



QSAR Modeling for Multi-Target Drug Discovery: Designing Simultaneous Inhibitors of Proteins in Diverse Pathogenic Parasites

Valeria V. Kleandrova¹, Luciana Scotti², Francisco Jaime Bezerra Mendonça Junior³, Eugene Muratov⁴, Marcus T. Scotti^{2*} and Alejandro Speck-Planche^{2*}

¹Laboratory of Fundamental and Applied Research of Quality and Technology of Food Production, Moscow State University of Food Production, Moscow, Russian Federation, ²Postgraduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, João Pessoa, Brazil, ³Laboratory of Synthesis and Drug Delivery, State University of Paraíba, João Pessoa, Brazil, ⁴Laboratory for Molecular Modeling, The UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

OPEN ACCESS

Edited by:

Simona Rapposelli,
University of Pisa, Italy

Reviewed by:

Roberta Rocca,
University of Catanzaro, Italy
Marco Tutone,
University of Palermo, Italy

*Correspondence:

Marcus T. Scotti
mtscotti@gmail.com
Alejandro Speck-Planche
alejspivanovich@gmail.com

Specialty section:

This article was submitted to
Medicinal and
Pharmaceutical Chemistry,
a section of the journal
Frontiers in Chemistry

Received: 28 November 2020

Accepted: 22 January 2021

Published: 10 March 2021

Citation:

Kleandrova VV, Scotti L,
Bezerra Mendonça Junior FJ,
Muratov E, Scotti MT and
Speck-Planche A (2021) QSAR
Modeling for Multi-Target Drug
Discovery: Designing Simultaneous
Inhibitors of Proteins in Diverse
Pathogenic Parasites.
Front. Chem. 9:634663.
doi: 10.3389/fchem.2021.634663

Parasitic diseases remain as unresolved health issues worldwide. While for some parasites the treatments involve drug combinations with serious side effects, for others, chemical therapies are inefficient due to the emergence of drug resistance. This urges the search for novel antiparasitic agents able to act through multiple mechanisms of action. Here, we report the first multi-target model based on quantitative structure-activity relationships and a multilayer perceptron neural network (mt-QSAR-MLP) to virtually design and predict versatile inhibitors of proteins involved in the survival and/or infectivity of different pathogenic parasites. The mt-QSAR-MLP model exhibited high accuracy (>80%) in both training and test sets for the classification/prediction of protein inhibitors. Several fragments were directly extracted from the physicochemical and structural interpretations of the molecular descriptors in the mt-QSAR-MLP model. Such interpretations enabled the generation of four molecules that were predicted as multi-target inhibitors against at least three of the five parasitic proteins reported here with two of the molecules being predicted to inhibit all the proteins. Docking calculations converged with the mt-QSAR-MLP model regarding the multi-target profile of the designed molecules. The designed molecules exhibited drug-like properties, complying with Lipinski's rule of five, as well as Ghose's filter and Veber's guidelines.

Keywords: artificial neural network, fragment, mt-QSAR, multilayer perceptron, multi-target inhibitor, parasites, virtual design, docking

INTRODUCTION

Parasitic diseases are dangerous and prevalent health issues, causing high morbidities and mortalities worldwide. Among them, malaria, Chagas' disease (ChD), African animal trypanosomiasis (AAT), and toxoplasmosis, deserve special attention. From one side, malaria (mainly caused by *Plasmodium falciparum*), although one of the oldest illnesses known by mankind, and yet it remains the deadliest parasitic disease, being responsible for 445,000 deaths and 216 million cases in 2016 (WHO, 2017). On the other hand, we have ChD and AAT, which are the consequences of the infections caused by *Trypanosoma cruzi* and several species belonging to *Trypanosoma* spp. (including *Trypanosoma*

brucei brucei), respectively; while ChD continues to threaten millions of people in Mexico, as well as Central and South America (Molyneux et al., 2017; Perez-Molina and Molina, 2018), AAT causes great economic losses due to its devastating mortality on livestock (Giordani et al., 2016; Amisigo et al., 2019). In contrast to malaria, ChD, and AAT, whose negative impacts are located in specific continental areas, toxoplasmosis (caused by *Toxoplasma gondii*) has a worldwide distribution, infecting humans as well as most warm-blooded animals including mammals and birds (Robert-Gangneux and Darde, 2012). In fact, in developed countries such as the United States, toxoplasmosis infects over a million people each year, where this illness is associated with an estimated cost of \$3 billion (Aguirre et al., 2019).

In terms of treatment, the parasitic diseases mentioned here present several factors in common that make their eradication a challenge. First, current antiparasitic drugs are associated with many side effects (Forsyth et al., 2016; Grabias and Kumar, 2016; Kwofie et al., 2016; Alday and Doggett, 2017; Buckner et al., 2017; Haeusler et al., 2018). Second, drug resistance has emerged among these parasitic organisms, and consequently, antiparasitic drugs are becoming (or have already become) less effective (Baker et al., 2013; Campos et al., 2017; Montazeri et al., 2018; Conrad and Rosenthal, 2019). Last, as a whole, parasite-parasite interactions are very complex and have been documented and recognized as phenomena that play a crucial role in epidemiology, disease severity, and evolution of parasite virulence (Hellard et al., 2015; Seppala and Jokela, 2016; Dallas et al., 2019; Karvonen et al., 2019).

Screening chemicals through experimental validation is without doubt the most reliable way of identifying antiparasitic agents. However, this trial-and-error approach currently constitutes a time- and cost-ineffective task since the chemical space to be experimentally screened is vast (10^{60} small to medium-sized organic compounds) (Jahnke and Erlanson, 2006). In contrast, computational approaches can accelerate the search for efficacious antiparasitic chemicals, which can later be experimentally validated. At the biomolecular level, many promising computational models and protocols have demonstrated to be essential in early drug discovery, serving as tools for the generation of inhibitors against proteins whose roles are important for the survival or virulence of any of the parasitic species mentioned above. For instance, in the field of malaria research, recent works have focused on the application of an integrative multi-kinase approach (Lima et al., 2019), the identification of malarial allosteric modulators combining molecular dynamics simulations and dynamic residue network analysis (Amusengeri et al., 2019), the ensemble of ligand-based computational models for virtual screening of falcipain-2 inhibitors (Alberca et al., 2019), and quantitative structure-activity relationships (QSAR) for the study of N-myristoyltransferase inhibitors (Santos-Garcia et al., 2018). Regarding ChD, a wide range of *in silico* approaches have been reported to discover protein inhibitors with a special focus on cruzipain (Palos et al., 2017; Dos Santos et al., 2018; Herrera-Mayorga et al., 2019). Following with AAT, several works have reported the use of computational tools to

accelerate the search for inhibitors of different targets (Latorre et al., 2016; Di Pisa et al., 2017; Kimuda et al., 2019; Zacharova et al., 2019). Finally, the importance of computer-aided drug discovery has also been evidenced in the identification of different protein inhibitors to tackle toxoplasmosis (Welsch et al., 2016; Rosada et al., 2019; Zhang et al., 2019).

However, despite the growing influence of the computational methods in antiparasitic research, at least one of the three following drawbacks remains. First, computational models use relatively small datasets of structurally related molecules. Second, they lack sufficiently clear physicochemical and/or structural information to guide the design of new and potent protein inhibitors. Last, computational models have been based on only one therapeutic target/protein. All this urges the development of advanced computational models, suggesting that the efforts of the scientific community to speed up the eradication of diseases caused by the aforementioned parasites should focus on the multi-target drug discovery paradigm (Ravikumar and Aittokallio, 2018). In this context, several research groups have emphasized the development of a series of multi-target QSAR (mt-QSAR) models to perform virtual screening of molecules at both biomolecular- and microorganism-based levels (Prado-Prado et al., 2008; Prado-Prado et al., 2010a; Prado-Prado et al., 2010b; Garcia et al., 2011). Yet, no mechanistic, physicochemical, or structural interpretations have been reported for these models.

Currently, there is no computational approach capable of designing and predicting multi-target inhibitors of proteins present in different parasitic species. An *in silico* tool with such capabilities could take advantage of the fact that many parasitic proteins/targets identified to date are conserved across parasitic species (Cowell and Winzeler, 2019); a multi-target computational model would be of great value in both filtering the chemical space in the search for versatile inhibitors against diverse parasitic proteins and guiding the fast and accurate generation of new and potent antiparasitic agents able to act through different mechanisms of action.

Considering all the aforementioned ideas, we introduce here the first mt-QSAR model based on multilayer perceptron network (mt-QSAR-MLP), providing the theoretical foundations for the prediction of chemicals with potential multi-target activity against five parasitic proteins, namely plasmepsin 2 and dihydroorotate dehydrogenase (*P. falciparum*), as well as cruzipain (*T. cruzi*), dihydrofolate reductase (*T. gondii*), and glycylopeptide N-tetradecanoyltransferase (*T. brucei brucei*). Also, we computationally demonstrate that a series of newly designed molecules are worth synthesizing in the future by considering a combination of four factors: 1) they were rationally designed by assembling different molecular fragments according to the physicochemical and structural interpretation of the mt-QSAR-MLP model, 2) they were predicted by the mt-QSAR-MLP as potent multi-target inhibitors of the parasitic proteins, 3) the results of the docking calculations also converges with the predictions from the mt-QSAR-MLP model regarding the multi-target profile of the designed molecules, and 4) the designed molecules were estimated to have good synthetic accessibility.

MATERIALS AND METHODS

Database and Calculation of the Molecular Descriptors

The chemical and biological data were extracted from ChEBML (Gaulton et al., 2012) and contained information regarding the inhibitory potency, i.e., the concentration required to cause 50% inhibition (IC_{50}) in any of the five parasitic proteins mentioned above. The dataset was curated in terms of removing all the molecules with missing features such as SMILES, values, units of activity, and duplicates. The present dataset was formed by 2,249 different molecules, and each of them was experimentally tested against only one parasitic protein. In the dataset each molecule was classified as active [$IA_i(tg) = 1$] or inactive [$IA_i(tg) = -1$], with $IA_i(tg)$ being a binary variable that indicated the inhibitory activity of i th molecule against a defined target/protein. Thus, a molecule was annotated as active if $IC_{50} \leq 800$ nM for Plasmeprin 2 (*P. falciparum*), $IC_{50} \leq 820$ nM for dihydroorotate dehydrogenase (*P. falciparum*), $IC_{50} \leq 890$ nM for cruzipain (*T. cruzi*), $IC_{50} \leq 250$ nM for dihydrofolate reductase (*T. gondii*), or $IC_{50} \leq 270$ nM for glycylopeptide N-tetradecanoyltransferase (*T. brucei brucei*). In any other case, the molecules were considered inactive. It should be pointed out that the cutoff values selected in this study comply with two important aspects. From one side, by being in the submicromolar range, they ensure the rigorous search for potent hits, a process which, in most drug discovery campaigns usually starts at the micromolar range (Anderson, 2003). On the other hand, in general terms, these cutoff values prevent any excessive imbalance between the number of molecules assigned as active and those labeled as inactive. Finally, the selected cutoffs maintain the number of molecules annotated as active as high as possible; this increases the chemical diversity, which is required when using the mt-QSAR-MLP model to rationally design new molecules. Notice that if a unified cutoff value of the inhibitory activity is selected, then, at least one of two situations will happen: 1) data involving on or more of the parasitic proteins will be considerably imbalanced (reduced chemical diversity among active molecules) which is detrimental to the predictive power of any model, or 2) even if a unified cutoff is set, it will remarkably decrease the rigor of the mt-QSAR-MLP model to search for (and/or design) potent and versatile inhibitors against several parasitic proteins.

The SMILES codes of all the molecules reported in the dataset were stored in a file of type *.smi. This file was converted to *.sdf using the program Standardizer v19.18.0 (ChemAxon, 1998–2019). During the conversion process, as the purpose was to obtain the connectivity table for each molecule, no standardization actions were applied. Following, the computer program QuBiLS-MAS v1.0 (Valdés-Martini et al., 2012; Valdes-Martini et al., 2017) used the file *.sdf as the input for the calculation of the molecular descriptors known as total and local atom-based quadratic indices. When doing so, QuBiLS-MAS v1.0 performed these calculations by considering theoretical aspects such as the algebraic form (quadratic), constrains

(atom-based), matrix form (mutual probability). The quadratic indices mentioned here considered all the elements of the mutual probability matrix, and they used the Manhattan distance as the aggregator operator. The reason to select quadratic indices is based on their wide applicability as reported in several works focused on computer-aided drug discovery (Marrero-Ponce et al., 2011; Medina Marrero et al., 2015; Speck-Planche et al., 2015). The quadratic indices can be calculated according to the following mathematical formalism:

$$TmpAq_k(x) = \sum_{i=1}^n \sum_{j=1}^n {}^k mp_{ij \cdot x_i \cdot x_j} \quad (1)$$

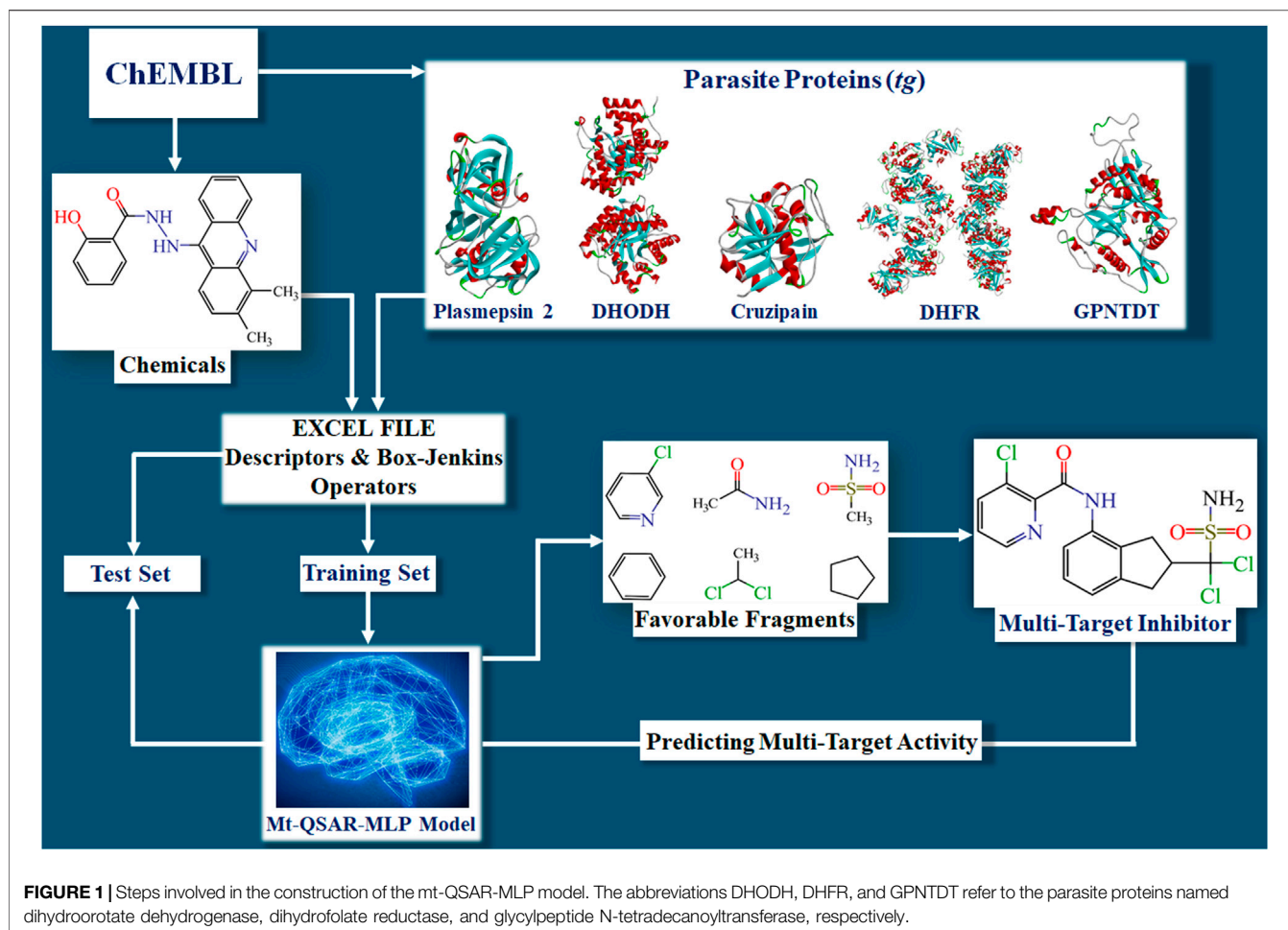
$$LmpAq_k(x)Z = \sum_{i=1}^n \sum_{j=1}^n {}^k mp_{ijz \cdot x_i \cdot x_j} \quad (2)$$

In Eqs. 1, 2, $TmpAq_k(x)$ and $LmpAq_k(x)Z$ represent the total and local atom-based quadratic indices of the mutual probability matrix, respectively. The symbol x refers to any atomic physicochemical property such as hydrophobicity (HYD), electronegativity (E), atomic weight (AW), polarizability (POL), polar surface area (PSA), or volume (V). It should be pointed out that while in Eq. 1 ${}^k mp_{ij}$ expresses the adjacency between any two atoms in a molecule, in Eq. 2, ${}^k mp_{ijz}$ has a similar meaning. Nevertheless, ${}^k mp_{ijz}$ depends on specific atoms types (Z) such as hydrogen bond acceptors, aliphatic and aromatic carbons, methyl groups, halogens, and heteroatoms. Both $TmpAq_k(x)$ and $LmpAq_k(x)Z$ describe a defined atom i and its chemical environment (formed by the j th neighbors) at the topological distance $d = k$.

The purpose here is to develop an mt-QSAR-MLP model as a computational tool able to predict inhibitory activity against dissimilar proteins present in diverse parasites. Thus, although the molecular descriptors calculated in Eqs. 1, 2 can characterize the chemical structure of the molecules, they will not be able to discriminate the structural and physicochemical information present in a molecule when this is tested against more than one target/protein. In this context, several works have applied an adaptation of the Box-Jenkins approach (used in time series analysis) to calculate multi-target molecular descriptors in a two-steps manner (Marzaro et al., 2011; Speck-Planche and Kleandrova, 2012a; Speck-Planche et al., 2012; Alonso et al., 2013; Speck-Planche et al., 2013; Romero Duran et al., 2014; Romero-Duran et al., 2016):

$$avgQI(tg) = \frac{1}{n(tg)} \times \sum_{a=1}^{n(tg)} QI_a \quad (3)$$

In Eq. 3, QI_a is any of the quadratic indices mentioned above. The symbol $avgQI(tg)$ represents the average of any quadratic index for all the molecules in the training set labeled as active and tested against the same parasite protein. Consequently, $n(tg)$ denotes the number of active molecules/cases (also present in the training set) that were assayed against the same protein. The second step applies the following formula:



$$DQI_a(tg) = \frac{QI_a - \text{avg}QI(tg)}{(QI_{MX} - QI_{MN}) \times \sqrt{p(tg)}} \quad (4)$$

In Eq. 4, $DQI_a(tg)$ is a multi-target descriptor and depends on the chemical structure of a molecule and the parasite protein against which that molecule was tested; this descriptor measures how much any molecule structurally deviates from a group of molecules assigned as active and assayed against the same protein. On the other hand, QI_{MX} and QI_{MN} are the maximum and minimum values of each quadratic index (in the training set), respectively. Last, $p(tg)$ is the *a priori* probability of finding a compound tested against a specific parasite protein; it is calculated as the ratio of the number of molecules in the training set assayed against a given protein to the total number of compounds present in the training set.

Building the Mt-QSAR-MLP Model

Developing the mt-QSAR-MLP occurred in different steps (Figure 1). First, the dataset containing the 2,249 molecules was split into training and test sets according to the following procedure. For each parasitic protein, the molecules were sorted according to their increasing IC_{50} values. Then, for each protein,

the first three molecules were assigned to the training set while the fourth molecule was assigned to the test set. Such a ratio of 3:1 was repeated in the whole dataset. Thus, the training set was employed to search for the best model and was formed by 1,691 molecules (75.19%), 788 considered as active and 903 annotated as inactive. The test set was meant to demonstrate the predictive power of the mt-QSAR-MLP model; this set contained 558 molecules (remaining 24.81% of the dataset), 259 assigned as active and 299 considered as inactive.

Second, it is known that the random forest (RF) is one of the most popular machine learning methods to obtain predictive models (Hastie et al., 2009). In this work, RF was used as a variable selection strategy. In this sense, and using the descriptors of the type $DQI_a(tg)$ as inputs, the RF package of the computer program software STATISTICA v13.5.0.17 (TIBCO-Software-Inc., 2018) was employed to perform multiple runs to find the best mt-QSAR-RF model. In doing so, we used default values for the different parameters in the RF algorithm [number of predictors: 259; number of trees: 100; subsample proportion: 0.5; seed for random number generator: 1; minimum number of cases: 56; minimum number in child node: 5; maximum number of levels: 10; maximum number of nodes: 100; cycles to calculate mean error: 10; percentage decrease in training error:

5]. While selecting the most influential descriptors (highest importance values) in the mt-QSAR-RF model, we conducted a correlation analysis for the molecular descriptors of the type $DQI_a(tg)$ by computing the Pearson's correlation coefficient (PCC) (Pearson, 1895); only the descriptors having pairwise correlation values in the interval $-0.7 < PCC < 0.7$ were chosen.

Artificial neural networks (ANNs) was used as the data analysis method to search for the best model, the architecture known as the multi-layer perceptron (MLP) were examined because of its popularity, accuracy, and relative ease of convergence. When training the MLP networks, we employed the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, setting the number of epochs to be 300. To obtain the most appropriate mt-QSAR-MLP model, several runs were performed using the ANNs package of STATISTICA v13.5.0.17 (TIBCO-Software-Inc., 2018) while inspecting the statistical indices known as accuracy [$Ac(\%)$] and Matthews' correlation coefficient (MCC) (Matthews, 1975), as well as sensitivity [$Sn(\%)$] and specificity [$Sp(\%)$] and their local counterparts [$Sn(\%)tg$] and [$Sp(\%)tg$]. It should be highlighted that while [$Sn(\%)$] and [$Sp(\%)$] give an idea of the global statistical quality (training set) and predictive power (test set) of the mt-QSAR-MLP model [$Sn(\%)tg$] and [$Sp(\%)tg$] provide similar information but depending on each of the five proteins reported in this work. Only the mt-QSAR-MLP model exhibiting the highest values of [$Sn(\%)$] [$Sp(\%)$] [$Sn(\%)tg$] and [$Sp(\%)tg$] was selected.

Molecular Docking

When performing docking calculations, we used the software Molegro Virtual Docker v6.0.1 (Thomsen and Christensen, 2006), employing the same protocol as recently reported in (Speck-Planche and Scotti, 2019). We retrieved all the crystallographic structures from the Protein Data Bank (PDB) (Berman et al., 2000). In doing so, we considered the PDB IDs 2BJU (Prade et al., 2005), 6I55 (Pippione et al., 2019), 1ME3 (Huang et al., 2003), and 4KY4 (Zaware et al., 2013) for the proteins plasmepsin 2 (*P. falciparum*), dihydroorotate dehydrogenase (*P. falciparum*), cruzipain (*T. cruzi*), and dihydrofolate reductase (*T. gondii*), respectively. These PDB files contained the aforementioned proteins complexed with their corresponding reference ligands. We validated the docking protocol by redocking each reference ligand into the active site of the protein for which the corresponding complex with that protein was experimentally reported. In the case of the protein glycylopeptide N-tetradecanoyltransferase (*T. brucei brucei*), no crystallographic structure has been reported to date. Therefore, we relied on homology modeling to create the 3D-structure of this protein. In this sense, we employed SWISS-MODEL (Waterhouse et al., 2018), which is fully automated protein homology modeling webserver. When performing homology modeling with SWISS-MODEL, we entered the UniprotID Q388H8, which corresponded to glycylopeptide N-tetradecanoyltransferase (*T. brucei brucei*). Then, SWISS-MODEL performed an automatic search for different proteins' amino acid sequences to use them as templates, selecting the most reliable model of the 3D-structure of glycylopeptide N-tetradecanoyltransferase (*T. brucei brucei*). Last all the

interactions for each ligand-protein complex were visualized by the Discovery Studio Visualizer v19.1 (BIOVIA, 2018).

RESULTS AND DISCUSSION

The Mt-QSAR-MLP Model

The best mt-QSAR-MLP model found by us had the profile MLP 9-27-2, which means that nine nodes [molecular descriptors of the type $DQI_a(tg)$] were used in the input layer, 27 nodes in the hidden layer with a logistic activation function, while in the output layer (based on a softmax function), the number two refers to the two possible categorical values (-1 and 1) of the variable of predicted inhibitory potency [$Pred_IA_i(tg)$]. Details regarding the different molecular descriptors used to build the mt-QSAR-MLP model appear in **Table 1**. At the same time, all the chemical and biological data can be gathered from **Supplementary Material S1**.

In terms of statistical quality, the mt-QSAR-MLP model exhibited $Ac(\%) = 84.68\%$, indicating that 1,432 out of 1,691 molecules were correctly classified. In the test set, also a good performance was achieved; 444 out of 558 molecules were correctly predicted [$Ac(\%) = 79.57\%$]. In addition to the values of accuracy mentioned here, relatively high values for the other statistical indices were obtained (**Table 2**). For instance, $Sn(\%)$ and $Sp(\%)$ had values higher than 83% in the training set, while for the test set, they exhibited values around 80%. Simultaneously, MCC took values higher than 0.59, and given their closeness to one (perfect performance) more than to zero (for a random predictor), it can be inferred that there is a strong correlation between the observed [$IA_i(tg)$] and predicted [$Pred_IA_i(tg)$] values of the inhibitory activity. For each molecule in the dataset information regarding its classification performed by the mt-QSAR-MLP model is reported in **Supplementary Material S1**.

Although any predictive model should have relatively high values of as $Sn(\%)$ and $Sp(\%)$, for the case of an mt-QSAR model, it is also important that the local sensitivities [$Sn(\%)tg$] and specificities [$Sp(\%)tg$] for each protein should also exhibit values as high as possible. In this context [$Sn(\%)tg$] and [$Sp(\%)tg$] were higher than 80% in the training set whereas, for the test set, values higher than 72% and 75% were computed for these two statistical indices, respectively. Details of the different [$Sn(\%)tg$] and [$Sp(\%)tg$] values are available in **Supplementary Material S2**. The only exception was [$Sp(\%)tg$] in the case of the protein dihydrofolate reductase (*T. gondii*) for which values of 69.92% and 65.82% for training and test sets, respectively. We attribute the wrong predictions to the fact that the molecular descriptors $DQI_a(tg)$ are not capable of considering all the differences in the chemical structures of the molecules which produce the corresponding changes in their inhibitory potency against the parasite proteins. This is another confirmation that the ability of the molecular descriptors reported to date to contain information on the complexity and diversity of the molecules is limited (Todeschini and Consonni, 2009). In any case, the joint analysis of the global statistical indices and the local sensitivities and specificities demonstrate the good statistical quality and predictive power of the mt-QSAR-MLP model.

TABLE 1 | Molecular descriptors and their definitions.

Molecular descriptor ^a	Definition
D[TmpAq2(HYD)]tg	Deviation of the total atom-based quadratic index (order 2) of the mutual probability matrix weighted by the <i>HYD</i> .
D[TmpAq5(PSA)]tg	Deviation of the total atom-based quadratic index (order 5) of the mutual probability matrix weighted by the <i>PSA</i> .
D[LmpAq6(POL)A]tg	Deviation of the local atom-based quadratic index (order 6) of the mutual probability matrix weighted by the <i>POL</i> . This descriptor considers only the hydrogen bond acceptors and their neighbor atoms
D[LmpAq0(PSA)A]tg	Deviation of the local atom-based quadratic index (order 0) of the mutual probability matrix weighted by the <i>PSA</i> . This descriptor considers only the presence of hydrogen bond acceptors
D[LmpAq6(PSA)A]tg	Deviation of the local atom-based quadratic index (order 6) of the mutual probability matrix weighted by the <i>PSA</i> . This descriptor considers only the hydrogen bond acceptors and their neighbor atoms
D[LmpAq6(POL)C]tg	Deviation of the local atom-based quadratic index (order 6) of the mutual probability matrix weighted by the <i>POL</i> . This descriptor considers only the aliphatic carbons and their neighbor atoms
D[LmpAq1(V)C]tg	Deviation of the local atom-based quadratic index (order 1) of the mutual probability matrix weighted by the atomic <i>V</i> . This descriptor considers only the aliphatic carbons and their adjacent atoms
D[LmpAq5(POL)P]tg	Deviation of the local atom-based quadratic index (order 5) of the mutual probability matrix weighted by the <i>POL</i> . This descriptor considers only the aromatic carbons and their neighbor atoms
D[LmpAq0(PSA)Y]tg	Deviation of the local atom-based quadratic index (order 0) of the mutual probability matrix weighted by the <i>PSA</i> . This descriptor considers only the presence of heteroatoms

^aAll these molecular descriptors depend on both the chemical structure of a molecule and the target/protein against which the molecule was experimentally assayed.

TABLE 2 | Internal quality and predictive performance of the mt-QSAR-MLP model.

SYMBOLS ^a	Training set	Test set
N_{Active}	788	259
CC_{Active}	676	208
Sn(%)	85.79%	80.31%
N_{Inactive}	903	299
CC_{Inactive}	756	236
Sp(%)	83.72%	78.93%
MCC	0.694	0.591

^a**N_{Active}**—Number of molecules annotated as active; **N_{Inactive}**—Number of molecules annotated as inactive; **CC_{Active}**—Molecules correctly classified as active; **CC_{Inactive}**—Molecules correctly classified as inactive; **Sn(%)**—Sensitivity (percentage of molecules correctly classified as active); **Sp(%)**—Specificity (percentage of molecules correctly classified as inactive); **MCC**—Matthews' correlation coefficient.

Applicability Domain

The assessment of the applicability domain (AD) of the mt-QSAR-MLP model was carried out by employing a modification of the descriptor space approach (Sahigara et al., 2012), which establishes that the maximum and minimum values of each molecular descriptor (in the training set) are the boundaries of the AD of a model. Here, we defined the maximum and minimum values of each $DQI_a(tg)$ descriptor in the mt-QSAR-MLP by considering only those molecules in the training set that were correctly classified (Speck-Planche, 2018). For each molecule present in the dataset, a local score of applicability domain for each of its $DQI_a(tg)$ descriptors was assigned. In this sense, if for a molecule, a given descriptor value was within the interval defined by the maximum and minimum values, the local score was equal to one; otherwise, the local score was equal to zero. This procedure was repeated for each $DQI_a(tg)$ descriptor in the mt-QSAR-MLP model. In the end, the sum of all the scores for each molecule was calculated, yielding the total score of the applicability domain (TSAD). Thus, as the mt-QSAR-MLP model was built from nine molecular descriptors, only the molecules with TSAD = 9

were considered to be within the AD (**Supplementary Material S3**).

Molecular Descriptors and Their Physicochemical and Structural Meanings

Interpreting any QSAR model is crucial for the understanding of the physicochemical properties and structural features that govern the enhancement (or the diminution) of the biological activity under study. To provide a more complete interpretation of the mt-QSAR-MLP model developed in this work, we have combined chemical reasoning, statistical aspects, a fragment-based analysis into a single explanation.

Chemical reasoning focuses on the fact that the $DQI_a(tg)$ descriptors employed to build the mt-QSAR-MLP model are characterized by two important elements. First, the topological distance $d = k$ [with k being the order of each $DQI_a(tg)$] expresses the number of bonds (without considering bond multiplicity) that exist between any two atoms in a molecule. Chemically speaking, by using this information, it is possible to know the regions in a molecule where atoms exhibiting certain physicochemical properties can be placed with respect to their neighbor atoms. Second, the $DQI_a(tg)$ descriptors also cover lower topological distances. For instance, a $DQI_a(tg)$ descriptor of order six will describe information at the topological distance equal to six but also at the topological distances of two and three. This is because $DQI_a(tg)$ descriptors also measure the degree of concentration of a physicochemical property at the topological distance $d \leq k$. Chemical reasoning will provide information in terms of the distributions of the atoms with different physicochemical properties throughout the entire structure of a molecule.

The statistical aspects focused on two elements, the relative importance of each quadratic index in the mt-QSAR-MLP. Such information was provided by carrying out a sensitivity analysis with the ANN package of STATISTICA v13.5.0.17. This permitted us to estimate the sensitivity values of the $DQI_a(tg)$ descriptors; the highest SVs corresponded to those which were the

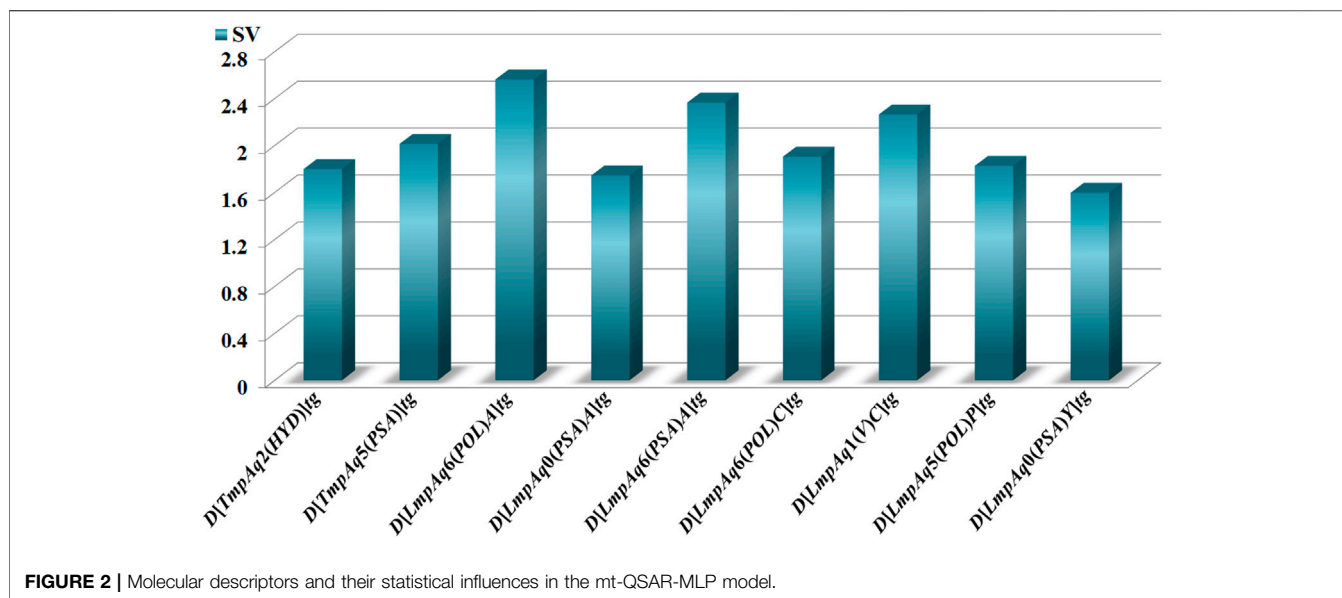


FIGURE 2 | Molecular descriptors and their statistical influences in the mt-QSAR-MLP model.

most influential in the mt-QSAR-MLP model (Figure 2). The other statistical element is the tendency of variation of the $DQI_a(tg)$ descriptors. We would like to emphasize that the model developed in this work is non-linear. Consequently, there is no equation from which the variation (increase or diminution) in the values of each $DQI_a(tg)$ descriptor can be determined. To solve this inconvenience, we applied the approach reported by Speck-Planche and co-workers (Speck-Planche and Kleandrova, 2012b; Speck-Planche, 2018; Speck-Planche, 2019). Basically, for each $DQI_a(tg)$ descriptor present in the mt-QSAR-MLP model, two average values were calculated: one for the molecules annotated as active and the other for the molecules assigned as inactive. It is important to highlight that the calculation of the two averages of each $DQI_a(tg)$ descriptor was carried out by considering only those molecules in the training set that were correctly classified by the mt-QSAR-MLP model. Comparing the two average values between each other offers the possibility of knowing how the value of a given $DQI_a(tg)$ descriptor should vary to enhance the biological effect under study, in this case, the multi-target activity against different parasite proteins. The class-based averages and the corresponding tendency of variation for each $DQI_a(tg)$ descriptor are reported in Table 3.

Regarding the fragment-based analysis, there is solid evidence that demonstrates that any topological (graph-based) descriptor calculated for a molecule can be expressed as the number of times in which different fragments (both connected and disconnected) appear in that molecule (Baskin et al., 1995). This means that the information content of any topological descriptor can be associated with a series of fragments. From a substructural point of view, the $DQI_a(tg)$ descriptors present in the mt-QSAR-MLP model constitute a class of topological descriptors, and therefore, while interpreting them, different fragments whose presence leads to favorable variations (responsible for increasing the inhibitory activity) of these $DQI_a(tg)$ descriptors can be

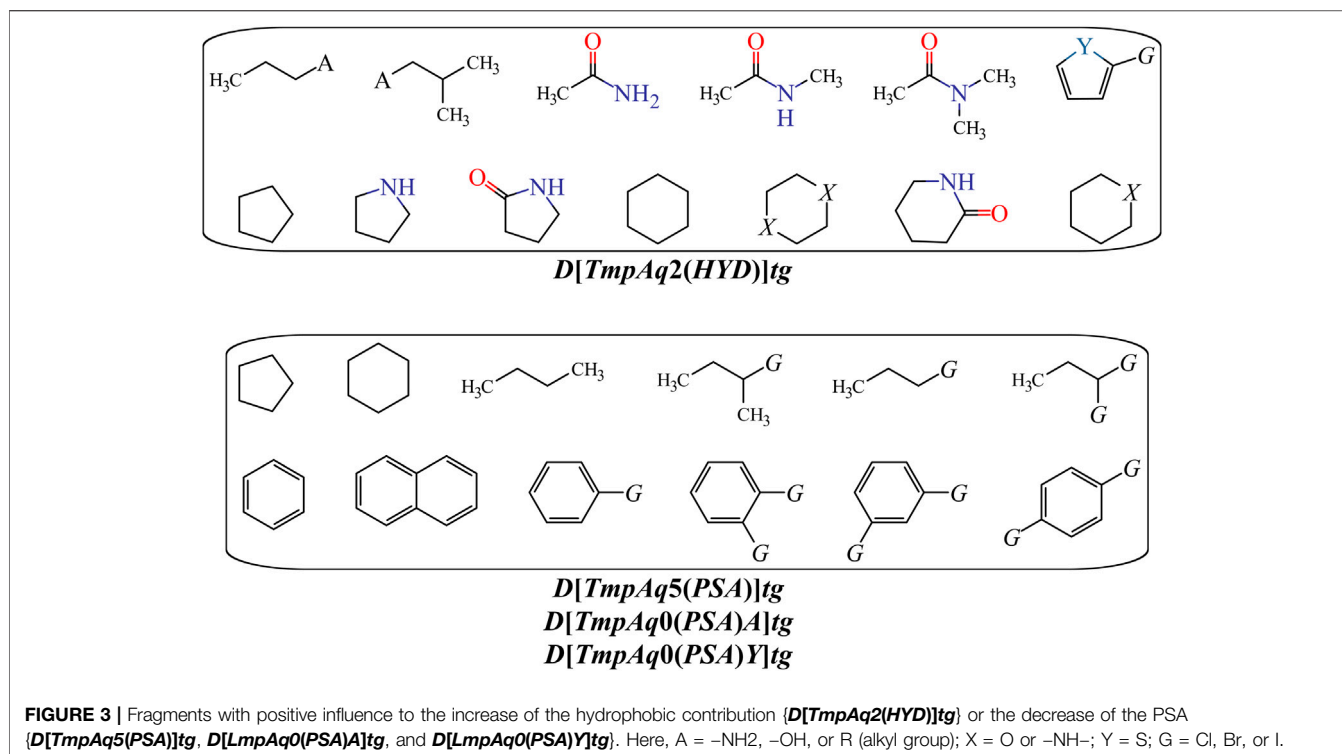
TABLE 3 | Tendencies of variation of the molecular descriptors in the mt-QSAR-MLP model according to the classes-based means' approach.

Symbol	Active	Inactive	Tendency ^a
$D[TmpAq2(HYD)]tg$	7.92×10^{-3}	-1.14×10^{-1}	Increase
$D[TmpAq5(PSA)]tg$	1.82×10^{-3}	3×10^{-2}	Decrease
$D[LmpAq6(POL)A]tg$	4.39×10^{-3}	-6.38×10^{-3}	Increase
$D[LmpAq0(PSA)A]tg$	-3.43×10^{-3}	9.63×10^{-2}	Decrease
$D[LmpAq6(PSA)A]tg$	3.09×10^{-3}	-5.63×10^{-3}	Increase
$D[LmpAq6(POL)C]tg$	-7.22×10^{-4}	2.93×10^{-2}	Decrease
$D[LmpAq1(V)C]tg$	2.32×10^{-3}	-5.77×10^{-2}	Increase
$D[LmpAq5(POL)P]tg$	4.84×10^{-3}	-4.53×10^{-2}	Increase
$D[LmpAq0(PSA)Y]tg$	6.14×10^{-4}	5.26×10^{-2}	Decrease

^aTendency—it indicates the type of variation (increase or diminution) of a descriptor in order to enhance the multi-target activity against the different parasite proteins.

extracted (Speck-Planche, 2019; Kleandrova and Speck-Planche, 2020).

We have $D[TmpAq2(HYD)]tg$ (the seventh most influential descriptor), which expresses the augmentation of the joint hydrophobic contribution (multiplication of the atomic hydrophobicity) of any two atoms placed at the topological distance of two. We would like to highlight that the atomic hydrophobicities used in this work are based on the hydrophobicity scale proposed by Ghose and co-workers (Ghose et al., 1998). According to this scale, aliphatic carbon atoms will have negative hydrophobicity values except for those of the type CHX3, CR2X2, CRX3, and CX4 (X is an electronegative atom such as O, N, S, P, Se, or any halogen). Nitrogen and oxygen atoms have also been reported to have negative hydrophobicity values; exceptions are pyrrolic nitrogen (or furan oxygen) atoms, nitrogen from amines (or oxygen from ethers) having attached two aromatic (or heteroaromatic) rings, and all the tertiary amines. That being said, it is clear that the presence of aliphatic amines and ethers, regardless of whether they are acyclic or cyclic) have favorable contributions to the



increase of *D*[*TmpAq2*(*HYD*)]*tg*. A non-exhaustive but useful list of suitable generic fragments is depicted in **Figure 3**.

In **Table 3** and **Figure 3**, we can see that the diminution of the *PSA* is governed by the descriptors *D*[*TmpAq5*(*PSA*)]*tg*, *D*[*LmpAq0*(*PSA*)*A*]*tg*, and *D*[*LmpAq0*(*PSA*)*Y*]*tg*, which are rank fourth, eighth, and ninth among the most significant descriptors, respectively. Particularly, *D*[*TmpAq5*(*PSA*)]*tg* considers the decrease of the *PSA* of any two atoms placed at the topological distance of five while *D*[*LmpAq0*(*PSA*)*A*]*tg* and *D*[*LmpAq0*(*PSA*)*Y*]*tg* indicate the reduction of the *PSA* depending on hydrogen bond acceptors and heteroatoms, respectively. Altogether, these three descriptors express that fragments containing aromatic rings (both unsubstituted and substituted) as well as aliphatic rings and chains are desirable for the favorable decrease of the *PSA*. An interesting fact is that in most of the molecules, the *PSA* strongly depends on the presence of nitrogen and oxygen atoms, which is characterized by both descriptors *D*[*LmpAq0*(*PSA*)*A*]*tg* and *D*[*LmpAq0*(*PSA*)*Y*]*tg*. Therefore, these two descriptors should correlate with each other. This, however, doesn't happen because *D*[*LmpAq0*(*PSA*)*Y*]*tg* also considers other atoms with *PSA* such as sulfur and phosphorus. In the database used to build the mt-QSAR-MLP model, there are many compounds with different functional groups containing sulfur, which is the main factor preventing the existence of a correlation between *D*[*LmpAq0*(*PSA*)*A*]*tg* and *D*[*LmpAq0*(*PSA*)*Y*]*tg*. In the end, the number of atoms with values of *PSA* different from zero should be kept as low as possible.

On the other hand, *DQI_a*(*tg*) descriptors such as *D*[*LmpAq1*(*V*)*C*]*tg*, *D*[*LmpAq6*(*POL*)*C*]*tg*, and *D*[*LmpAq5*(*POL*)*P*]*tg* describe the importance of controlling the steric factors

(**Figure 4**). Thus, *D*[*LmpAq1*(*V*)*C*]*tg* expresses the increase of property *V* of any two atoms (at least one of them being an aliphatic carbon) placed at the topological distance of one. This is the third most significant descriptor and its value can be increased by augmenting the number of aliphatic carbons in the molecule. In case that the number of aliphatic carbons is low, these atoms should be attached to others with relatively high bulkiness (e.g., Cl, Br, and I). In terms of the number of aliphatic carbons that should exist in the molecules, the descriptor *D*[*LmpAq6*(*POL*)*C*]*tg* constrains *D*[*LmpAq1*(*V*)*C*]*tg*. This is because *D*[*LmpAq6*(*POL*)*C*]*tg* (ranked fifth in terms of importance) involves the decrease of the *POL* of any two atoms (one of them being an aliphatic carbon) placed at the topological distance equal to six. Consequently, to decrease the value of this descriptor, the number of aliphatic carbons should be kept to a minimum, and/or the atoms placed at the topological distance of six (or lower) with respect to these aliphatic carbons should be preferably low-polarizability atoms such as fluorine, oxygen, and in less degree, nitrogen.

In the case of the descriptor *D*[*LmpAq5*(*POL*)*P*]*tg* (the sixth most influential descriptor), this characterizes the augmentation of the *POL* of any two atoms (at least one of them must be an aromatic carbon) which are placed at the topological distance of five. The value of this molecular descriptor can be increased by raising the number of aromatic carbons and/or placing bulky atoms such as halogens (except for fluorine) at topological distances of five or three with respect to the aromatic carbons.

Finally, **Figure 5** depicts different types of fragments; some of them have a positive influence on *D*[*LmpAq6*(*POL*)*A*]*tg* while others favorably augment the value of *D*[*LmpAq6*(*PSA*)*A*]*tg*. In this sense, the descriptor *D*[*LmpAq6*(*POL*)*A*]*tg* is the most

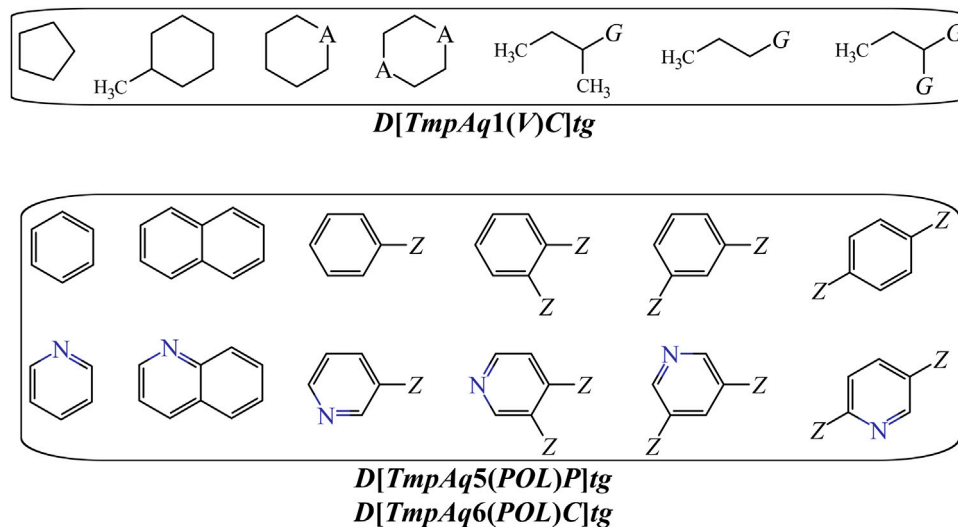


FIGURE 4 | Substructures exhibiting with positive contributions to the desirable increase of the V ($D[\text{LmpAq1}(\text{V})\text{C}]_{\text{tg}}$) or the favorable variation of the POL ($D[\text{LmpAq5}(\text{POL})\text{P}]_{\text{tg}}$ and $D[\text{LmpAq6}(\text{POL})\text{C}]_{\text{tg}}$). Here, A = $-\text{CH}_2-$, $-\text{NH}-$, O, or S; G = Cl, Br, or I; Z = any group lacking aliphatic carbons.

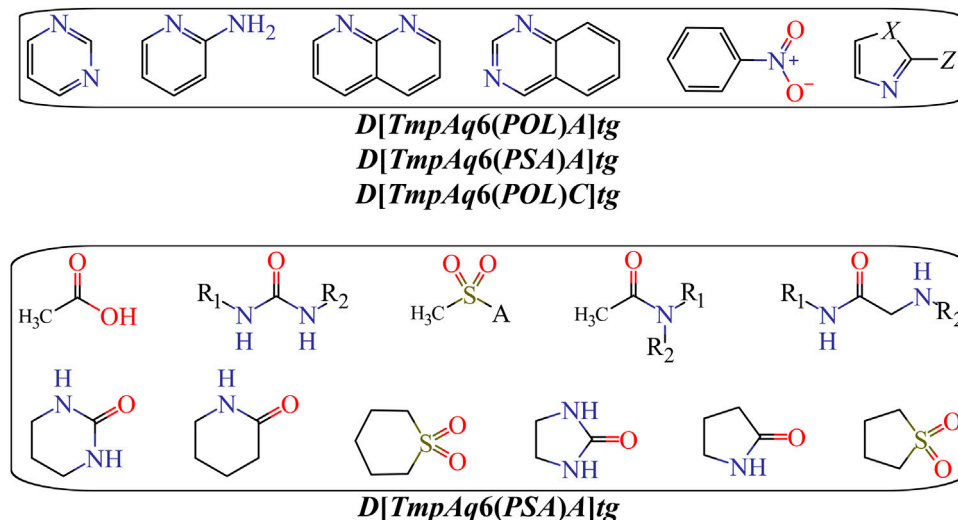


FIGURE 5 | Fragments whose presence positively increase the POL ($D[\text{LmpAq6}(\text{POL})\text{A}]_{\text{tg}}$) or the PSA ($D[\text{LmpAq6}(\text{PSA})\text{A}]_{\text{tg}}$). Here, A = $-\text{NH}_2$, $-\text{OH}$, or R (alkyl group); X = $-\text{NH}-$, O, or S; Z = Cl, Br, I, or $-\text{SH}$; substituents R_1 and/or R_2 can be H, alkyl or aryl groups.

important descriptor in the mt-QSAR-MLP model and represents the increase of the *POL* of any two atoms (at least one of them must be a hydrogen bond acceptor) placed at the topological distance equal to six but also lower distances such as two or three. As most of the atoms able to act as hydrogen bond acceptors (N, O, and F) have very low *POL*, then, their neighbor atoms at the aforementioned topological distance should have high polarizabilities (e.g., Cl, Br, I, S, an aromatic carbon, or pyridinic nitrogen). On the other hand, the descriptor $D[\text{LmpAq6}(\text{PSA})\text{A}]_{\text{tg}}$ follows the same line of thinking in terms of the topological distance and the type of

atoms involved. Nevertheless, $D[\text{LmpAq6}(\text{PSA})\text{A}]_{\text{tg}}$ focuses on the augmentation of the *PSA*, being the second most influential descriptor.

We would like to point out that although each $DQI_a(\text{tg})$ descriptor offers information regarding a defined physicochemical property combined with a specific structural aspect, it must not be expected that the infinitely (seemingly desirable) variation in the values of the $DQI_a(\text{tg})$ descriptors will conduct to an increase in the inhibitory activity. Noticed that, as explained above, some $DQI_a(\text{tg})$ descriptors are constrained by

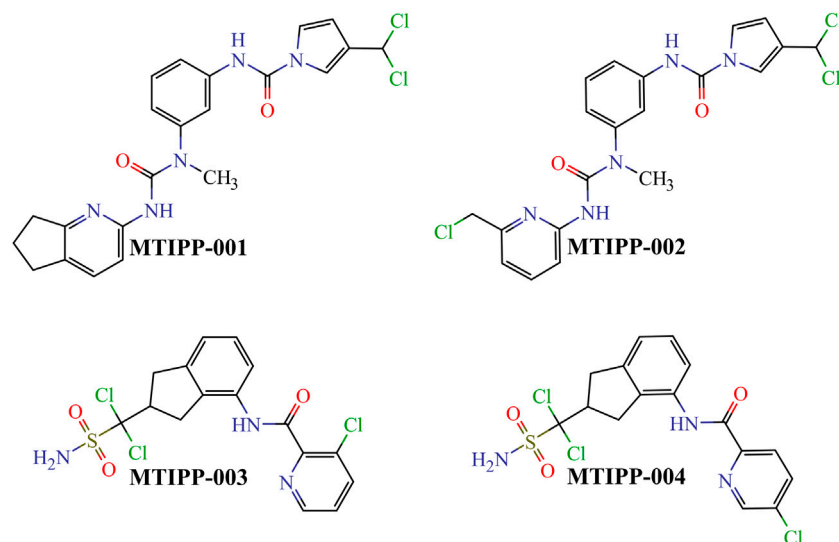


FIGURE 6 | Chemical structures of the molecules designed and predicted by using the mt-QSAR-MLP model.

others. Therefore, only the joint interpretation of the $DQI_a(tg)$ descriptors in the mt-QSAR-MLP model will provide how, through the introduction of certain molecular fragments, these descriptors can vary harmoniously so a molecule will comply with the structural requirements needed to exhibit multi-target activity against the five proteins reported in this study. The joint interpretation of the descriptors in the mt-QSAR-MLP model indicates that the aromatic and heteroaromatic rings (at least two) can be present in any region. Aliphatic chains and rings (including their heteroatom-based counterparts) can also appear in different parts of a molecule but preferably attached to both aromatic (or heteroaromatic) rings and bulky atoms (e.g., Cl, Br, I, S, and P). Halogens must also be kept in the periphery of the molecules. At least two functional groups containing atoms capable of acting as hydrogen bond acceptors (or donors) must be present, being also close (topological distance lower than 6) to the aforementioned bulky atoms and/or attached to aromatic carbons; if two or more polar functional groups formed by at least two atoms are present, they must be as distant as possible one from the other.

Virtual Design of Multi-Target Inhibitors Against Parasitic Proteins

Here, we experimented by following a series of guidelines reported recently, which enable the virtual design of new molecules with multi-target activity (Kleandrova et al., 2016; Speck-Planche et al., 2016; Speck-Planche and Cordeiro, 2017a; Speck-Planche and Cordeiro, 2017b; Speck-Planche, 2018; Speck-Planche, 2019; Speck-Planche and Scotti, 2019). The purpose of the experiment was to demonstrate that although the presence of certain fragments is important for the appearance and/or enhancement of the multi-target

activity, how these fragments are connected between each other will principally define whether a molecular can simultaneously inhibit different parasite proteins.

By rigorously following the joint interpretation of the $DQI_a(tg)$ descriptors in the mt-QSAR-MLP model, we designed four molecules belonging to two different chemical families (Figure 6). While doing so, we assembled the molecules by connecting or fusing different molecular fragments considered to positively contribute to the desirable variations in the values of the $DQI_a(tg)$ descriptors.

We would like to emphasize that when referring to the potential inhibitory activity of any of the designed molecules against a given parasitic protein, we do it always by considering the corresponding cutoff (IC_{50}) value of inhibitory activity reported in this work. Thus, according to the results of the predictions performed by the mt-QSAR-MLP mode (Table 4), all the designed molecules were predicted as multi-target inhibitors of at least three of the five parasite proteins reported in this study. All the predictions fell within the applicability domain of the mt-QSAR-MLP except for those belonging to the inhibitory activity of the molecules MTIPP-001 and MTIPP-002 against the protein glycyIpeptide N-tetradecanoyltransferase (*T. brucei brucei*). More details regarding the calculated $DQI_a(tg)$ descriptors, the predictions of the designed molecules, and the assessment of the applicability for these can be found in Supplementary Material S4.

If we inspect MTIPP-001 and MTIPP-002, it will be easy to see the remarkable similarity between their chemical structures. The only difference is that the cyclopentane moiety fused with the pyridinic ring in MTIPP-001 is replaced by the chloromethyl moiety in MTIPP-002. Yet, this small change is responsible for the differences in the multi-target profiles of these two molecules. Although the aforementioned replacement leads to a detrimental decrease of the value of the descriptor $D[LmpAq1(V)C]tg$ (which benefits from the

TABLE 4 | Molecules designed and the predictions of their multi-target profiles against the parasite proteins.

ID	Pred_IA _i (tg) ^{a,b}				
	Plasmepsin 2 (<i>P. falciparum</i>)	DHODH (<i>P. falciparum</i>)	Cruzipain (<i>T. cruzi</i>)	DHFR (<i>T. gondii</i>)	GPNTDT (<i>T. brucei brucei</i>)
MTIPP-001	1	1	-1	1	-1
MTIPP-002	1	1	1	1	-1
MTIPP-003	1	1	1	1	1
MTIPP-004	1	1	1	1	1

^aPred_IA_i(tg)—Predicted value of the categorical variable of inhibitory activity (IA_i(tg)).

^bThe abbreviations DHODH, DHFR, and GPNTDT stand for the proteins named dihydroorotate dehydrogenase, dihydrofolate reductase, and glycopeptide N-tetradecanoyltransferase, respectively.

increment of aliphatic carbons), it desirably increases the values of *D[LmpAq6(POL)A]tg* and *D[LmpAq5(POL)P]tg*, also favorably decreasing *D[LmpAq6(POL)C]tg*. These three *DQI_a(tg)* descriptors account for the fact while MTIPP-001 has been predicted to inhibit three proteins, its analog MTIPP-002 may be able to inhibit four of these biomolecular targets.

In contrast to MTIPP-001 and MTIPP-002, the molecules MTIPP-003 and MTIPP-004 present sulfonamide moiety which has a considerably higher PSA than any of the other functional groups. Furthermore, in MTIPP-003 and MTIPP-004, both fragments sulfonamide and amide are closer to the chlorines. This arrangement of atoms, which also includes the correct positioning of the aliphatic portions with respect to both aromatic carbons and chlorines particularly causes the dramatic (favorable) increment of the values of the descriptors *D[LmpAq6(POL)A]tg*, *D[LmpAq6(PSA)A]tg*, and *D[LmpAq1(V)C]tg*; these are the top three *DQI_a(tg)* descriptors, exhibiting the highest influence/discriminatory power in the mt-QSAR-MLP model. Consequently, these *DQI_a(tg)* descriptors are the main responsible for the fact that MTIPP-003 and MTIPP-004 were predicted as multi-target inhibitors against the five parasite proteins reported in this work.

Considering their potential multi-target activity, the designed molecules were searched in different databases such as ChEMBL and ZINC (Irwin and Shoichet, 2005). The aim here was to check if these molecules are reported in the scientific literature. When searching for similar compounds, the similarity cutoff was ≥ 0.7 . Under this condition, all the designed molecules seem to be new, as no results of similar molecules were found.

Docking Calculations Suggest the Multi-Target Potential of the Designed Molecules

As depicted in **Table 5**, each protein was docked against four organic compounds. The first of them is the reference ligand, which forms the crystallized complex with the protein. The second organic chemical is present in both the ChEMBL database and our dataset used to build the mt-QSAR-MLP model; the experimental IC₅₀ value of that organic chemical is equal to the activity cutoff selected for each protein. Notice that these ChEMBL organic chemicals offer a point of comparison to estimate the inhibitory activity of any query

molecule by considering the different cutoffs of activity associated with the parasitic proteins. We also docked the designed molecules MTIPP-002 and MTIPP-004, which belong to different chemical families. In the case of MTIPP-002, we selected it over its analog MTIPP-001 because the former was predicted by the mt-QSAR-MLP model to inhibit 4 out of 5 parasitic proteins; MTIPP-001 was predicted as active against only three proteins. We also chose MTIPP-004 over its analog MTIPP-003 because MTIPP-004 was predicted slightly better against two of the parasitic proteins according to their posterior probabilities; in one protein, MTIPP-003 was predicted better than MTIPP-004 and for the other two remaining proteins, MTIPP-003 and MTIPP-004 had the same value of predicted probabilities (see **Supplementary Material S4**).

In general, the preliminary results of the docking calculations depicted in **Table 5** converge with the results of the predictions performed by the mt-QSAR-MLP model regarding the multi-target profile of the designed molecules. We, however, observed divergencies in the proteins plasmepsin 2 (MTIPP-002 and MTIPP-004 suggested as inactive) and dihydroorotate dehydrogenase (MTIPP-004 indicated as inactive). This comes from the fact that the molecular docking calculations and QSAR modeling (e.g., the mt-QSAR-MLP model developed here) ‘catch’ different physicochemical and structural information regarding the chemical diversity and complexity of the molecules when inhibiting proteins. Therefore, these two computational techniques can be used in a complementary manner to study the biological profiles of the molecules in the context of protein inhibition.

Another detail that can be extracted from **Table 5** is that for the molecule MTIPP-002, the mt-QSAR-MLP and the docking calculations converge in 4 out of 5 parasitic proteins in the sense that this designed molecule is a multi-target inhibitor. Notice that the binding energy for MTIPP-002 is lower than those of the ChEMBL organic chemicals. Similar behavior occurs for the case of MTIPP-004 in 3 out of 5 parasitic proteins. Interestingly, regardless of the protein, the docking calculations suggest that, according to the energy values, MTIPP-002 is more active than MTIPP-004 although the latter was predicted by the mt-QSAR-MLP model to inhibit the five parasitic proteins while MTIPP-002 was predicted as active against only four proteins. However, there is no contradiction because while the docking calculations can be used to compare if one molecule is more active than the other, the

TABLE 5 | Results from the docking calculations.

PDB ID ^a	Protein (Organism) ^b	Ligand ID ^c	Energy (kcal/mol)	RMSD ^d
2BJU	Plasmeprin 2 (<i>P. falciparum</i>)	IH4	-172.78	0.25
		CHEMBL3264802	-153.7	-
		MTIPP-002	-150.21	-
		MTIPP-004	-103.99	-
6I55	DHODH (<i>P. falciparum</i>)	DZB	-110.4	0.67
		CHEMBL1784557	-80.8	-
		MTIPP-002	-120.5	-
		MTIPP-004	-70.9	-
1ME3	Cruzipain (<i>T. cruzi</i>)	P10	-151.93	0.33
		CHEMBL565866	-95.77	-
		MTIPP-002	-145.7	-
		MTIPP-004	-96.48	-
4KY4	DHFR (<i>T. gondii</i>)	1UE	-123.8	0.15
		CHEMBL145528	-103.4	-
		MTIPP-002	-150.9	-
		MTIPP-004	-114.2	-
HM_6QD9	GPNTDT (<i>T. brucei brucei</i>)	CHEMBL3959734	-67.8	-
		MTIPP-002	-111.7	-
		MTIPP-004	-110.22	-

^aThe ID written as "HM_6QD9" indicates that the 3D structure of the protein was obtained via homology modeling by using the protein's amino acid sequence from the PDB ID 6QD9 as the template.

^bThe abbreviations DHODH, DHFR, and GPNTDT stand for the proteins named dihydroorotate dehydrogenase, dihydrofolate reductase, and glycylopeptide N-tetradecanoyltransferase, respectively.

^cFor each protein (PDB ID), the ligands are ordered in the following manner: **1**) the reference ligand, **2**) compound from the dataset used to build the mt-QSAR-MLP model and whose IC₅₀ value is equal to the cutoff of activity selected for each protein, **3**) ligand belonging to the first chemical family of designed molecules, and **4**) ligand belonging to the second chemical family of designed molecules. ^dRMSD is the root-mean-square deviation of the atomic coordinate during the redocking of the reference ligands.

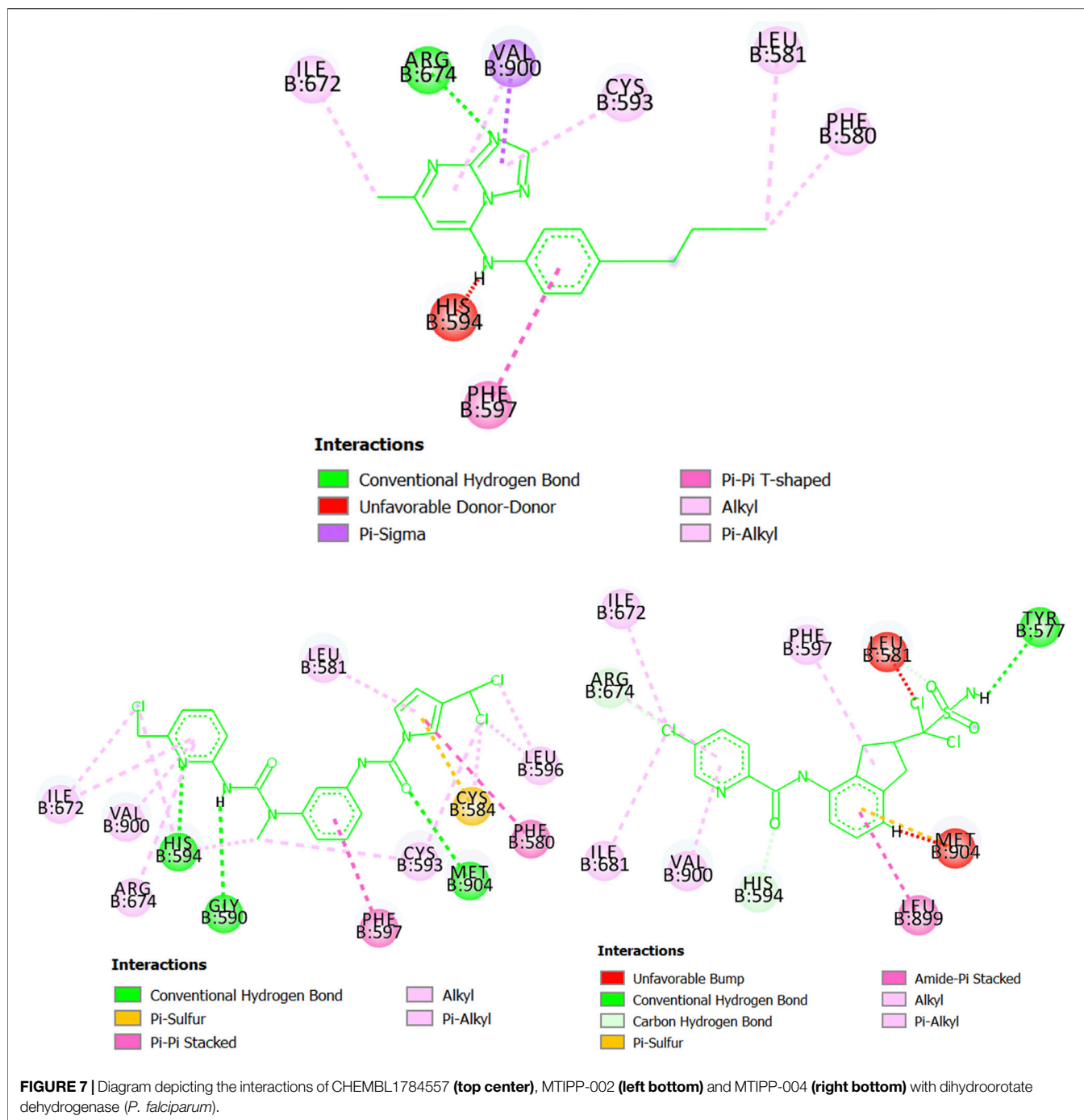
mt-QSAR-MLP model only predicts if a molecule will be active or inactive against a protein by considering a defined cutoff value.

At the structural level, the results from the docking calculations are provided in **Supplementary Material S5**, which illustrates the different protein-ligands interactions. Thus, here, for the case of each parasitic protein, we will compare the designed molecules MTIPP-002 and MTIPP-004 with ChEMBL chemicals in terms of the strength and number of interactions that help explain the results obtained in **Table 5**. In doing so, we will focus only on the parasitic proteins where the docking calculations converge (either partially or totally) with the predictions performed by the mt-QSAR-MLP model. These proteins are dihydroorotate dehydrogenase (*P. falciparum*), as well as cruzipain (*T. cruzi*), dihydrofolate reductase (*T. gondii*), and glycylopeptide N-tetradecanoyltransferase (*T. brucei brucei*). Our objective is to demonstrate that the designed molecules MTIPP-002 and MTIPP-004 are more active than the corresponding chemicals represented in **Table 5**.

We would like to highlight that in some cases, some unfavorable interactions were observed (marked in red color in the upcoming figures). We do not discard the possibility that these interactions may be associated with the computational algorithm employed to perform the docking calculations but we prefer a plausible phenomenological explanation. This is related to the fact that none of the ChEMBL chemicals present sufficiently optimized structures to effectively interact with the different amino acids in the binding site of each parasitic proteins. Nevertheless, these ChEMBL chemicals have experimental IC₅₀ values in the submicromolar range against their corresponding parasitic proteins. On the other

hand, the molecules MTIPP-002 and MTIPP-004 were designed as potential multi-target inhibitors. This means that because of the very different physicochemical and structural characteristics of the binding sites of the parasitic proteins, it is very probable that they will not cause strong inhibition as in the case of a specific/mono-target inhibitor. However, as MTIPP-002 and MTIPP-004 were designed to inhibit most of the parasitic proteins at the submicromolar range, this could translate into a much higher inhibition of the growth of the parasitic species when compared with a mono-target inhibitor. Following, we will discuss the interactions that mainly contribute to the stability/instability of the different protein-ligand complexes.

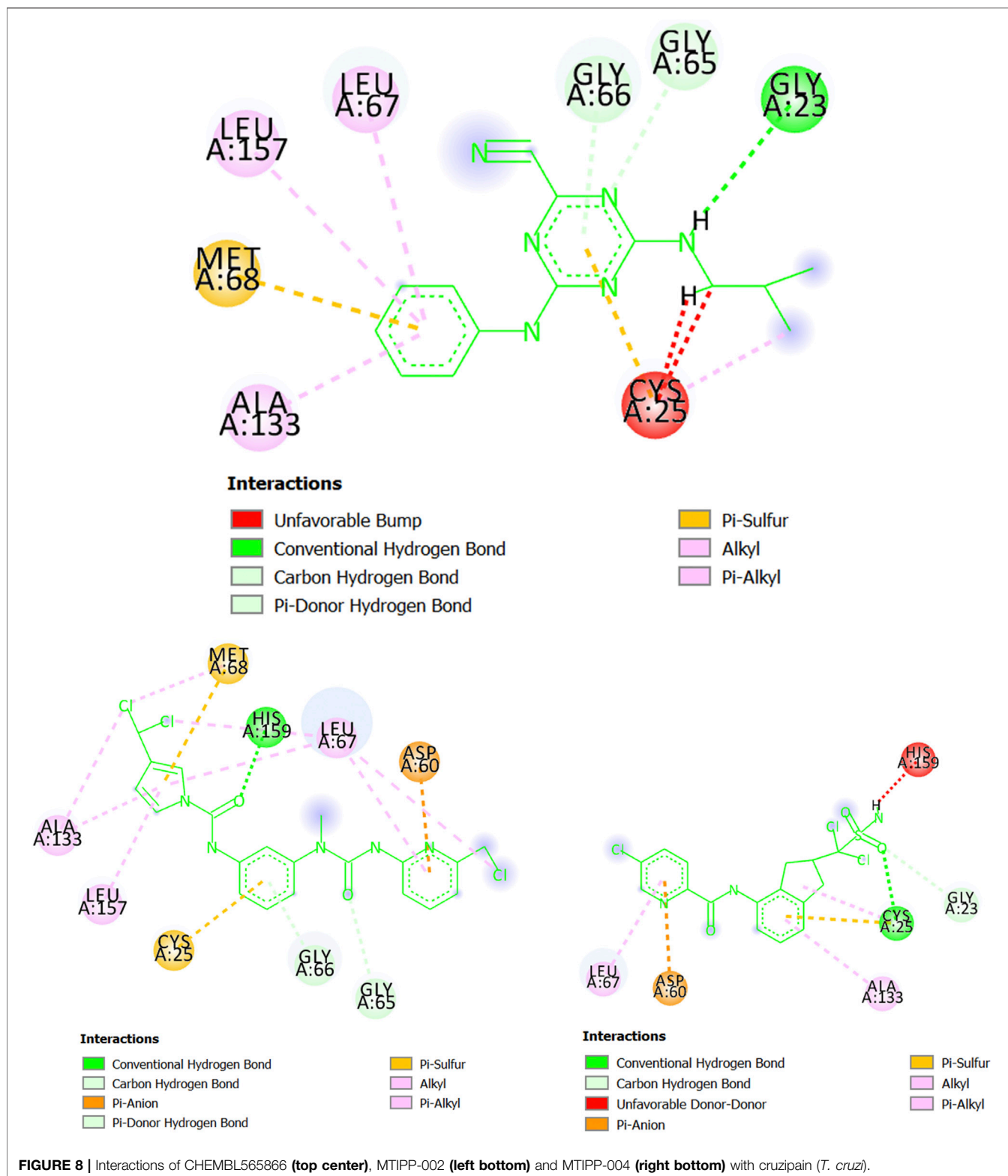
Starting with the protein dihydroorotate dehydrogenase (*P. falciparum*), it can be seen that one of the pyridinic nitrogen atoms in the five-membered ring of the molecule CHEMBL1784557 (experimental IC₅₀ = 820 nM) interacts with the residue Arg674 via hydrogen bond (**Figure 7**). Also, the ring itself makes contact with Val900 and Cys593 through pi-sigma and pi-alkyl interactions, respectively. The benzene ring of CHEMBL1784557 also contributes to stabilizing the complex by interacting with Phe507 (pi-pi T-shaped). Other interactions involve the two methyl groups of CHEMBL1784557. Yet, due to proximity, a repulsion-based interaction takes place between the secondary amine of CHEMBL1784557 and the residue His594, which is detrimental to the stability of the protein-ligand complex. In contrast, the molecule MTIPP-002 forms a hydrogen bond with His594, as well as with Gly590 and Met904. These hydrogen bonds together with pi-pi stacked (Phe580 and Phe597) and pi-sulfur (Cys584) are the main contributors to the stability of the complex formed by MTIPP-002 and



dihydroorotate dehydrogenase (*P. falciparum*), which has lower energy than that formed by the same protein and CHEMBL1784557. This suggests that the inhibitory potency of MTIPP-002 is greater than that of CHEMBL1784557. In the case of the molecule MTIPP-004, it forms a hydrogen bond with Tyr577, as well as other interactions such as pi-sulfur (Met904), amide-pi stacked (Leu899). Despite these and other several interactions such as alkyl, pi-alkyl, carbon-hydrogen bonds, MTIPP-004 presents two bumps due to steric hindrance with Leu581 and Met904, which decreases the stability of the complex,

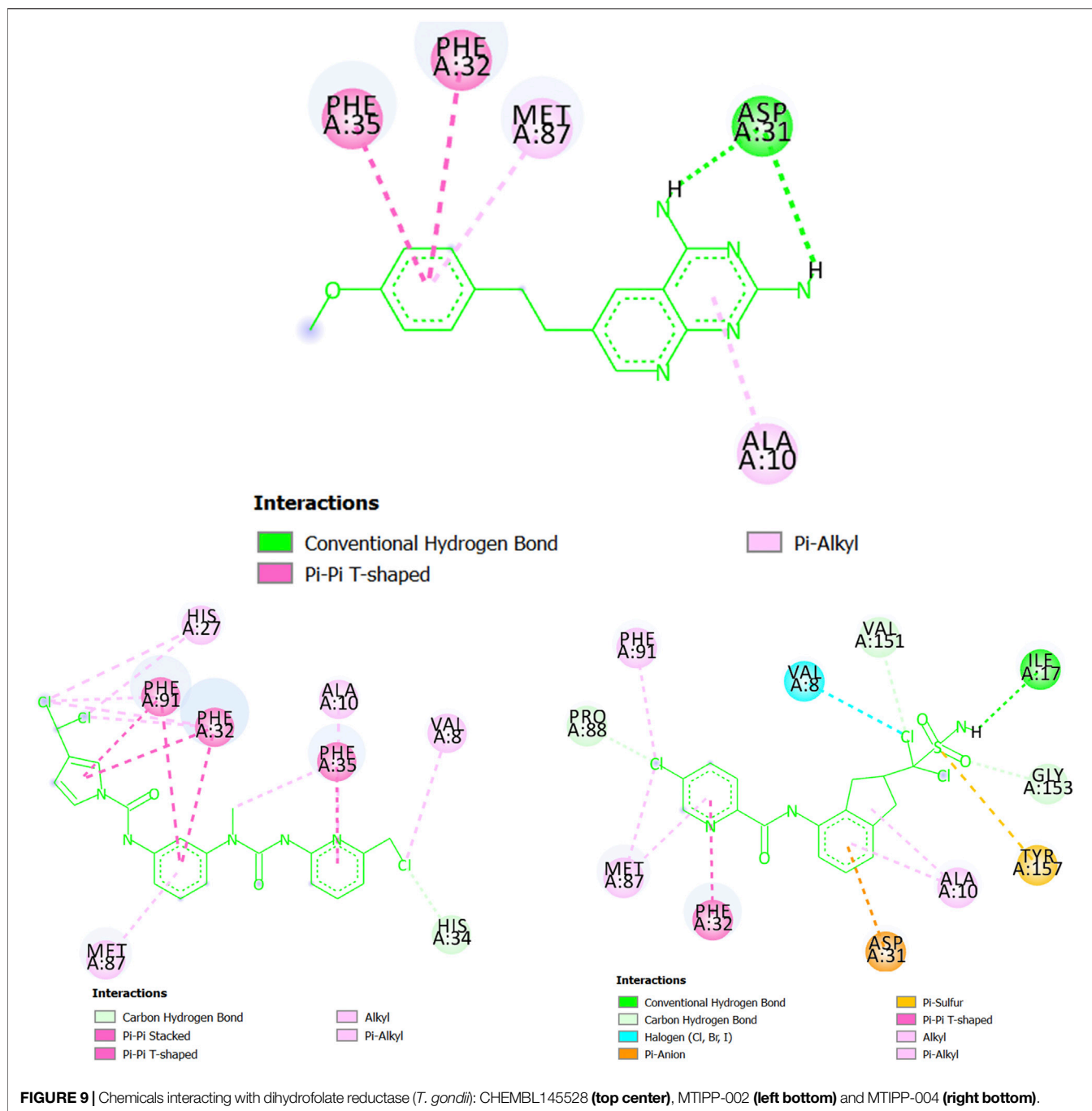
thus diminishing the ability of MTIPP-004 to strongly inhibit dihydroorotate dehydrogenase (*P. falciparum*) at the submicromolar concentration of 820 nM (activity cutoff based on IC_{50}).

Regarding cruzipain (*T. cruzi*), we can observe in **Figure 8** that the molecule CHEMBL565866 (experimental IC_{50} = 890 nM) forms a hydrogen bond with the residue Gly23 while also having pi-sulfur interactions with Met68 and Cys25. Simultaneously, CHEMBL565866 is involved in other interactions such as pi-alkyl, pi-donor hydrogen bond, and carbon-hydrogen bond. This



molecule has a double repulsive interaction with Cys25, which considerably decreases the stability of the complex formed between CHEMBL565866 and cruzipain (*T. cruzi*). In this context, the molecule MTIPP-002 has several interactions,

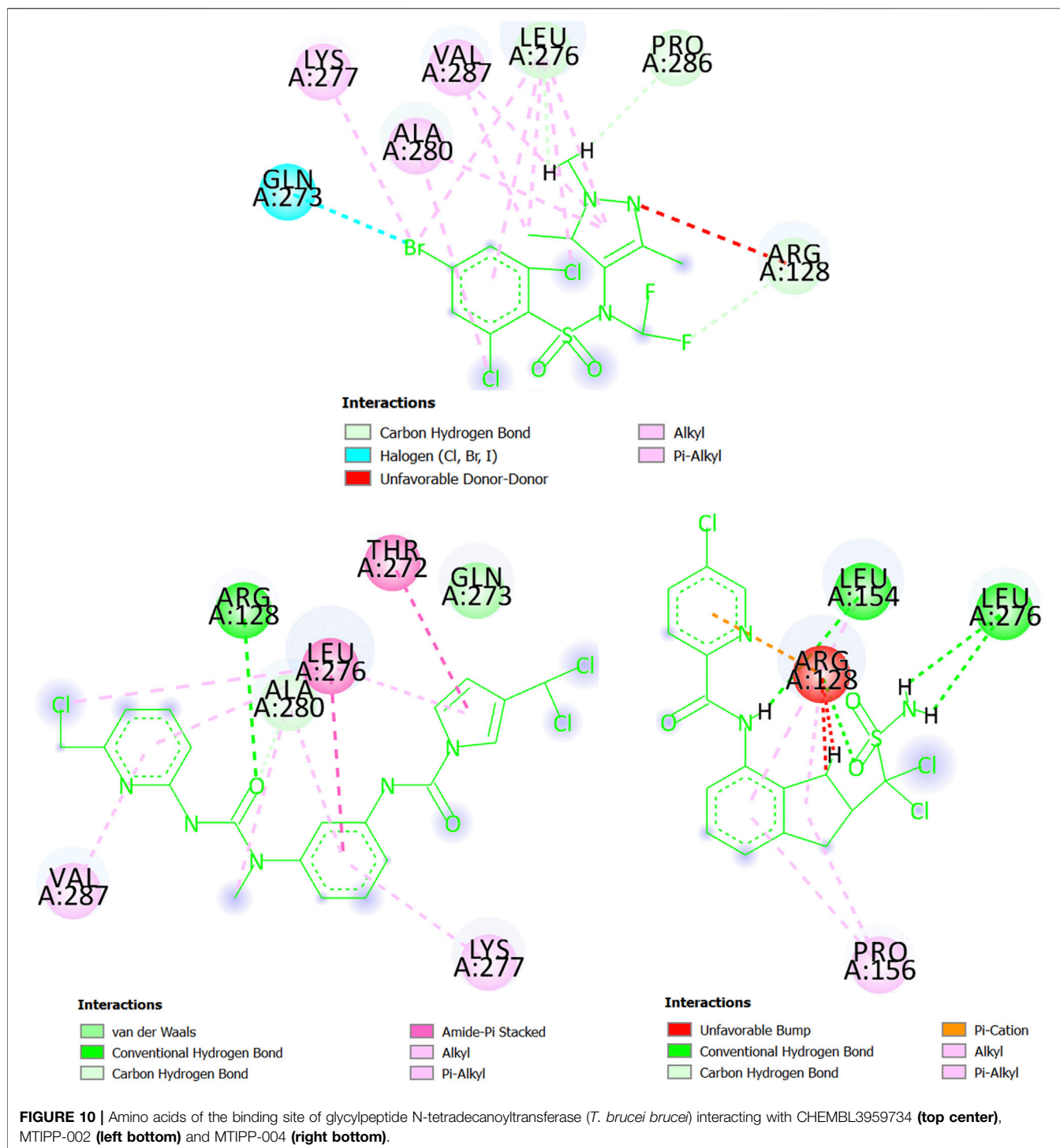
including those with the residues His159 (hydrogen bond), Asp60 (pi-anion), and Cys25 and Met68 (both pi-sulfur) which are the main energetic contributors to the complex stability. These interactions, together with those involving the



three chlorine atoms, the pyrrolic ring, and other moieties, indicate that MTIPP-002 should have a higher inhibitory potency (lower IC_{50} value) than CHEMBL565866. In the case of MTIPP-004, despite having an unfavorable (donor-donor) contact with His159, it greatly compensates by interacting with the residues Cys25 (hydrogen bond, pi-sulfur, and alkyl-alkyl), Asp60 (pi-anion), Leu67, and Ala133 (pi-alkyl), as well as Gly23 (carbon-hydrogen bond). For the case of the complex cruzipain-MTIPP-004, these interactions lead to an energy value lower than that estimate for the complex cruzipain-

CHEMBL565866. This suggests that MTIPP-004 should have $IC_{50} \leq 890$ nM.

Following with the protein dihydrofolate reductase (*T. gondii*), in **Figure 9** we have several regions of CHEMBL145528 (experimental $IC_{50} = 250$ nM) interacting via a double hydrogen bond (Asp31), pi-alkyl associations (Ala10 and Met87), and pi-pi T-shaped configurations (Phe32 and Phe35). Notice, however, that the two hydrogen bonds with Asp31 lack directionality (see the 3D view in **Supplementary Material S5**), and therefore they may be relatively weak. Interestingly, MTIPP-



002, despite lacking hydrogen bond, forms a huge number of hydrophobic interactions such as pi-pi stacked and pi-pi T-shaped configurations with the amino acids Phe32, Phe35, and Phe91. In this sense, it has been experimentally demonstrated that the presence of simultaneous pi-interactions of a molecule with the residues Phe32 and Phe91 are essential in achieving inhibitory potency (IC_{50}) at the submicromolar range (Welsch

et al., 2016). On the other hand, there is also a great number of alkyl-alkyl and pi-alkyl interactions where the amino acids Val8, Ala10, His27, and Met87 participate; His34 is involved in a carbon-hydrogen bond. A key aspect of the interactions of MTIPP-002 with the different amino acids is that most of them seem highly directional, which, for the complex formed by this molecule and dihydrofolate reductase (*T. gondii*), yield an

energy value lower than the complexes formed by the same protein with either CHEMBL145528 or MTIPP-004. Therefore, MTIPP-002 should have an $IC_{50} \leq 250$ nM. For the case of MTIPP-004, this molecule should also be expected to exhibit $IC_{50} \leq 250$ nM since the stability of the complex MTIPP-004-dihydrofolate reductase (*T. gondii*) is favored over that of the complex CHEMBL145528-dihydrofolate reductase (*T. gondii*). This is due to the presence of adequate interactions of MTIPP-004 with the amino acids Ile17 (hydrogen bond), Val8 (halogen bond), Asp31 (pi-anion), Phe32 (pi-pi T-shaped), and Tyr157 (pi-sulfur). There are also other favorable interactions involving alkyl groups (either from MTIPP-004 or the amino acids) and carbon-hydrogen bonds.

Last, we have the glycylopeptide N-tetradecanoyltransferase (*T. brucei brucei*) whose 3D structure was created by using homology modeling. The steps and results of the homology modeling for this protein can be found in **Supplementary Material S6**. The 3D structure of glycylopeptide N-tetradecanoyltransferase (*T. brucei brucei*) is now freely available at <https://swissmodel.expasy.org/repository/uniprot/Q388H8> as part of the SWISS-MODEL Repository (Bienert et al., 2017), which is a database of annotated 3D protein structure models. In **Figure 10**, the pi-alkyl interactions (and others involving alkyl groups or halogens) between the chemical CHEMBL3959734 (experimental $IC_{50} = 270$ nM) and glycylopeptide N-tetradecanoyltransferase prevail, particularly with the residues Leu276, Ala280, and Val287 (Lys277 participates in less degree). There are also three carbon-hydrogen bonds (Arg128, Leu276, and Pro286) and a halogen bond with the residue Gln273. In any case, Arg128 is present in an unfavorable donor-donor interaction with CHEMBL3959734. At the same time, we can deduct from **Figure 10** that both MTIPP-002 and MTIPP-004, form more stable complexes with glycylopeptide N-tetradecanoyltransferase (*T. brucei brucei*) than CHEMBL3959734. From one side, MTIPP-002 forms a hydrogen bond with Arg128, the same amino acid the unfavorable influences the stability of the complex formed by CHEMBL3959734 and glycylopeptide N-tetradecanoyltransferase (*T. brucei brucei*). Besides, MTIPP-002 exhibits relatively strong amide-pi stacked interactions with the residues Thr272 and Leu276; the latter also participates in pi-alkyl (together with Ala280 and Val287) and other interactions based on the presence alkyl group (together with Ala280). Other non-covalent interactions can also be observed. All these interactions point out to the direction of considering MTIPP-002 to have $IC_{50} \leq 270$ nM. On the other hand, in the case of MTIPP-004, there is detrimental energetic contribution because of the repulsion with Arg128 although the same residue favorably present in a hydrogen bond, and pi-cation, pi-alkyl, and alkyl-alkyl interactions. Anyway, MTIPP-004 counterbalances by forming other three hydrogen bonds, one with Leu154 and two with Leu276; Leu154 also participates in a pi-alkyl interaction together with Pro156 (also involved in an alkyl-alkyl interaction). All these interactions help to explain why the energy value obtained for MTIPP-002 and MTIPP-004 are very similar when interacting with s for the complexes of glycylopeptide N-tetradecanoyltransferase (*T. brucei brucei*). Consequently, MTIPP-004 is also expected to have $IC_{50} \leq 270$ nM.

DRUGLIKENESS AND SYNTHETIC ACCESSIBILITY

We examined the four designed molecules in terms of their compliance with Lipinski's rule of five (Lipinski et al., 2001), the Ghose's filter (Ghose et al., 1999), and the Veber's rule (Veber et al., 2002). These guidelines are based on the estimation of a series of physicochemical properties that permit to analyze of the druglikeness of any molecule, in particular, their capacity to exhibit a good oral bioavailability. The physicochemical properties were calculated by the program AlvaDesc v1.0.14 (Alvascience-Srl, 2019) and included the number of hydrogen bond donors (HBD), the number of hydrogen bond acceptors (HBA), the molecular weight (MW), the logarithm of the partition coefficient octanol/water ($\log P$), the number of atoms (nAT), the molar refractivity (MR), the number of rotatable bonds (RBN), and the PSA. A report of these properties for the designed molecules can be found in **Supplementary Material S7**; the physicochemical properties of the molecules designed here are in agreement with Lipinski's rule of five and the other variants. We also employed the webserver SwissADME to estimate the synthetic accessibility of the designed molecules. In this sense, SwissADME predicts the synthetic accessibility score (SAS), which ranges from 1 (easily synthesizable) to 10 (difficult to synthesize). The SAS values for the designed molecules ranges from 3.23 to 3.48 (**Supplementary Material S7**). Considering the closeness to 1 of these SAS values, it can be deduced that the designed molecules should be relatively easy to synthesize.

CONCLUDING REMARKS

A more efficient eradication of many parasitic diseases can in principle be achieved with the use of multi-target inhibitors. The fast search of such a class of antimicrobial therapeutics depends in great part on the power and accuracy of modern computational tools. The mt-QSAR-MLP built in this work model represents an advance in early drug discovery against parasitic diseases because with this *in silico* tool and the theoretical support provided by the molecular docking calculations, it is possible to rationally design potential antiparasitic agents by simultaneously inhibiting diverse targets involved in the virulence and/or survival of several pathogenic parasites. The present report confirms the promising applications of the mt-QSAR approaches, which can be extended to many therapeutic areas.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**SupplementaryMaterial**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

AS-P and MTS conceptualized the study, supervised the data analysis, contributed to the interpretation of the molecular descriptors, designed the molecules, interpreted the docking calculations, and contributed to the writing of the manuscript. VVK curated the database, performed the data analysis, and contributed to the writing of the manuscript. LS performed the docking calculations. LS, FJBMJr, and EM calculated the molecular descriptors and contributed to their interpretation; they also contributed to the writing of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The present work was carried out with the support of the Brazilian National Council for Scientific and Technological

Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico–CNPq) under the grant numbers 308590/2017–1, 309648/2019–0, and 431254/2018–4.

ACKNOWLEDGMENTS

The authors also acknowledge all the experimental scientists working in the fields of parasitic research, whose works have served as inspiration for the compilation of the dataset that led to the creation of the present mt-QSAR-MLP model.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2021.634663/full#supplementary-material>.

REFERENCES

- Aguirre, A. A., Longcore, T., Barbieri, M., Dabritz, H., Hill, D., Klein, P. N., et al. (2019). The one health approach to toxoplasmosis: epidemiology, control, and prevention strategies. *EcoHealth* 16, 378–390. doi:10.1007/s10393-019-01405-7
- Alberca, L. N., Chuguransky, S. R., Álvarez, C. L., Talevi, A., and Salas-Sarduy, E. (2019). In silico guided drug repurposing: discovery of new competitive and non-competitive inhibitors of falcipain-2. *Front. Chem.* 7, 534. doi:10.3389/fchem.2019.00534
- Alday, P. H., and Doggett, J. S. (2017). Drugs in development for toxoplasmosis: advances, challenges, and current status. *Drug Des. Dev. Ther.* 11, 273–293. doi:10.2147/DDDT.S60973
- Alonso, N., Caamaño, O., Romero-Duran, F. J., Luan, F., D S Cordeiro, M. N., Yañez, M., et al. (2013). Model for high-throughput screening of multitarget drugs in chemical neurosciences: synthesis, assay, and theoretic study of rasagiline carbamates. *ACS Chem. Neurosci.* 4, 1393–1403. doi:10.1021/cn400111n
- Alvascience-Srl (2019). AlvaDesc (software for molecular descriptor calculation). v1.0.14. Available at: <https://www.alvascience.com/>. (Accessed November 12, 2020).
- Amisigo, C. M., Antwi, C. A., Adjimani, J. P., and Gwira, T. M. (2019). In vitro anti-trypanosomal effects of selected phenolic acids on *Trypanosoma brucei*. *PLoS One* 14, e0216078. doi:10.1371/journal.pone.0216078
- Amusengeri, A., Astl, L., Lobb, K., Verkhivker, G. M., and Tastan Bishop, O. (2019). Establishing computational approaches towards identifying malarial allosteric modulators: a case study of *Plasmodium falciparum* Hsp70s. *Int. J. Mol. Sci.* 20, 5574. doi:10.3390/ijms20225574
- Anderson, A. C. (2003). The process of structure-based drug design. *Chem. Biol.* 10, 787–797. doi:10.1016/j.chembiol.2003.09.002
- Baker, N., de Koning, H. P., Mäser, P., and Horn, D. (2013). Drug resistance in African trypanosomiasis: the melarsoprol and pentamidine story. *Trends Parasitol.* 29, 110–118. doi:10.1016/j.pt.2012.12.005
- Baskin, I. L., Skvortsova, M. I., Stankevich, I. V., and Zefirov, N. S. (1995). On the basis of invariants of labeled molecular graphs. *J. Chem. Inf. Comput. Sci.* 35, 527–531. doi:10.1021/ci00025a021
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein Data Bank. *Nucleic Acids Res.* 28, 235–242. doi:10.1093/nar/28.1.235
- Bienert, S., Waterhouse, A., de Beer, T. A., Tauriello, G., Studer, G., Bordoli, L., et al. (2017). The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.* 45, D313–D319. doi:10.1093/nar/gkw1132
- BIOVIA (2018). *Discovery studio visualizer. v19.1, vélizy-villacoublay, France: Dassault Systèmes SE.*
- Buckner, F. S., Moore, A. C., and Cetron, M. S. (2017). “Chapter 26 - Chagas disease,” in *The travel and tropical medicine manual*. 5th Edn, Editors C. A. Sanford, P. S. Pottinger, and E. C. Jong (Edinburgh, London, New York, Oxford, Philadelphia, St Louis, Sydney, Toronto: Elsevier), 371–381.
- Campos, M. C., Phelan, J., Francisco, A. F., Taylor, M. C., Lewis, M. D., Pain, A., et al. (2017). Genome-wide mutagenesis and multi-drug resistance in American trypanosomes induced by the front-line drug benznidazole. *Sci. Rep.* 7, 14407. doi:10.1038/s41598-017-14986-6
- ChemAxon (1998-2019). Standardizer. v19.18.0 Budapest, Hungary. Available at: <https://www.chemaxon.com>. (Accessed November 7, 2020).
- Conrad, M. D., and Rosenthal, P. J. (2019). Antimalarial drug resistance in Africa: the calm before the storm? *Lancet Infect. Dis.* 19, e338–e351. doi:10.1016/S1473-3099(19)30261-0
- Cowell, A. N., and Winzeler, E. A. (2019). Advances in omics-based methods to identify novel targets for malaria and other parasitic protozoan infections. *Genome Med.* 11, 63. doi:10.1186/s13073-019-0673-3
- Dallas, T. A., Laine, A. L., and Ovaskainen, O. (2019). Detecting parasite associations within multi-species host and parasite communities. *Proc. Biol. Sci.* 286, 20191109. doi:10.1098/rspb.2019.1109
- Di Pisa, F., Landi, G., Dello Iacono, L., Pozzi, C., Borsari, C., Ferrari, S., et al. (2017). Chroman-4-One derivatives targeting pteridine reductase 1 and showing anti-parasitic activity. *Molecules* 22, 426. doi:10.3390/molecules22030426
- Dos Santos, A. M., Cianni, L., De Vita, D., Rosini, F., Leitão, A., Laughton, C. A., et al. (2018). Experimental study and computational modelling of cruzain cysteine protease inhibition by dipeptidyl nitriles. *Phys. Chem. Chem. Phys.* 20, 24317–24328. doi:10.1039/c8cp03320j
- Forsyth, C. J., Hernandez, S., Olmedo, W., Abuhamidah, A., Traina, M. I., Sanchez, D. R., et al. (2016). Safety profile of nifurtimox for treatment of Chagas disease in the United States. *Clin. Infect. Dis.* 63, 1056–1062. doi:10.1093/cid/ciw477
- García, I., Fall, Y., Gómez, G., and González-Díaz, H. (2011). First computational chemistry multi-target model for anti-Alzheimer, anti-parasitic, anti-fungi, and anti-bacterial activity of GSK-3 inhibitors *in vitro*, *in vivo*, and in different cellular lines. *Mol. Divers.* 15, 561–567. doi:10.1007/s11030-010-9280-3
- Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., et al. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, D1100–D1107. doi:10.1093/nar/gkr777
- Ghose, A. K., Viswanadhan, V. N., and Wendoloski, J. J. (1998). Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental Methods: an analysis of ALOGP and CLOGP methods. *J. Phys. Chem.* 102, 3762–3772. doi:10.1021/jp980230o
- Ghose, A. K., Viswanadhan, V. N., and Wendoloski, J. J. (1999). A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* 1, 55–68. doi:10.1021/cc9800071

- Giordani, F., Morrison, L. J., Rowan, T. G., DE Koning, H. P., and Barrett, M. P. (2016). The animal trypanosomiasis and their chemotherapy: a review. *Parasitology* 143, 1862–1889. doi:10.1017/S0031182016001268
- Grabias, B., and Kumar, S. (2016). Adverse neuropsychiatric effects of antimalarial drugs. *Expet Opin. Drug Saf.* 15, 903–910. doi:10.1080/14740338.2016.1175428
- Hausler, I. L., Chan, X. H. S., Guérin, P. J., and White, N. J. (2018). The arrhythmogenic cardiotoxicity of the quinoline and structurally related antimalarial drugs: a systematic review. *BMC Med.* 16, 200. doi:10.1186/s12916-018-1188-2
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). “Random forests,” in *The elements of statistical learning: data mining, inference, and prediction*. Editors T. Hastie, R. Tibshirani, and J. Friedman (Springer New York, New York, NY), 587–604.
- Hellard, E., Fouchet, D., Vavre, F., and Pontier, D. (2015). Parasite-parasite interactions in the wild: how to detect them? *Trends Parasitol.* 31, 640–652. doi:10.1016/j.pt.2015.07.005
- Herrera-Mayorga, V., Lara-Ramirez, E. E., Chacon-Vargas, K. F., Aguirre-Alvarado, C., Rodriguez-Paez, L., Alcantara-Farfan, V., et al. (2019). Structure-based virtual screening and in vitro evaluation of new trypanosoma cruzi cruzain inhibitors. *Int. J. Mol. Sci.* 20, 1742. doi:10.3390/ijms20071742
- Huang, L., Brinen, L. S., and Ellman, J. A. (2003). Crystal structures of reversible ketone-Based inhibitors of the cysteine protease cruzain. *Bioorg. Med. Chem.* 11, 21–29. doi:10.1016/S0968-0896(02)00427-3
- Irwin, J. J., and Shoichet, B. K. (2005). ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 45, 177–182. doi:10.1021/ci049714+
- Jahnke, W., and Erlanson, D. A. (2006). *Fragment-based approaches in drug discovery*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA.
- Karvonen, A., Jokela, J., and Laine, A. L. (2019). Importance of sequence and timing in parasite coinfections. *Trends Parasitol.* 35, 109–118. doi:10.1016/j.pt.2018.11.007
- Kimuda, M. P., Laming, D., Hoppe, H. C., and Tastan Bishop, Ö. (2019). Identification of novel potential inhibitors of pteridine reductase 1 in trypanosoma brucei via computational structure-based approaches and in vitro inhibition assays. *Molecules* 24, 142. doi:10.3390/molecules24010142
- Kleandrova, V. V., Ruso, J. M., Speck-Planche, A., and Dias Soeiro Cordeiro, M. N. (2016). Enabling the discovery and virtual screening of potent and safe antimicrobial peptides. Simultaneous prediction of antibacterial activity and cytotoxicity. *ACS Comb. Sci.* 18, 490–498. doi:10.1021/acscmbosci.6b00063
- Kleandrova, V. V., and Speck-Planche, A. (2020). PTML modeling for alzheimer's disease: design and prediction of virtual multi-target inhibitors of GSK3B, HDAC1, and HDAC6. *Curr. Top. Med. Chem.* 20, 1657–1672. doi:10.2174/1568026620666200607190951
- Kwofie, K. D., Tung, N. H., Suzuki-Ohashi, M., Amoa-Bosompem, M., Adele, R., Sakyiamah, M. M., et al. (2016). Antitrypanosomal activities and mechanisms of action of novel tetracyclic iridoids from *Morinda lucida* benth. *Antimicrob. Agents Chemother.* 60, 3283–3290. doi:10.1128/AAC.01916-15
- Latorre, A., Schirmeister, T., Kesselring, J., Jung, S., Johé, P., Hellmich, U. A., et al. (2016). Dipeptidyl nitroalkenes as potent reversible inhibitors of cysteine proteases rhodesain and cruzain. *ACS Med. Chem. Lett.* 7, 1073–1076. doi:10.1021/acsmchemlett.6b00276
- Lima, M. N. N., Cassiano, G. C., Tomaz, K. C. P., Silva, A. C., Sousa, B. K. P., Ferreira, L. T., et al. (2019). Integrative multi-kinase approach for the identification of potent antiplasmodial hits. *Front. Chem.* 7, 773. doi:10.3389/fchem.2019.00773
- Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 46, 3–26. doi:10.1016/S0169-409X(00)00129-0
- Marrero-Ponce, Y., Siverio-Mota, D., Gálvez-Llompert, M., Recio, M. C., Giner, R. M., García-Domènech, R., et al. (2011). Discovery of novel anti-inflammatory drug-like compounds by aligning in silico and in vivo screening: the nitroindazolinone chemotype. *Eur. J. Med. Chem.* 46, 5736–5753. doi:10.1016/j.ejmech.2011.07.053
- Marzaro, G., Chilin, A., Guiotto, A., Uriarte, E., Brun, P., Castagliuolo, I., et al. (2011). Using the TOPS-MODE approach to fit multi-target QSAR models for tyrosine kinases inhibitors. *Eur. J. Med. Chem.* 46, 2185–2192. doi:10.1016/j.ejmech.2011.02.072
- Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* 405, 442–451. doi:10.1016/0005-2795(75)90109-9
- Medina Marrero, R., Marrero-Ponce, Y., Barigye, S. J., Echeverría Díaz, Y., Acevedo-Barrios, R., Casañola-Martín, G. M., et al. (2015). QuBiLS-MAS method in early drug discovery and rational drug identification of antifungal agents. *SAR QSAR Environ. Res.* 26, 943–958. doi:10.1080/1062936X.2015.1104517
- Molyneux, D. H., Savioli, L., and Engels, D. (2017). Neglected tropical diseases: progress towards addressing the chronic pandemic. *Lancet* 389, 312–325. doi:10.1016/S0140-6736(16)30171-4
- Montazeri, M., Mehrzadi, S., Sharif, M., Sarvi, S., Tanzifi, A., Aghayan, S. A., et al. (2018). Drug resistance in *Toxoplasma gondii*. *Front. Microbiol.* 9, 2587. doi:10.3389/fmicb.2018.02587
- Palos, I., Lara-Ramirez, E. E., Lopez-Cedillo, J. C., Garcia-Perez, C., Kashif, M., Bocanegra-García, V., et al. (2017). Repositioning FDA drugs as potential cruzain inhibitors from trypanosoma cruzi: virtual screening, in vitro and in vivo studies. *Molecules* 22, 1015. doi:10.3390/molecules22061015
- Pearson, K. (1895). Notes on regression and inheritance in the case of two parents. *Proc. Roy. Soc. Lond.* 58, 240–242. doi:10.1098/rspl.1895.0041
- Pérez-Molina, J. A., and Molina, I. (2018). Chagas disease. *Lancet* 391, 82–94. doi:10.1016/S0140-6736(17)31612-4
- Pippione, A. C., Sainas, S., Goyal, P., Fritzson, I., Cassiano, G. C., Giraud, A., et al. (2019). Hydroxazole scaffold-based Plasmodium falciparum dihydroorotate dehydrogenase inhibitors: synthesis, biological evaluation and X-ray structural studies. *Eur. J. Med. Chem.* 163, 266–280. doi:10.1016/j.ejmech.2018.11.044
- Prade, L., Jones, A. F., Boss, C., Richard-Bildstein, S., Meyer, S., Binkert, C., et al. (2005). X-ray structure of plasmepsin II complexed with a potent achiral inhibitor. *J. Biol. Chem.* 280, 23837–23843. doi:10.1074/jbc.M501519200
- Prado-Prado, F. J., González-Díaz, H., de la Vega, O. M., Ubeira, F. M., and Chou, K. C. (2008). Unified QSAR approach to antimicrobials. Part 3: first multi-tasking QSAR model for input-coded prediction, structural back-projection, and complex networks clustering of antiprotozoal compounds. *Bioorg. Med. Chem.* 16, 5871–5880. doi:10.1016/j.bmc.2008.04.068
- Prado-Prado, F. J., García-Mera, X., and González-Díaz, H. (2010a). Multi-target spectral moment QSAR versus ANN for antiparasitic drugs against different parasite species. *Bioorg. Med. Chem.* 18, 2225–2231. doi:10.1016/j.bmc.2010.01.068
- Prado-Prado, F. J., Ubeira, F. M., Borges, F., and González-Díaz, H. (2010b). Unified QSAR & network-based computational chemistry approach to antimicrobials. II. Multiple distance and triadic census analysis of antiparasitic drugs complex networks. *J. Comput. Chem.* 31, 164–173. doi:10.1002/jcc.21292
- Ravikumar, B., and Aittokallio, T. (2018). Improving the efficacy-safety balance of polypharmacology in multi-target drug discovery. *Expet Opin. Drug Discov.* 13, 179–192. doi:10.1080/17460441.2018.1413089
- Robert-Gangneux, F., and Dardé, M. L. (2012). Epidemiology of and diagnostic strategies for toxoplasmosis. *Clin. Microbiol. Rev.* 25, 264–296. doi:10.1128/CMR.05013-11
- Romero Durán, F. J., Alonso, N., Caamaño, O., García-Mera, X., Yañez, M., Prado-Prado, F. J., et al. (2014). Prediction of multi-target networks of neuroprotective compounds with entropy indices and synthesis, assay, and theoretical study of new asymmetric 1,2-rasagiline carbamates. *Int. J. Mol. Sci.* 15, 17035–17064. doi:10.3390/ijms150917035
- Romero-Durán, F. J., Alonso, N., Yañez, M., Caamaño, O., García-Mera, X., and González-Díaz, H. (2016). Brain-inspired cheminformatics of drug-target brain interactome, synthesis, and assay of TVP1022 derivatives. *Neuropharmacology* 103, 270–278. doi:10.1016/j.neuropharm.2015.12.019
- Rosada, B., Bekier, A., Cytarska, J., Plaziński, W., Zavyalova, O., Sikora, A., et al. (2019). Benzo[b]thiophene-thiazoles as potent anti-Toxoplasma gondii agents: design, synthesis, tyrosinase/tyrosine hydroxylase inhibitors, molecular docking study, and antioxidant activity. *Eur. J. Med. Chem.* 184, 111765. doi:10.1016/j.ejmech.2019.111765
- Sahigara, F., Mansouri, K., Ballabio, D., Mauri, A., Consonni, V., and Todeschini, R. (2012). Comparison of different approaches to define the applicability

- domain of QSAR models. *Molecules* 17, 4791–4810. doi:10.3390/molecules17054791
- Santos-García, L., de Mecnas Filho, M. A., Musilek, K., Kuca, K., Ramalho, T. C., and da Cunha, E. F. F. (2018). QSAR study of N-myristoyltransferase inhibitors of antimalarial agents. *Molecules* 23, 2348 doi:10.3390/molecules23092348
- Seppälä, O., and Jokela, J. (2016). Do coinfections maintain genetic variation in parasites? *Trends Parasitol.* 32, 930–938. doi:10.1016/j.pt.2016.08.010
- Speck-Planche, A. (2018). Combining ensemble learning with a fragment-based topological approach to generate new molecular diversity in drug discovery: in silico design of Hsp90 inhibitors. *ACS Omega* 3, 14704–14716. doi:10.1021/acsomega.8b02419
- Speck-Planche, A., and Cordeiro, M. N. D. S. (2017a). Fragment-based in silico modeling of multi-target inhibitors against breast cancer-related proteins. *Mol. Divers.* 21, 511–523. doi:10.1007/s11030-017-9731-1
- Speck-Planche, A., and Dias Soeiro Cordeiro, M. N. (2017b). Speeding up early drug discovery in antiviral research: a fragment-based in silico approach for the design of virtual anti-hepatitis C leads. *ACS Comb. Sci.* 19, 501–512. doi:10.1021/acscmbosci.7b00039
- Speck-Planche, A., and Kleandrova, V. V. (2012a). In silico design of multi-target inhibitors for C-C chemokine receptors using substructural descriptors. *Mol. Divers.* 16, 183–191. doi:10.1007/s11030-011-9337-y
- Speck-Planche, A., and Kleandrova, V. V. (2012b). QSAR and molecular docking techniques for the discovery of potent monoamine oxidase B inhibitors: computer-aided generation of new rasagiline bioisosteres. *Curr. Top. Med. Chem.* 12, 1734–1747. doi:10.2174/1568026611209061734
- Speck-Planche, A., Kleandrova, V. V., Luan, F., and Cordeiro, M. N. (2012). A ligand-based approach for the in silico discovery of multi-target inhibitors for proteins associated with HIV infection. *Mol. Biosyst.* 8, 2188–2196. doi:10.1039/c2mb25093d
- Speck-Planche, A., Kleandrova, V. V., Luan, F., and Cordeiro, M. N. (2013). Multi-target inhibitors for proteins associated with Alzheimer: in silico discovery using fragment-based descriptors. *Curr. Alzheimer Res.* 10, 117–124. doi:10.2174/1567205011310020001
- Speck-Planche, A., Kleandrova, V. V., Ruso, J. M., and Cordeiro, M. N. (2016). First multitarget chemo-bioinformatic model to enable the discovery of antibacterial peptides against multiple Gram-positive pathogens. *J. Chem. Inf. Model.* 56, 588–598. doi:10.1021/acs.jcim.5b00630
- Speck-Planche, A. (2019). Multicellular target QSAR model for simultaneous prediction and design of anti-pancreatic cancer agents. *ACS Omega* 4, 3122–3132. doi:10.1021/acsomega.8b03693
- Speck-Planche, A., and Scotti, M. T. (2019). BET bromodomain inhibitors: fragment-based in silico design using multi-target QSAR models. *Mol. Divers.* 23, 555–572. doi:10.1007/s11030-018-9890-8
- Speck-Planche, A., Kleandrova, V. V., and Cordeiro, M. N. D. S. (2015). Chemoinformatics in antibacterial drug discovery: simultaneous modeling of anti-enterococci activities and ADMET profiles through the use of probabilistic quadratic indices. *Int. Electron Conf. Synth. Org. Chem.* e003. doi:10.3390/ecsoc-19-e003
- Thomsen, R., and Christensen, M. H. (2006). MolDock: a new technique for high-accuracy molecular docking. *J. Med. Chem.* 49, 3315–3321. doi:10.1021/jm051197e
- TIBCO Software-Inc. (2018). STATISTICA (data analysis software system). v13.5.0.17 Palo Alto, California, USA. Available at: <http://tibco.com>. (Accessed November 8, 2020).
- Todeschini, R., and Consonni, V. (2009). *Molecular descriptors for chemoinformatics*. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA.
- Valdés-Martini, J. R., García-Jacas, C. R., Marrero-Ponce, Y., Silveira Vaz 'd Almeida, Y., and Morell, C. (2012). QUBILS-MAS: free software for molecular descriptors calculator from quadratic, bilinear and linear maps based on graph-theoretic electronic-density matrices and atomic weightings. v1.0 Villa Clara, Cuba. Available at: <http://tomocomd.com/>. (Accessed November 7, 2020).
- Valdés-Martini, J. R., Marrero-Ponce, Y., García-Jacas, C. R., Martínez-Mayorga, K., Barigye, S. J., Vaz d'Almeida, Y. S., et al. (2017). QuBiLS-MAS, open source multi-platform software for atom- and bond-based topological (2D) and chiral (2.5D) algebraic molecular descriptors computations. *J. Cheminf.* 9, 35. doi:10.1186/s13321-017-0211-5
- Veber, D. F., Johnson, S. R., Cheng, H. Y., Smith, B. R., Ward, K. W., and Kopple, K. D. (2002). Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 45, 2615–2623. doi:10.1021/jm020017n
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. doi:10.1093/nar/gky427
- Welsch, M. E., Zhou, J., Gao, Y., Yan, Y., Porter, G., Agnihotri, G., et al. (2016). Discovery of potent and selective leads against *Toxoplasma gondii* dihydrofolate reductase via structure-based design. *ACS Med. Chem. Lett.* 7, 1124–1129. doi:10.1021/acsmchemlett.6b00328
- WHO (2017). World malaria report 2017 “ISBN 978-92-4-156552-3”. Available at: <https://www.who.int/malaria/publications/world-malaria-report-2017/report/en/>. (Accessed November 21, 2020).
- Zacharova, M. K., Tulloch, L. B., Gould, E. R., Fraser, A. L., King, E. F., Menzies, S. K., et al. (2019). Structure-based design, synthesis and biological evaluation of bis-tetrahydropyran furan acetogenin mimics targeting the trypanosomatid F1 component of ATP synthase. *Eur. J. Org. Chem.* 2019, 5434–5440. doi:10.1002/ejoc.201900541
- Zaware, N., Sharma, H., Yang, J., Devambatta, R. K., Queener, S. F., Anderson, K. S., et al. (2013). Discovery of potent and selective inhibitors of *Toxoplasma gondii* thymidylate synthase for opportunistic infections. *ACS Med. Chem. Lett.* 4, 1148–1151. doi:10.1021/ml400208v
- Zhang, P., Jia, L., Tian, Y., Xi, L., Duan, R., Chen, X., et al. (2019). Discovery of potential *Toxoplasma gondii* CDPK1 inhibitors with new scaffolds based on the combination of QSAR and scaffold-hopping method with *in-vitro* validation. *Chem. Biol. Drug Des.* 95, 476–484. doi:10.1111/cbdd.13603

Conflict of Interest: The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Copyright © 2021 Kleandrova, Scotti, Bezerra Mendonça Junior, Muratov, Scotti and Speck-Planche. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.