids. Thus, we had demonstrated that the use of 2-pyrimidinyl directing group led to C-2 metalation of pyrrole using $Pd(OAc)_2$ as the pre-catalyst in toluene, which were acylated with aldehydes in the presence of TBHP as the oxidant and pivalic acid as the additive. The procedure could be efficiently extended to a series of aldehydes with different substitution patterns on the aromatic ring, obtaining 2-acylpyrroles **5aa−an** (Scheme 1), though diacylation could not be completely avoided. However, under the same experimental conditions, the use of 3-methyl-2-pyridinyl directing group led to the formation of 2-acylpyrrole derivatives **5ba−bs** in moderate to good yields, except when electron-withdrawing substituents were present in the aromatic ring (Scheme 1).

3.1.4. Antileishmanial Activity Pre-Clinical Assay. The 2-(hetero)aroylpyrrole derivatives **5a** and **5b** were tested against *L. amazonensis* and *L. donovani*, which are responsible for the two main clinical forms of this neglected tropical disease, cutaneous and visceral leishmaniasis, respectively (Table 2). We performed *in vitro* promastigote and *in vitro* intracellular amastigote susceptibility assays (IC_{50}) and cytotoxicity assays (CC_{50}) on the J774 cell line of macrophages using miltefosine as the drug of reference (see Materials and Methods), and the corresponding selectivity indexes (SI) were calculated. Detailed information of the biological activity of the more interesting compounds, including the compound code, concentration, repeated measures of biological activity, average values, can be found in the Supporting Information 2. This file also contains the graphic representations of dose−effect curves for these compounds and the drug of reference, miltefosine.

The performance of each *N*-pyrimidin-2-yl acylated pyrrole **5a** was compared with that of the corresponding *N*-(3-methylpyridin-2-yl) derivative **5b** (Table 2). The bioactivities of some compounds of both series compare well in terms of activity and selectivity against *L. amazonensis* promastigotes. The aromatic substitution pattern of the acyl group plays an important role in the antileishmanial activity of these pyrrole derivatives. In some cases, we observed similar trends in the bioactivity profile for pyrimidine derivatives **5a** and the corresponding pyridines **5b**. For example, the 4-^{t}butylphenyl pyrrolyl methanones **5ac**/**5bc** and the 3,5-disubstituted phenyl pyrrolyl methanones **5aj**/**5bj**, with electron-donating (MeO) substituents, showed IC_{50} in a similar micromolar range to miltefosine (Table 2, entries 3 *vs* 17 and 10 *vs* 22). The parallel behavior was maintained also for trisubstituted derivatives **5ak**/**5bk**, which were both inactive under our bioassay conditions (Table 2, entries 11 *vs* 23). However, there were significant differences in the 2-(hetero)aroylpyrroles derivatives with halogenated aromatic rings. In particular, in the pyridine series, **5bd** (R = F) was found to be more active and selective than the drug of reference (miltefosine) (IC_{50} = 16.87 ± 0.73 μM, SI > 10.67), while the corresponding pyrimidine derivative **5ad** (R = F) was inactive (Table 2, entry 17 *vs* entry 4). It also should be pointed out that compound **5an**, where the phenyl ring had been changed to a naphthyl ring, showed similar activity to the drug of reference with better selectivity (Table 2, entry 14).

The same set of 2-(hetero)aroylpyrroles **5a,b** was also tested on promastigotes forms of *L. donovani* (Table 2). All compounds were considerably less active and selective than miltefosine. Halogenated pyridine derivatives **5bd−5bf** presented the best profiles, **5bd** again being the most active and selective of all 2-acylpyrroles (IC_{50} = 7.78 ± 0.27 μM and SI > 23.15). However, it should be highlighted that all tested pyrrole

https://doi.org/10.1021/acs.jcim.2c00731
J. Chem. Inf. Model. 2022, 62, 3928−3940

Table 4. Relative Biological Activity Score with Respect to Miltefosine

assay conditions c_j[a]				relative variation in biological activity score[b]						
c_0	c_1	c_2	c_3	$\Delta f(v_{ij})\%_{calc}$ values						
property (units)	target protein	cell line	assay org.	**5am**	**5af**	**5bc**	**5bd**	**5bp**	n_j	$f(v_{ij})_{ref}$
Top species predicted with model										
IC_{50} (nM)	MD	MD	*L. amazonensis*	1.70	−0.06	2.09	0.30	2.02	724	0
IC_{50} (nM)	MD	THP-1	*H. sapiens*	3.17	−0.12	3.89	0.56	3.77	546	0
Inh. (%)	MD	MD	*L. donovani*	3.17	−0.12	3.89	0.56	3.77	2654	0.01
IC_{50} (u)	MD	MD	*L. aethiopica*	3.13	−0.12	3.84	0.56	3.73	81	0.82
EC_{50} (nM)	MD	MD	*L. major*	3.13	−0.11	3.84	0.56	3.72	482	0.01
GI (%)	MD	MD	*L. tropica*	3.00	−0.11	3.68	0.53	3.57	279	0.51
IC_{50} (nM)	MD	MD	*L. chagasi*	2.79	−0.10	3.42	0.50	3.31	85	0
Top cell lines predicted with model										
ratio	MD	J774.A1	*MD*	1.32	−0.05	1.62	0.23	1.57	84	0.91
SI	MD	L6	*MD*	3.39	−0.12	4.16	0.60	4.03	181	0.89
IC_{50} (u)	MD	J774.A1	*L. donovani*	3.38	−0.12	4.15	0.60	4.02	144	0.82
SI	MD	J774	*MD*	3.23	−0.12	3.97	0.58	3.85	65	0.91
IC_{50} (nM)	MD	THP-1	*H. sapiens*	3.15	−0.12	3.87	0.56	3.75	637	0
SI	MD	LLC-MK2	*MD*	3.02	−0.11	3.71	0.54	3.60	61	0.91
SI	MD	Vero	*MD*	2.98	−0.11	3.65	0.53	3.54	51	0.91
SI	MD	HepG2	*MD*	1.41	−0.05	1.72	0.25	1.67	63	0.91
IC_{50} (u)	MD	J774.A1	*L. donovani*	3.38	−0.12	4.15	0.60	4.02	144	0.82
Top target proteins predicted with model										
K_i (nM)	P07382	MD	*L. major*	3.75	−0.14	4.60	0.67	4.46	48	0.01
IC_{50} (nM)	O61059	MD	*L. mexicana*	3.29	−0.12	4.03	0.59	3.91	372	0
IC_{50} (nM)	P11166	MD	*L. mexicana*	3.27	−0.12	4.01	0.58	3.89	13,658	0
IC_{50} (nM)	O97467	MD	*L. mexicana*	3.27	−0.12	4.01	0.58	3.89	13,642	0
IC_{50} (nM)	Q0GKD7	MD	*L. major*	2.95	−0.11	3.62	0.53	3.51	76	0
Inh. (%)	P39050	MD	*L. donovani*	2.81	−0.10	3.44	0.50	3.34	53	0.01
Act. (%)	Q27686	MD	*L. mexicana*	2.36	−0.09	2.90	0.42	2.81	62	0.1
K_i (nM)	Q01782	MD	*L. major*	1.52	−0.06	1.86	0.27	1.81	46	0.01
Inh. (%)	E9BF75	MD	*L. donovani*	1.32	−0.05	1.62	0.24	1.58	68,577	0.01

[a]Act.(%) = activity (%), Inh. (%) = inhibition (%), GI (%) = growth inhibition (%), Cyt. (%) = cytotoxicity (%),SI = selectivity index = ratio CC_{50}/IC_{50}, IC_{50} (u) = IC_{50} (ug mL^{-1}). [b]$\Delta f(v_{ij})\%_{calc} = 100\cdot[f(v_{ij})_{calc} - f(v_{mtfj})_{calc}]/f(v_{mtfj})_{calc}$, $f(v_{ij})_{calc}$ = calculated biological activity score of *i*th compound, $f(v_{mtfj})_{calc}$ = calculated biological activity score of drug of reference miltefosine, MD = missing data.

derivatives were less toxic than miltefosine with values of concentration of the compound that produces 50% reduction of cell viability (cytotoxic concentration, CC_{50}) in the range 87−401 μM in J774 cells. This is a promising result, taking into account high toxicity (low selectivity) of marketed available drugs.[3]

Then, one compound of each series was further tested *in vitro* on *L. amazonensis* and *L. donovani* amastigotes (Table 3). Pyrimidine derivative **5bc** showed good performance with an activity similar to miltefosine and better selectivity (IC_{50} = 60.55 ± 7.88 μM, SI > 5.19) against *L. amazonensis*. Nevertheless, pyridine derivative **5bc** presented bad results in terms of activity and selectivity (IC_{50} = 153.27 ± 9.11 μM, SI > 1.99).

3.1.5. IFPTML-Based Computational Screening of New Compounds. For this predictive study, we selected 28 compounds previously synthesized by our group (see structures on Scheme 1), whose *in vitro* biological activity (IC_{50} values) *vs* two *Leishmania* species (*L. donovani*[42] and *L. amazonensis*[43]) and cytotoxicity *vs* one cell line (J774 line of BALB/c mice macrophages[44]) has been carried out (Tables 2 and 3). However, there are >20 clinically relevant *Leishmania* species, such as *L. major*,[45] *L. mexicana*,[46] *L. aethiopica*,[47] *L. braziliensis*, *L. amazonensis*, *L. donovani*,[48] *L. infantum*,[49] *etc.* Therefore, it could be very interesting to know (a) other parameters (K_i, K_m, etc.) of *in vitro* biological activity *vs* specific target proteins and (b) the cytotoxicity of these compounds *vs* other human and animal cell lines such as Jurkat,[50] Vero,[51] THP-1,[52]HEK293,[53] HeLa,[54] HL-60,[55] Sf9,[56] *etc.* Consequently, we decided to use our multioutput IFPTML model to perform an in-depth computational screening of the biological activity of these compounds in all the biological assays space. Thus, we ran a computational screening experiment involving calculation of 20,704 activity scores for 29 compounds (28 compounds + miltefosine as reference) in 647 different preclinical assays. These 647 preclinical assays of reference present unique combinations of the biological assay conditions c_0 = parameter (K_i, IC_{50}, K_m, *etc.*), c_1 = target protein, c_2 = cell line (J774, HeLa, HL60, *etc.*), c_3 = organism (*L. major*, *L. mexicana*, *etc.*), or c_4 = organism stage of development.

The following steps were performed. First, DRAGON software[25] was used to calculate the entries of the vector of molecular descriptors for each compound. Next, the values of the molecular descriptors D_{ki} were substituted into the model, obtaining as output the scores of biological activity $f(v_{ij})_{calc}$ for the *i*th compound in the *j*th assay. Finally, the scores of biological activity $f(v_{ij})_{calc}$ were expressed in terms of relative deviation $\Delta f(v_{ij})\%_{calc} = 100\cdot[f(v_{ij})_{calc} - f(v_{mtfj})_{calc}]/f(v_{mtfj})_{calc}$. These relative scores express the deviation of the *i*th query compound from the reference, miltefosine (mtf). The predictions show, in general, a higher relative biological activity

https://doi.org/10.1021/acs.jcim.2c00731
J. Chem. Inf. Model. 2022, 62, 3928−3940

score $\Delta f(v_{ij})\%_{calc}$ for the compound series **5b** than for series **5a** compared to the drug of reference (miltefosine). Specifically, compounds **5bc** and **5bp** show 1−4 fold higher relative biological activity score values compared to miltefosine. The prediction is consistent with our experimental findings for IC_{50} activity assays *vs L. amazonensis* and *L. donovani* and selectivity index in the J774 cell line (Table 4).

Table 4 summarizes the results of scores of biological activity $f(v_{ij})_{calc}$ calculated for some of the 28 compounds (**5am**, **5af**, **5bc**, **5bd**, and **5bp**), considering the most relevant organisms, cell lines, and target proteins. Compound **5bd** was predicted with a positive value $\Delta f(v_{ij})\%_{calc} = 0.30$ for IC_{50} *vs L. amazonensis* promastigotes. It means that the model predicts this compound with high probability to be in the same range of activity than miltefosine. This result is in agreement with the experimental findings reported in the previous section, $IC_{50} = 16.87$ μM for compound **5bd** and $IC_{50} = 30.67$ for miltefosine *vs L. amazonensis* promastigotes. Compound **5bc** is also predicted with a positive value of $\Delta f(v_{ij})\%_{calc} = 2.09$ for IC_{50} *vs L. amazonensis* promastigotes, which also matches with our experimental finding ($IC_{50} = 30.87$ μM for compound **5bc** *vs* $IC_{50} = 30.67$ for miltefosine in *L. amazonensis* promastigotes assay). The same trend has been observed for other compounds (e.g., **5am**) that were also predicted with positive $\Delta f(v_{ij})\%_{calc}$ values approximately in the same range as miltefosine. Interestingly, compound **5af** is predicted with values $\Delta f(v_{ij})\%_{calc}$ lower than miltefosine (negative values of $\Delta f(v_{ij})\%_{calc}$) both *vs L. amazonensis* and *L. donovani* promastigotes, as observed in the experimental results (see Table 2).

In addition, the scores of biological activity of these compounds *vs* different cell lines were predicted. For this series, the cutoff for the scores of biological activity was 1.62 and 50 for n_j. Only one assay per cell line has been shown. First, we focused on the selectivity index (SI = ratio CC_{50}/IC_{50}) of the compounds *vs* J774 cell lines because they were the experimental lines used. Specifically, compound **5bc** has a value of $\Delta f(v_{ij})\%_{calc} = 1.62$ for cell line J774.A1, which means that this compound is expected to show a similar-to-higher probability than miltefosine of presenting positive SI, in agreement with our experimental results. In fact, compound **5bc** was found to have a SI >10.17, which is approximately 6 times the value for miltefosine with SI = 1.80. The model was able to reproduce, in general, the trends on SI values for all the compounds of both series for cell lines J774 and/or J774.A1. Similar results of positive SI were predicted with the IFPTML model for other cell lines not experimentally tested here. The higher values were calculated for cell lines: L6, J774.A1, J774 THP-1, LLC-MK2, Vero, and HepG2. This points to these lines as interesting targets for further testing the safety of these compounds in the future.

Finally, those proteins with the higher increase in biological activity score from the reference miltefosine were selected. In addition, proteins were filtered by the number of assays (n_j) reported in the ChEMBL dataset for each protein. The proteins with higher n_j are the most studied and probably the most relevant due to the increased attention they are receiving. To select the cases that include the most relevant proteins, the cutoff for the scores of biological activity was $\Delta f(v_{ij})\%_{calc} = 1.62$ and the cutoff for $n_j = 45$. Only one assay per protein is shown. According to the results obtained with the IFPTML model, the most plausible target proteins are the following: vifunctional dihydrofolate reductase-thymidylate synthase (P07382),[57] glucose transporter (O61059),[58] solute carrier family 2 facilitated glucose transporter member 1 (P11166),[59] hexose transporter 1 (Q0GKD7),[60] farnesyl pyrophosphate synthase (O97467),[61] trypanothione reductase (P39050),[62] pyruvate kinase (Q27686),[63] pteridine reductase 1 (Q01782),[64] and methionine-tRNA ligase (E9BF75).[65] After a detailed inspection of all compounds active *vs* these proteins in our ChEMBL database, no significant similarity was found between the previously reported compounds and the 2-acylpyrrole derivatives tested here. Consequently, 2-acylpyrrole derivatives could be considered a new class of antileishmanial lead compounds that deserve further investigation. The detailed results of the computational study are released in Supporting Information 3.

4. CONCLUSIONS

We have shown that SOFT.PTML is a useful tool for developing predictive models for drug discovery. The software implements IFPTML algorithms in a user-friendly interface without the need to rely on multiple software to run the different stages (IF, PT, and ML) of the algorithm. More importantly, SOFT.PTML can process complex datasets with big data features (high volume, multiple outputs, multiple target proteins, cell lines, pathogen species, missing data, *etc.*). Specifically, the use of this software has been illustrated by processing a very large ChEMBL dataset (>145,000 cases) from preclinical assays against different *Leishmania* species. Among the different ML algorithms (SVM, LOGR, and RF) explored, the best model was IFPTML-LOGR, which estimates the probability with which multiple parameters (IC_{50}, CC_{50}, SI, *etc.*) of a new compound get to a desired level in pre-clinical assays with high specificity and sensitivity (80−98%) in both training and validation series. This result demonstrates that the all-in-one strategy implemented in SOFT.PTML is capable of reproducing the results obtained from the multisoftware strategy, using a single program with a user-friendly interface that makes the work noticeably easier and faster. The pre-clinical assays studied involve different *Leishmania* species and cell lines, as well as multiple target proteins. The use of the new tool has been illustrated in a practical case study, the 2-acylpyrrole derivatives. The *in vitro* evaluation of the leishmanicidal activity of 2-acylpyrrole series **5a** and **5b** against visceral (*L. donovani*) and cutaneous (*L. amazonensis*) leishmaniasis revealed that all tested 2-acylpyrroles showed very low cytotoxicity, $CC_{50} > 100$ μg/mL in J774 cells (highest tested dose). This is an important feature as drug toxicity is one of the main limitations of current chemotherapy for leishmaniasis. In particular, **5bd** ($IC_{50} = 16.87$ μM, SI > 10.67) was approximately 6-fold more potent and selective than the drug of reference (miltefosine) in *L. amazonensis* promastigote assays. These results point to 2-acylpyrroles as a new class of lead compounds worthy of further optimization as antileishmanial hits.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.2c00731.

ChEMBL dataset used to train and validate the model, compounds codes, SMILE codes, preclinical assay conditions, observed values, predicted classifications, probabilities, *etc.* (XLSX)

Detailed information of the biological activity of the more interesting compounds, including the compound code, concentration, repeated measures of biological activity, average values; graphic representations of dose−effect

https://doi.org/10.1021/acs.jcim.2c00731
J. Chem. Inf. Model. 2022, 62, 3928−3940

curves for these compounds and the drug of reference, miltefosine. (XLSX)

Results of the computational study of the new series of 2-acylpyrrole derivatives (XLSX).

AUTHOR INFORMATION

Corresponding Authors

Humberto González-Díaz − *Departamento de Química Orgánica e Inorgánica, Facultad de Ciencia y Tecnología, Universidad del País Vasco / Euskal Herriko Unibertsitatea UPV/EHU, 48080 Bilbao, Spain; BIOFISIKA. Basque Center for Biophysics CSIC-UPV/EHU, 48940 Bilbao, Spain; IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain;* orcid.org/0000-0002-9392-2797; Email: humberto.gonzalezdiaz@ehu.es

Nuria Sotomayor − *Departamento de Química Orgánica e Inorgánica, Facultad de Ciencia y Tecnología, Universidad del País Vasco / Euskal Herriko Unibertsitatea UPV/EHU, 48080 Bilbao, Spain;* orcid.org/0000-0003-3079-6380; Email: nuria.sotomayor@ehu.es

Esther Lete − *Departamento de Química Orgánica e Inorgánica, Facultad de Ciencia y Tecnología, Universidad del País Vasco / Euskal Herriko Unibertsitatea UPV/EHU, 48080 Bilbao, Spain;* orcid.org/0000-0001-8624-6842; Email: esther.lete@ehu.es

Authors

Carlos Santiago − *Departamento de Química Orgánica e Inorgánica, Facultad de Ciencia y Tecnología, Universidad del País Vasco / Euskal Herriko Unibertsitatea UPV/EHU, 48080 Bilbao, Spain*

Bernabé Ortega-Tenezaca − *Department of Computer Science and Information Technologies, University of A Coruña (UDC), 15071 A Coruña, Spain*

Iratxe Barbolla − *Departamento de Química Orgánica e Inorgánica, Facultad de Ciencia y Tecnología, Universidad del País Vasco / Euskal Herriko Unibertsitatea UPV/EHU, 48080 Bilbao, Spain; BIOFISIKA. Basque Center for Biophysics CSIC-UPV/EHU, 48940 Bilbao, Spain*

Brenda Fundora-Ortiz − *Departamento de Química Orgánica e Inorgánica, Facultad de Ciencia y Tecnología, Universidad del País Vasco / Euskal Herriko Unibertsitatea UPV/EHU, 48080 Bilbao, Spain*

Sonia Arrasate − *Departamento de Química Orgánica e Inorgánica, Facultad de Ciencia y Tecnología, Universidad del País Vasco / Euskal Herriko Unibertsitatea UPV/EHU, 48080 Bilbao, Spain;* orcid.org/0000-0003-2601-5959

María Auxiliadora Dea-Ayuela − *Departamento de Farmacia, Facultad de Ciencias de la Salud, Universidad CEU Cardenal Herrera, 46115 Alfara del Patriarca, Valencia, Spain*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.2c00731

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

The Supporting Information including all the data used in this paper is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.xxxxxx. The software SOFT.PTML used in this paper is also available upon request to corresponding author at humberto.gonzalezdiaz@ehu.eus. Not proprietary data is reported on this work. Materials and Methods describe all the theory at a level that allows a person skilled in the art could implement the method.

ACKNOWLEDGMENTS

Ministerio de Ciencia e Innovación (PID2019-104148GB-I00) and Gobierno Vasco (IT1558-22) are gratefully acknowledged for their financial support. I.B. wishes to thank Fundación Biofísica Bizkaia/Biofisika Bizkaia Fundazioa (FBB) for a postdoctoral grant funded by BERC Basque Government program. Technical and human support provided by Servicios Generales de Investigación SGIker (UPV/EHU, MINECO, GV/EJ, ERDF and ESF) is also acknowledged.

REFERENCES

(1) World Health Organization (WHO): Global leishmaniasis surveillance: 2019-2020, a baseline for the 2030 roadmap. https://www.who.int/publications/i/item/who-wer9635-401-419 (accessed 07/07/2022)

(2) Gupta, O.; Pradhan, T.; Bhatia, R.; Monga, V. Recent advancements in anti-leishmanial research: Synthetic strategies and structural activity relationships. *Eur. J. Med. Chem.* **2021**, No. 113606.

(3) Brindha, J.; Balamurali, M. M.; Kaushik, C. An Overview on the Therapeutics of Neglected Infectious Diseases-Leishmaniasis and Chagas Diseases. *Front. Chem.* **2021**, *9*, No. 622286.

(4) Jones, N. G.; Catta-Preta, C. M. C.; Lima, A. P. C. A.; Mottram, J. C. Genetically Validated Drug Targets in *Leishmania*: Current Knowledge and Future Prospects. *ACS Infect. Dis.* **2018**, *4*, 467−477.

(5) Nagle, A. S.; Khare, S.; Kumar, A. B.; Supek, F.; Buchynskyy, A.; Mathison, C. J. N.; Chennamaneni, N. K.; Pendem, N.; Buckner, F. S.; Geb, M. H.; Moleteni, V. Recent Developments in Drug Discovery for Leishmaniasis and Human African Trypanosomiasis. *Chem. Rev.* **2014**, *114*, 11305−11347.

(6) Vitaku, E.; Smith, D. T.; Njardarson, J. T. Analysis of the structural diversity, substitution patterns, and frequency of nitrogen heterocycles among U.S. FDA approved pharmaceuticals. *J. Med. Chem.* **2014**, *57*, 10257−10274.

(7) Papireddy, K.; Smilkstein, M.; Kelly, J. X.; Salem, S. M.; Alhamadsheh, M.; Haynes, S. W.; Challis, G. L.; Reynodls, K. A. Antimalarial Activity of Natural and Synthetic Prodiginines. *J. Med. Chem.* **2011**, *54*, 5296−5306.

(8) Parra, L. L. L.; Bertonha, A. F.; Severo, I. R. M.; Aguiar, A. C. C.; de Souza, G. E.; Oliva, G.; Guido, R. V. C.; Grazzia, N.; Costa, T. R.; Miguel, D. C.; Gadelha, F. R.; Ferreira, A. G.; Hajdu, E.; Romo, D.; Berlinck, R. G. S. Isolation, Derivative Synthesis, and Structure-Activity Relationships of Antiparasitic Bromopyrrole Alkaloids from the Marine Sponge *Tedania brasiliensis*. *J. Nat. Prod.* **2018**, *81*, 188−202.

(9) Zhang, Z.; Ray, S.; Imlay, L.; Callaghan, L. T.; Niederstrasser, H.; Mallipeddi, P. L.; Posner, B. A.; Wetzel, D. M.; Phillips, M. A.; Smith, M. W. Total synthesis of (+)-spiroindimicin A and congeners unveils their antiparasitic activity. *Chem. Sci.* **2021**, *12*, 10388−10394.

(10) Albino, S. L.; da Silva, J. M.; de C Nobre, M. S.; de M e Silva, Y.; Santos, M. B.; de Araujo, R. S.; do CA de Lima, M.; Schmitt, M.; de Moura, R. O. Bioprospecting of Nitrogenous Heterocyclic Scaffolds with Potential Action for Neglected Parasitosis: A Review. *Curr. Pharm. Des.* **2020**, *26*, 4112−4150.

(11) Allocco, J. J.; Donald, R.; Zhong, T.; Lee, A.; Tang, Y. S.; Hendrickson, R. C.; Liberator, P.; Nare, B. Inhibitors of casein kinase 1 block the growth of *Leishmania major* promastigotes *in vitro*. *Int. J. Parasitol.* **2006**, *36*, 1249−1259.

(12) Baiocco, P.; Poce, G.; Alfonso, S.; Cocozza, M.; Porretta, G. C.; Colotti, G.; Biava, M.; Moraca, F.; Botta, M.; Yardley, V.; Fiorillo, A.; Lantella, A.; Malatesta, F.; Ilari, A. Inhibition of *Leishmania infantum* Trypanothione Reductase by Azole-Based Compounds: a Comparative

Analysis with Its Physiological Substrate by X-ray Crystallography. *ChemMedChem* **2013**, *8*, 1175−1183.
(13) Rizvi, S. U. F.; Siddiqui, H. L.; Parvez, M.; Ahmad, M.; Siddiqui, W. A.; Yasinzai, M. M. Antimicrobial and Anti-leishmanial Studies of Novel (2*E*)-3-(2-Chloro-6-methyl/methoxyquinolin-3-yl)-1-(Aryl)-prop-2-en-1-ones. *Chem. Pharm. Bull.* **2010**, *58*, 301−306.
(14) Vieira, P. S.; Souza, T. D. A. C. B.; Honorato, R. V.; Zanphorlin, L. M.; Severiano, K. L.; Rocco, S. A.; de Oliveira, A. H. C.; Cordeiro, A. T.; Oliveira, P. S. L.; de Giuseppe, P. O.; Murakami, M. T. Pyrrole-indolinone SU11652 targets the nucleoside diphosphatekinase from *Leishmania* parasites. *Biochem. Biophys. Res. Commun.* **2017**, *488*, 461−465.
(15) Santiago, C.; Rubio, I.; Sotomayor, N.; Lete, E. Selective Pd^{II} Catalyzed Acylation of Pyrrole with Aldehydes. Application to the Synthesis of Celastramycin Analogues and Tolmentin. *Eur. J. Org. Chem.* **2020**, 4284−4295.
(16) Barbolla, I.; Hernández-Suárez, L.; Quevedo-Tumailli, V.; Nocedo-Mena, D.; Arrasate, S.; Dea-Ayuela, M. A.; González-Díaz, H.; Sotomayor, N.; Lete, E. Palladium-mediated synthesis and biological evaluation of C-10b substituted dihydropyrrolo[1,2-*b*]-isoquinolines as anti-leishmanial agents. *Eur. J. Med. Chem.* **2021**, *220*, No. 113458.
(17) Simón-Vidal, L.; García-Calvo, O.; Oteo, U.; Arrasate, S.; Lete, E.; Sotomayor, N.; González-Díaz, H. Perturbation-Theory and Machine Learning (PTML) Model for High-Throughput Screening of Parham Reactions: Experimental and Theoretical Studies. *J. Chem. Inf. Model.* **2018**, *58*, 1384−1396.
(18) Díez-Alarcia, R.; Yañez-Pérez, V.; Muneta-Arrate, I.; Arrasate, S.; Lete, E.; Meana, J. J.; González-Díaz, H. Big Data Challenges Targeting Proteins in GPCR Signaling Pathways; Combining PTML-ChEMBL Models and [35S]GTPgammaS Binding Assays. *ACS Chem. Neurosci.* **2019**, *10*, 4476−4491.
(19) Nocedo-Mena, D.; Cornelio, C.; Camacho-Corona, M. R.; Garza-González, E.; Waksman, N. H.; Arrasate, S.; Sotomayor, N.; Lete, E.; González-Díaz, H. Modeling Antibacterial Activity with Machine Learning and Fusion of Chemical Structure Information with Microorganism Metabolic Networks. *J. Chem. Inf. Model.* **2019**, *59*, 1109−1120.
(20) Santana, R.; Zuluaga, R.; Gañán, P.; Arrasate, S.; Onieva, E.; González-Díaz, H. Predicting coated-nanoparticle drug release systems with perturbation-theory machine learning (PTML) models. *Nanoscale* **2020**, *12*, 13471−13483.
(21) Ambure, P.; Halder, A. K.; González-Díaz, H.; Cordeiro, M. N. D. S. QSAR-Co: An Open Source Software for Developing Robust Multitasking or Multitarget Classification-Based QSAR Models. *J. Chem. Inf. Model.* **2019**, *59*, 2538−2544.
(22) Ortega-Tenezaca, B.; González-Díaz, H. IFPTML mapping of nanoparticle antibacterial activity vs. pathogen metabolic networks. *Nanoscale* **2021**, *13*, 1318−1330.
(23) Ortega-Tenezaca, B.; Quevedo-Tumailli, V.; Bediaga, H.; Collados, J.; Arrasate, S.; Madariaga, G.; Munteanu, C. R.; Cordeiro, M. N. D. S.; González-Díaz, H. PTML Multi-Label Algorithms: Models, Software, and Applications. *Curr. Top. Med. Chem.* **2020**, *20*, 2326−2337.
(24) Martínez-Arzate, S. G.; Tenorio-Borroto, E.; Barbabosa-Pliego, A.; Díaz-Albiter, H. M.; Vázquez-Chagoyán, J. C.; González-íaz, H. PTML Model for Proteome Mining of B-Cell Epitopes and Theoretical-Experimental Study of Bm86 Protein Sequences from Colima, Mexico. *J. Proteom. Res.* **2017**, *16*, 4093−4103.
(25) Mauri, A.; Consonni, V.; Pavan, M.; Todeschini, R. Dragon software: An easy approach to molecular descriptor calculations. *Match* **2006**, *56*, 237−248.
(26) Evans, D. A. History of the Harvard ChemDraw Project. *Angew. Chem., Int. Ed.* **2014**, *53*, 11140−11145.
(27) Bilbao-Ramos, P.; Galiana-Roselló, C.; Dea-Ayuela, M. A.; González-Alvarez, M.; Vega, C.; Rolón, M.; Pérez-Serrano, J.; Bolás-Fernández, F.; González-Rosende, M. E. Nuclease activity and ultrastructural effects of new sulfonamides with anti-leishmanial and trypanocidal activities. *Parasitol. Int.* **2012**, *61*, 604−613.
(28) Dea-Ayuela, M. A.; Bilbao-Ramos, P.; Bolás-Fernández, F.; González-Cardenete, M. A. Synthesis and anti-leishmanial activity of C7- and C12-functionalized dehydroabietylamine derivatives. *Eur. J. Med. Chem.* **2016**, *121*, 445−450.
(29) Bilbao-Ramos, P.; Sifontes-Rodríguez, S.; Dea-Ayuela, M. A.; Bolás-Fernández, F. A fluorometric method for evaluation of pharmacological activity against intracellular *Leishmania* amastigotes. *J. Microbiol. Meth.* **2012**, *89*, 8−11.
(30) Galiana-Roselló, C.; Bilbao-Ramos, P.; Dea-Ayuela, M. A.; Rolón, M.; Vega, C.; Bolás-Fernández, F.; García-España, E.; Alfonso, J.; Coronel, C.; González-Rosende, M. E. *In vitro* and *in vivo* anti-leishmanial and trypanocidal studies of new *N*-benzene- and *N*-naphthalenesulfonamide derivatives. *J. Med. Chem.* **2013**, *56*, 8984−8998.
(31) Cortes, C.; Vapnik, V. N. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273−297.
(32) Urbanowicz, R. J.; Moore, J. H. Learning Classifier Systems: A Complete Introduction, Review, and Roadmap. *J. Artif. Evol. Appl.* **2009**, 736398.
(33) Bzdok, D.; Altman, N.; Krzywinski, M. Statistics versus Machine Learning. *Nat. Methods* **2018**, *15*, 233−234.
(34) Hoffman, R.; Minkin, V. I.; Carpenter, B. K. Ockham's Razor and Chemistry. *Int. J. Phil. Chem.* **1997**, *3*, 3−28.
(35) Rabinowitz, M.; Myers, L.; Banjevic, M.; Chan, A.; Sweetkind-Singer, J.; Haberer, J.; McCann, K.; Wolkowicz, R. Accurate prediction of HIV-1 drug response from the reverse transcriptase and protease amino acid sequences using sparse models created by convex optimization. *Bioinformatics* **2006**, *22*, 541−549.
(36) Kumar, P.; Dutta, S.; Kumar, S.; Bahadur, V.; Van der Eycken, E. V.; Vimaleswaran, K. S. V.; Parmar, V. S.; Singh, B. K. Aldehydes: magnificent acyl equivalents for direct acylation. *Org. Biomol. Chem.* **2020**, *18*, 7987−8033.
(37) Gensch, T.; Hopkinson, M. N.; Glorius, F.; Wencel-Delord, J. Mild metal-catalyzed C−H activation: examples and concepts. *Chem. Soc. Rev.* **2016**, *45*, 2900−2936.
(38) Santiago, C.; Sotomayor, N.; Lete, E. Pd(II)-Catalyzed C-H Acylation of (Hetero)arenes-Recent Advances. *Molecules* **2020**, *25*, 3247.
(39) Hunjan, M. K.; Panday, S.; Gupta, A.; Bhaumik, J.; Das, P.; Laha, J. K. Recent Advances in Functionalization of Pyrroles and their Translational Potential. *Chem. Rec.* **2021**, *21*, 715−780.
(40) Sun, X.; Dong, X.; Liu, H.; Liu, Y. Recent Progress in Palladium-Catalyzed Radical Reactions. *Adv. Synth. Catal.* **2021**, *363*, 1527−1558.
(41) Joshi, A.; Iqbal, Z.; Jat, J. L.; De, S. R. Pd(II)-Catalyzed Chelation-Induced C(sp2)-H Acylation of (Hetero)Arenes Using Toluenes as Aroyl Surrogate. *ChemistrySelect* **2021**, *6*, 12383−12406.
(42) Blanchard, P.; Dodic, N.; Fourrery, J. L.; Lawrence, F.; Mouna, A. M.; Robert-Gero, M. Synthesis and biological activity of sinefungin analogs. *J. Med. Chem.* **1991**, *34*, 2798−2803.
(43) Magaraci, F.; Jimenez, C.; Rodrigues, C.; Rodrigues, J. C. F.; Vianna Braga, M.; Yardley, V.; de Luca-Fradley, K.; Croft, S. L.; de Souza, W.; Ruiz-Perez, L. M.; Urbina, J.; Gonzalez Pacanowska, D.; Gilbert, I. H. Azasterols as inhibitors of sterol 24-methyltransferase in *Leishmania* species and *Trypanosoma cruzi*. *J. Med. Chem.* **2003**, *46*, 4714−4727.
(44) Hu, L.; Arafa, R. K.; Ismail, M. A.; Wenzler, T.; Brun, R.; Munde, M.; Wilson, W. D.; Nzimiro, S.; Samyesudhas, S.; Werbovetz, K. A.; Boykin, D. W. Azaterphenyl diamidines as anti-leishmanial agents. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 247−251.
(45) Booth, R. G.; Selassie, C. D.; Hansch, C.; Santi, D. V. Quantitative structure-activity relationship of triazine-antifolate inhibition of *Leishmania* dihydrofolate reductase and cell growth. *J. Med. Chem.* **1987**, *30*, 1218−1224.
(46) Bressi, J. C.; Verlinde, C. L.; Aronov, A. M.; Shaw, M. L.; Shin, S. S.; Nguyen, L. N.; Suresh, S.; Buckner, F. S.; Van Voorhis, W. C.; Kuntz, I. D.; Hol, W. G.; Gelb, M. H. Adenosine analogues as selective inhibitors of glyceraldehyde-3-phosphate dehydrogenase of *Trypanosomatidae* via structure-based drug design. *J. Med. Chem.* **2001**, *44*, 2080−2093.

https://doi.org/10.1021/acs.jcim.2c00731
J. Chem. Inf. Model. 2022, 62, 3928−3940

(47) Mäntylä, A.; Garnier, T.; Rautio, J.; Nevalainen, T.; Vepsälainen, J.; Koskinen, A.; Croft, S. L.; Järvinen, T. Synthesis, *in Vitro* Evaluation, and Anti-leishmanial Activity of Water-Soluble Prodrugs of Buparvaquone. *J. Med. Chem.* **2004**, *47*, 188−195.
(48) Del Olmo, E.; García Armas, M.; López-Pérez, J. L.; Muñoz, V.; Deharo, E.; San Feliciano, A. Leishmanicidal activity of some stilbenoids and related heterocyclic compounds. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2123−2126.
(49) Girault, S.; Grellier, P.; Berecibar, A.; Maes, L.; Mouray, E.; Lemière, P.; Debreu, M.-A.; Davioud-Charvet, E.; Sergheraert, C. Antimalarial, Antitrypanosomal, and Antileishmanial Activities and Cytotoxicity of Bis(9-amino-6-chloro-2-methoxyacridines): Influence of the Linker. *J. Med. Chem.* **2008**, *43*, 2646−2654.
(50) Gamage, S. A.; Figgitt, D. P.; Wojcik, S. J.; Ralph, R. K.; Ransijn, A.; Mauel, J.; Yardley, V.; Snowdon, D.; Croft, S. L.; Denny, W. A. Structure-activity relationships for the anti-leishmanial and antitrypanosomal activities of 1′-substituted 9-anilinoacridines. *J. Med. Chem.* **1997**, *40*, 2634−2642.
(51) Da Silva, L. E.; Joussef, A. C.; Kramer Pacheco, L.; Gaspar da Silva, D.; Steindel, M.; Andrade Rebelo, R.; Schmidt, B. Synthesis and *in vitro* evaluation of leishmanicidal and trypanocidal activities of *N*-quinolin-8-yl-arylsulfonamides. *Bioorg. Med. Chem.* **2007**, *15*, 7553−7560.
(52) Coelho, A. C.; Messier, N.; Ouellette, M.; Cotrim, P. C. Role of the ABC transporter PRP1 (ABCC7) in pentamidine resistance in *Leishmania* amastigotes. *Antimicrob. Agents Chemother.* **2007**, *51*, 3030−3032.
(53) Changtam, C.; de Koning, H. P.; Ibrahim, H.; Sajid, M. S.; Gould, M. K.; Suksamrarn, A. Curcuminoid analogs with potent activity against *Trypanosoma* and *Leishmania* species. *Eur. J. Med. Chem.* **2010**, *45*, 941−956.
(54) Marhadour, S.; Marchand, P.; Pagniez, F.; Bazin, M.-A.; Picot, C.; Lozach, O.; Ruchaud, S.; Antoine, M.; Meijer, L.; Rachidi, N.; Le Pape, P. Synthesis and biological evaluation of 2,3-diarylimidazo[1,2-*a*]pyridines as anti-leishmanial agents. *Eur. J. Med. Chem.* **2012**, *58*, 543−556.
(55) Gehrke, S. S.; Pinto, E. G.; Steverding, D.; Pleban, K.; Tempone, A. G.; Hider, R. C.; Wagner, G. K. Conjugation to 4-aminoquinoline improves the anti-trypanosomal activity of Deferiprone-type iron chelators. *Bioorg. Med. Chem.* **2013**, *21*, 805−813.
(56) Syrjänen, L.; Vermelho, A. B.; de Almeida-Rodrigues, I.; Corte-Real, S.; Saonen, T.; Pan, P.; Vullo, D.; Parkkila, S.; Capasso, C.; Supuran, C. T. Cloning, characterization, and inhibition studies of a *β*-carbonic anhydrase from *Leishmania donovani chagasi*, the protozoan parasite responsible for leishmaniasis. *J. Med. Chem.* **2013**, *56*, 7372−7381.
(57) Ferrari, S.; Morandi, F.; Motiejunas, D.; Nerini, E.; Henrich, S.; Luciani, R.; Venturelli, A.; Lazzari, S.; Calò, S.; Gupta, S.; Hannaert, V.; Michels, P. A. M.; Wade, R. C.; Costi, M. P. Virtual Screening Identification of Nonfolate Compounds, Including a CNS Drug, as Antiparasitic Agents Inhibiting Pteridine Reductase. *J. Med. Chem.* **2011**, *54*, 211−221.
(58) Burchmore, R. J.; Landfear, S. M. Differential regulation of multiple glucose transporter genes in *Leishmania Mexicana*. *J. Biol. Chem.* **1998**, *273*, 29118−29126.
(59) Mueckler, M.; Caruso, C.; Baldwin, S. A.; Panico, M.; Blench, I.; Morris, H. R.; Allard, W. J.; Lienhard, G. E.; Lodish, H. F. Sequence and structure of a human glucose transporter. *Science* **1985**, *229*, 941−945.
(60) Ghosh, S.; Chan, J. M. W.; Lea, C. R.; Meints, G. A.; Lewis, J. C.; Tovian, Z. S.; Flessner, R. M.; Loftus, T. C.; Bruchhaus, I.; Kendrick, H.; Croft, S. L.; Kemp, R. G.; Kobayashi, S.; Nozaki, T.; Oldfield, E. Effects of Bisphosphonates on the Growth of *Entamoeba histolytica* and *Plasmodium* Species *in Vitro* and *in Vivo*. *J. Med. Chem.* **2004**, *47*, 175−187.
(61) Louw, A. I. Molecular characterisation of facilitated transport proteins of the human malaria parasite. *S. Afr. J. Sci.* **1998**, *94*, 275−277.
(62) Shukla, A. K.; Patra, S.; Dubey, V. K. Iridoid glucosides from *Nyctanthes arbortristis* result in increased reactive oxygen species and cellular redox homeostasis imbalance in *Leishmania* parasite. *Eur. J. Med. Chem.* **2012**, *54*, 49−58.
(63) Nowicki, M. W.; Tulloch, L. B.; Worralll, L.; McNae, I. W.; Hannaert, V.; Michels, P. A. M.; Fothergill-Gilmore, L. A.; Walkinshaw, M. D.; Turner, N. J. Design, synthesis and trypanocidal activity of lead compounds based on inhibitors of parasite glycolysis. *Bioorg. Med. Chem.* **2008**, *16*, 5050−5061.
(64) Tulloch, L. B.; Martini, V. P.; Iulek, J.; Huggan, J. K.; Lee, J. H.; Gibson, C. L.; Smith, T. K.; Suckling, C.; Hunter, W. N. Structure-Based Design of Pteridine Reductase Inhibitors Targeting African Sleeping Sickness and the Leishmaniases. *J. Med. Chem.* **2010**, *53*, 221−229.
(65) Downing, T.; Imamura, H.; Decuypere, S.; Clark, T. G.; Coombs, G. H.; Cotton, J. A.; Hilley, J. D.; de Doncker, S.; Maes, I.; Mottram, J. C.; Quail, M. A.; Rijal, S.; Sanders, M.; Schönian, G.; Stark, O.; Sundar, S.; Vanaerschot, M.; Hertz-Fowler, C.; Dujardin, J.-C.; Berriman, M. Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res.* **2011**, *21*, 2143−2156.

https://doi.org/10.1021/acs.jcim.2c00731
J. Chem. Inf. Model. 2022, 62, 3928−3940