



# Ensemble learning application to discover new trypanothione synthetase inhibitors

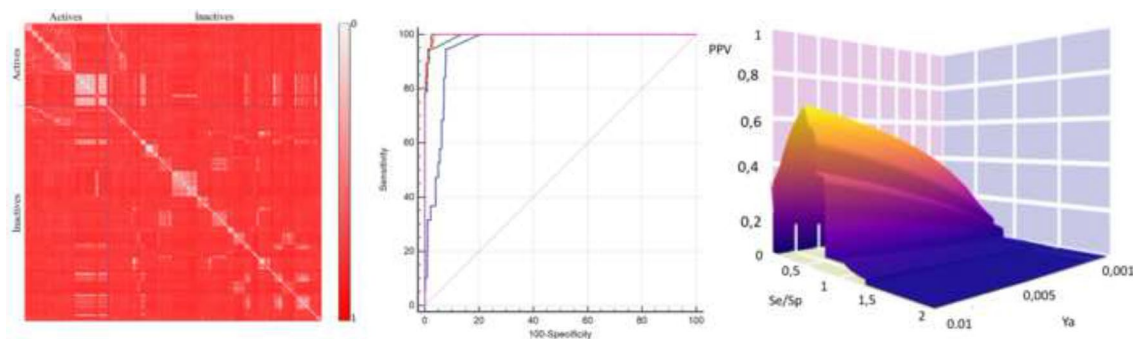
Juan I. Alice<sup>1,2</sup> · Carolina L. Bellera<sup>1,2</sup> · Diego Benítez<sup>3</sup> · Marcelo A. Comini<sup>3</sup> · Pablo R. Duchowicz<sup>4</sup> · Alan Talevi<sup>1,2</sup>

Received: 12 April 2021 / Accepted: 24 June 2021 / Published online: 15 July 2021  
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

## Abstract

Trypanosomatid-caused diseases are among the neglected infectious diseases with the highest disease burden, affecting about 27 million people worldwide and, in particular, socio-economically vulnerable populations. Trypanothione synthetase (TryS) is considered one of the most attractive drug targets within the thiol-polyamine metabolism of trypanosomatids, being unique, essential and druggable. Here, we have compiled a dataset of 401 *T. brucei* TryS inhibitors that includes compounds with inhibitory data reported in the literature, but also in-house acquired data. QSAR classifiers were derived and validated from such dataset, using publicly available and open-source software, thus assuring the portability of the obtained models. The performance and robustness of the resulting models were substantially improved through ensemble learning. The performance of the individual models and the model ensembles was further assessed through retrospective virtual screening campaigns. At last, as an application example, the chosen model-ensemble has been applied in a prospective virtual screening campaign on DrugBank 5.1.6 compound library. All the *in-house* scripts used in this study are available on request, whereas the dataset has been included as supplementary material.

## Graphic abstract



**Keywords** Ensemble learning · Machine learning · QSAR · *Trypanosoma cruzi* · Chagas disease · Trypanothione synthetase

✉ Alan Talevi  
alantalevi@gmail.com

<sup>1</sup> Laboratorio de Investigación y Desarrollo de Bioactivos (LIDeB), Facultad de Ciencias Exactas, Universidad Nacional de La Plata (UNLP), La Plata, Buenos Aires, Argentina

<sup>2</sup> Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), CCT La Plata, La Plata, Argentina

<sup>3</sup> Group Redox Biology of Trypanosomes, Institut Pasteur Montevideo, Montevideo, Uruguay

<sup>4</sup> Facultad de Ciencias Exactas, Universidad Nacional de La Plata (UNLP)-Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Instituto de Investigaciones Físicoquímicas Teóricas y Aplicadas (INIFTA), La Plata, Argentina

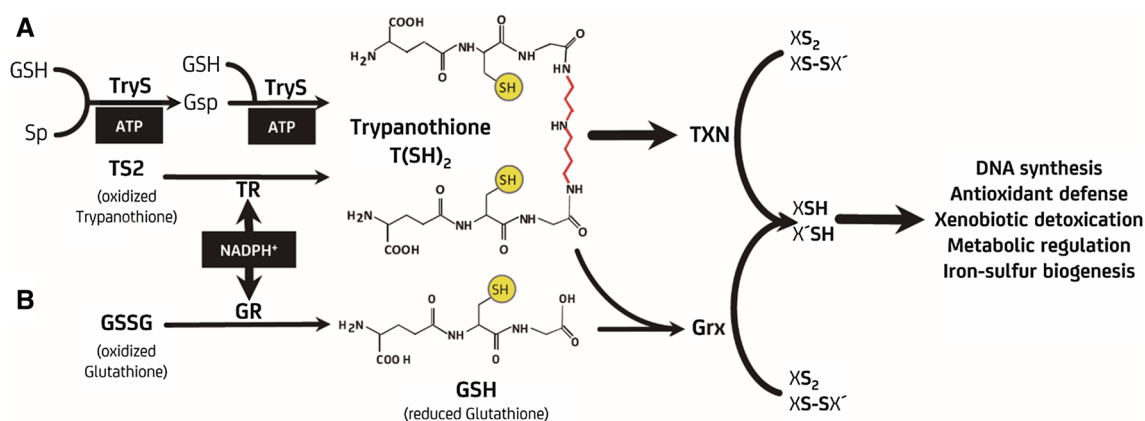
## Introduction

Protozoa from the trypanosomatid lineage (which include leishmaniasis, Chagas disease and African sleeping sickness) affect more than 27 million people worldwide and are the etiologic agents of some of the neglected diseases with the highest disease burden [1]. Their treatments are exclusively dependent on chemotherapy since no vaccines are available. They primarily affect populations living in conditions of socio-economic vulnerability and limited access to healthcare facilities. For example, only 10% of the 6–7 million people suffering from Chagas disease are diagnosed, and less than 1% have access to treatment [2]. The available treatment relies on only two drugs, benznidazole and nifurtimox, which were developed more than 50 years ago. Both drugs are primarily used for treating acute cases, as their efficacy in the chronic stage of the disease is limited, or at least controversial [3–5]. Altogether, this highlights the historical lack of investment in novel therapeutic solutions to fight neglected conditions.

The thiol-polyamine metabolism of trypanosomatids has long been regarded as a suitable pharmacological target because of its indispensable role in redox homeostasis and the uniqueness of several of its molecular components [6]. In these organisms, the low-molecular mass thiol-polyamine conjugate trypanothione (bisglutathionyl spermidine) [7] takes over most redox functions carried out by glutathione in other living organisms [8, 9]

(Fig. 1). Trypanothione and the enzyme in charge of its synthesis, namely trypanothione synthetase (TryS), are absent in mammals, which turns the latter an attractive drug target for the development of selective therapeutic agents against trypanosomatid-caused conditions. In this regard, the essentiality and druggability of TryS have been genetically and chemically verified in the major pathogenic species [10–12]. Moreover, metabolic control analysis positions TryS among the pathway components of the trypanothione-dependent redox system most vulnerable to therapeutic intervention [13]. This study predicted that a reduction of 50% in the trypanothione synthesis flux can be achieved by partial (63%) inhibition of TryS.

TryS inhibitors have been reported using experimental approaches that included the development of substrate-like and mechanistic-based inhibitors [14–16] as well as small to large random or scaffold-focused screening campaigns [17, 18]. So far, there are no reports on the identification of TryS ligands/inhibitors using computational or structure-based strategies. Both approaches are commonly applied in the field of medicinal chemistry and represent powerful tools that may speed-up and increase the chances of identifying novel molecular entities with drug-like properties and TryS inhibitory activity. A consistent *in silico* screening strategy demands the development of robust analytical tools to mine chemical libraries and predict with high confidence the inhibitory potential of a molecule against a specific molecular target.



**Fig. 1** Schematic representation of the low-molecular mass thiol-dependent redox systems of trypanosomatids and mammals. **a** The low-molecular mass thiol-dependent redox system of trypanosomatids depends on bisglutathionylspermidine (trypanothione), which is stepwise synthesized by trypanothione synthetase (TryS) at the expense of ATP and the substrates glutathione (GSH), spermidine (Sp) and monoglutathionylspermidine (Gsp). The NADPH<sup>+</sup>-dependent enzyme trypanothione reductase (TR) recycles reduced trypanothione (T(SH)<sub>2</sub>) from its oxidized/disulfide form (TS<sub>2</sub>). The oxidoreductase trypanoperoxidase (TXN) and, to a minor extent, glutaredoxin (Grx) receive reducing equivalents from T(SH)<sub>2</sub>. Reduced TXN

and Grx catalyze the reduction of several protein and non-protein homo- or hetero-disulfides (XS<sub>2</sub> and XS-SX<sup>•</sup>) that fulfil different and indispensable functions in the cells. **b** Mammals lacks TryS and, hence, trypanothione, and their low-molecular mass thiol-dependent redox system relies on the use of the tripeptide glutathione. Glutathione is maintained in its reduced form (GSH) by the action of the NADPH<sup>+</sup>-dependent enzyme glutathione reductase (GR), which is absent in trypanosomatids. Grx uses GSH as redox cofactor to catalyze the oxidoreduction of different protein targets engaged in the regulation of important cellular processes

Thus, as a first step toward the identification of new TryS inhibitors, we report the development and validation of TryS ligand-based models. The individual models were subsequently combined through ensemble learning to improve their robustness and predictive ability. Noteworthy, the models were inferred from a dataset combining compounds with activity data retrieved from the literature and from results of TryS assays measured in house. A prospective screening campaign on the DrugBank database was performed applying the model ensemble that showed the best performance.

## Materials and methods

### Dataset collection

A dataset of 401 compounds previously tested against *Trypanosoma brucei* TryS was collected: One hundred and forty-one of such compounds were retrieved from the literature [10, 12, 14–22], whereas the data on the remaining 260 compounds were acquired in house, as described earlier [23]. Standardized molecular representations of the dataset compounds were initially obtained using Instant Jchem's Standardizer (v 16.10.10.0) (ChemAxon), applying the following commands: *Clean 2D*, *Aromatize*, *Strip salts*, *Clear stereo*, *Remove absolute stereo*, *Remove solvents*, and *Add explicit hydrogens*. The compounds were flagged as either ACTIVE

or INACTIVE based on their experimentally observed inhibitory data. Those that inhibited TryS by more than 50% at concentrations  $\leq 30 \mu\text{M}$  were assigned to the ACTIVE class; if not, they were assigned to the INACTIVE class. Considering such criterion, the dataset comprised 111 active compounds and 290 inactive compounds. The diversity of the entire dataset and within each class is graphically displayed in the heatmap in Fig. 2. The heatmap exhibits, for each possible pair of compounds, the Tanimoto distances computed using ECFP<sub>4</sub> molecular fingerprints. The heatmap was built using Gitoools v. 2.3.1 [24] and Tanimoto distances were calculated using ScreenMD—Molecular Descriptor Screening v. 17.3.27.0 (ChemAxon). The dataset has been included as Supplementary Material (Data Sheet 1).

### Molecular descriptors calculation

A set containing 66,128 conformation-independent descriptors of the TryS datasets was calculated using various freely available and open-source software packages, as described below, with compounds provided in MDL sdf format.

The Pharmaceutical Data Exploration Laboratory (PaDEL)-Descriptor (v. 2.20) [25] was utilized to compute the values of 17,536 0D-2D molecular descriptors and fingerprint types, using the options *Standardize nitro groups* and *Detect aromaticity*. Mold<sup>2</sup> [26] calculates 777 1D-2D

**Fig. 2** Dissimilarity heatmap for the dataset compounds. Red areas indicate the pair of compounds under comparison exhibits high dissimilarity, whereas white areas indicate the opposite



molecular descriptors. The ISIDA/Fragmentor [27] counts atom types and linear structural fragments ranging from 2- to 6-atom lengths. In the present study, 2,487 atom and fragment types were computed. Finally, Molecular Descriptors from Local Vertex Invariants (MD-LOVIs) (v. 1.0) were applied to compute molecular descriptors from local vertex invariants (LOVIs) [28]. The selected options were: *total* (global) and *local* (fragment-type) indexes, with atoms labels: chemical properties—atomic number (Z), polarizability (P), van der Waals volume (VW), Pauling's electronegativity (E), AlogP (L), atomic mass (A), covalent radius (R), total polar surface area (T), charge (C), molar refraction (M); vertex degree (N), bond vertex degree (BD), intrinsic state (I), eccentric connectivity (Y), electrotopological state (S), Kupchik's vertex degree (KU), Li's vertex degree (LI), Hu-Xu's vertex degree (HX), Ivanciuc vertex degree (IN), Alikhanidi vertex degree (Alk), and distance counts (DC). No standardized invariants were used. The *Aggregation operators* chosen were: norms (metrics)—Euclidean distance (N2); mean first moment — arithmetic mean (M); statistics—standard deviation (DE); classical algorithms — autocorrelation (AC(i)), gravitational (GI(i)), Kier-Hall connectivity (CN(i)), total sum k lags (TS(i)), total information content (TI), mean information content (MI(i)), standardized information content (SI), electrotopological state (ES), Ivanciuc Balaban type (IB). For the local indices, the following local types were kept: heteroatoms (HT), C-atoms (CB), halogens (HL), H-atoms acceptor (HA), H-atoms donor (HD), methyl group (MD), unsaturated bonds (IS), aliphatic atoms (LA), aromatic atoms (RA), group\_lagk (GL) (topological distances 1–8, with all the above-mentioned group types HT-RA, cut-off 1). These options led to 48,400 molecular descriptors of the LOVIs type.

## Molecular descriptors pruning

The 66,128 molecular descriptors computed were analyzed to eliminate collinear descriptors with redundant information. Thus, highly correlated descriptor pairs were identified (descriptors pairs with maximum squared correlation coefficient  $R_{ij}^{2\max} = 1$ ), and only the most interpretable variable from each pair was maintained for further analysis. Moreover, descriptors with scarce information content (those with constant and near-constant values) and descriptors with missing values were also eliminated. This procedure led to a final descriptor pool comprising 44,209 linearly independent descriptors.

## Dataset partition

The molecular set was partitioned into training, validation, and test sets by means of the Balanced Subsets Method (BSM) [29]. The training set (train) includes compounds that are used for model calibration purposes, the validation set (val) comprises instances used for internal model validation, while the test set (test) comprises compounds completely independent from the calibration procedure, which help to assess the true predictive power of the obtained QSAR.

The BSM approach is a sampling procedure that was developed by our group to ensure that balanced, representative subsets are derived from the dataset, in such a way that training, validation and test instances are not chosen randomly but provide similar structure–activity relationships within each subset. It is based on k-means cluster analysis (k-MCA) [30]: in essence, k-MCA creates k groups of compounds in such manner that the compounds within a given cluster are very similar in terms of a distance metric (here, Euclidean distance), whereas compounds in different clusters are very dissimilar.

## Variable subset selection

The Replacement Method (RM) [31] has been devised in our group as an efficient variable selection approach to build multiple linear regression models from training samples, by searching for a subset comprising  $d$  descriptors through a large pool of  $D$  descriptors ( $d$  being much lower than  $D$ ). In the current study, the RM technique has been modified in order to be applicable to classification problems. Therefore, the highest value of the non-error rate ( $NER$ ) or average sensitivity [32] has been searched in the training set. The quality of the models built by this method is close to that of the models obtained through a full search of all the possible combinations of  $d$  independent variables from a pool of  $D$  variables.

The fundamental principle underlying the RM is that one can approach the maximum of  $NER_{train}$  by sensibly considering the relative errors of the coefficients of the least-squares model provided by the subset of  $d$  descriptors. This is to say, the global maximum of  $NER_{train}(d)$  will be pursued in a subspace of  $D!/ [d!(D-d)!]$  possible classifiers.

All the Octave [33] programmed algorithms used for this study have been developed *in-house* and are available to the reader upon request.

## Retrospective screening campaigns

By conducting thorough simulated ranking experiments, Truchon and Bayly [34] have formerly proven that the area under the receiver operating characteristic curve (AUC ROC), a commonly used enrichment metric to evaluate and compare the performance of virtual screening methods, depends on the proportion of active compounds in the screened library, and that the standard deviation of this metric tends to a constant value when a small proportion of active compounds (or yield of active compounds,  $Y_a$ ) is present in the screened library ( $Y_a$  below 0.05 seems to provide robust results). What is more, a small  $Y_a$  avoids or at least limits the saturation effect. Accordingly, a large number of decoys (about 1,000 or above) and a small  $Y_a$  contribute to a controlled statistical behavior.

Therefore, a retrospective virtual screening experiment was implemented for further validation of the individual classifiers (which were built as described in the preceding subsections). A library for such retrospective screening campaign, which will be termed DUDE-1, was obtained by seeding 18 known active compounds among 913 decoys retrieved from the Directory of Useful Decoys Enhanced [35]), a broadly used benchmarking application used to obtain putative inactive compounds. Besides assessing the performance of the individual classifiers, DUDE-1 was used to train classifier ensembles (see next subsection). A second retrospective screening library (DUDE-2), obtained in a similar manner than DUDE-1 and comprising 18 active compounds and 872 decoys was utilized to independently estimate the performance of the best model ensemble.

## Ensemble learning

Generally speaking, it is recognized that classifier ensembles (meta-classifiers) tend to enhance the predictivity and robustness in comparison with individual (single model) classifiers [36, 37] and it may be useful to mitigate the impact of noisy data [38].

In the present study, the best individual classifiers were hence systematically combined using the AUC ROC in the first retrospective screen (DUDE-1) as criterion of performance. To choose the optimal number of individual classifiers to be included in the ensemble, the performance of systematic combinations of the 2 to 10 best-performing individual classifiers was assessed (the two best-performing classifiers were assembled, then the three best-performing

classifiers, and so on up to a total of 10 classifiers). Four schemes have been used to get a combined score: Average Score (AVE); Average Ranking (RANK), MIN operator (which considers the minimum score across the scores of the individual classifiers included in an ensemble) and Average Voting (VOT). Voting was computed as formerly described by Zhang and Muegge [39]. AUC ROCs were computed with the pROC package [40] and Delong's method was applied for the statistical comparison of the AUC ROC using the open-source application Rocker [41]. BEDROC ( $\alpha = 20$ ) and the enrichment factor in the top-ranked 1% of the libraries (EF1%) were also computed.

## Use of positive predictive value surfaces to guide the choice of a score cut-off value

A practical concern in any virtual screening campaign relates to estimating the probability that a predicted *in silico* hit will corroborate its activity when subject to experimental testing (such probability is termed Positive Predictive Value, *PPV*, also known as *precision*). Still, prospective estimation of such probability is not feasible owing to its dependency on the yield of active compounds of the screened library, which is *a priori* unknown in a prospective (real) virtual screening experiment:

$$PPV = \frac{Se \times Y_a}{Se Y_a + (1 - Sp)(1 - Y_a)} \quad (1)$$

where *Se* symbolizes the sensitivity associated to a given score cut-off value, and *Sp* denotes the specificity. As in previous reports (see, for instance, [42] and [43]), we applied Eq. (1) to build *PPV* surfaces. In order to decide on an optimal score cut-off value to later select hits in prospective virtual screening experiments, 3D plots displaying the interplay between the *Se/Sp* ratio,  $Y_a$  and *PPV* were generated for the best individual classifier and for every classifier ensemble. Using the library built for the retrospective screening experiment, *Se* and *Sp* were calculated in all the range of possible score cut-off values. Because controlled statistical behavior has been observed for libraries of about 1,000 compounds or more and  $Y_a$  below 0.05, we may reasonably assume that the ROC curve and its derived metrics will be similar when applying the classifiers or meta-classifiers to screen other chemical libraries with similar features. Bearing in mind that in prospective virtual screening experiments  $Y_a$  is ignored beforehand but always low, the hypothetical yield of active compounds has been varied between 0.1% and 1%. *PPV* graphs were obtained via the R package *plotly*. *PPV* surfaces were visually inspected to select a score cut-off associated to the required *PPV* range.

## Prospective virtual screening

The best model ensemble was applied to screen 10,759 compounds from DrugBank 5.1.6, an online database containing extensive information about approved, withdrawn, experimental and investigational small drugs and biologics, as well as nutraceuticals and illicit drugs [44]. DrugBank is commonly utilized for computer-guided drug repurposing campaigns. The screened compounds were pre-processed using Standardizer 16.10.10.0 (ChemAxon). The following actions were applied to obtain standardized representations of the molecular structure for the subsequent in silico screen: (1) *Strip salts*; (2) *Remove Solvents*; (3) *Clear Stereo*; (4) *Remove Absolute Stereo*; (5) *Aromatize*; (6) *Neutralize*; (7) *Add Explicit Hydrogens*; and (8) *Clean 2D*.

Whether each screened compound belonged or not to the applicability domain of the models included in the best ensemble was estimated through the extent of extrapolation approach [45], with the warning leverage fixed at  $2k/n$ ,  $k$  being the number of descriptors in the model and  $n$  being the number of training set examples.

## Results and discussion

The BSM procedure led to a dataset partition with 74:73:254 representative compounds in the training, validation, and test sets, respectively. The training and validation sets were sampled in such a way that they contain the same number of active and inactive compounds. Thus, active compounds were in a 37:37:37 proportion, while inactive compounds are in a 37:36:217 proportion.

Equation (2) presents the best-performing individual classifier (AUC ROC = 0.954, standard deviation = 0.0099):

$$\text{Class} = -1.63 + 0.49 * \text{Frag43} + 1.13 * \text{Frag295} + 0.07 * \text{ATSC5p} + 2.05 * \\ 1\text{CN} - \text{NE}(\text{DE}) - \text{E} - \text{WH} - \text{HT} + 2.05 * 2\text{TS} - \text{NE}(\text{DE}) - \text{R} - \text{WH} - \text{GL}_{1\_2\_3\_4\_5\_6\_7\_8} \quad (2)$$

Table 1 includes the classification results for the best models established in the present QSAR study through the RM technique; here, different classification measures are reported: the non-error rate (*NER*), accuracy (*Acc*) and the Matthews correlation coefficient (*MCC*).

The best model is highlighted in Table 1 (M7), which had the best performance in the validation set for the *NER* parameter. This QSAR classification model involves five conformation-independent molecular descriptors selected out of 44,209 variables: two fragment descriptors: *Frag43*, the count of C–N–C\*C\*C–Cl fragments (where \* denotes aromatic), and *Frag295*, the count of C\*N\*N–C–C–N fragments; a 2D-autocorrelation descriptor: *ATSC5p*, the centered Broto-Moreau autocorrelation-lag 5/weighted by polarizabilities; and two algebraic MD-LOVIs indices: *1CN-NE(DE)-E-WH-HT* and *2TS-NE(DE)-R-WH-GL<sub>1\_2\_3\_4\_5\_6\_7\_8</sub>*.

Both *Frag43* and *Frag295* descriptors are intuitive for their structural interpretation.

The *ATSC5p* descriptor is an autocorrelation function between selected atom pairs of the molecule, with the main purpose of capturing the degree of interaction between atoms at a given topological distance (lag). The contributions to this bidimensional autocorrelation are obtained by considering at lag 5 the atomic polarizabilities involved.

The two MD-LOVIs local descriptors represent an alternative strategy that generalizes the traditional method of obtaining a topological descriptor by summation of the local vertex invariants. The first one is *1CN-NE(DE)-E-WH-HT*, which describes heteroatoms, and the second one is *2TS-NE(DE)-R-WH-GL<sub>1\_2\_3\_4\_5\_6\_7\_8</sub>*, describing group lags 1–8. The hydrogen-filled graph is used as molecular representation in both cases. The Pauling's electronegativity (*E*) and the covalent radius (*R*) are used, respectively, as

**Table 1** QSAR classification results for the TryS inhibitors dataset. The best model appears in bold

Model	<i>d</i>	<i>NER<sub>train</sub></i>	<i>NER<sub>val</sub></i>	<i>NER<sub>test</sub></i>	<i>Acc<sub>train</sub></i>	<i>Acc<sub>val</sub></i>	<i>Acc<sub>test</sub></i>	<i>MCC<sub>train</sub></i>	<i>MCC<sub>val</sub></i>	<i>MCC<sub>test</sub></i>
M1	2	0.89	0.89	0.87	0.89	0.89	0.86	0.78	0.79	0.60
M2	3	0.91	0.89	0.87	0.91	0.89	0.86	0.81	0.79	0.60
M3	3	0.89	0.89	0.83	0.89	0.89	0.84	0.78	0.79	0.54
M4	4	0.96	0.90	0.86	0.96	0.90	0.82	0.92	0.82	0.56
M5	4	0.97	0.89	0.85	0.97	0.89	0.82	0.95	0.79	0.54
M6	4	0.96	0.89	0.85	0.96	0.89	0.82	0.92	0.79	0.55
M7	<b>5</b>	<b>0.95</b>	<b>0.93</b>	<b>0.90</b>	<b>0.95</b>	<b>0.93</b>	<b>0.89</b>	<b>0.89</b>	<b>0.86</b>	<b>0.66</b>
M8	5	0.95	0.92	0.89	0.95	0.92	0.89	0.89	0.84	0.66
M9	5	0.96	0.90	0.87	0.96	0.90	0.87	0.92	0.81	0.62
M10	6	0.97	0.90	0.82	0.97	0.90	0.84	0.95	0.81	0.53

atomic weights for characterizing the nature of atoms. For the first descriptor, the Kier-Hall connectivity of order 1 (1CN) operator is first applied on the atomic weights of the heteroatoms, while for the second descriptor the total sum of 2 lags (2TS) operator is applied. The standard deviation is finally used as aggregation operator of these no-standardized LOVIs. Noteworthy, all the regression coefficients are above zero (positive terms), indicating that all the features described above directly contribute to the inhibitory activity against TryS. For instance, the counts of C–N–C\*C\*C–Cl and C\*N\*N–C–C–N directly correlate with inhibitory activity, indicating that the presence of aromatic heterocycles containing nitrogen atoms (e.g., pyridine or pyrrole rings) are favorable to activity. Similarly, inhibitory activity is seemingly potentiated by the present of atoms with relatively high product of their correspondent atomic polarizability at a topological distance of five, as suggested by *ATSC5p*.

Table 2 shows the AUC ROC of the single best-performing model and the model ensembles obtained by systematically combining 2–10 models using the AVE, RANK, MIN and VOT operators (first retrospective virtual screening, DUDE-1). The standard deviation estimated by bootstrapping is also included. As expected, the performance of the model ensembles is, in all cases, significantly better ( $p < 0.05$ ) than the one of the best individual model. This seems to confirm the ability of ensemble learning to enhance predictivity. On the other hand, the robustness is also apparently improved through ensemble learning, as judged from the comparatively lower standard deviations

of the best-performing ensembles compared with the one of the best individual model.

By jointly considering the performance (based on the AUC ROC metric), the standard deviation of such metric and the principle of parsimony, we chose the 4-model ensembles for further validation (note that these ensembles behave practically identically or, in some cases, better than the 5-, 6- and 7-model ensembles, but with a smaller number of combined models). The principle of parsimony or Occam's razor indicates that, given two solutions of similar performance, the simplest one should be kept.

After selecting the 4-model ensembles for further analysis, we relied on PPV surface analysis to decide on an optimal score threshold for forthcoming prospective virtual screening applications. With the support of PPV surfaces, the relationship between the *Se/Sp* ratio and the PPV or precision (i.e., the probability that the predicted activity of an in silico hit will be verified when subject to experimental validation) can be visually (or, eventually, mathematically) optimized across a pertinent range of *Ya* values. Note that, from a pragmatical viewpoint, PPV is possibly the most relevant metric when realizing virtual screening experiments, as it allows deciding what number of in silico hits should be acquired and tested to expect one true, verified experimental hit. For such analysis, we have considered that the association between the *Se/Sp* and the score values of the model ensembles observed in the first retrospective screen would remain approximately unmodified when conducting screens on other libraries. This strong assumption is not necessarily

**Table 2** Performance of the best individual classifier and the classifier ensembles in the first retrospective screening experiment (DUDE-1)

	MIN		AVE		RANK		VOT	
	AUC ROC	sd	AUC ROC	Sd	AUC ROC	Sd	AUC ROC	Sd
Best individual model	0.954	0.010	0.954	0.010	0.954	0.010	0.954	0.010
<i>E-2</i>	0.993	0.003	0.99	0.003	0.996	0.002	0.996	0.002
<i>E-3</i>	0.990	0.006	0.987	0.006	0.996	0.002	0.996	0.002
<i>E-4</i>	<b>0.994</b>	<b>0.004</b>	<b>0.997</b>	<b>0.002</b>	<b>0.997</b>	<b>0.002</b>	<b>0.997</b>	<b>0.002</b>
<i>E-5</i>	0.994	0.004	0.997	0.002	0.992	0.004	0.993	0.004
<i>E-6</i>	0.993	0.004	0.994	0.003	0.988	0.005	0.987	0.006
<i>E-7</i>	0.995	0.003	0.995	0.003	0.990	0.005	0.986	0.007
<i>E-8</i>	0.971	0.028	0.993	0.003	0.991	0.004	0.986	0.008
<i>E-9</i>	0.971	0.027	0.992	0.003	0.991	0.004	0.987	0.006
<i>E-10</i>	0.962	0.036	0.988	0.004	0.988	0.005	0.986	0.006

The AUC ROC and its correspondent standard deviation (sd) are shown for the best individual model and for each model ensemble. The 4-model ensembles (in bold) displayed, in general, the best performance (highest AUC ROC with comparatively low sd). The 2- to 7-model ensembles showed excellent performance (AUC ROC above 0.99) and low dispersion ( $sd < 0.010$ ), no matter which combination scheme was utilized. Remarkably, the performance dropped from the combination of 8 models and beyond, with a considerable raise in the standard deviation of the metric.

true. Since the AUC ROC values obtained for the retrospective screen are, however, undeniably high (above 0.99, which indicates a nearly ideal performance), while on the other hand, the yield of actives (0.02) and size ( $\approx 1000$  molecules) of the library favor a controlled statistical behavior [33], our assumption is sensible in the present setting (also note the extremely low standard deviations obtained by bootstrapping for the ensemble models). The resulting PPV surfaces are shown in Fig. 3. All the combination schemes resulted in similar surfaces (similar shape and height) except for AVE, which seems to provide smaller PPVs than the other operators.

The 4-model ensembles were further validated through our second retrospective virtual screening experiment (DUDE-2). Table 3 shows the correspondent results, in comparison with the best-performing individual model. Note that, in good agreement with previous reports [42, 43, 46], the rather conservative MIN operator (which assigns as the ensemble output the lowest score given by the individual models comprised by the ensemble) seems to exhibit the best and most robust performance, either in terms of average performance across the whole ranking (as reflected by AUC ROC) or in terms of early recognition (quantified by BEDROC and EF1%). The comparisons of the AUC ROCs for the best individual model and the 4-model ensembles for both the first and second retrospective screen are shown in Fig. 4. Table 4 displays the optimal score cut-off value

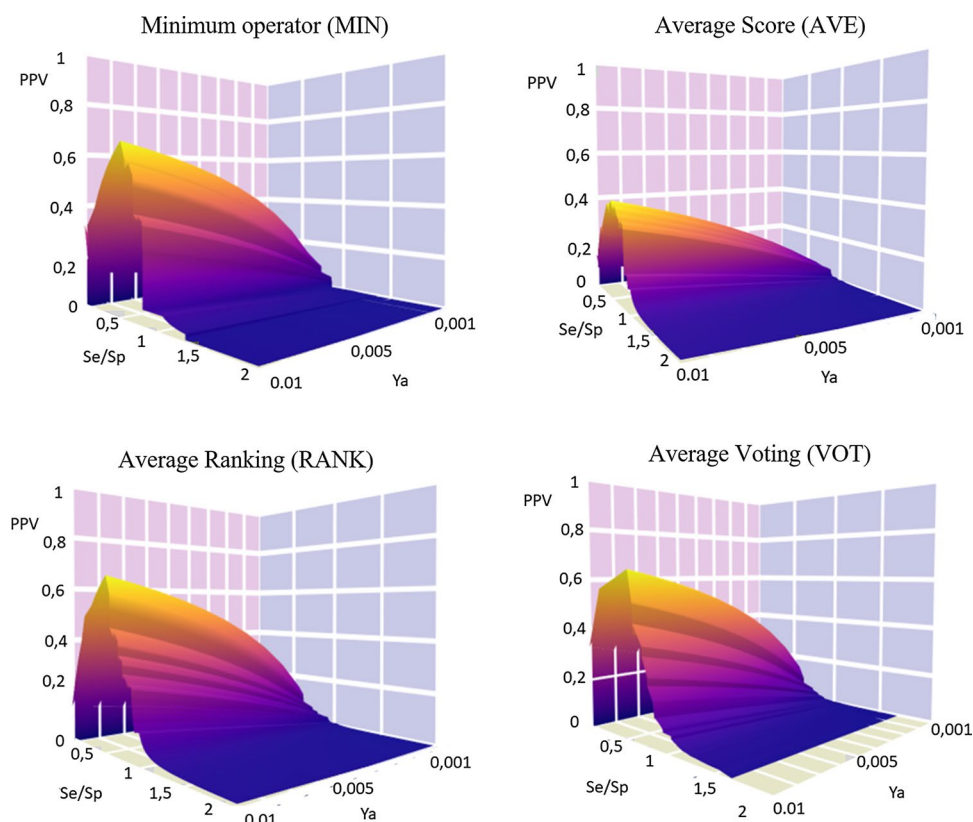
**Table 3** Performance of the best individual classifier and the 4-classifier ensembles in the second retrospective screening experiment (DUDE-2)

	AUC ROC	Sd	EF1%	BEDROC (alpha=20.0)	Sd
Best individual model	0.899	0.028	9.51	0.438	0.0496
MIN-4	0.958	0.020	35.20	0.814	0.0451
AVE-4	0.948	0.025	21.19	0.760	0.0436
RANK-4	0.943	0.030	35.20	0.817	0.0409
VOT-4	0.947	0.028	35.20	0.795	0.0467

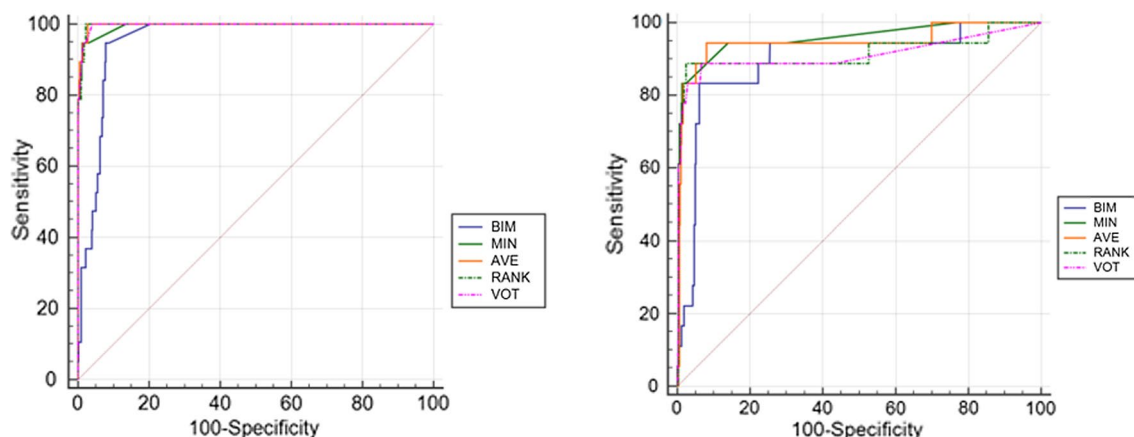
selected for the 4-model ensembles for each combination scheme, and the associated *PPV* ( $Y_a=0.01$ ) obtained from the second retrospective screen. It can be observed that the *PPV* associated to the optimal cut-off value for the 4-model ensemble obtained with the MIN operator is higher than the ones achieved with the other schemes. These theoretic values suggest that, if a prospective screening were implemented and the yielding of active compounds was 1%, about one in three in silico hits would be a true, confirmed hits at the experimental validation step.

Table 5 shows the ten top-ranked in silico hits emerging from the prospective virtual screen of the DrugBank database. It can be appreciated that the compounds in the

**Fig. 3** PPV surfaces of the 4-model ensembles







**Fig. 4** AUC ROCs for the best individual model (BIM) and the 4-model ensembles obtained through the four combination schemes: minimum (MIN), average score (AVE), average raking (RANK) and

voting (VOT). The curves for the first retrospective screen are shown on the left, and the results for the second retrospective screen are shown on the right

**Table 4** Score cut-off values chosen from visual inspection of the PPV surfaces, and associated Se/Sp ratio and PPV (for  $Y_a=0.01$ )

Operator	Score cut-off	Se/Sp	PPV
MIN	-0.42	0.85	0.35
AVE	0.23	0.85	0.30
RANK	<121	0.91	0.24
VOT	5.25	0.80	0.22

table display several of the features that, according to the best individual model (Eq. 2) are favorable for inhibitory activity on TryS. For example, heterocyclic rings with one or more nitrogen atoms, occurrence of pairs of highly polarizable atoms at topological distances of five (e.g., Cl-S, C-S, C-Cl), etc. According to the analysis of the PPV surface for the 4-model MIN ensemble, these compounds present a PPV above 0.35 for a  $Y_a=0.01$ , which suggests that upon experimental assays at least one third of these in silico hits are expected to confirm inhibitory activity against TryS. Most of them, except one, display high probability of being orally bioavailable through passive diffusion, as they meet at least three of the four propositions of Lipinski's rule of five. The hits belong to a diversity of therapeutic classes, including antivirals, diuretics, anti-obesity drugs, and others. Noteworthy, most of the listed hits display limited freedom-to-operate, as, with one exception (the off-patent drug azosemide), they are experimental or investigational drugs that

belong to proprietary chemical libraries. Furthermore, not all the hits are identically attractive as potential treatments against Chagas disease. Azosemide, for example, is a loop diuretic which can improve cardiac output and congestive symptoms [47], one of the typical cardiac manifestations of Chagas disease; this therapeutic effect could add to its trypanocidal action, if confirmed. In contrast, rimonabant is a poor repurposing prospect, as it was promptly removed from the markets where it got approval due to severe mood disorders [49].

## Conclusions

Linear classifiers to identify TryS inhibitors were derived using the Replacement Method and a pool of molecular descriptors entirely computed through publicly available and open-source software, which maximizes the portability of the obtained models (furthermore, all the in-house scripts used are available on request). Remarkably, a substantial proportion of the dataset used here (65%) corresponds to highly standardized in-house acquired data of inhibitory activity against *T. brucei* TryS, which can be regarded as low noise data. Whereas the individual models obtained showed good performance at the validation and retrospective screening steps, they were considerably outperformed by ensemble learning. The so derived meta-classifiers displayed not only improved enrichment metrics, but also a

**Table 5** List of the ten top-ranked hits from the DrugBank database, according to the predictions of the best 4-model ensemble (only compounds belonging to the applicability domain of at least three of the four models that comprise the ensemble were considered). As the MIN operator assigns the lowest score from the individual classifiers that compose the ensemble, the remaining three models have assigned a higher score than the one tabulated. The number of propositions in the Lipinski rule that are met by each hit is indicated

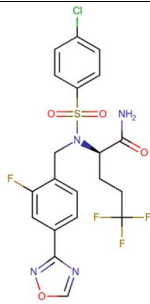
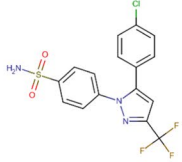
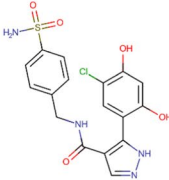
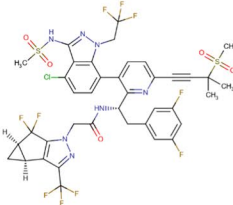
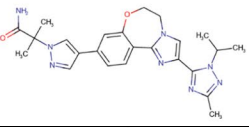
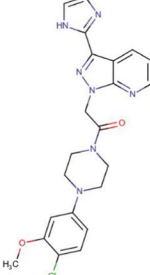
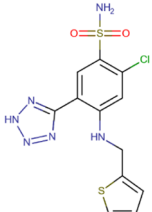
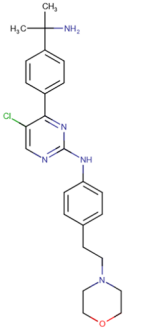
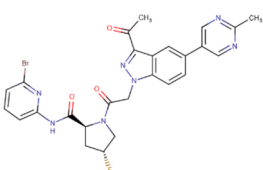
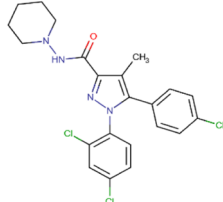
Compound	Structure	Score	Status	Lipinski rules
Avagacestat		0.38	Investigational	3/4
SC-236		0.20	Investigational	4/4
N-[4-(aminosulfonyl)benzyl]-5-(5-chloro-2,4-dihydroxyphenyl)-1H-pyrazole-4-carboxamide		0.20	Experimental	4/4
Lenacapavir		0.20	Investigational	2/4
Taselisib		0.18	Investigational	4/4
CCX-354		0.18	Investigational	4/4

Table 5 (continued)

Azosemide		0.17	Approved	4/4
4-[4-(1-Amino-1-Methylethyl)Phenyl]-5-Chloro-N-[4-(2-Morpholin-4-Ylethyl)Phenyl]Pyrimidin-2-Amine		0.08	Experimental	4/4
Danicopan		0.06	Investigational	3/4
Rimonabant		0.04	Withdrawn	3/4

more robust behavior according to the standard deviation in the enrichment metrics estimated by bootstrapping. The score cut-off values of the best-performing model combinations were rationally optimized through inspection of Positive Predictive Value surfaces; the optimized score threshold was then applied here in the illustrative prospective screening of DrugBank, a database commonly used for computer-aided drug repurposing implementations. It is worth mentioning that our models are only predictive of inhibitory effects on TryS and should thus be complemented with other *in silico* filters or *in vitro* assays related to pharmaceutically relevant properties, such as drug bioavailability (e.g., Lipinski rules, models to predict P-glycoprotein efflux liability, etc.).

Computer-aided drug discovery represents a key strategy for the identification of new active scaffolds in a cost- and time-efficient manner, which is particularly relevant when seeking for novel therapeutic solutions for neglected conditions, such as trypanosomatid-caused diseases.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11030-021-10265-9>.

**Acknowledgements** Juan I. Alice, Carolina L. Bellera, Pablo R. Duchowicz and Alan Talevi thank the National University of La Plata (UNLP) and the Argentinean National Council of Scientific and Technical Research Council (CONICET). The present work was funded by The National Agency of Scientific and Technological Promotion (ANPCyT PICT 2017-0643 and PICT 2016-02056), UNLP (National University of La Plata, 11/X785), and the CONICET PIP11220130100311 project. Diego Benítez and Marcelo A. Comini belong to the Uruguayan National System of Researchers (SNI) from ANII and thank the support of FOCEM (Fondo para la Convergencia Estructural del Mercosur, COF 03/11).

## Declarations

**Conflict of interest** The authors have declared no conflict of interest.

## References

- Nussbaum K, Honek J, Cadmus CM, Efferth T (2010) Trypanosomatid parasites causing neglected diseases. *Curr Med Chem* 17:1594–1617. <https://doi.org/10.2174/092986710790979953>

2. Chatelain E (2016) Chagas disease research and development: Is there light at the end of the tunnel? *Comput Struct Biotechnol J* 15:98–103. <https://doi.org/10.1016/j.csbj.2016.12.002>
3. Muñoz C, Zulantay I, Apt W, Ortiz S, Schijman AG, Bisio M, Ferrada V, Herrera C, Martínez G, Solari A (2013) Evaluation of nifurtimox treatment of chronic Chagas disease by means of several parasitological methods. *Antimicrob Agents Chemother* 57:4518–4523. <https://doi.org/10.1128/AAC.00227-13>
4. Hasslocher-Moreno AM, Saraiva RM, Sangesis LHC, Xavier SS, de Sousa AS, Costa AR, de Holanda MT, Veloso HH, Mendes FSNS, Costa FAC, Boia MN, Brasil PEAA, Carneiro FM, da Silva GMS, Mediano MFF (2020) Benzimidazole decreases the risk of chronic Chagas disease progression and cardiovascular events: A long-term follow up study. *EClinicalMedicine* 31:100694. <https://doi.org/10.1016/j.eclinm.2020.100694>
5. Pecoul B, Batista C, Stobbaerts E, Ribeiro I, Vilasanjuan R, Gascon J, Pinazo MJ, Moriana S, Gold S, Pereiro A, Navarro M, Torrico F, Bottazzi ME, Hotez PJ (2016) The BENEFIT Trial: where do we go from here? *PLoS Negl Trop Dis* 10:e0004343. <https://doi.org/10.1371/journal.pntd.0004343>
6. Talevi A, Carrillo C, Comini M (2019) The Thiol-polyamine metabolism of *Trypanosoma cruzi*: molecular targets and drug repurposing strategies. *Curr Med Chem* 26:6614–6635. <https://doi.org/10.2174/0929867325666180926151059>
7. Fairlamb AH, Blackburn P, Ulrich P, Chait BT, Cerami A (1985) Trypanothione: a novel bis(glutathionyl)spermidine cofactor for glutathione reductase in trypanosomatids. *Science* 227:1485–1487. <https://doi.org/10.1126/science.3883489>
8. Fairlamb AH, Cerami A (1992) Metabolism and functions of trypanothione in the Kinetoplastida. *Annu Rev Microbiol* 46:695–729. <https://doi.org/10.1146/annurev.mi.46.100192.003403>
9. Comini MA, Guerrero SA, Haile S, Menge U, Lünsdorf H, Flohé L (2004) Validation of *Trypanosoma brucei* trypanothione synthetase as drug target. *Free Radic Biol Med* 36:1289–1302. <https://doi.org/10.1016/j.freeradbiomed.2004.02.008>
10. Torrie LS, Wyllie S, Spinks D, Oza SL, Thompson S, Harrison JR, Gilbert IH, Wyatt PG, Fairlamb AH, Frearson JA (2009) Chemical validation of trypanothione synthetase: a potential drug target for human trypanosomiasis. *J Biol Chem* 25(284):36137–36145. <https://doi.org/10.1074/jbc.M109.045336>
11. Sousa AF, Gomes-Alves AG, Benítez D, Comini MA, Flohé L, Jaeger T, Passos J, Stuhlmann F, Tomás AM, Castro H (2014) Genetic and chemical analyses reveal that trypanothione synthetase but not glutathionylspermidine synthetase is essential for *Leishmania infantum*. *Free Radic Biol Med* 73:229–238. <https://doi.org/10.1016/j.freeradbiomed.2014.05.007>
12. Benítez D, Medeiros A, Fiestas L, Panozzo-Zenere EA, Maiwald F, Prousis KC, Roussaki M, Calogeropoulou T, Detsi A, Jaeger T, Šarlauskas J, Peterlin Mašič L, Kunick C, Labadie GR, Flohé L, Comini MA (2016) Identification of novel chemical scaffolds inhibiting trypanothione synthetase from pathogenic trypanosomatids. *PLoS Negl Trop Dis* 10:e0004617. <https://doi.org/10.1371/journal.pntd.0004617>
13. Olin-Sandoval V, González-Chávez Z, Berzunza-Cruz M, Martínez I, Jasso-Chávez R, Becker I, Espinoza B, Moreno-Sánchez R, Saavedra E (2012) Drug target validation of the trypanothione pathway enzymes through metabolic modelling. *FEBS J* 279:1811–1833. <https://doi.org/10.1111/j.1742-4658.2012.08557.x>
14. Daunes S, D'Silva C, Kendrick H, Yardley V, Croft SL (2001) QSAR study on the contribution of log P and E(s) to the in vitro antiprotozoal activity of glutathione derivatives. *J Med Chem* 44:2976–2983. <https://doi.org/10.1021/jm000502n>
15. Liñares GE, Ravaschino EL, Rodríguez JB (2006) Progresses in the field of drug design to combat tropical protozoan parasitic diseases. *Curr Med Chem* 13:335–360. <https://doi.org/10.2174/092986706775476043>
16. Spinks D, Torrie LS, Thompson S, Harrison JR, Frearson JA, Read KD, Fairlamb AH, Wyatt PG, Gilbert IH (2012) Design, synthesis and biological evaluation of *Trypanosoma brucei* trypanothione synthetase inhibitors. *Chem Med Chem* 7:95–106. <https://doi.org/10.1002/cmdc.201100420>
17. Zimmermann S, Oufir M, Leroux A, Krauth-Siegel RL, Becker K, Kaiser M, Brun R, Hamburger M, Adams M (2013) Cynaropicrin targets the trypanothione redox system in *Trypanosoma brucei*. *Bioorg Med Chem* 21:7202–7209. <https://doi.org/10.1016/j.bmc.2013.08.052>
18. D'Silva C, Daunes S (2000) Structure-activity study on the in vitro antiprotozoal activity of glutathione derivatives. *J Med Chem* 43:2072–2078. <https://doi.org/10.1021/jm990259w>
19. Koch O, Jäger T, Flohé L, Selzer PM (2013) Inhibition of trypanothione synthetase as a therapeutic concept. In: Jäger T, Koch O, Flohé L (eds) *Trypanosomatid Diseases*. Wiley, Weinheim, pp 429–443
20. Leroux AE, Krauth-Siegel RL (2016) Thiol redox biology of trypanosomatids and potential targets for chemotherapy. *Mol Biochem Parasitol* 206:67–74. <https://doi.org/10.1016/j.molbiopara.2015.11.003>
21. Oza SL, Chen S, Wyllie S, Coward JK, Fairlamb AH (2008) ATP-dependent ligases in trypanothione biosynthesis—kinetics of catalysis and inhibition by phosphinic acid pseudopeptides. *FEBS J* 275:5408–5421. <https://doi.org/10.1111/j.1742-4658.2008.06670.x>
22. Stuhlmann FD, Jäger TD, Flohé LPD, Schinzer DPD (2007) N5-substituted benzo-2,3-lazepino-4,5-blindol-6-ones for treating tropical diseases. European Union Patent No. EP1757607A1.
23. Perez-Llamas C, Lopez-Bigas N (2011) Gitoools: analysis and visualisation of genomic data using interactive heat-maps. *PLoS ONE* 6:e19541. <https://doi.org/10.1371/journal.pone.0019541>
24. PaDEL 2.20 (Pharmaceutical Data Exploration Laboratory), <http://www.yapcwsoft.com>, last accessed 1 April 2021
25. Hong H, Xie Q, Ge W, Qian F, Fang H, Shi L, Su Z, Perkins R, Tong W (2008) Mold<sup>2</sup>, molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. *J Chem Inf Model* 48:1337–1344. <https://doi.org/10.1021/ci800038f>
26. ISIDA/Fragmentor (2017) Laboratoire de Chémoinformatique, Chimie de la Matière Complexe (SMS UMR 7140), Université de Strasbourg, France, <http://complex-matter.unistra.fr/equipes-de-recherche/laboratoire-de-chemoinformatique/home>, last accessed 1 April 2021
27. Martínez-López Y, Marrero-Ponce Y, Barigye SJ, Teran E, Martínez-Santiago O, Zambrano CH, Torres FJ (2020) When global and local molecular descriptors are more than the sum of its parts: simple, but not simpler? *Mol Divers* 24:913–932. <https://doi.org/10.1007/s11030-019-10002-3>
28. Aranda JF, Bacelo DE, Leguizamón Aparicio MS, Ocsachoque MA, Castro EA, Duchowicz PR (2017) Predicting the bioconcentration factor through a conformation-independent QSPR study. *SAR&QSAR Environ Res* 28(9):749–763. <https://doi.org/10.1080/1062936X.2017.1377765>
29. Kaufman L, Rousseeuw PJ (2005) *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York, USA
30. Duchowicz PR, Castro EA, Fernández FM (2006) Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR Studies. *MATCH Commun Math Comput Chem* 55:179–192
31. Ballabio D, Grisoni F, Todeschini R (2018) Multivariate comparison of classification performance measures. *Chemom Intel Lab Syst* 174:33–44. <https://doi.org/10.1016/j.chemolab.2017.12.004>
32. Octave 5.2.0, <https://www.gnu.org/software/octave>, last accessed 1 April 2021

33. Truchon JF, Bayly CI (2007) Evaluating virtual screening methods: Good and bad metrics for the “early recognition” problem. *J Chem Inf Model* 47:488–508. <https://doi.org/10.1021/ci600426e>
34. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK (2012) Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem* 55:6582–6594. <https://doi.org/10.1021/jm300687e>
35. Carbonneau MA, Granger E, Raymond AJ, Gagnon G (2016) Robust multiple-instance learning ensembles using random subspace instance selection. *Pattern Recognit* 58:83–99. <https://doi.org/10.1016/j.patcog.2016.03.035>
36. Li Y, Chen W (2020) A comparative performance assessment of ensemble learning for credit scoring. *Mathematics* 8:1756. <https://doi.org/10.3390/math8101756>
37. Wang G, Sun J, Ma J, Xu K, Gu J (2014) Sentiment classification: the contribution of ensemble learning. *Decis Support Syst* 57:77–93. <https://doi.org/10.1016/j.dss.2013.08.002>
38. Zhang Q, Muegge I (2006) Scaffold hopping through virtual screening using 2D and 3D similarity descriptors: ranking, voting, and consensus scoring. *J Med Chem* 49:1536–1548. <https://doi.org/10.1021/jm050468i>
39. Robin X, Turck N, Hainard A et al (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. <https://doi.org/10.1186/1471-2105-12-77>
40. Lähti S, Niinivehmas S, Pentikäinen OT (2016) Rocker: Open source, easy-to-use tool for AUC and enrichment calculations and ROC visualization. *J Cheminform* 8:45. <https://doi.org/10.1186/s13321-016-0158-y>
41. Alberca LN, Chuguransky SR, Álvarez CL, Talevi A, Salas-Sarduy E (2019) In silico guided drug repurposing: discovery of new competitive and non-competitive inhibitors of falcipain-2. *Front Chem* 7:534. <https://doi.org/10.3389/fchem.2019.00534>
42. BÉlgamo JA, Alberca LN, Pórfido JL, Romero FNC, Rodríguez S, Talevi A, Córscico B (2020) Franchini GR (2020) Application of target repositioning and in silico screening to exploit fatty acid binding proteins (FABPs) from *Echinococcus multilocularis* as possible drug targets. *J Comput Aided Mol Des* 34:1275–1288. <https://doi.org/10.1007/s10822-020-00352-8>
43. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 46:D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>
44. Tropsha A, Gramatica P, Gombar V (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR Models. *QSAR Comb Sci* 22: 69–77. <https://doi.org/10.1002/qsar.200390007>
45. Alberca LN, Sbaraglini ML, Morales JF, Dietrich R, Ruiz MD, Pino Martínez AM, Miranda CG, Fraccaroli L, Alba Soto CD, Carrillo C, Palestro PH, Talevi A (2018) Cascade ligand- and structure-based virtual screening to identify new trypanocidal compounds inhibiting putrescine uptake. *Front Cell Infect Microbiol* 8:173. <https://doi.org/10.3389/fcimb.2018.00173>
46. Masuyama T, Tsujino T, Origasa H, Yamamoto K, Akasaka T, Hirano Y, Ohte N, Daimon T, Nakatani S, Ito H (2012) Superiority of long-acting to short-acting loop diuretics in the treatment of congestive heart failure. *Circ J* 76:833–842. <https://doi.org/10.1253/circj.cj-11-1500>
47. Sam AH, Salem V, Ghatei M (2011) Rimonabant: from RIO to Ban. *J Obes* 2011:432607. <https://doi.org/10.1155/2011/432607>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.