Results in Chemistry 4 (2022) 100256

Contents lists available at ScienceDirect

Results in Chemistry

journal homepage: www.sciencedirect.com/journal/results-in-chemistry

QSAR predictions on antichagas fenarimols

Pablo R. Duchowicz [a,*], Silvina E. Fioressi [b], Daniel E. Bacelo [b,*]

[a] Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA), CONICET, UNLP, Diag. 113 y 64,C.C. 16 Sucursal 4, 1900 La Plata, Argentina
[b] CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Belgrano, Villanueva 1324, CP 1426 Buenos Aires, Argentina

ARTICLE INFO

Keywords:
Chagas disease
Quantitative structure-activity relationships
Fenarimol compounds
Molecular descriptors

ABSTRACT

A useful QSAR model was developed to predict the antichagas activity for 760 fenarimol analogues obtained from the ChEMBL database, which are considered as very active and selective inhibitors of *Trypanosoma cruzi*. Various molecular descriptor programs provided a large number of 67,116 non-conformational molecular descriptors that were analyzed through multivariable linear regressions and the Replacement Method technique. Through THESE descriptors, the quantification of the structure–activity relationship achieves an acceptable statistical quality for compounds having experimental activity. The present work provides a prospective guide for predicting the inhibitory activity against *T. cruzi* of structurally-related fenarimol compounds.

Introduction

Chagas disease is endemic in Latin America affecting an estimated 8 million people [1] and is also present in Europe and in other developed countries such as Japan and USA [2,3]. The causative agent of Chagas is the protozoan parasite *Trypanosoma cruzi* (*T. cruzi*), which may infect humans as well as other mammals [4]. The parasite is vectored by several triatomine species and is also transmitted congenitally, through blood transfusions and organ transplantation from infected donors, and also by oral ingestion of contaminated food. The disease evolves from an initial acute stage to a chronic illness with high mortality [5]. The progress to the chronic phase takes years and occurs as parasites travel from the bloodstream to the heart and/or the gastrointestinal tract. Once there, the parasite affects their muscular cells causing extensive organ damage, being cardiac involvements most frequent and severe than gastrointestinal involvements.

People suffering Chagas disease need continuous medical therapy and require specialized and intensive healthcare [6]. Only two drugs are currently available for the treatment of *T. cruzi* infection: Benznidazole and Nifurtimox. Benznidazole is a nitroimidazole that is effective when administered during the acute phase of the disease [7]. It is generally well tolerated, without dangerous side effects, but its clinical efficacy in patients with chronic illness is limited and is contraindicated during pregnancy [8]. On the other hand, Nifurtimox is a less used drug because it causes severe side effects, making difficult to complete a continuous treatment and therefore, increasing the development of drug resistance [9].

It is clear that there is an urgent need for safe and efficacious new drug treatments for Chagas disease. Because it is a complex, long lasting disease that goes through different phases affecting different organs, it is very difficult, expensive and time consuming to find new potential candidates through experimental research only. The current access to a great diversity of libraries and data from pharmaceutical companies and the public domain offers a good opportunity for *in silico* methods to play a crucial role in the discovery of new strategies to prevent and treat Chagas. Recently, Andricopulo et al. published a QSAR study on 363 structurally diverse compounds with inhibitory activity towards *T. cruzi* through a artificial neural networks (ANNs) analysis [10]. The models found exhibited a good predictive ability for the test set and were used to delineate the physicochemical profile of 50 fragments that they found had a positive effect on the biological activity of the studied compounds.

In the last years, several azole-derived antifungals have been proposed as potential new drugs for treatment of *T. cruzi* infections. Clinical trials for Posaconazole and Ravuconazole demonstrate that those antifungals are safe but have poor sustained efficacy after one year of the end of treatment [11,12]. Fenarimol (Fig. 1), a nontoxic fungicide, is identified as a *T. cruzi* inhibitor with IC_{50} of 350 nM by Keenan and coworkers [13–16]. Several Fenarimol analogues are also proposed as potential inhibitors with low nM IC_{50} in *T. cruzi* whole cell *in vitro* assay. Searching for new compounds with increased inhibitory activity with the fenarimol basic structure, Costa et al. [17] propose a QSAR model using the 32 fenarimol derivatives reported in reference [13] and 70 quantum chemical descriptors. Based on their four descriptors model, ten new compounds with high predicted inhibitory activity were

* Corresponding authors.
E-mail addresses: pabloducho@gmail.com (P.R. Duchowicz), dbacelo@aol.com (D.E. Bacelo).

https://doi.org/10.1016/j.rechem.2021.100256
Received 15 September 2021; Received in revised form 14 November 2021; Accepted 29 November 2021
Available online 2 December 2021

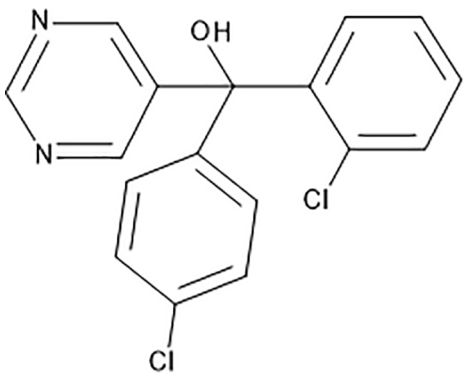


Fig. 1. Fenarimol.

proposed in this work. In other study, de Souza et al. [18] performed a 2D and 3D QSAR study on 77 compounds with molecular structures similar to those proposed by Keenan [13,16]. In this study, all the compounds included in the data set posses three aromatic rings, with variations in the substituents and also in some of the aromatic substructures. Both 2D and 3D QSAR models proposed showed statistically good performance with good correlation ability, and the authors conclude by analyzing the 2D model that all three aromatic rings are essential for conserving the inhibitory activity of the type of compounds studied. Just recently, Cotuá et al. published a QSAR study using 52 fenarimol compounds [19] also based on Keenan's work [15]. Derivatives of piperazine analogues of fenarimol with amide, sulfonamide, aromatic, carbamate, and carbonate substituents were used to construct a three descriptors QSAR model with good predictive performance.

Since Chagas is a disease associated with poverty, low-cost drugs are needed and the first step to achieve rapid and inexpensive development of a new drug is to theoretically test potential candidates. The Drugs for Neglected Diseases (DNDi) initiative has selected a large number of fenarimol derivatives as inhibitors of *T. cruzi* and these results were made public in the ChEMBL database [20] to be used in QSAR research intended at enhancing activity against *T. cruzi* and develop compounds suitable for in vivo characterization. Given the great potential of fenarimol compounds to effectively treat Chagas disease, with the aim of perform a comprehensive QSAR analysis that includes all the molecular information available in the literature, we performed a QSAR study on a great number of such analogues obtained from the ChEMBL database. In order to find a simple and easy-to-interpret QSAR model, only conformation-independent molecular descriptors were considered. The programs used to calculate the molecular descriptors were selected according to the Organization for Economic Cooperation and Development (OECD) criteria (calculation accuracy, ease of access, free availability and recognition by the scientific community) [21,22].

Materials and methods

The antichagas activity of a large set of fenarimol analogues was recopilated from the ChEMBL database version 23 [20]. After curation of the data to remove duplicate compounds and compounds with missing IC90 values, a molecular data set consisting of 760 active and selective inhibitors of T. cruzi with their corresponding bioactivity values (CHEMBL1863512) was obtained. These compounds are analogues of fenarimol, a plant fungicide studied by Keeman and coworkers [13–16]. In order to afford a more uniform data distribution the IC_{90} values were converted to decimal logarithmic units. The $logIC_{90}$ values of the compounds range in the interval (0.47, 5.00). Specific details of the studied compounds are provided in the Supplementary Material (Table 1S).

First of all, the molecules were written in SMILES notation and afterwards non-conformational molecular descriptors were calculated for considering the most important structural features affecting the inhibitory activity. The molecular descriptor programs were chosen as recommended by the OECD [22]. The Pharmaceutical Data Exploration Laboratory (PaDEL)-Descriptor (v. 2.20) [23] calculated 17,536 0D-2D molecular descriptors and fingerprint types, with molecular structures in MDL mol (V2000) format and selected options: standardize nitro groups and detect aromaticity. Chemical file format conversion into MDL sdf format was performed with Open Babel for Windows. Mold^2 [24] calculated 777 1D-2D molecular descriptors. The ISIDA/Fragmentor [25] counting molecular substructural fragments ranging from 1 to 6 atoms length, led to 1,084 constitutional descriptors. Finally, Molecular Descriptors from Local Vertex Invariants (MD-LOVIs) (v. 1.0) [26] was applied to compute 48,400 molecular descriptors obtained from local vertex invariants (LOVIs). The selected options used in this program were: total (global) and local (fragment-type) indexes, with atom labels: chemical properties - atomic number, van der Waals volume, polarizability, atomic mass, covalent radius, Pauling's electronegativity; physical properties – total polar surface area, AlogP, molar refraction, charge; vertex degree (vd) – valence degree, eccentric connectivity, electrotopological state, Kupchik's vd, intrinsic state, bond vd, Li's vd, Hu-Xu's vd, Alikhanidi vd, Ivanciuc vd, and distance counts. For the local indices, the following local types are kept: heteroatoms, C-atoms, halogens, H-atoms acceptor, H-atoms donor, methyl group, unsaturated bonds, aliphatic atoms, aromatic atoms, group_lagk (topological distances 1–8, with all the above mentioned group types HT-RA, cutoff 1). No standardized invariants were used. The Aggregation operators chosen were: Norms (metrics) – euclidean distance; Means (1st moment) – arithmetic mean; Statistics – standard deviation; Classical algorithms – autocorrelation, gravitational, Kier-Hall connectivity, total sum k lags, total information content, mean information content, standardized information content, electrotopological state, Ivanciuc Balaban type.

Among the computed 67,797 total descriptors, highly correlated descriptor pairs were identified and the most interpretable variable from each pair was kept. Moreover, descriptors with scarce information content and descriptors with missing values were removed, thus leading to a final reduced pool of 40,207 molecular descriptors.

The molecular set of 760 antichagasic agents was partitioned into training (train), validation (val), and test sets in similar proportions (about 33%) by means of the Balanced Subsets Method (BSM) (Aranda et al., 2017). The training set (253 compounds) is used for model calibration, the validation set for partially assessing the model's predictive capability (254 compounds), and a test set for assessing the true predictive power of the obtained QSAR (253 compounds). The BSM approach is based on k-Means Cluster Analysis (k-MCA) [27], and is a sampling procedure developed by our group to ensure that balanced (representative) subsets are derived from the dataset, in such a way that similar structure–activity relationships are established among the three sets.

Multivariable linear regression models were established through the Replacement Method (RM) variable subset selection technique [28]. The best molecular descriptors were searched with RM among the 40,208 available ones (including the best flexible descriptor), that lead to the smallest value for the root mean square error (*RMSE*) or the standard deviation (*S*) in the training set. Apart from using a test set, the model was further validated by means of theoretical validation parameters as proposed in the literature [29–31]. All the Octave programmed algorithms [32] used in the present study have been developed in-house and are available upon request.

Results and discussion

The best linear regression model found by RM through the analysis of 40,208 molecular descriptors is represented by Eq. (1). In the first place, this model was searched by decreasing *RMSE* in the training set, and in the second place by also decreasing *RMSE* in the validation set, in such a

way that both *RMSE*s achieve close values. The test set is never seen during the model building and the *RMSE* for this set is calculated only after the model is established. Seven non-conformational descriptors (with meanings described below) are able to predict the inhibitory concentration, as expressed by Eq. (1) together with the statistical quality achieved:

$$logIC_{90} = 0.70\,d_1 + 0.86\,d_2 - 0.79\,d_3 + 3.57\,d_4 - 0.62\,d_5 - 1.53\,d_6 + 0.06\,d_7 + 5.84 \quad (1)$$

$N_{train} = 253,\ R^2_{train} = 0.74,\ \mathrm{RMSE_{train}} = 0.44,\ \mathrm{R^2_{ijmax}} = 0.32$

$R^2_{loo} = 0.72,\ RMSE_{loo} = 0.46,\ R^2_{rand} = 0.03,\ RMSE_{rand} = 0.86\ (500,000\ \text{cases})$

$N_{val} = 254,\ R^2_{val} = 0.56,\ RMSE_{val} = 0.52$

$N_{test} = 253,\ R^2_{test} = 0.49,\ RMSE_{test} = 0.52$

The presence of seven molecular descriptors is necessary in Eq. (1) to predict the antichagas activity, which is justified through the maximum correlation coefficient between descriptor pairs of $R^2_{ijmax} = 0.32$ and the maximum variance inflation factor $VIF_{max} = 1.24$ [33]. These results suggest that no serious structural information overlapping exists between the descriptors, and that multicollinearity among them is absent. Table 1 includes the squared correlation matrix for this QSAR model.

In Eq. (1), the signs of both the regression coefficients and the numerical descriptors determine the magnitude of the predicted activity. The lower is the predicted IC_{90} arising from these contributions, the higher is the predicted antichagas activity for a given fenarimol compound. Table 1S also includes the $logIC_{90}$ predictions for the 760 fenarimols, indicating the molecules that take part of the validation and test sets according to the BSM technique. Table 2S provides the numerical values for the molecular descriptors.

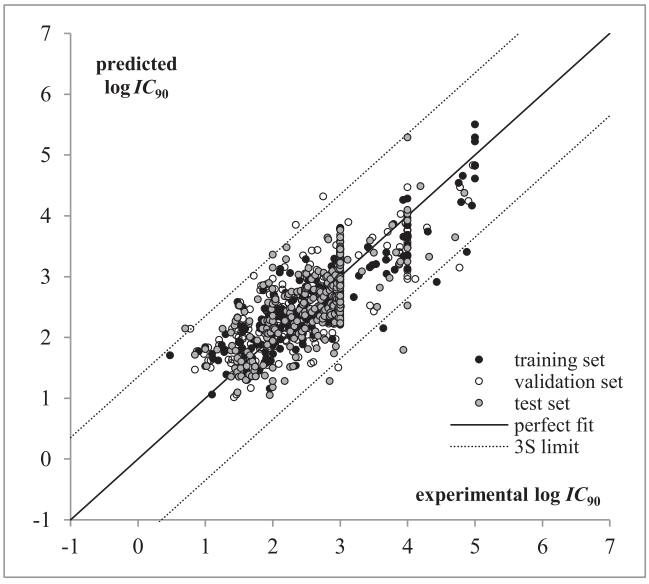
A graphical representation of Eq. (1) is given by Fig. 2A; only three outliers out of 253 training set compounds have high residuals, greater than three times S_{train} ($o3$). The seven chosen descriptors tend to provide a straight line trend, capturing the general biochemical behavior in the structurally diverse data set.

The proposed QSAR model acceptably predicts the biological activity for 253 fenarimols belonging to the test set not considered during the model development ($RMSE_{test} = 0.52$). The reliability of these QSAR predictions are estimated through the model's applicability domain (AD), which defines the molecules that are not predicted as extrapolations [34]. Fig. 2B plots the standardized residual of Eq. (1) as a function of the leverage value, revealing that most of the test set compounds belong to the AD, with leverages falling under the warning leverage limit of 0.0949. Some test set compounds exceed this warning value, although standing close to a training or validation compound.

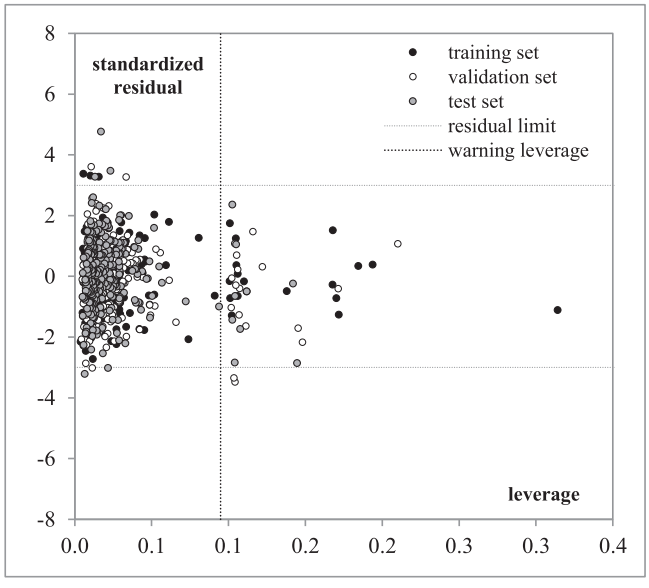
The theoretical validation process of the leave-one-out (loo) cross validation technique [29] succeeded in predicting one molecule excluded at a time from the training set, achieving $R^2_{loo} > 0.5$ which reveals the stability of Eq. (1). The MAE (mean absolute error) based criteria [31] achieves a 'moderate' predictive capacity for this relationship. In the case of considering the whole test set, $MAE(100\%) =$

Table 1
Squared correlation matrix for the QSAR model (Eq. (1)). The last column includes the variance inflation factor for each descriptor.

Eq. (1)	d_1	d_2	d_3	d_4	d_5	d_6	d_7	*VIF*
d_1	1.00	1.00 10^{-3}	0.01	3.50 10^{-3}	2.50 10^{-4}	1.83 10^{-7}	4.27 10^{-3}	1.00
d_2		1.00	0.01	0.01	2.72 10^{-3}	4.50 10^{-5}	0.04	1.00
d_3			1.00	1.05 10^{-4}	0.32	0.02	9.47 10^{-5}	1.17
d_4				1.00	0.04	0.02	0.07	1.03
d_5					1.00	0.03	0.05	1.24
d_6						1.00	0.10	1.02
d_7							1.00	1.09



A



B

Fig. 2. A. Experimental and QSAR predicted antichagas activity of fenarimol analogues. B. Standardized residual as a function of the leverage.

0.41, $Q^2_{F1}(100\%) = 0.44$ and $Q^2_{F2}(100\%) = 0.43$, while omitting 5% of the test set (13 compounds having high residuals), $MAE(95\%) = 0.36$, $Q^2_{F1}(95\%) = 0.58$ and $Q^2_{F2}(95\%) = 0.56$.

Finally, it was proven that Eq. (1) does not involve chance correlation through Y-Randomization [30], as $R^2_{rand} < R^2_{train}$ and $RMSE_{rand} > RMSE_{train}$. In this essay, the activity value was randomized (scrambled) for each analyzed case, while maintaining the selected descriptors (the regression coefficients are re-computed). The number of tested cases was 1,000,000 and the average R^2_{rand} and $RMSE_{rand}$ values are reported.

From the analysis of the seven molecular descriptors appearing in Eq. (1), three are constitutional descriptors, d_1–d_3, while four of them are topological indices, d_4–d_7. Descriptor d_1 indicates the presence of the atom nearest neighbor pattern C(:C)(:C)(:C), with bond aromaticity being denoted by ":" (Pubchem fingerprint number 385). Descriptor d_2 is the presence of the SMART substructure [!#1]c1[cH][cH][cH][cH]c1Cl (Klekota-Roth fingerprint number 1583). Descriptor d_3 is the minimum atom-type E-State > CH–. Descriptor d_4 is a total (global) descriptor that considers all the substructure contributions, while the Hu-Xu's vertex degree characterizes heteroatoms. The contributions from the mean information content are first derived, while the arithmetic mean is finally used as aggregation operator of these no-standardized LOVIs. Descriptor d_5 is a local index of the H-atoms acceptor, where atoms are differentiated through the Pauling's electronegativity and the aggregation operator used is the standard deviation. Descriptor d_6 is another local index for the aromatic atoms characterized though the Ivanciuc vertex degree, where the standardized information content is the aggregation operator. Descriptor d_7 is the optimized flexible molecular descriptor based on the number of paths of length 3 which are starting in a reference vertex of the structure.

Analysis of the descriptors shows that aromaticity plays a key role in the antichagas activity of the fenarimol analogs. Furthermore, the presence of heteroatoms could be an important molecular characteristic to consider when designing new potential candidates for drug development. Although the chemical interpretation of a constitutional descriptor is considered quite direct, the meaning of the selected topological descriptors remains hidden behind their algebraic definition. Nevertheless, both types of descriptors are useful and result important for predicting the antichagas activity of fenarimol compounds.

Conclusions

A multivariable linear regression QSAR model is able to predict the antichagas activity for 760 fenarimol analogues, compounds considered as very active and selective inhibitors. Seven descriptors acceptably quantify this structure–activity relationship, succeeding in generalizing and characterizing the biochemical behavior of the studied structurally diverse data set. Now, the derived QSAR can be applied to predict the antichagas activity of structurally similar fenarimol derivatives falling within the model's applicability domain. The model can be useful for the design of new compounds with potential for the development of novel and selective drugs, with improved efficiency for the treatment of chagas disease.

Research involving human participants and/or animals

Research does not involve human participants and/or animals.

CRediT authorship contribution statement

Pablo R. Duchowicz: Conceptualization, Methodology, Software, Writing - review & editing. **Silvina E. Fioressi:** Conceptualization, Methodology, Data curation, Writing - review & editing. **Daniel E. Bacelo:** Conceptualization, Methodology, Data curation, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank the financial support provided by the National Research Council of Argentina (CONICET) PIP11220130100311 project and to the Ministerio de Ciencia, Tecnología e Innovación Productiva for the electronic library facilities.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rechem.2021.100256.

References

[1] WHO (World Health Organization) Weekly Epidemiological Record. Chagas disease in Latin America: an epidemiological update based on 2010 estimates, 2015. <https://www.who.int/wer/2015/wer9006.pdf>. (accessed 06/04 2021).
[2] G. Turabelidze, A. Vasudevan, C. Rojas-Moreno, S.P. Montgomery, M. Baker, D. Pratt, S. Enyeart, Autochthonous Chagas disease—Missouri, 2018, Morbidity Mortality Wkly. Rep. 69 (7) (2020) 193–195.
[3] H.B. Tanowitz, L.M. Weiss, S.P. Montgomery, Chagas disease has now gone global, PLoS Negl. Trop. Dis. 5 (4) (2011), e1136, https://doi.org/10.1371/journal.pntd.0001136.
[4] A.C. Meyers, J.C. Purnell, M.M. Ellis, L.D. Auckland, M. Meinders, S.A. Hamer, Nationwide exposure of US working dogs to the chagas disease parasite, Trypanosoma cruzi, Am. J. Trop. Med. Hyg. 102 (5) (2020) 1078–1085, https://doi.org/10.4269/ajtmh.19-0582.
[5] J.A. Pérez-Molina, I. Molina, Chagas disease, The Lancet 391 (10115) (2018) 82–94, https://doi.org/10.1016/S0140-6736(17)31612-4.
[6] M.C.P. Nunes, W. Dones, C.A. Morillo, J.J. Encina, A.L. Ribeiro, Cardiology, chagas disease: an overview of clinical and epidemiological aspects, J. Am. Coll. Cardiol. 62 (9) (2013) 767–776, https://doi.org/10.1016/j.jacc.2013.05.046.
[7] A.L.S.S. de Andrade, F. Zicker, R.M. de Oliveira, S.A. e Silva, A. Luquetti, L. R. Travassos, I.C. Almeida, S.S. de Andrade, J.G. de Andrade, C.M. Martelli, Randomised trial of efficacy of benznidazole in treatment of early Trypanosoma cruzi infection, The Lancet 348 (9039) (1996) 1407–1413, https://doi.org/10.1016/s0140-6736(96)04128-1.
[8] S. Meymandi, S. Hernandez, S. Park, D.R. Sanchez, C. Forsyth, Treatment of Chagas disease in the United States, Curr. Treat. Opt. Infect. Dis. 10 (3) (2018) 373–388, https://doi.org/10.1007/s40506-018-0170-z.
[9] C. Muñoz, I. Zulantay, W. Apt, S. Ortiz, A.G. Schijman, M. Bisio, V. Ferrada, C. Herrera, G. Martínez, A. Solari, Evaluation of nifurtimox treatment of chronic Chagas disease by means of several parasitological methods, Antimicrob. Agents Chemother. 57 (9) (2013) 4518–4523, https://doi.org/10.1128/AAC.00227-13.
[10] A.S. de Souza, L.L. Ferreira, A.S. de Oliveira, A.D. Andricopulo, Quantitative structure-activity relationships for structurally diverse chemotypes having anti-trypanosoma cruzi activity, Int. J. Mol. Sci. 20 (11) (2019) 2801, https://doi.org/10.3390/ijms20112801.
[11] B.P. Olivieri, J.T. Molina, S.L. de Castro, M.C. Pereira, C.M. Calvet, J.A. Urbina, T. C. Araújo-Jorge, A comparative study of posaconazole and benznidazole in the prevention of heart damage and promotion of trypanocidal immune response in a murine model of Chagas disease, Int. J. Antimicrob. Agents 36 (1) (2010) 79–83, https://doi.org/10.1016/j.ijantimicag.2010.03.006.
[12] E. Chatelain, Chagas disease drug discovery: toward a new era, J. Biomol. Screen. 20 (1) (2015) 22–35, https://doi.org/10.1177/1087057114550585.
[13] M. Keenan, M.J. Abbott, P.W. Alexander, T. Armstrong, W.M. Best, B. Berven, A. Botero, J.H. Chaplin, S.A. Charman, E. Chatelain, Analogues of fenarimol are potent inhibitors of Trypanosoma cruzi and are efficacious in a murine model of Chagas disease, J. Med. Chem. 55 (9) (2012) 4189–4204, https://doi.org/10.1021/jm2015809.
[14] T.Y. Hargrove, Z. Wawrzak, P.W. Alexander, J.H. Chaplin, M. Keenan, S. A. Charman, C.J. Perez, M.R. Waterman, E. Chatelain, G.I. Lepesheva, Complexes of Trypanosoma cruzi sterol 14α-demethylase (CYP51) with two pyridine-based drug candidates for Chagas disease: structural basis for pathogen selectivity, J. Biol. Chem. 288 (44) (2013) 31602–31615, https://doi.org/10.1074/jbc.M113.497990.
[15] M. Keenan, P.W. Alexander, H. Diao, W.M. Best, A. Khong, M. Kerfoot, R. A. Thompson, K.L. White, D.M. Shackleford, E. Ryan, Design, structure–activity relationship and in vivo efficacy of piperazine analogues of fenarimol as inhibitors of Trypanosoma cruzi, Biorg. Med. Chem. 21 (7) (2013) 1756–1763, https://doi.org/10.1016/j.bmc.2013.01.050.
[16] M. Keenan, J.H. Chaplin, P.W. Alexander, M.J. Abbott, W.M. Best, A. Khong, A. Botero, C. Perez, S. Cornwall, R.A. Thompson, K.L. White, D.M. Shackleford, M. Koltun, F.C.K. Chiu, J. Morizzi, E. Ryan, M. Campbell, T.W. von Geldern,

I. Scandale, E. Chatelain, S.A. Charman, Two analogues of fenarimol show curative activity in an experimental model of Chagas disease, J. Med. Chem. 56 (24) (2013) 10158–10170, https://doi.org/10.1021/jm401610c.
[17] C.H. Costa, A.R. Oliveira, M.J. Pereira, A.F. Figueiredo, J.E. Ferreira, R. M. Miranda, Quantum chemistry, quantitative structure-activity relationship and molecular docking study on fenarimol derivatives with biological activity against chagas disease, J. Comput. Theor. Nanosci. 12 (10) (2015) 3309–3321, https://doi.org/10.1166/jctn.2015.4119.
[18] A.S. De Souza, L.G. Ferreira, A.D. Andricopulo, 2D and 3D QSAR studies on a series of antichagasic fenarimol derivatives, Int. J. Quant. Struct.-Prop. Relat. 2 (1) (2017) 44–63, https://doi.org/10.4018/IJQSPR.2017010104.
[19] V. Cotuá, S. LLinás, O. Cotes, Virtual Screening Based on QSAR and Molecular Docking of Possible Inhibitors Targeting Chagas CYP51, Journal of Chemistry 2021 (2021), https://doi.org/10.1155/2021/6640624.
[20] D. Mendez, A. Gaulton, A.P. Bento, J. Chambers, M. De Veij, E. Félix, M. P. Magariños, J.F. Mosquera, P. Mutowo, M. Nowotka, ChEMBL: towards direct deposition of bioassay data, Nucleic Acids Res. 47 (D1) (2019) D930–D940, https://doi.org/10.1093/nar/gky1075.
[21] OECD Guidance Document On The Validation Of (Quantitative) Structure-Activity Relationship [(Q)Sar] Models. <http://www.oecd.org/dataoecd/55/35/38130292.pdf>. (accessed 5 April 2021).
[22] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci. 26 (5) (2007) 694–701.
[23] C.W. Yap, PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints, J. Comput. Chem. 32 (7) (2011) 1466–1474, https://doi.org/10.1002/jcc.21707.
[24] H. Hong, Q. Xie, W. Ge, F. Qian, H. Fang, L. Shi, Z. Su, R. Perkins, W. Tong, Mold2, molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics, J. Chem. Inf. Model. 48 (7) (2008) 1337–1344, https://doi.org/10.1021/ci800038f.
[25] F. Ruggiu, G. Marcou, V. Solov'ev, D. Horvath, A. Varnek, ISIDA fragmentor 2017. <http://complex-matter.unistra.fr/equipes-de-recherche/laboratoire-de-chemoinformatique/home>. (accessed April 13 2020).
[26] Y. Martínez-López, Y. Marrero-Ponce, S.J. Barigye, E. Teran, O. Martínez-Santiago, C.H. Zambrano, F.J. Torres, When global and local molecular descriptors are more than the sum of its parts: Simple, But Not Simpler? Mol. Divers. 24 (4) (2020) 913–932, https://doi.org/10.1007/s11030-019-10002-3.
[27] L. Kaufman, P.J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, John Wiley & Sons, 2009.
[28] P.R. Duchowicz, E.A. Castro, F.M. Fernández, Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR studies, MATCH Commun. Math. Comput. Chem 55 (2006) 179–192.
[29] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graphics Modell. 20 (4) (2002) 269–276, https://doi.org/10.1016/s1093-3263(01)00123-1.
[30] S. Wold, L. Eriksson, S. Clementi, Statistical validation of QSAR results, in: D.H.v. d. Waterbeemd (Ed.), Chemometric Methods in Molecular Design, VCH Verlagsgesellschaft mbH, Weinheim, 2008, pp. 309–338.
[31] K. Roy, R.N. Das, P. Ambure, R.B. Aher, Be aware of error measures. Further studies on validation of predictive QSAR models, Chemom. Intell. Lab. Syst. 152 (2016) 18–33, https://doi.org/10.1016/j.chemolab.2016.01.008.
[32] J.W. Eaton, D. Bateman, S. Hauberg, R. Wehbring, GNU Octave Version 5.1.0 2020. <https://www.gnu.org/software/octave/>. (accessed March 10, 2020).
[33] K. Roy, P.P. Roy, Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA, G/PLS and ANN techniques, Eur. J. Med. Chem. 44 (7) (2009) 2913–2922, https://doi.org/10.1016/j.ejmech.2008.12.004.
[34] D. Gadaleta, G.F. Mangiatordi, M. Catto, A. Carotti, O. Nicolotti, Applicability domain for QSAR models: where theory meets reality, Int. J. Quant. Struct.-Prop. Relat. 1 (1) (2016) 45–63, https://doi.org/10.4018/IJQSPR.2016010102.