Comparison of Multiple Linear Regressions and Neural Networks based QSAR models for the design of new antitubercular compounds

ARTICLE *in* EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY · OCTOBER 2013

Impact Factor: 3.45 · DOI: 10.1016/j.ejmech.2013.10.029 · Source: PubMed

3 AUTHORS:

Cristina Ventura
University of Lisbon
20 PUBLICATIONS **146** CITATIONS
SEE PROFILE

Diogo ARS Latino
Eawag: Das Wasserforschungs-Institut des …
41 PUBLICATIONS **145** CITATIONS
SEE PROFILE

Filomena Martins
University of Lisbon
56 PUBLICATIONS **506** CITATIONS
SEE PROFILE

European Journal of Medicinal Chemistry 70 (2013) 831–845

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Comparison of Multiple Linear Regressions and Neural Networks based QSAR models for the design of new antitubercular compounds

Cristina Ventura [a,b], Diogo A.R.S. Latino [c,d], Filomena Martins [a,*]

[a] Centro de Química e Bioquímica (CQB), Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade de Lisboa, Ed. C8, Campo Grande, 1749-016 Lisboa, Portugal
[b] Instituto Superior de Educação e Ciências, Alameda das Linhas de Torres 179, 1750 Lisboa, Portugal
[c] REQUIMTE and CQFB, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Monte de Caparica, 2829-516 Caparica, Portugal
[d] Centro de Ciências Moleculares e Materiais (CCMM), Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade de Lisboa, Ed. C8, Campo Grande, 1749-016 Lisboa, Portugal

ARTICLE INFO

Article history:
Received 9 May 2012
Received in revised form
26 July 2013
Accepted 11 October 2013
Available online 23 October 2013

Keywords:
QSARs
Neural Networks
Multiple Linear Regressions
Antitubercular activity
Hydrazides

ABSTRACT

The performance of two QSAR methodologies, namely Multiple Linear Regressions (MLR) and Neural Networks (NN), towards the modeling and prediction of antitubercular activity was evaluated and compared. A data set of 173 potentially active compounds belonging to the hydrazide family and represented by 96 descriptors was analyzed. Models were built with Multiple Linear Regressions (MLR), single Feed-Forward Neural Networks (FFNNs), ensembles of FFNNs and Associative Neural Networks (AsNNs) using four different data sets and different types of descriptors. The predictive ability of the different techniques used were assessed and discussed on the basis of different validation criteria and results show in general a better performance of AsNNs in terms of learning ability and prediction of antitubercular behaviors when compared with all other methods. MLR have, however, the advantage of pinpointing the most relevant molecular characteristics responsible for the behavior of these compounds against *Mycobacterium tuberculosis*. The best results for the larger data set (94 compounds in training set and 18 in test set) were obtained with AsNNs using seven descriptors (R^2 of 0.874 and *RMSE* of 0.437 against R^2 of 0.845 and *RMSE* of 0.472 in MLRs, for test set). Counter-Propagation Neural Networks (CPNNs) were trained with the same data sets and descriptors. From the scrutiny of the weight levels in each CPNN and the information retrieved from MLRs, a rational design of potentially active compounds was attempted. Two new compounds were synthesized and tested against *M. tuberculosis* showing an activity close to that predicted by the majority of the models.

1. Introduction

In the World Health Organization (WHO) 2012 report on Global Tuberculosis Control, the relevant numbers for 2011 were 8.7 million new cases worldwide and 1.4 million deaths [1]. About 3.7% of all new tuberculosis (TB) infections are resistant to at least one known anti-TB drug and cases of multidrug-resistant (MDR-TB) and extensively drug-resistant tuberculosis (XDR-TB) are believed to be far higher than reported [2,3]. WHO estimated that in 2011 there was a prevalence of 310.000 cases of MDR-TB among notified TB patients, of which 9% were XDR-TB [3]. The increasing mobility of populations, the long demanding, and sometimes wrong, treatments and the poor patient compliance, giving rise to an alarming emergence of MDR- and XDR-TB strains, along with observed drug–drug interactions in immunocompromised individuals such as those co-infected, for instance, with HIV or diabetes [4], make

Abbreviations: AA, antitubercular activity; *AAE*, absolute average error; *AE*, average error; ANN, Artificial Neural Network; AsNNs, associative neural networks; CPNNs, Counter-Propagation Neural Network; EnsFFNNs, Ensembles of Feed-Forward Neural Networks; FFNNs, Feed-Forward Neural Networks; HIV, human immunodeficiency virus; kNN, *k*-Nearest Neighbor; LMO, leave-many-out; LOO, leave-one-out; MDR-TB, multidrug-resistant tuberculosis; MIC, minimum inhibitory concentration; MLR, Multiple Linear Regression; MLT, machine learning techniques; *M. tuberculosis*, *Mycobacterium tuberculosis*; NNs, Neural Networks; Q^2, cross-validated correlation coefficient; QSARs, quantitative structure–activity relationships; QSPRs, quantitative structure–property relationships; RFs, random forests; *RMSE*, root mean squared error; *SD*, standard deviation; *SL*, significance level; SOMs, self-organizing maps; TB, tuberculosis; TDR-TB, totally drug-resistant tuberculosis; WHO, World Health Organization; XDR-TB, extensively drug-resistant tuberculosis.

* Corresponding author. Tel.: +351 21 7500870; fax: +351 21 7500088.
E-mail addresses: cventura@isec.universitas.pt (C. Ventura), latino@fc.ul.pt (D.A.R.S. Latino), filomena.martins@fc.ul.pt (F. Martins).

0223-5234/$ – see front matter © 2013 Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.ejmech.2013.10.029

the fight against TB a very serious global health issue. To make the scenario even worse, researchers in India have just identified what they call a completely drug-resistant form of TB (TDR-TB), resistant to all first- and second-line current drugs [5], after previously documented cases in Italy [6] and Iran [7]. Some authors suggest, however, that this TDR-TB is just a deadlier development of the highly resistant forms of TB and not itself a new entity.

On the other hand, the fact that the needed funding for TB care and control in low- and middle-income countries is around US$ 8 billion per year and that the funding gap is approximately US$ 3 billion per year [3], makes it very clear that, in addition to an undeniable social burden, the economic load associated with this disease is very substantial.

These very significant facts have led to an increasing awareness of the scientific community towards the need of developing new, structurally modified, active antitubercular drugs. In order to succeed in the design of these new compounds it would be of relevance to be able to establish a clear-cut relationship between structure and activity. Several efforts, in general not very successful, have been made towards the understanding of the mechanisms of action of known active compounds against *Mycobacterium tuberculosis* (*M. tuberculosis*). Although a number of studies have appeared in the literature in these last few years [8–10], the level of knowledge about these drug-*M. tuberculosis* interactions is still rather limited. The establishment of quantitative relationships between structural features and measured biological activity, the so-called QSARs, extensively used in the field of Medicinal Chemistry, is one of the most useful and elegant strategies in the quest for new bioactive molecules. Various methodologies and mathematical tools have been developed throughout the years with the purpose of modeling and predicting all sort of biological behaviors and they have been applied with considerable levels of success in the rational design and synthesis of different new drugs [11–13].

Comprehensive QSAR models can be achieved through the application of various supervised and/or unsupervised statistical and machine learning techniques (MLT). From these, Multiple Linear Regressions (MLR) [14,15], Neural Networks (NN) [16–18], Decision Trees [19], Random Forests (RF) [20], Partial Least Squares (PLS) [21,22], Principal Components Analysis [23] and Support Vector Machines (SVM) [24] are among the most successful methodologies. The extra-thermodynamic approach using Multiple Linear Regression analyses (MLR), which has been launched more than 40 years ago by Hansch et al. [14,15], has indeed been very successful in the fields of Medicinal Chemistry and Pharmaceutical Sciences [11,25]. This extra-thermodynamic approach when applied to the study of pharmacological action is based on the premise that the biological activity of a series of similar drugs, taken as having the same mechanism of action, can be described as a linear combination of molecular properties or descriptors identified as relevant for the process under study.

Another especially successful methodology introduced some years ago and which simulates the processing of information by biological neurons in the human nervous system is the so-called Artificial Neural Networks (ANNs) [16–18]. ANNs overcame some of the frailties of the MLR approach in the design of new drugs, especially because of their ability to deal with non-linear relationships of high complexity when performing input–output transformations. However, to our knowledge, there are in the literature just a few applications of ANNs to the study of antitubercular activity [26–29]. Yet, more important than using each of these (or other) methodologies *per se*, is the possibility of combining them to improve our ability to interpret and predict biological behaviors.

Within our research group, we have started a systematic analysis of a large data set of potentially active compounds against *M. tuberculosis*, with the purpose of establishing robust and predictive QSAR models [30,31]. The first aim of this work is not to explore the potentialities of various MLT in the modeling and prediction of biological activity but rather to compare the performance of two particular QSAR methodologies, MLRs and NNs, in the modeling of the antitubercular activity of a series of compounds with a hydrazide functionality. A second important goal is the rational design of new potentially active compounds based on the information provided by the best found models.

2. Materials and methods

It is widely accepted that the development of any classical QSAR model is based on the following postulates: (i) the molecular structure is responsible for the observed activity (or property); and (ii) the established model can only be applied to compounds belonging to the same physicochemical–structural–biological space, *i.e.*, a model should only be used to make predictions within its applicability domain [32–36].

On the other hand, the success of a QSAR model in terms of its interpretative and predictive abilities depends on an adequate data pre-processing and on rigorous validation procedures [34,35,37–41].

2.1. Biological data

In this paper we have analyzed 173 compounds with the hydrazide functionality, which are potentially active against *M. tuberculosis*, and whose structures have been given in Ref. [30]. Experimental activities, measured as minimum inhibitory concentrations, MIC, were extracted from Ref. [42] (136 compounds tested against the BCG strain of *M. tuberculosis*) and from Refs. [43–46] (37 compounds tested against the H37Rv strain). An additional set of 22 experimental MIC values for compounds used to make a lateral external validation of the developed models, also tested against the H37Rv strain, was collected from the Novartis public access database [47].

MIC values determined in this work against H37Rv were conducted using the BACTEC 460 system [48].

2.2. Descriptors

Molecular structures were drawn with ChemDraw Ultra, 11.0.1 [49] and saved as SMILES strings (available as Supporting information in Table S1).

Different types of descriptors, *e.g.*, physicochemical, steric, geometrical, energetic and electronic, were calculated for the entire data set of 173 hydrazide derivatives. Abraham's descriptors were calculated using the *Absolv* program [50,51]. Partition coefficients were collected either from Leo and Hansch's database as clog *P* [52] or calculated as log $P_{w/oct}$ using Molecular Modeling Pro Plus (MMP^{+}) [53] software. Some clog *P* and molar refractivity values were calculated using ChemDraw Ultra 11.0.1 package [49], making use of specific algorithms derived from fragment-based methods developed by the *Medicinal Chemistry Project* and *BioByte* [54]. Descriptors requiring 3D structures were calculated using MMP^{+}. 3D geometries were optimized with the molecular mechanics MM2 method incorporated in the referenced software [53]. Partial charges were calculated using the Del Re method and dipole moments by a combination of PEOE and Huckel methods, also included in the software.

The list of the 96 calculated descriptors is available in Table S2 of Supporting information. The numerical values of descriptors for all 173 analyzed compounds are also given as Supporting information (Table S3), as well as the values of descriptors for compounds used in the developed models along with clog *P* or MIC values taken from literature (Table S4).

2.3. Data sets and pre-processing procedures

With the objective of analyzing the physicochemical and biological behaviors of hydrazide compounds and, in particular, of the sub-family of isoniazid (INH) derivatives, given that INH is still the most potent drug to fight TB, we have tested various data sets. First, the initial data set of 173 compounds, represented by the calculated 96 descriptors, was pre-processed for outliers' detection. This procedure was repeated for each one of the 96 descriptors and was based on the analysis of dispersion plots and inter-quartile intervals for the identification of soft and extreme outliers [55]. For methods requiring normalized data, each input and output variable was linearly normalized between 0.1 and 0.9 on the basis of the corresponding training set space.

After this pre-processing, the initial data set was divided into 6 different data sets, each one partitioned into a training and a test set – Table 1. These data sets are not mutually exclusive.

The division of the data in these six data sets (see Table S1) and their subsequent partition was done so that it always lead to homogeneous and representative data sets, with appropriate sizes in terms of number of descriptors (or adjustable parameters) for MLR (or NN) development (or training) and prediction. The choice of compounds for training and test sets took into consideration the biological activities values and the distribution of the compounds in a Kohonen Self-Organizing Map (please refer to Section 2.5.4) in order to guarantee a similar variability in both sets.

An additional external data set composed of 22 compounds selected from the 283 hydrazide derivatives included in the Novartis TB public access database, was also considered [47]. First, compounds were selected by structure, being the filter the O=C–NHNH2 group and were retained if they had a similarity with this structural feature of at least 40%. Concurrently, it was used a MIC cut-off value of less than 400 μg mL^{-1}. 41 compounds complied with these requisites. However, they were structurally very diverse from the training set compounds. Therefore, in a second attempt, the structural filter was the INH core structure, for the same similarity level and MIC limit. From the compounds returned from the database, a sub-set of 22 compounds was randomly chosen so that it had a similar dimension to previously defined test sets, particularly test set of data set 6, and in such a way that the corresponding chemical space of the descriptors and the log (1/MIC) values fell into the training set range values of the data set used to establish the best models (please refer to Table S5 of Supporting information).

2.4. Selection of descriptors

A first pre-selection of the most relevant set of descriptors to model antitubercular activity was performed using Random Forests (RFs) [20,56]. A Random Forest is an ensemble of unpruned regression trees created by using bootstrap samples of the training data set and random subsets of variables to define the best split at each node. It is a high-dimensional nonparametric method that works well on large numbers of variables. RFs predictions are made by an average of the predictions from individual trees in the forest. It has been shown that this method is very accurate in a variety of applications [31,57]. Additionally, the performance is internally assessed with the prediction error for the objects left out in the bootstrap procedure. In this work, RFs were grown with the R program version [58], using the Random Forest library [59]. The number of trees in the forest was set to 1000, and the number of variables tested for each split was set to default (square root of the number of variables).

RFs also present a number of attractive features for QSAR applications, namely the fact that the algorithm is not affected by correlated descriptors since it uses random samples to build each tree in the forest. Another feature is an in-bound procedure for descriptor selection which provides measures of importance for each descriptor in the model [19]. Here the measurement of the most relevant descriptors to model the antitubercular activity was performed using the "Mean Decrease Accuracy" value. The importance of a variable was measured by the increase in the error occurring when the values of the variable were randomly permuted. A pre-selection of the 30 most relevant descriptors to model the antitubercular activity of the compounds under study (identified in bold in Table S2 of Supporting information) was performed with the objects of the training sets.

Table 1
Composition of the data sets.

Dataset	Type	Number of compounds		
		Total	Training set	Test set
1	Hydrazide derivatives	137	98	39
2	Isoniazid derivatives	43	24	19
3	Hydrazide derivatives	59	30	29
4	Isoniazid derivatives	33[a]	22	11
5	Isoniazid derivatives	33[a]	22	11
6	Hydrazide derivatives	112[b]	94	18

[a] Data sets 4 and 5 have the same number of objects but different compositions.
[b] The difference for the total number of compounds (173) in data set 6 which leads to the best models (models 21 and 24) correspond to outliers which were detected and removed, according to the procedure described in Sections 2.3 and 2.6.1 (51 in training set and 10 in test set).

2.5. Statistical and machine learning techniques (MLT)

2.5.1. Multiple Linear Regressions (MLR)

In the context of the present study, a Multiple Linear Regression (MLR) is a supervised method that aims at establishing a mathematical relationship between a property of a given system and a set of molecular characteristics or descriptors that encode chemical information, being expressed as follows:

$$Y = AX + \xi \quad (1)$$

where ξ is a $n \times 1$ residuals vector, X is a known $n \times k$ matrix of descriptors, A is a $k \times 1$ vector of adjusted parameters and Y is a $n \times 1$ vector of the response variable related with either the activity or other system property.

The success of a MLR approach in the development of interpretative and predictive model equations requires that data should be representative and homogeneous, the absence of redundancy among descriptors, a ratio between number of compounds in data set and number of descriptors of at least three and the fulfillment of strict validation procedures [34,35,37,38].

For the purpose of this work, we have used the Data Analysis add-in, available in Microsoft Excel. After pre-selection of the 30 most relevant descriptors with RFs, these descriptors were subsequently used in the search for the best structure–property/activity Multiple Linear Regressions, through a forward-stepping procedure beginning with an equation containing only a single descriptor and adding up terms, one at a time, and testing every C_p^m combination, where m is the number of descriptors in each equation and p varies from 1 to m. The introduction of each descriptor was carefully examined for its contribution to the model's goodness according to strict statistical criteria to ensure the reliability of the analyses (*e.g.*, t statistics, model's determination coefficient, R^2, model's F statistics and standard deviation of the fit, SD_{fit}) until the best fit statistics was obtained. Finally, we also tested the best found models with the inclusion of other descriptors (namely, c, d_4 and d_K) that in previous studies [30] had shown to be important to the description of the antitubercular activity of isoniazid derivatives.

2.5.2. Feed-Forward Neural Networks (FFNNs) and Ensembles of Feed-Forward Neural Networks (EnsFFNNs)

Feed Forward Neural Networks, FFNNs [18], were defined with a number of input neurons corresponding to the number of descriptors for each data set, one hidden layer of neurons, and one output neuron. In the input layer and in the hidden layer, an extra neuron (called bias) with a constant value of one was also added. A FFNN converts a series of input data into one or more output data. All the neurons in one layer are connected to all the neurons of the next layer. These connections are associated with weights that represent the strength of the connection. All neurons perform the same basic operations: a) obtain input signals from above neurons; b) convert these signals to a *net* input signal; c) and transform the net signal into an output signal by application of a transfer function (in this case, a sigmoidal function).

FFNNs were trained with the ASNN program from Igor Tetko [60] to predict log (1/MIC), taking as input the mentioned selected descriptors for different data sets. Corrections were performed on the weights during the training (learning) using the Levenberg–Marquardt algorithm [61,62]. The number of neurons in the hidden layer was optimized for each case, in a range from one to five. Before the training, the whole training set was randomly partitioned into a learning set with 50% of the objects and a validation set with the other 50%. Full cross-validation of the entire training set was performed using the leave-one-out method (LOO). Each input and output variable was linearly normalized between 0.1 and 0.9 on the basis of the training set space. The maximum number of iterations used in the training was set to 5000. The training was stopped when there was no further improvements in the root mean squared error (*RMSE*) for the validation set. After the training, results were calculated for the learning set, the validation set, LOO, and an external test set.

The next step was to obtain the final prediction of an object from an ensemble of networks. An EnsFFNNs is made of several independently trained FFNNs that contribute to a single prediction [63,64]. The final prediction for an object, in our case the log (1/MIC) value, was the average of the outputs from all FFNNs of the ensemble. This methodology smoothes random fluctuations in the individual predictions of individual FFNNs. In our experiments ensembles of 50 networks were used.

2.5.3. Associative Neural Networks (AsNNs)

An AsNN [60,65] is a combination of a memory-less (ensemble of Feed-Forward Neural Networks, EnsFFNNs) and a memory-based (*k*-Nearest Neighbor, kNN, [66]) methods. The EnsFFNNs is combined with a *memory* into a so-called Associative Neural Network (AsNN). In this work, the memory consisted of a list of molecules, represented by their selected descriptors, and their corresponding log (1/MIC) value. The AsNN scheme was employed for composing a prediction of the log (1/MIC) value from a) the outputs from the EnsFFNNs and b) the data in the memory. More details concerning the full procedure of AsNNs to obtain a final prediction, in this case of log (1/MIC), can be found in the literature [67]. The experiments here described were performed with the ASNN program [60].

2.5.4. Counter-Propagation Neural Networks (CPNNs)

A Kohonen Self-Organizing Map, Kohonen SOM [18,68] is an unsupervised method that maps multidimensional objects, on the basis of their descriptors, into a 2D surface (a grid of neurons). This mapping is able to reveal similarities between objects: similar objects are mapped into the same or neighbor neurons in the map. Each neuron of the map consists in so many weights as there are input variables (object descriptors). In a Counter-Propagation Neural Network, CPNN, the input layer, consisting of a Kohonen SOM, is linked to an output layer of neurons. The input layer (Kohonen SOM) and the output layer are aligned. The output data, in our experiments the antitubercular activity expressed as log (1/MIC), are stored in the output layer that acts as a look-up table.

Before the training starts, the weights take random values calculated using as parameters the mean and the standard deviation of the corresponding descriptor in the input data set. In each cycle of the training, each individual object (in our experiments each hydrazide derivative) is mapped into the neuron with the most similar weights compared to its descriptors – called the central neuron, or winning neuron. The weights of the winning neuron are then corrected to make them even more similar to the properties of the presented object. The neurons in the neighborhood of the winning neuron also have their weights corrected. The extension of the correction depends on the topological distance to the winning neuron – larger adjustments are applied to weights of neurons closer to the winning neuron. The objects of the training set are iteratively fed to the map, and the weights corrected, until a pre-defined number of cycles is attained.

CPNNs with toroidal topology were used in the experiments presented in this work. A linear decreasing triangular scaling function was used in the training with an initial learning rate of 0.1 and an initial learning span of 3 and 7. The winning neuron was selected using the minimum Euclidean distance between the input vector and the neuron weights. In most of the experiments, 300 cycles were used to perform the training, with the learning span and the learning rate linearly decreasing until zero. CPNNs were implemented throughout this study with an in-house developed Java application derived from the JATOON Java applets [69,70].

2.6. Validation procedures

Any QSAR model should, in principle, meet two important goals: i - it should be able to interpret a given phenomenon, pinpointing the key factors that shape a certain response; and ii - it should be capable of making accurate predictions. The development of reliable and predictive QSAR models strongly depends on the application of rigorous validation procedures [30,34,35,37,39,71].

2.6.1. Internal validation

Internal validation was performed to guarantee the model's robustness to describe the data set used to establish it (training set). In this work, the validation procedure was initiated with a set of tests to confirm the model's linearity which was achieved by analyzing the residuals and the normal probability plots. Traditional statistical criteria such as the determination coefficient, R^2, the standard deviation of the fit, SD_{fit}, the F statistics, and the significance level (SL) of each adjusted parameter (parameters were kept if $SL > 95\%$) were likewise used to assess the reliability of the model. Intercorrelations among descriptors were also assessed. The data set was then scrutinized for outliers, using two criteria: the conventional measure $|Y_{\text{calc}} - Y_{\text{exp}}| > 2SD$ and a more refined measure known as the Cook's distance [72–75]. Outliers were carefully inspected to extract any relevant information from them and sequentially removed from the original training sets. Models were refitted after each removal until stable training sets were attained being the above mentioned criteria re-evaluated each time. Internal validation was further achieved by the leave-many-out (LMO) cross validation test, leading to a cross-validated correlation coefficient, Q^2:

$$Q^2 = 1 - \frac{\sum_{i=1}^{\text{training}} \left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{\text{training}} \left(Y_i - \overline{Y}_i\right)^2} \qquad (2)$$

where Y_i, $\hat{Y}_i$ e $\overline{Y}_i$ are the measured, predicted and averaged (over the whole data set) values of the dependent variable, respectively. A $Q^2 > 0.6$ was taken as a good measure of the robustness of the model. However, high Q^2 values are considered a necessary, but not a sufficient, condition for a model to have a high predictive power [37,71]. *RMSE* values were also calculated.

2.6.2. External validation

In order to evaluate the true predictive ability of a QSAR model it is mandatory to use an external validation procedure, that is, it is essential to compare predicted and experimental activities for an independent data set (a test set) of compounds not used in the development of the model. In this work, the test sets used for external validation were chosen so that they satisfied the same variability requisites as the training sets in terms of both activities and descriptors' space and, moreover, it was ensured that every compound in each test set was close to at least one compound in the corresponding training set.

The predictive power of each MLR and NN QSAR model was assessed on the basis of complementary validation criteria and following the strictest threshold values proposed to date in the literature [76]. Firstly, it was estimated an external Q^2 (Q^2_{ext} (F1) [77,78]) value:

$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{test}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{test}\left(Y_i - \overline{Y}_{training}\right)^2} \quad (3)$$

where Y_i *e* $\hat{Y}_i$ are, respectively, the measured and predicted (over the test set) values, and $\overline{Y}_{training}$ is the averaged value of the dependent variable for the training set. A severe cut-off value of 0.70 has been recently suggested as an unambiguous value to accept a model as being predictive [76].

The models' predictive ability was also checked by the r^2_m metrics [79–82], which does not consider the differences between $Y_{i(test)}$ and $\overline{Y}_{training}$, thus avoiding an alleged overestimation of the quality of predictions due to a wide *Y*-range:

$$r^2_m = R^2\left(1 - \sqrt{\left(R^2 - R^2_0\right)}\right) \quad (4)$$

In this expression R^2 is the test set regression determination coefficient and R^2_0 is the corresponding value for the regression that passes through the origin [79–82]. Instead of r^2_m it is, however, more common to use $\overline{r}^2_m$, which results from the average between observed *vs.* predicted and predicted *vs.* observed r^2_m values for test set objects and also their (absolute) difference, Δr^2_m, which should be as close as possible to zero. Also in the case of this validation parameter, a new and more restrictive value of 0.65 has been recommended [76] to substitute the commonly used threshold of 0.50 [81,82] for the acceptance of good QSAR models.

A further "measure" of external predictivity given by the concordance correlation coefficient, *CCC*, was also considered:

$$CCC = \frac{2\sum_{i=1}^{n_{ext}}\left(Y_i - \overline{Y}\right)\left(\hat{Y}_i - \overline{\hat{Y}}\right)}{\sum_{i=1}^{n_{ext}}\left(Y_i - \overline{Y}\right)^2 + \sum_{i=1}^{n_{ext}}\left(\hat{Y}_i - \overline{\hat{Y}}\right)^2 + n_{ext}\left(\overline{Y} - \overline{\hat{Y}}\right)^2} \quad (5)$$

where Y_i and $\hat{Y}_i$ correspond to the abscissa and ordinate values of the plot of experimental *vs.* predicted values, n is the number of compounds, and $\overline{Y}$ and $\overline{\hat{Y}}$ correspond to the averages of abscissa and ordinate values, respectively.

This parameter measures both precision (*i.e.*, how far the observations are from the fitting line) and accuracy (*i.e.*, how far the regression line deviates from the line with slope equal to 1 and passing through the origin, that is the concordance line), and any departure of the regression line from the concordance line gives a *CCC* value smaller than 1. The advantage of *CCC* is that there is no training set information involved in its definition and, therefore, it can be taken as a true external validation measure [78], being 0.85 the proposed cut-off value [76].

The predictive capacity of each model was further evaluated by computing various measures of fit between predicted and experimental values, namely, the average error (*AE*), the absolute average error (*AAE*) and the already referred *RMSE* [30]. The external validation of test set compounds still complied with the following criteria [37,71]:

$$R^2 > 0.6; \frac{R^2 - R^2_0}{R^2} < 0.1 \;;\; 0.85 < m < 1.15$$

where m is the slope of the regression between predicted and experimental values and the remaining parameters have the same meaning as above.

Finally, in order to preclude the possibility of having high statistical parameter values and yet models poorly predictable, scatter plots of predicted *vs.* experimental values for test set compounds were concomitantly visually inspected to spot possible deviations in the process of checking a model's predictive ability [76].

2.6.3. Randomization

To remove the possibility of chance correlation in the developed MLR and NN QSAR models, a randomization experiment was carried out for each model. A *Y*-randomization technique was therefore implemented for the best models by scrambling twenty five times the *Y*-column values of log (1/MIC) while keeping the descriptor matrix unchanged. Each randomized model was then used to make predictions for the test set and the corresponding determination coefficient, R^2, and Root Mean Squared Error, *RMSE*, were computed. A considerable decrease in R^2 and an increase in *RMSE*, in comparison with the non-randomized model, clearly emphasize the statistical reliability and robustness of the original model. To better assess if the difference between the original R^2 and the average R^2 for the randomized modelsR^2_r, can be considered significant, an additional statistical parameter, the Todeschini ${}^cR^2_p$ parameter, was calculated according to the following expression [83]:

$${}^cR^2_p = R\sqrt{\left(R^2 - R^2_r\right)} \quad (6)$$

QSAR models that present ${}^cR^2_p$ values higher that 0.5 are considered robust and reliable and not simply an outcome of chance correlation.

3. Results and discussion

3.1. Comparison between MLR and NN models

3.1.1. Lipophilicity

In early stages of drug development, it is very important to be able to model the behavior of certain physicochemical properties such as lipophilicity which is usually considered to determine the ability of a compound to penetrate a biological membrane. In addition, this property affects drug absorption, bioavailability, metabolism and toxicity, and drug-receptor interactions, and may hence influence significantly the activity of a given compound or family of compounds. Its prediction is therefore of crucial importance to anticipate the viability of new compounds as drug-like candidates. From the available measures of lipophilicity, the

partition coefficient between octanol and water is one of the most frequently used parameters in QSAR studies.

We have thus begun our study by analyzing the performance of MLR and NN models in describing and predicting the lipophilicity of hydrazide (data set 1) and isoniazid (data set 2) derivatives. The same training and test sets were always used when different QSPR models were compared. The best found models are shown in Table 2 (predicted values for these test set compounds are available in Table S4 of Supporting information).

Polanski et al. [84] have stated that in general log *P* is used to describe biological transport phenomena, having consequently an optimum value and so, according to these authors, its behavior could not be modeled by a linear function. However, our results clearly show that linear regression models such as model 1 and 5 in Table 2 are able to describe reasonably well the lipophilic behavior of these hydrazide derivatives in terms of the basicity of the solute (*B*), its polarity/polarizability (*S*) and its McGowan characteristic molar volume (*V*). Nevertheless, non-linear NN models developed with the same descriptors show an even higher predictive ability than MLRs for both data sets. Among these, AsNNs and EnsFFNNs models have, as expected, the best predictive capability, presenting the smallest prediction errors. Our MLT models are therefore suitable to model and predict lipophilicity, as measured by *c*log *P*, for this family of compounds.

3.1.2. Antitubercular activity

In a previous work [30] we have shown that the antitubercular activity of these same compounds with a hydrazide functionality was well described by the basicity of the solute (*B*) and its McGowan characteristic molar volume (*V*) – (model 9 in Table 3). Results in Table 3 show, additionally, that all four MLR and NN models, based on these two descriptors have very similar predictive abilities. We notice, however, that in the chemical space of the corresponding training set, *B* and *V* have a nonnegligible intercorrelation ($R^2 = 0.58$) and the use of both descriptors can be considered redundant. In fact, *V* alone explains almost all of the variability in log (1/MIC).

We have also demonstrated in the referenced work [30] that in the case of isoniazid derivatives (see Fig. 1) electronic and geometrical descriptors such as dipole moment (μ), Mulliken charge on the external hydrazide N atom (*c*) and interatomic distance between the external hydrazide N atom and the first atom of the substituent (d_4), lead to a MLR model with a significant predictive power (model 13 in Table 4).

We have also reported [30] that steric factors as the length and volume of the substituents, described, respectively, by the Verloop parameters *L* and B_5, and the 2D distance (d_K) between the external hydrazide N atom of the hydrazide functionality and the N atom of the pyridinic ring were also relevant to interpret log (1/MIC) values and gave rise likewise to a MLR model with a high predictive power (model 17 in Table 5).

The statistical quality of MLR results expressed in Tables 3–5 show that electronic, geometrical and steric properties are important and critical to explain the antitubercular activity of both hydrazide and isoniazid derivatives, in contrast to their physicochemical properties. Independently of the set of descriptors that best explain the variability in log (1/MIC), the three MLR model equations are able to predict log (1/MIC) values for these classes of compounds to around 0.30 log units.

Applying the same methodology as before, that is, testing the various NN approaches for the same set of descriptors, reveals once more that, regardless of the nature of the descriptors involved, NNs always have a higher predictive ability than the corresponding MLR models, and this is particularly evident in the case of AsNNs (Tables 4 and 5). Also, as expected, the ensembles of networks (EnsFFNNs) give, in general, better predictions than individual FFNN models.

Giving the considerable ability of Artificial Neural Networks to deal with a high number of variables simultaneously, we decided to investigate the influence of increasing the number of descriptors on the prediction capability of NN models.

For that purpose, we used the 30 most-relevant descriptors selected by Random Forests and derived the best MLR model by a forward stepwise procedure and rigorous statistical criteria. The best found descriptors were the following: three energetic descriptors, the highest occupied molecular orbital (*HOMO*), the lowest unoccupied molecular orbital (*LUMO*), and the 1,4-interaction energy (1,4-IE), two geometry related descriptors, the molecular length (*ML*) and the surface area (*SA*) and two steric descriptors, the molar refractivity (*MR*) and an attempt to define the effective volume of the whole substituent, measured by the Verloop parameter B_5. Then, with these descriptors and the same training and test sets we established the corresponding NN models. Results for these test sets are presented in Table 6.

A close inspection of the statistical parameters in Table 6 reveals again a clear superiority of NN models, in particular of AsNNs, in the prediction of the antitubercular activity of hydrazide derivatives (within ±0.37 log units). An incautious comparison between models 20 (Table 5) and 24 (Table 6) could give the wrong impression that model 20 is a better model to predict log (1/MIC).

However, we recall that model 24, as all others presented in Table 6, allows the prediction of antitubercular activity with good statistical quality for a more diverse test set of compounds (we emphasize that in this case both training and test sets correspond to a larger number of compounds with a higher chemical dissimilarity – both isoniazid and hydrazide derivatives are included).

Fig. 2 shows the predicted *vs.* experimental values of log (1/MIC), for model 24. Predicted values for all test set compounds of

Table 2
Statistical coefficients for test set predictions of *c*log *P* by four statistical and machine learning techniques for two different data sets. QSPR models developed with three descriptors (*B*, *S*, *V*).

Model	Technique	Data set	R^2	R_0^2	*SD*	*F*	*AE*	*AAE*	Q^2_{ext}	*RMSE*	$\bar{r}^2_m$	Δr^2_m	*CCC*
1	MLRs	1	0.776	0.763	0.596	128	0.191	0.547	0.788	0.672	0.677	0.020	0.867
2	FFNNs		0.832	0.831	0.469	183	0.075	0.430	0.820	0.578	0.808	0.022	0.892
3	EnsFFNNs		0.849	0.847	0.441	208	0.109	0.479	0.831	0.560	0.800	0.019	0.898
4	AsNNs		0.850	0.846	0.447	209	0.125	0.480	0.831	0.559	0.788	0.018	0.900
5	MLRs	2	0.871	0.824	0.575	115	0.253	0.570	0.855	0.672	0.709	0.054	0.915
6	FFNNs		0.849	0.825	0.585	95	0.113	0.513	0.836	0.694	0.742	0.050	0.903
7	EnsFFNNs		0.907	0.903	0.493	166	0.031	0.395	0.903	0.533	0.860	0.027	0.946
8	AsNNs		0.913	0.912	0.498	179	−0.002	0.363	0.913	0.507	0.896	0.023	0.953

Techniques: MLRs – Multiple Linear Regressions; FFNNs – Feed Forward Neural Networks; EnsFFNNs – Ensemble of Feed Forward Neural Networks; AsNNs – Associative Neural Networks.

Table 3
Statistical coefficients for test set predictions of log (1/MIC) by four statistical and machine learning techniques. QSAR models developed with two descriptors (*B*, *V*).

Model	Technique	Data set	R^2	R_0^2	*SD*	*F*	*AE*	*AAE*	Q^2_{ext}	*RMSE*	$\overline{r}_m^2$	Δr_m^2	*CCC*
9	MLRs	3	0.774	0.679	0.291	93	−0.074	0.294	0.764	0.347	0.614	0.158	0.859
10	FFNNs		0.778	0.687	0.298	95	−0.092	0.290	0.761	0.348	0.621	0.155	0.864
11	EnsFFNNs		0.776	0.577	0.251	93	−0.073	0.304	0.739	0.364	0.528	0.197	0.829
12	ASNNs		0.776	0.577	0.251	93	−0.073	0.304	0.739	0.364	0.528	0.197	0.829

Techniques: MLRs – Multiple Linear Regressions; FFNNs – Feed Forward Neural Networks; EnsFFNNs – Ensemble of Feed Forward Neural Networks; AsNNs – Associative Neural Networks.

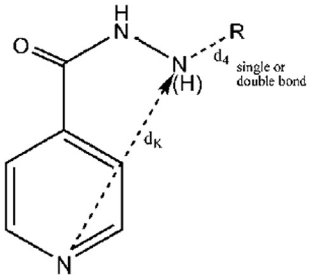

Fig. 1. Representation of the geometrical descriptors used in the isoniazid derivatives' models.

data sets 3–6, as computed by each model referred in Tables 3–6, are available in Table S4 of Supporting Information.

3.2. Importance of descriptors within different QSAR models

The importance of the seven descriptors used to build the MLR, EnsFFNNs and AsNNs models (data set 6) was further analyzed using the randomization technique. After the models were built, the first column corresponding to the first descriptor used in the model (Verloop parameter B_5) was scrambled, keeping the remaining descriptor matrix and *Y*-column unchanged. The new matrix was submitted to the previous trained models and the results compared to the original ones in terms of absolute average error, *AAE*. The increment in *AAE* using the descriptor values scrambled is a measure of the importance of that descriptor in the model. Higher increments correspond to higher importance of the descriptors in the model. This procedure was performed for all seven descriptors. Results are presented in Fig. 3.

An analysis of Fig. 3 indicates that the most important descriptor for the three MLT is the molar refractivity, *MR*, and the less important is the energy of the lowest unoccupied molecular orbital, *LUMO*. The three most important descriptors are the molar refractivity, *MR*, the surface area, *SA* and the molecular length, *ML*. Worth of notice is the fact that the three most important descriptors are the same regardless of the MLT used, in spite of the different ways used by each method to "learn" the data and model the antitubercular activity by giving a certain relative importance to each descriptor. The same happens on the side of the less important descriptors, being the global ranking the same for all three models. It is interesting to note that the most sensitive method to the scrambling of descriptors is MLR with an *AAE* increment in some cases 3 or 4-fold higher than that of AsNNs. From the three methods, AsNNs are the less sensitive, perhaps due to the fact that a prediction from an AsNN is composed from the output from the EnsFFNNs and the data in the memory (training data) that was used to make a correction in the EnsFFNNs prediction.

3.3. Assessment of validity, accuracy and predictive ability

As mentioned before, all QSAR/QSPR models were evaluated on the basis of rigorous statistical criteria. We highlight the figures of merit of the external validation procedures applied to test sets (particularly those from Tables 4–6), namely, the average error *AE*, the absolute average error *AAE*, the cross validation coefficient, Q^2_{ext}, the Root Mean Squared Error, *RMSE*, the $\overline{r}_m^2$ and Δr_m^2 metrics and *CCC*. The latter three metrics as well as Q^2_{ext} exceed in every case the recommended threshold values referred in Section 2.6.2.

In order to compare the accuracy of the predictions made by the chosen MLR, EnsFFNNs and AsNNs models (Table 6), we have scrutinized the magnitude of the differences between predicted and experimental values in each case. The absolute error of predictions for the test set of data set 6 varies between 0 and 1.0 log (1/MIC) units, as shown in Fig. 4. The distribution of compounds in a given range of absolute errors is almost similar for all three methodologies. In fact they predict the same number of compounds (6) with an absolute error lower than 0.2 and a similar number (6 or 7) with an absolute error between 0.4 and 1.0.

Fig. 5 shows a chart of log (1/MIC) values predicted by MLR and AsNNs models for all test set compounds of data set 6, and their comparison with experimental values. If one considers a compound as potentially active when log (1/MIC) $>$ 0, it is remarkable to realize that only one test set compound, compound Te12, is wrongly predicted as active, being all other compounds correctly predicted as active or inactive by both models. In spite of some differences already observed in the prediction accuracy between MLR and AsNNs (Fig. 4), the trends spotted in Fig. 5 put in evidence that MLR and AsNNs models agree in 100% of the cases when it comes to the prediction of a compound as active or inactive, being the only, common, misclassification compound Te12, just mentioned.

MLR and AsNNs models exhibit for the test set of data set 6 an *AAE* of 0.382 and 0.345, a *RMSE* of 0.472 and 0.437, a $\overline{r}_m^2$ of 0.694 and 0.702, a Δr_m^2 of 0.057 and 0.051 and a *CCC* of 0.888 and 0.910,

Table 4
Statistical coefficients for test set predictions of log (1/MIC) by four statistical and machine learning techniques. QSAR models developed with three descriptors (μ, *c*, d_4).

Model	Technique	Data set	R^2	R_0^2	*SD*	*F*	*AE*	*AAE*	Q^2_{ext}	*RMSE*	$\overline{r}_m^2$	Δr_m^2	*CCC*
13	MLRs	4	0.835	0.833	0.273	46	−0.053	0.294	0.817	0.326	0.788	0.012	0.888
14	FFNNs		0.838	0.837	0.257	47	−0.042	0.311	0.810	0.333	0.809	0.013	0.879
15	EnsFFNNs		0.885	0.884	0.196	70	−0.048	0.299	0.816	0.327	0.851	0.010	0.875
16	AsNNs		0.930	0.921	0.169	120	−0.079	0.234	0.877	0.268	0.839	0.005	0.922

Techniques: MLRs – Multiple Linear Regressions; FFNNs – Feed Forward Neural Networks; EnsFFNNs – Ensemble of Feed Forward Neural Networks; AsNNs – Associative Neural Networks.

Table 5
Statistical coefficients for test set predictions of log (1/MIC) by four statistical and machine learning techniques. QSAR models developed with three descriptors (L, B_5, d_K).

Model	Technique	Data set	R^2	R_0^2	SD	F	AE	AAE	Q^2_{ext}	$RMSE$	$\bar{r}^2_m$	Δr^2_m	CCC
17	MLRs	5	0.896	0.858	0.256	78	0.131	0.222	0.867	0.272	0.740	0.040	0.931
18	FFNNs		0.885	0.883	0.223	69	−0.018	0.242	0.867	0.270	0.861	0.026	0.919
19	EnsFFNNs		0.946	0.937	0.150	158	0.017	0.166	0.913	0.218	0.862	0.017	0.946
20	AsNNs		0.935	0.935	0.185	129	−0.014	0.159	0.928	0.198	0.926	0.015	0.960

Techniques: MLRs – Multiple Linear Regressions; FFNNs – Feed Forward Neural Networks; EnsFFNNs – Ensemble of Feed Forward Neural Networks; AsNNs – Associative Neural Networks.

Table 6
Statistical coefficients for test set predictions of log (1/MIC) by four statistical and machine learning techniques. QSAR models developed with seven descriptors (1,4-*IE*, *HOMO*, *LUMO*, *ML*, *SA*, *MR*, B_5).

Model	Technique	Data set	R^2	R_0^2	SD	F	AE	AAE	Q^2_{ext}	$RMSE$	$\bar{r}^2_m$	Δr^2_m	CCC
21	MLRs	6	0.845	0.824	0.386	87	0.235	0.382	0.794	0.472	0.694	0.057	0.888
22	FFNNs		0.846	0.802	0.392	88	0.298	0.402	0.768	0.504	0.639	0.060	0.876
23	EnsFFNNs		0.860	0.840	0.345	98	0.243	0.363	0.802	0.467	0.713	0.053	0.884
24	AsNNs		0.874	0.846	0.374	111	0.234	0.345	0.824	0.437	0.702	0.051	0.910

Techniques: MLRs – Multiple Linear Regressions; FFNNs – Feed Forward Neural Networks; EnsFFNNs – Ensemble of Feed Forward Neural Networks; AsNNs – Associative Neural Networks.

respectively (Table 6). The latter metrics are well above the most precautionary cut-off values [76]. On the other hand, a scatter plot of predicted *vs.* experimental log (1/MIC) values for this test set – Fig. 6 – does not show any systematic deviation, thus further assessing the quality of the predictions.

Despite the good results obtained regarding test set predictions, Table 7 discloses some relatively high errors for a few compounds. Three examples of compounds predicted with errors higher than 0.5 are presented in Table 8: compound Te01 which has the smallest log (1/MIC), compound Te12 which was the compound wrongly predicted as active, and compound Te02, randomly chosen from the remaining compounds with errors larger than 0.5. Table 8 shows the predictions made by MLR and AsNN models in the three cases and for each of them the most similar compounds in the training set. The Euclidean distance between the descriptor vectors for each test set example and the compounds from the training set was used as a measure of similarity.

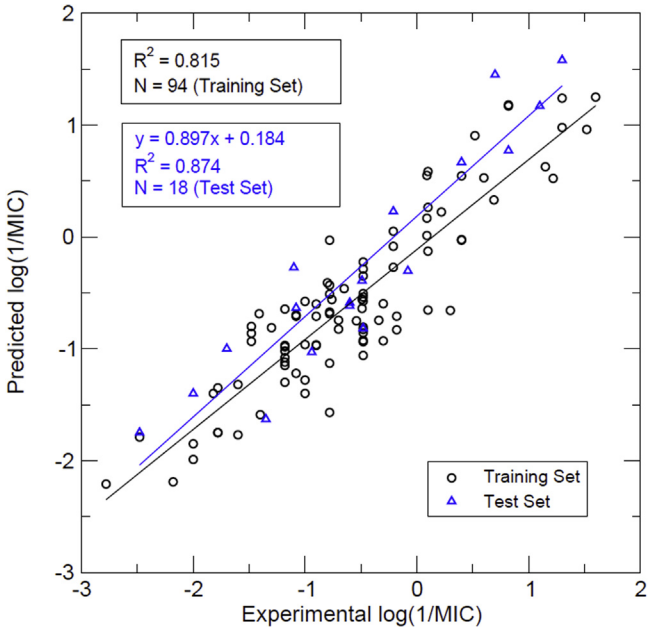

Fig. 2. Predicted *vs.* experimental AsNNs values for log (1/MIC) (data set 6, model 24).

The search for the most similar compounds to Te01 in the training set, in terms of Euclidean distance, leads to a significant range in this parameter (from 0.15 to 0.30) and shows that the training set compounds present higher experimental values of log (1/MIC), in reasonable agreement with predicted values, even if quite different from the experimental value. The same happens for compound Te02 that has an experimental log (1/MIC) of −2.00 but higher predicted values, namely, −1.66 and −1.40. The five most similar compounds in the training set (degree of similarity ranging from 0.21 to 0.24) have experimental log (1/MIC) between −1.18 and −1.48, again in better conformity with predicted values. Finally, compound Te12, as already referred, is the only inactive compound that was wrongly predicted as active. After the analysis of the five most similar compounds in the training set, one can see that two of them are active compounds, Tr09 and Tr10, with log (1/MIC) of 0.40 and 0.40, respectively, very close to the predicted values for the test set compound. Tr12 has a higher Euclidean distance and although its experimental value is a bit farther away from the predicted values, it is also active. Therefore, it is not strange that Te12 has been predicted as active like the majority of the similar training set compounds.

Fig. 7 reveals the molecular structures of the compounds analyzed in Table 8.

The first row, corresponding to the analysis of compound Te01, shows very clearly that the most similar training set compounds in terms of descriptors' similarity display very diverse structures

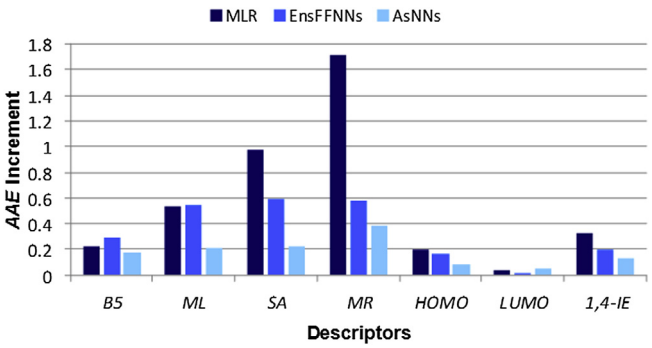

Fig. 3. Increment in absolute average error, *AAE*, for test set compounds with descriptor values scrambled for each one of the seven descriptors (data set 6, models 21, 23 and 24).

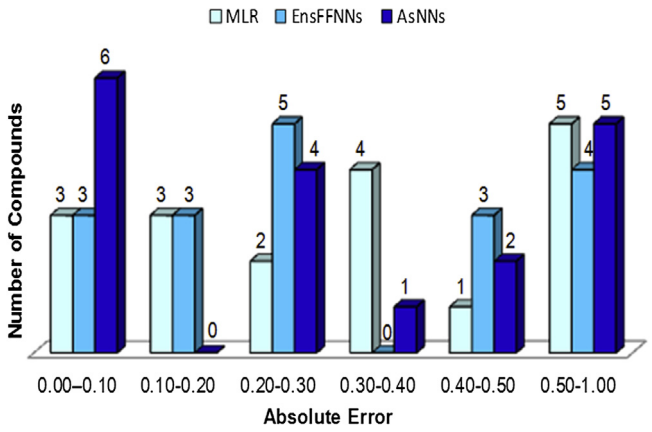

Fig. 4. Assessment of number of test set compounds predicted by MLR (model 21), EnsFFNNs (model 23) and AsNNs (model 24) in each absolute error interval (data set 6).

(hydrazides, isonicotinoyl hydrazides, benzoic acid hydrazides and furan-2-carboxylic acid hydrazides). This is the reason why Euclidean distances to the test set compound are so different from each other and why the two models make fairly good predictions, *vis-à-vis* their most similar compounds in the training set, and not so good when one compares the predicted values with the experimental log (1/MIC) for Te01. A similar situation was found for Te02 with perhaps even more diversity than in the previous case. The five most similar compounds in terms of descriptors' vectors present quite different structures (benzoic acid hydrazides, furan-2-carboxylic acid hydrazides, hydrazides and isonicotinoyl hydrazides). In this case, Euclidean distances are unexpectedly similar, but the ability of the models to predict the test set compound's activity is, as previously, also dependent on how much different is Te02 from the closest compounds in the training set. Compound Te12 is from the three examples the one where the most similar compounds in terms of descriptors' vectors correspond indeed to similar structures – all structures are isonicotinoyl hydrazones and the average Euclidean distance is the smallest of all three cases. Structurally speaking, the most similar compounds to Te12 are compounds Tr08, Tr11 and Tr12, whose experimental log (1/MIC) values are in fact either identical or very close to the Te12 experimental value.

Results on Fig. 5 and Tables 7 and 8 show that, in spite of the very good performance of MLR and AsNNs models in the great majority of the cases, there are a few circumstances where different structures exhibit very similar activities and others where very similar structures present different activities. In some of the former, different structures might be encoded by similar descriptors and as a result the methods are able to "learn" and model the relation between structure and activity. In the latter, the modeling process becomes very hard, and the selected descriptors used to establish the models are not able to explain, for the analyzed cases, the whole variability in the activity.

A consensus model resulting from averaging predictions for all four models – MLR, FFNNs, EnsFFNNs and AsNNs – in Table 6 was also tried in order to assess if a hybrid approach would lead to an enhanced performance [31,35,85,86]. However, in this case, the consensus model did not do better than the individual AsNN model. The reason for this has to do, on one hand, with the fact that the AsNN model itself already produces the most accurate predictions and, on the other hand, because neither MLR nor AsNNs models show strong fluctuations in the predictions. As already seen in Fig. 5 they have indeed a very similar ability to predict a compound as active or inactive, with just one exception, as pointed out before. Also, EnsFFNNs and AsNNs are by nature methods that already smooth random fluctuations in the individual predictions of FFNNs and therefore any fluctuations that could possibly be overcome by a consensus approach are again not an issue for these models.

To further validate models 21–24 in Table 6 as regards to the removal of any possibility of chance correlation, a *Y*-randomization technique was performed as explained in Section 2.6.3. After every randomization, each model was used to make predictions for the test set of data set 6, giving always rise to a considerably lower R^2 and a systematically higher *RMSE* when compared to the non-randomized models, thus supporting the statistical reliability/robustness of the original models. In order to confirm whether the difference between R^2 and R_r^2 was significant, values of ${}^cR_p^2$ were calculated and are presented in Table 9.

It can be easily seen that ${}^cR_p^2$ values are markedly above the threshold value of 0.5 and all models are therefore robust and are not the result of chance only.

Additionally, the two best models (models 21 and 24 in Table 6) were used to predict the antitubercular activity of an external compounds' collection of 22 INH derivatives retrieved from the Novartis TB public access database [47]. The figures of merit on Table 10 show that our models have also a good predictive ability for compounds totally outside the initial compounds' matrix but falling into the same applicability domain (predicted values for these test set compounds are available in Table S4 of Supporting information). In fact, values for Q_{ext}^2 for both MLR and AsNNs models are higher than the referred most demanding threshold value (0.707 and 0.734, respectively) and $\overline{r}_m^2$ values, although below 0.65, are well above 0.5, the formerly accepted value for this metric, being 0.582 and 0.617, respectively. Also, *CCC* is very close in both cases to the recommended 0.85 value.

Moreover, a scatter plot of predicted *vs.* experimental log (1/MIC) values for the Novartis test set – Fig. 8 – does not show again (as for test set of data set 6) any systematic deviation between predicted and experimental data, which, together with all the used validation criteria, reinforces the predictive ability of our QSAR models.

The biological data used in this work were collected from several sources in the literature to guarantee that a data set of convenient size could be used for the QSAR analysis. However, these data refer to two different strains (BCG and H37Rv) and this could cast some doubts on whether the two strains would be giving the same information on the biological activity of these compounds. Even though none of our models which were built using, indistinctly, MIC values for both strains, have ever shown any systematic tendency/deviation in MICs ascribable to this fact, in order to circumvent the possibility of attributing some of the described behaviors to data inconsistency as a result of merging diverse data, we have re-done models 21 and 24 considering in the training set only compounds with MIC values against BCG (81 compounds, 3 outliers) – see new division of strains in Table S6 of Supporting information. We have then used the new models (models 21a and 24a) to predict the activity of test set compounds which included only compounds with MIC values against the H37Rv strain (22 compounds, 6 outliers). Results are presented in Table S7 of Supporting information. Models' 21a and 24a predictive ability towards the independent test set with just H37Rv MIC values is very similar to that of models using BCG and H37Rv strains altogether. Hence, at least in this case, there seems to be no reason to take these two strains as different "end-points" for modeling purposes.

3.4. Design of new potentially active compounds

The ultimate goal of any QSAR model applied to the study of tuberculosis is the understanding of the factors responsible for the

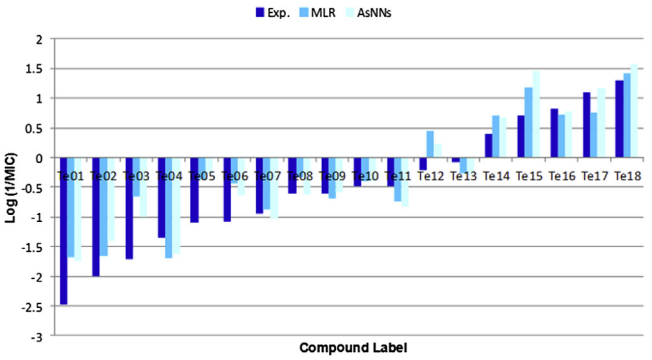

Fig. 5. Comparison between experimental log (1/MIC) values and predictions made by MLR and AsNNs models for test set (data set 6, models 21 and 24).

mechanism of action of available drugs against *M. tuberculosis*. This knowledge of an underlying relationship between structural characteristics and measurable activity will help in the design of new compounds with potentially high antitubercular activity and, will, hopefully, also contribute to better comprehend the phenomenon of bacterial resistance to existing drugs.

Having established and externally validated the MRL and AsNNs models, we have used them to assist us in the design of new potentially active compounds. The selection of compounds to be synthesized was not done solely on the basis of a simple ranking of activities for a set of known compounds submitted to the models but was also based on a meticulous analysis of individual features that clearly seemed to improve the probability of a given compound to be active. The modeling power of these individual features was evaluated from: a) the information retrieved from MLR models; and b) the inspection of weights, at their different levels, in CPNNs.

Our best MLR models [30] clearly show that the most relevant factors to the antitubercular activity of isoniazid derivatives are steric characteristics related, on one hand, with the size and volume of the substituent (L, B_5) and, on the other hand, with the distance parameter described by d_K (see Fig. 1 and model 17). L and d_K enhance the antitubercular activity, whereas the substituent's volume has the opposite effect. Furthermore, electronic aspects reflected by μ and c, which were shown to be activity promoters

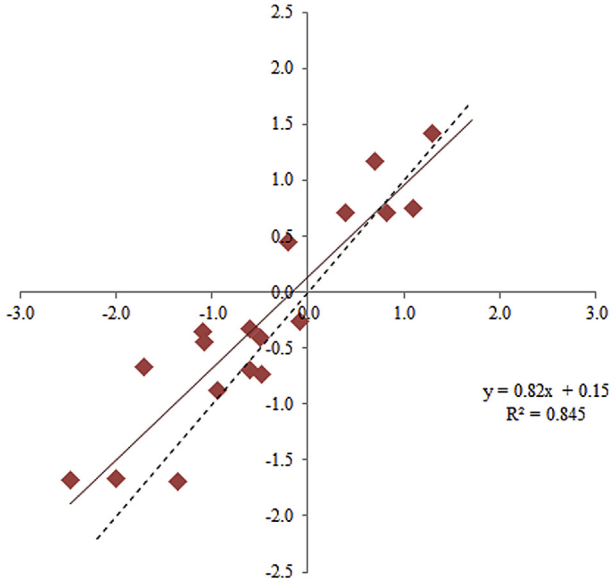

Fig. 6. Scatter plot of log $(1/MIC)_{pred}$ *vs.* log $(1/MIC)_{exp}$ for test set of data set 6.

Table 7
Comparison between experimental log (1/MIC) values and predictions made by MLR ($R^2 = 0.845$, $Q^2_{ext} = 0.794$) and AsNNs ($R^2 = 0.874$, $Q^2_{ext} = 0.824$) models for test set (dataset 6, models 21 and 24).

Label	Log (1/MIC)		
	Exp.	MLR	AsNNs
Te01	−2.48	−1.68	−1.75
Te02	−2.00	−1.66	−1.40
Te03	−1.70	−0.66	−1.00
Te04	−1.35	−1.69	−1.63
Te05	−1.10	−0.36	−0.27
Te06	−1.08	−0.44	−0.64
Te07	−0.94	−0.88	−1.03
Te08	−0.60	−0.32	−0.62
Te09	−0.60	−0.69	−0.59
Te10	−0.49	−0.40	−0.39
Te11	−0.48	−0.74	−0.82
Te12	−0.21	0.45	0.23
Te13	−0.08	−0.26	−0.30
Te14	0.40	0.71	0.67
Te15	0.70	1.18	1.45
Te16	0.82	0.72	0.77
Te17	1.10	0.76	1.17
Te18	1.30	1.42	1.58

and interatomic distance, d_4, which was seen as having a negative impact upon activity, are also of great importance to the biological action of these compounds (model 13).

Moreover, the MLR model developed in this work and which involves seven descriptors (model 21), reveals that compounds with high values of the descriptors molecular length (*ML*), surface area (*SA*), 1,4-interaction energy (1,4-*IE*) and energy of the highest molecular orbital, *HOMO*, should have their antitubercular activity favored. In addition, the most active compounds should concurrently present low values of B_5, molar refractivity (*MR*) and energy of the lowest unoccupied molecular orbital (*LUMO*).

For the second part of the analysis, CPNNs were trained with exactly the same data sets used to develop the MLRs and AsNNs models. CPNNs models were not used here to predict compounds' activities but just to assist in the evaluation of individual features to help in the design of the new compounds. Fig. 9 shows the weights of two CPNNs at their three levels of descriptors for data sets 4 and 5, respectively.

A visual analysis of the weight levels confirms and supports the analysis from MLR models. In the CPNN trained with data set 4 using as descriptors the dipole moment, μ, the Mulliken charge, c, and the distance d_4 (first row), good correlations can be observed

Table 8
Examples of compounds with errors higher than 0.5 and comparison with the most similar compounds in training set in terms of Euclidean distance.

Label	Log (1/MIC)			Most similar compounds		
	Exp.	MLRs	AsNNs	Label	Euc. Dist.	Exp. log (1/MIC)
Te01	−2.48	−1.68	−1.75	Tr01	0.15	−1.78
				Tr02	0.25	−2.00
				Tr03	0.27	−1.48
				Tr04	0.28	−1.48
				Tr05	0.30	−1.18
Te02	−2.00	−1.66	−1.40	Tr03	0.21	−1.48
				Tr05	0.21	−1.18
				Tr06	0.23	−1.18
				Tr07	0.23	−1.41
				Tr04	0.24	−1.48
Te12	−0.21	0.45	0.23	Tr08	0.16	−0.21
				Tr09	0.19	0.40
				Tr10	0.21	0.40
				Tr11	0.22	−0.21
				Tr12	0.24	0.09

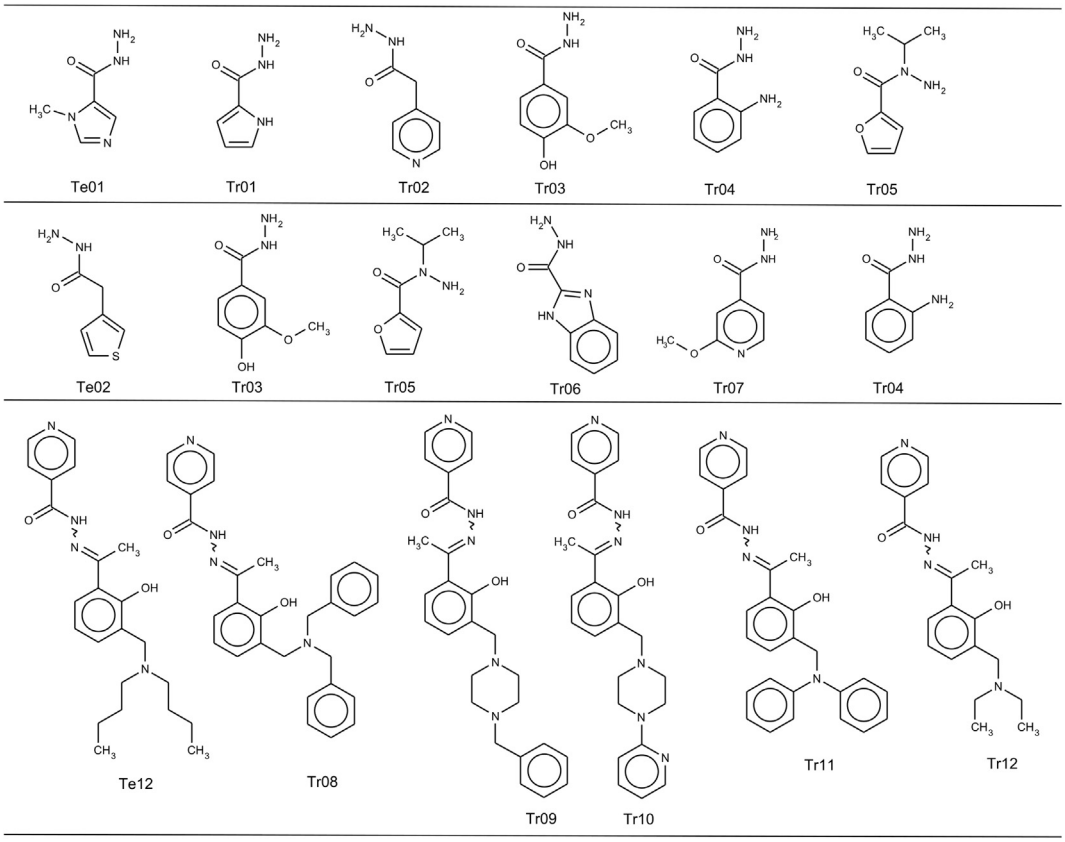

Fig. 7. Molecular structures of the examples analyzed in Table 8.

Table 9
Results of randomization tests for QSAR models 21–24 from Table 6.

Models	R^2	R_r^2	$^cR_p^2$
MLRs	0.845	0.230	0.702
FFNNs	0.846	0.221	0.718
EnsFFNNs	0.859	0.303	0.676
AsNNs	0.874	0.276	0.708

Techniques: MLRs – Multiple Linear Regressions; FFNNs – Feed Forward Neural Networks; EnsFFNNs – Ensemble of Feed Forward Neural Networks; AsNNs – Associative Neural Networks.

between the weight levels corresponding to Mulliken charge and, some regions, of dipole moment and the output layer, log (1/MIC). That is, regions corresponding to higher values of log (1/MIC) in the output layer correspond to regions of high values of weights in the Mulliken charge level 2 and also, in a small region of the map, in the dipole moment level 1. On the other hand, there is an inverse correlation between d_4 and the output layer (regions of higher values of d_4 correspond to regions of low values of log (1/MIC)).

The second row shows a CPNN trained with data set 5 and using as descriptors the Verloop parameter, *L*, the Verloop parameter B_5, and the distance d_K. The analysis of the weight levels in the regions of high values of log (1/MIC) in the output layer shows good correlations between the output layer and the weights level corresponding to d_K and, in some degree, also between the output layer and the weights level corresponding to the Verloop parameter *L*. The weight level corresponding to the Verloop parameter B_5 shows, on the contrary, an inverse correlation with the output layer, log (1/MIC), in regions of higher log (1/MIC).

Fig. 10 shows the weight levels of a CPNN trained with data set 6 using seven descriptors. Regarding the regions of the output layer with high values of weights (corresponding to high values of log (1/MIC)), an analysis of the regions in the various weight levels reveals that the levels corresponding to molecular length, *ML*, and the energy of the highest molecular orbital, *HOMO*, show the best correlations with the output layer. Conversely, the weight level corresponding to the energy of the lowest unoccupied molecular orbital, *LUMO*, shows an inverse correlation with the output layer.

The result of the analysis of the MLRs and CPNNs models discussed in this section is summarized in Table 11. This table shows, in a qualitative way, how the magnitude of each descriptor promotes the antitubercular activity (a plus sign indicates that high values of the descriptor promote activity, whereas a minus sign indicates that low values of the descriptor also enhances activity). From the

Table 10
Statistical coefficients for the predictions of log (1/MIC) for the external compounds collection retrieved from Novartis database using MLRs and AsNNs. QSAR models developed with seven descriptors (1,4-*IE*, *HOMO*, *LUMO*, *ML*, *SA*, *MR*, B_5).

Model	Technique	Data set	R^2	R_0^2	*SD*	*F*	*AE*	*AAE*	Q^2_{ext}	*RMSE*	$\bar{r}_m^2$	Δr_m^2	*CCC*
21	MLRs	NOVARTIS Database (22 INH derivatives)	0.655	0.611	0.328	38	−0.041	0.317	0.707	0.356	0.582	0.130	0.802
24	AsNNs		0.729	0.721	0.322	54	0.109	0.288	0.734	0.339	0.617	0.090	0.840

Techniques: MLRs – Multiple Linear Regressions; AsNNs – Associative Neural Networks.

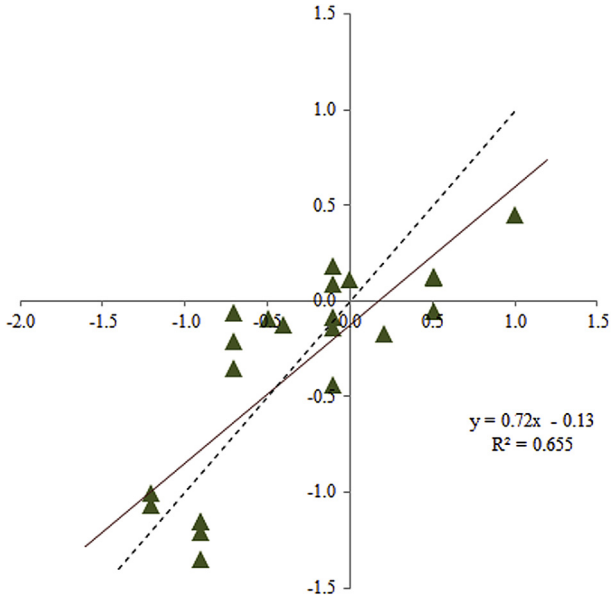

Fig. 8. Scatter plot of log (1/MIC)$_{pred}$ *vs.* log (1/MIC)$_{exp}$ for test set of the Novartis data set.

scrutiny of the individual features which worked as guidelines, a rational design of potentially active compounds was therefore performed. However, it is quite difficult to obtain compounds which obey, simultaneously, to all features. The various designed compounds were submitted to the previously established models (MLR and AsNNs) and their antitubercular activity predicted. A ranking of the best 30 compounds based on their predicted activity was assembled and two compounds (Fig. 11, compounds A and B), predicted with high activity by the majority of the models, were directly selected to be synthesized. The synthesis was achieved by the acylation of isoniazid with the appropriated acid chloride using a *N*-methylmorpholine as catalyst. The purity of the compounds was assessed by GC or high-performance liquid chromatography (HPLC) and was in both cases ≥98%. Compounds were fully characterized by ^{1}H and ^{13}C NMR, IR, and HRMS analysis [87]. The experimental antitubercular activity was then evaluated *in vitro* against the H37Rv strain of *M. tuberculosis*.

A few other compounds have likewise been tested *in silico* and appear also potentially promising. They have therefore been also selected for synthesis and further evaluation is now being carried out.

As mentioned before, it is not easy to design compounds for which all descriptors' values promote activity. In the two last columns of Table 11, we show, for compounds A and B, which descriptors agree in each case with the models' predictions.

The two new hydrazide molecules, A and B, were in fact designed so that most of the properties identified as promoting activity were enhanced. For compound A descriptors μ, *c*, *L*, d_K, *ML*, *MR* and *LUMO* present values which promote activity. For compound B the descriptors responsible for an increase in activity are descriptors *c*, d_4, B_5, d_K, *MR* and *LUMO*.

In Table 12 the difference between *in vitro* and predicted activities for the two synthesized compounds is computed, according to each model.

As can be observed in Table 12, in models developed with descriptors *L*, B_5 and d_K, MLRs appear to be more accurate than AsNNs in the prediction of MIC values for the two synthesized compounds, exactly in the opposite direction of what was expected from statistical reasoning solely based on the analysis of Table 5. On the contrary, for the other two sets of descriptors, models derived from associative neural networks, which had already been referred as

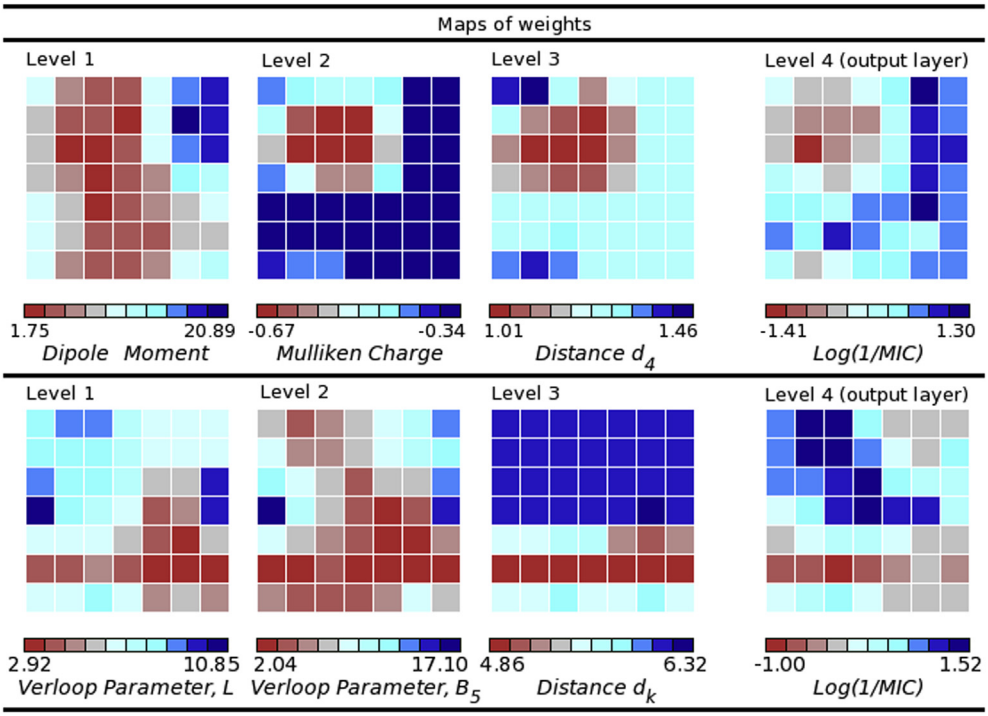

Fig. 9. Representation of weights in two trained CPNNs at different levels. First row-CPNN trained with data set 4 with 3 input layers (first three levels) and one output layer. Level 1 corresponds to dipole moment, level 2 to Mulliken charge, level 3 to distance d_4 and level 4 to log (1/MIC) (output layer). Second row – CPNN trained with data set 5 with three input layers (first three levels) and one output layer. Level 1 corresponds to Verloop parameter *L*, level 2 to Verloop parameter B_5, level 3 to distance d_K and level 4 to log (1/MIC) (output layer). The colored bar represents the weight scale for each layer. Low values of weights are represented by red and high values by blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

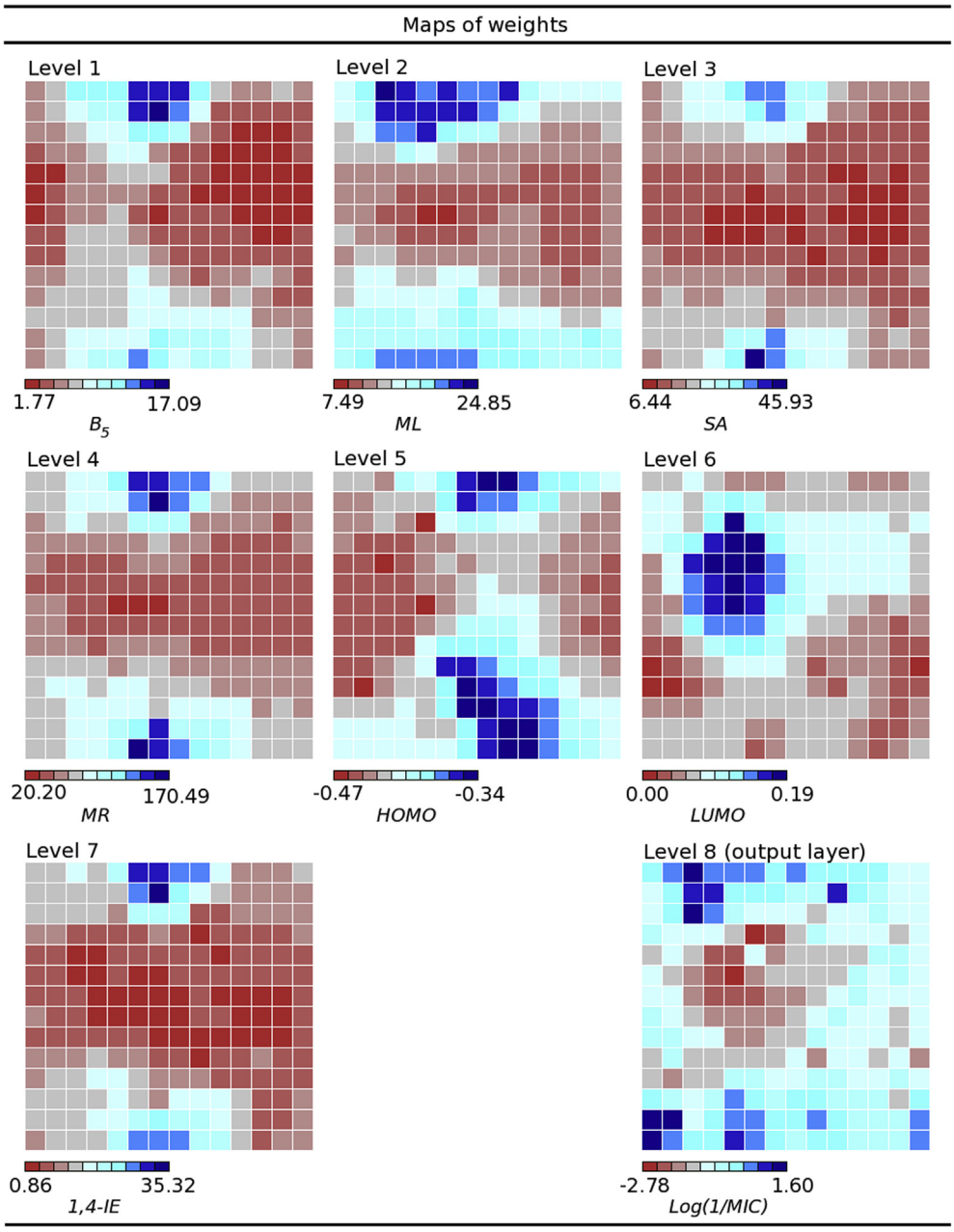

Fig. 10. Representation of weights in a trained CPNN at different levels. CPNN trained with data set 6 with 7 input layers and one output layer. Level 1 corresponds to Verloop parameter B_5, level 2 to molecular length, *ML*, level 3 to surface area, *SA*, level 4 to molar refractivity, *MR*, level 5 to *HOMO* energy, level 6 to *LUMO* energy, level 7 to 1,4-interaction energy, and level 8 to log (1/MIC) (output layer). The colored bar represents the weight scale for each layer. Low values of weights are represented by red and high values by blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

having a better predictive ability, as measured by various external validation parameters (Tables 4 and 6), provide indeed much better predictions than MLR models. One exception is, however, detected and that is the case involving descriptors μ, c and $d_4$where the MLR model delivers again the best prediction for compound B.

Compound B is wrongly predicted as inactive by both MLR and AsNN models built with seven descriptors. However, the poor prediction of its activity (computed with an error of −1.771 and −1.507 in log (1/MIC), respectively) is not totally unexpected. In fact, from Table 11 one can see that only 3 descriptors contribute to the enhancement of activity of this compound, and from these just one (*MR*) was identified in Fig. 3 as one of the three most important descriptors responsible for the high predictive ability of these models. Nonetheless, this compound was chosen to be synthesized since it was predicted as active by all other four models and this was later confirmed by the experimental MIC value.

These observations show very clearly how important it is in this type of studies to go a step further in the evaluation of the quality of a model by determining the experimental biological activities of the rationally designed compounds, after all the ultimate, the real test, to any QSAR model.

4. Conclusions

The antitubercular activity of a large set of hydrazide derivatives was modeled by two different QSAR methodologies: MLRs and NNs.

A comparison of the statistical performance of the four MLT (MLRs, FFNNs, EnsFFNNs and AsNNs) used, revealed that NNs, in

Table 11
Qualitative analysis of the descriptors' magnitude, + or −, that promote antitubercular activity as given by MLR and CPNN models, and matching of this analysis with the actual features of the two selected compounds.

Descriptor	Descriptor's magnitude[a]		Matching[b]	
	MLR	CPNN	Cpd A	Cpd B
μ	+	+	√	−−−
c	++	++	√	√
d_4	−−	+−	−−−	√
L	+	+−	√	−−−
B_5	−	−−	−−−	√
d_K	++	++	√	√
B_5	−	−−	−−−	√
ML	+	+	√	−−−
SA	+	−	−−−	−−−
MR	−	−	√	√
HOMO	++	+	−−−	−−−
LUMO	−−	−−	√	√
1,4-IE	+	−	−−−	−−−

[a] Descriptor's magnitude promoting AA: (+) high values and (−) low values.
[b] Matching between models' predictions and features of selected compounds.

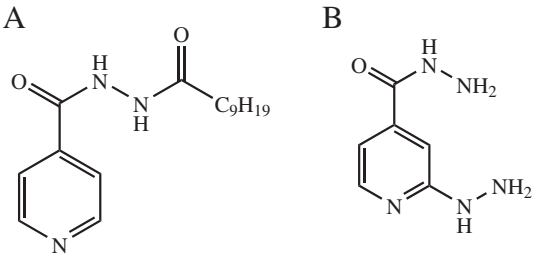

Fig. 11. Compounds A and B rationally designed, synthesized and biologically tested against *M. tuberculosis*.

particular AsNNs, had consistently better predictive abilities than the corresponding MLRs, for independent test sets.

From all studied models, those with seven descriptors were analyzed and discussed in more detail. The relative importance of each descriptor in the model's predictive power was assessed on the basis of the impact of their variation upon scrambling. Although MRLs appeared to be more sensitive to descriptors' scrambling, the ranking of the most and the least important descriptors were exactly the same for both MLRs and NNs. Also the distribution of compounds in different ranges of errors (of prediction) was very similar in both methodologies. Worth of notice is also the fact that both MLRs and AsNNs made the right classification of a compound as active or inactive in approximately 94% of the cases.

With the purpose of designing a few potentially active compounds to be synthesized, MLR and CPNN models corresponding to data sets 4–6 were analyzed. The inspection of weights at the different levels (descriptors) of the three trained CPNNs, confirmed and supported the conclusions derived from MLRs on the molecular characteristics that either increase or decrease antitubercular activity. Based on the information from these two independent approaches, it was possible and reliable to propose the rational synthesis of two new potentially active compounds whose activity was experimentally evaluated against *M. tuberculosis* (H37Rv) and corroborated that predicted by the great majority of the established models.

In conclusion, the integrated use of various MLT and the concomitant use of various validation metrics and the visual inspection of scatter plots of predicted *vs.* experimental data allowed the building of robust and highly predictive QSAR models that can be used efficiently in the design of new active antitubercular agents.

Table 12
Difference between experimental and predicted values in log (1/MIC) and MIC (μg/mL), respectively, for the synthesized compounds.

Descriptors	Compound	Experimental-predicted	
		MLR	AsNNs
L, B_5, d_K	A	0.398 (−0.051)	−0.353 (0.105)
	B	0.117 (−0.028)	−0.586 (0.342)
μ, c, d_4	A	0.318 (−0.044)	−0.162 (0.038)
	B	−0.070 (0.021)	0.229 (−0.049)
B_5, *ML*, *SA*, *MR*, *HOMO*, *LUMO*, 1,4-*IE*	A	−0.485 (0.172)	0.035 (−0.006)
	B	−1.771 (6.954)	−1.507 (3.735)

Acknowledgments

The authors are grateful to F. Rodrigues and M.J. Sousa from INSRJ (Porto) for the preliminary MIC determinations of the two new compounds and to S. Santos, M. Reis and S. Vitorino (FCUL) for their synthesis. The authors also wish to thank Prof. João Aires-de-Sousa for his helpful comments and suggestions. Financial support from Fundação para a Ciência e a Tecnologia, Portugal, under project FCT/PTDC/QUI/67933/2006 and grants BPD/20743/2004 (C. Ventura) and BPD BPD/63192/2009 (D. Latino), as well as project PEst-OE/QUI/UI0612/2013, is greatly appreciated. Anonymous Reviewers are gratefully acknowledged for helpful suggestions that contributed to improve the paper.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ejmech.2013.10.029.

References

[1] Global Tuberculosis Control 2012, WHO, 2012. Report.
[2] N.S. Shah, A. Wright, G.-H. Bai, L. Barrera, F. Boulahbal, N. Martín-Casabona, F. Drobniewski, C. Gilpin, M. Havelková, R. Lepe, R. Lumb, B. Metchock, F. Portaels, M.F. Rodrigues, S. Rüsch-Gerdes, A. Van Deun, V. Vincent, K. Laserson, C. Wells, J.P. Cegielski, Worldwide emergence of extensively drug-resistant tuberculosis, Emerg. Inf. Dis. 13 (2007) 380–387.
[3] Tuberculosis Global Facts 2012, WHO, 2012.
[4] A. Koul, E. Amoult, N. Lounis, J. Guillemont, K. Andries, The challenge of new drug discovery for tuberculosis, Nature 469 (2011) 483–490.
[5] Z.F. Udwadia, R.A. Amale, K.K. Ajbani, C. Rodrigues, Totally drug-resistant tuberculosis in India, Clin. Inf. Dis. (December 21, 2011), http://dx.doi.org/10.1093/cid/cir889.
[6] G.B. Migliori, G. De Iaco, G. Besozzi, R. Centis, D.M. Cirillo, First tuberculosis cases in Italy resistant to all tested drugs, Eurosurveillance 12 (2007) 1–3.
[7] A.A. Velayati, M.R. Masjedi, P. Farnia, P. Tabarsi, J. Ghanavi, A.H. ZiaZarifi, S.E. Hoffner, Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran, Chest 136 (2009) 419–426.
[8] L. Ballell, R.A. Field, K. Duncan, R.J. Young, New small-molecule synthetic antimycobacterial, Antimicrob. Agents Chemother. 49 (2005) 2153–2163.
[9] J.S. Blanchard, Molecular mechanisms of drug resistance in *Mycobacterium tuberculosis*, Annu. Rev. Biochem. 65 (1996) 215–239.
[10] L.A. Kayukova, K.D. Praliev, Search for new drugs – main directions in the search for new antituberculous drugs, Pharm. Chem. J. 34 (2000) 11–18.
[11] C.G. Wermuth, The Practice of Medicinal Chemistry, third ed., Elsevier Academic Press, Amsterdam, 2008.
[12] A.J. Kurup, C-QSAR: a database of 18,000 QSARs and associated biological and physical data, Comput. Aided Mol. Des. 17 (2003) 187–196.
[13] D.A. Winkler, The role of quantitative structure–activity relationships (QSAR) in biomolecular discovery, Brief Inf. 3 (2002) 73–86.
[14] C. Hansch, A quantitative approach to biochemical structure–activity relationships, Acc. Chem. Res. 2 (1969) 232–239.
[15] C. Silipo, C. Hansch, Correlation analysis. Its application to the structure–activity relation of triazines inhibiting dihydrofolate reductase, J. Am. Chem. Soc. 97 (1975) 6849–6861.
[16] T. Aoyama, Y. Suzuki, H. Ichikawa, Neural networks applied to structure–activity relationships, J. Med. Chem. 33 (1990) 905–908.
[17] J. Zupan, Introduction to artificial neural networks (ANN) methods: what they are and how to use them, Acta Chim. Slov. 41 (1994) 327–352.

[18] J. Zupan, J. Gasteiger, Neural Networks in Chemistry and Drug Design, second ed., Wiley-VCH, Weinheim, 1999.
[19] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Chapman & Hall/CRC, Boca Raton, Florida, 1984.
[20] L. Breiman, Random forests, Machine Learn. 45 (2001) 5–32.
[21] L. Erikson, H. Antti, E. Holmes, E. Johansson, T. Lundstedt, J. Shockor, S. Wold, Partial least squares (PLS) in cheminformatics, in: J. Gasteiger, J. Engel (Eds.), Handbook of Chemoinformatics, vol. 3, Wiley-VCH, New York, 2003.
[22] S. Wold, M. Sjöstrom, L. Eriksson, PLS-Regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130.
[23] S. Wold, K. Esbensen, P. Geladi, Principal components analysis, Chemom. Intell. Lab. Syst. 2 (1987) 37–52.
[24] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, 2000.
[25] C. Hansch, A. Leo, Exploring QSAR. Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington DC, 1995.
[26] K.-J. Schaper, M. Pickert, A.W. Frahm, Substituted xanthones as antimycobacterial agents, Arch. Pharm. Pharm. Med. Chem. 332 (1999) 91–102.
[27] E.E. Oruç, S. Rollas, F. Kandemirli, N. Shvets, A.S. Dimoglo, 1,3,4-Thiadiazole derivatives. Synthesis, structure elucidation and structure-antituberculosis activity relationship investigation, J. Med. Chem. 47 (2004) 6760–6767.
[28] F. Macaev, G. Rusu, S. Pogrebnoi, A. Gudima, E. Stingaci, L. Vlad, N. Shvets, F. Kandemirli, A. Dimoglo, R. Reynolds, Synthesis of novel 5-aryl-2-thio-1,3,4-oxadiazoles and the study of their structure-anti-mycobacterial activities, Bioorg. Med. Chem. 13 (2005) 4842–4850.
[29] K.E. Hevener, D.M. Ball, J.K. Buolamwini, R.E. Lee, Quantitative-structure–activity relationship studies of nitrofuranyl antitubercular agents, Bioorg. Med. Chem. 16 (2008) 8042–8053.
[30] C. Ventura, F. Martins, Application of quantitative structure–activity relationships to the modeling of antitubercular compounds. 1. The hydrazide family, J. Med. Chem. 51 (2008) 612–624.
[31] V. Kovalishyn, J. Aires-de-Sousa, C. Ventura, R.E. Leitão, F. Martins, QSAR modeling of antitubercular activity of diverse organic compounds, Chemom. Intell. Lab. Syst. 107 (1) (2011) 69–74.
[32] A.R. Katritzky, R. Petrukhin, D. Tatham, S. Basak, E. Benfenati, M. Karelson, U. Maran, Interpretation of quantitative structure–property and –activity relationships, J. Chem. Inf. Comput. Sci. 41 (2001) 679–685.
[33] J. Jaworska, N. Nikolova-Jeliazkova, T. Aldenberg, QSAR applicability domain estimation by projection of the training set in descriptor space: a review, ATLA 33 (2005) 445–459.
[34] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci. 26 (2007) 694–701.
[35] A. Tropsha, Best practices for QSAR model development, validation and exploitation, Mol. Inf. 29 (2010) 476–488.
[36] F. Sahigara, K. Mansouri, D. Ballabio, A. Mauri, V. Consonni, R. Todeschini, Comparison of different approaches to define the applicability domain of QSAR models, Molecules 17 (2012) 4791–4810.
[37] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models, QSAR Comb. Sci. 22 (2003) 69–77.
[38] L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell, P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification – and regression-based QSARs, Environ. Health Perspect. 111 (2003) 1361–1375.
[39] J.D. Walker, L. Carlsen, J. Jaworska, Improving opportunities for regulatory acceptance of QSARs: the importance of model domain, uncertainty, validity and predictability, QSAR Comb. Sci. 22 (2003) 346–350.
[40] N. Nikolova, J. Jaworska, Approaches to measure chemical similarity – a review, QSAR Comb. Sci. 22 (2003) 1006–1026.
[41] V. Consonni, D. Ballabio, R. Todeschini, Comments on the definition of the Q^2 parameter for QSAR validation, J. Chem. Inf. Model. 49 (2009) 1669–1678.
[42] G. Klopman, D. Fercu, J. Jacob, Computer-aided study of the relationship between structure and antituberculosis activity of a series of isoniazid derivatives, Chem. Phys. 204 (1996) 181–193.
[43] J.K. Seydel, K.-J. Schaper, E. Wempe, H.P. Cordes, Mode of action and quantitative structure–activity correlations of tuberculostatic drugs of the isonicotinic acid hydrazide type, J. Med. Chem. 19 (1976) 483–492.
[44] S. Mohamad, P. Ibrahim, A. Sadikun, Susceptibility of *Mycobacterium tuberculosis* to isoniazid and its derivative, 1-isonicotinyl-2-nonanoyl hydrazine: investigation at cellular level, Tuberculosis 84 (2004) 56–62.
[45] A. De Logu, V. Onnis, B. Saddi, C. Congiu, M.L. Schivo, M.T. Cocco, Activity of a new class of isonicotinoylhydrazones used alone and in combination with isoniazid, rifampicin, ethambutol, para-aminosalicylic acid and clofazimine against *Mycobacterium tuberculosis*, J. Antimicrob. Chemother. 49 (2002) 275–282.
[46] D. Sriram, P. Yogeeswari, K. Madhu, Synthesis and *in vivo* antimycobacterial activity of isonicotinoyl hydrazones, Bioorg. Med. Chem. Lett. 15 (2005) 4502–4505.
[47] https://www.collaborativedrug.com/pages/public_access, (accessed 08.12).
[48] M.A. Morgan, C.D. Horstmeier, D.R. DeYoung, G.D. Roberts, Comparison of radiometric method (BACTEC) and conventional culture media for recovery of mycobacteria from smear negative specimens, J. Clin. Microbiol. 18 (1983) 384–388.
[49] ChemDraw Ultra, Version 11.0.1, ©1986–2007, CambridgeSoft.
[50] J.A. Platts, D. Butina, M.H. Abraham, A. Hersey, Estimation of molecular linear free energy relation descriptors by a group contribution approach, J. Chem. Inf. Comput. Sci. 39 (1999) 835–845.
[51] M.H. Abraham, A. Ibrahim, A.M. Zissimos, Y.H. Zhao, J. Comer, D.P. Reynolds, Application of hydrogen bonding calculations in property based drug design, DDT 7 (2002) 1056–1063.
[52] A. Leo, C. Hansch, D. Elkins, Partition coefficients and their uses, Chem. Rev. 71 (1971) 525–616.
[53] Molecular Modeling Pro Plus, Version 6.2.5., www.chemistry-software.com.
[54] http://www.biobyte.com/bb/index.html, (accessed 01.12).
[55] R. Peck, C. Olsen, J.L. Devore, Introduction to Statistics & Data Analysis, fourth ed., Brooks/Cole, Boston, 2012.
[56] V. Svetnik, A. Liaw, C. Tong, J.C. Culberson, R.P. Sheridan, B.P. Feuston, Random forest: a classification and regression tool for compound classification and QSAR modeling, J. Chem. Inf. Comput. Sci. 43 (2003) 1947–1958.
[57] D.A.R.S. Latino, J. Aires-de-Sousa, Assignment of EC numbers to enzymatic reactions with MOLMAP reaction descriptors and random forests, J. Chem. Inf. Model. 49 (2009) 1839–1846.
[58] R Development Core Team, R: A Language and Environment for Statistical Computing, Vienna, 2004.
[59] Fortran original by L. Breiman and A. Cutler, Report by A. Liaw and M. Wiener, http://www.stat.berkeley.edu/users/breiman/, 2004.
[60] I.V. Tetko, Associative neural network, Neural Process. Lett. 1 (6) (2002) 187–199.
[61] W.H. Press, S.A. Teukolsky, W.T. Vetterlung, B.P. Flannery, Numerical Recipes in C, second ed., Cambridge University Press, New York, 1994.
[62] A.J. Shepherd, Second Order Methods for Neural Networks, Springer-Verlag, London, 1997.
[63] D.K. Agrafiotis, W. Cedeno, V.S. Lobanov, On the use of neural network ensembles in QSAR and QSPR, J. Chem. Inf. Comput. Sci. 42 (2002) 903–911.
[64] T.G. Dietterich, Ensemble learning, in: M.A. Arbib (Ed.), The Handbook of Brain Theory and Neural Networks, MIT Press, Cambridge, MA, 2002, pp. 405–408.
[65] I.V. Tetko, Neural network studies. 4. Introduction to associative neural networks, J. Chem. Inf. Comput. Sci. 42 (2002) 717–728.
[66] B. Dasarthy, Nearest Neighbor (NN) Norms, IEEE Computer Society Press, Washington, DC, 1991.
[67] D.A.R.S. Latino, F.F.M. Freitas, J. Aires-de-Sousa, F.M.S.S. Fernandes, Neural networks to approach potential energy surfaces: application to a molecular dynamics simulation, Int. J. Quan. Chem. 107 (2007) 2120–2132.
[68] T. Kohonen, Self-Organization and Associative Memory, Springer, Berlin, 1988.
[69] J. Aires-de-Sousa, JATOON: Java tools for neural networks, Chemom. Intell. Lab. Syst. 61 (2002) 167–173.
[70] JATOON applets, http://www.dq.fct.unl.pt/staff/jas/jatoon/.
[71] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graph. Model. 20 (2002) 269–276.
[72] S. Chatterjee, A.S. Hadi, Influential observations, high leverage points, and outliers in linear regression, Stat. Sci. 1 (1986) 379–416.
[73] C. Kim, B.E. Storer, Reference values for Cook's distance, Comm. Stat. Theory Methods 25 (1996) 691–708.
[74] C. Kim, Y. Lee, B.U. Park, Cook's distance in local polynomial regression, Stat. Prob. Lett. 54 (2001) 33–40.
[75] J.A. Díaz-García, G.A. González-Farías, Note on the Cook' distance, J. Stat. Plan. Inf. 120 (2004) 119–136.
[76] N. Chirico, P. Gramatica, Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection, J. Chem. Inf. Model. 52 (2012) 2044–2058.
[77] Organization for Economic Co-Operation and Development (OECD). Environment Health and Safety Publications Series on Testing and Assessment, 69, Guidance Document on the Validation of (Quantitative) Structure–Activity Relationship [(Q)SAR] Models 2007, http://www.oecd.org/fr/env/ess/risques/guidancedocumentsandreportsrelatedtoqsars.htm, (accessed 04.13).
[78] N. Chirico, P. Gramatica, Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient, J. Chem. Inf. Model. 51 (2011) 2320–2335.
[79] P.P. Roy, K. Roy, On some aspects of variable selection for partial least squares regression models, QSAR Comb. Sci. 27 (2008) 303–313.
[80] P.P. Roy, S. Paul, I. Mitra, K. Roy, On two novel parameters for validation of predictive QSAR models, Molecules 14 (2009) 1660–1701.
[81] P.K. Ojha, I. Mitra, R.N. Das, K. Roy, Further exploring r^2_m metrics for validation of QSPR models, Chemom. Intell. Lab. Syst. 107 (2011) 194–205.
[82] K. Roy, I. Mitra, S. Kar, P.K. Ojha, R.N. Das, H. Kabir, Comparative studies on some metrics for external validation of QSPR Models, J. Chem. Inf. Model. 52 (2012) 396–408.
[83] I. Mitra, A. Saha, K. Roy, Exploring quantitative structure–activity relationship studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants, Mol. Simul. 36 (2010) 1067–1079.
[84] J. Polanski, A. Bak, R. Gieleciak, T. Magdziarz, Modeling robust QSAR, J. Chem. Inf. Model. 46 (2006) 2310–2318.
[85] P. Gramatica, P. Pilutti, E. Papa, Validated QSAR prediction of OH tropospheric degradation of VOCs: splitting into training-test sets and consensus modeling, J. Chem. Inf. Comput. Sci. 44 (2004) 1794–1802.
[86] J. Li, B. Lei, H. Liu, S. Li, X. Yao, M. Liu, P. Gramatica, QSAR study of malonyl-CoA decarboxylase inhibitors using GA-MLR and a new strategy of consensus modeling, J. Comput. Chem. 29 (2008) 2636–2647.
[87] C. Ràfols, E. Bosch, R. Ruiz, K.J. Box, M. Reis, C. Ventura, S. Santos, M.E. Araújo, F. Martins, Acidity and hydrophobicity of several new potential antitubercular drugs: isoniazid and benzimidazole derivatives, J. Chem. Eng. Data 57 (2012) 330–338.