Journal of Molecular Graphics and Modelling 32 (2012) 32–38

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Predictive QSAR modeling of phosphodiesterase 4 inhibitors

Vasyl Kovalishyn*, Vsevolod Tanchuk, Larisa Charochkina, Ivan Semenuta, Volodymyr Prokopenko

Institute of Bioorganic Chemistry and Petrochemistry, National Academy of Sciences of Ukraine, Kyiv-94, 02660 Murmanska, 1, Ukraine

ARTICLE INFO

Article history:
Received 23 June 2011
Received in revised form 4 October 2011
Accepted 5 October 2011
Available online 14 October 2011

Keywords:
PDE-4 inhibitors
Drug design
QSAR
Artificial neural networks
Random forest

ABSTRACT

A series of diverse organic compounds, phosphodiesterase type 4 (PDE-4) inhibitors, have been modeled using a QSAR-based approach. 48 QSAR models were compared by following the same procedure with different combinations of descriptors and machine learning methods. QSAR methodologies used random forests and associative neural networks. The predictive ability of the models was tested through leave-one-out cross-validation, giving a Q^2 = 0.66–0.78 for regression models and total accuracies Ac = 0.85–0.91 for classification models. Predictions for the external evaluation sets obtained accuracies in the range of 0.82–0.88 (for active/inactive classifications) and Q^2 = 0.62–0.76 for regressions. The method showed itself to be a potential tool for estimation of IC_{50} of new drug-like candidates at early stages of drug development.

1. Introduction

The cyclic nucleotide phosphodiesterases (PDEs) are intracellular enzymes that catalyze the hydrolysis of 3′,5′-cyclic nucleotides, such as cyclic adenosine monophosphate (cAMP) and cyclic guanosine monophosphate (cGMP), to their corresponding 5′-nucleotide monophosphates [1]. There are at least 11 families of PDEs. They play a central role in a variety of intracellular signaling events by controlling cellular concentrations of cyclic nucleotides. Their classification is based upon a variety of criteria including substrate specificity, inhibition potency, enzyme kinetics, amino acid sequence, cellular and tissue distribution. Much effort is now devoted to development of isoenzyme selective phosphodiesterase inhibitors [2–5]. PDE-4 modulates levels of cAMP-regulating several cellular physiological processes such as leukocyte responses including the proinflammatory actions of monocytes, T-cells and neutrophils, airway and vascular smooth muscle constriction, gene transcription through cAMP response elements, protein phosphorylation via cAMP-dependent protein kinase A (PKA), and cyclic nucleotide gated ion channels. There is a great potential for pharmacological intervention in a variety of disorders by influencing PDE-4, which is involved in pathological processes. Selective inhibitors of PDE-4 have potential use in treatment of a range of major diseases – psoriasis, asthma, osteoporosis, chronic obstructive pulmonary disease, Alzheimer's disease, Parkinson's disease, schizophrenia, depression and others [6–10]. The selective inhibition of PDE-4 generates significant functional effects and PDE-4 inhibitors are under development to provide therapeutics for the treatment of the mentioned above diseases.

The search of novel PDE-4 inhibitors can be boosted by creation of computational models of biological activity based on chemical structure. This approach, called QSAR (quantitative structure-activity relationship), became an important tool for automated pre-virtual screening, combinatorial library design and data mining [11]. Many different molecular descriptors and computational techniques including multiple linear regression analysis (MLRA), partial least squares (PLSs) regression, support vector machines (SVMs), random forests (RFs), artificial neural networks (ANNs) have been applied to analyze the relationship between chemical structure and biological activity [12–18]. Three-dimensional quantitative structure-activity relationship (3D-QSAR) techniques such as comparative molecular field analysis (CoMFA) and comparative molecular similarity analysis (CoMSIA), are often used in modern drug design. The methods are especially useful because they help to understand drug–receptor interaction. It has been shown that these computational techniques can help to design novel, more potent PDE-4 inhibitors by revealing the mechanism of drug–receptor interaction [17–19]. QSAR models based on a new multi-conformational structure-based pharmacophore key method have been also proposed [20,21]. This method calculates molecular descriptors based on the matching of a molecule's pharmacophore features with those of the target binding pocket. However, many of these models are specially designed for good interpretation. Interpretation involves dissecting the model to learn more about the interaction of a set of molecules with a target. The interpretive models are based on data of very high quality, descriptors easily transforming into molecular properties familiar to synthetic chemists, and mapping methods making the

* Corresponding author. Tel.: +380 445732592; fax: +380 445732552.
E-mail address: vkovalishyn@yahoo.com (V. Kovalishyn).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.10.001

form of the model and the contribution of each descriptor to the model clear. Predictive models represent another approach. They often involve much larger training sets, computationally efficient descriptors and less obvious mapping methods. Predictive methods are aimed at providing the best estimate of a property for a molecule not yet synthesized and to assist in prioritizing lead discovery. Thus, the major goal of the present study was to create a vast database of PDE-4 inhibitors from different classes of compounds and use a variety of machine learning techniques to develop robust QSAR models for virtual screening of vendor's libraries aimed at identifying potential PDE-4 inhibitors.

In the present paper we exploited two different chemometrics methods to make connections between structural properties and PDE-4 inhibition. These methods included artificial neural networks (ANNs) and random forest (RF) classifier. ANNs are a group of methods that are widely used in drug design to study QSAR. This approach is able to elucidate structure-activity relationships and take into account any non-linear character of these relationships [22]. Thus, this method can be of significant interest in QSAR studies. RF also offers some unique features that make it suitable for QSAR tasks [23]. These include built-in estimation of prediction accuracy, measures of descriptor importance, ignore irrelevant descriptors, measure of similarity between molecules and others (see Section 2 in details).

2. Materials and methods

2.1. Dataset

In the current study, a diverse series of PDE-4 inhibitors with known IC_{50} values were collected from different published papers [1–10,12,17,18,24–30]. The detailed structures and the corresponding bioactivities of the compounds and full list of publications are documented in Supplementary Materials. The range of IC_{50} values for these 1015 compounds ranging from 0.05 nM to 660 μM. All compounds were divided into two classes: active (426 with $IC_{50} \leq 100$ nM) and inactive molecules (589 with $IC_{50} > 100$ nM).

All molecules were processed by the Chemaxon standardizer [31]. The 2D coordinates of atoms were recalculated, molecules were neutralized, mesomerized and aromatized. Data set was then filtered from duplicates. Then, the dominant protonation state of molecules at pH 7.4 was calculated using the major microspecies plugin [31]. The 3D structures were generated via Chemaxon standardize [31] from the SMILES notation available for each compound, and stored in SDF format. Using DRAGON Software [32], more than a thousand descriptors were calculated as the original variables, including structure, spatial, electronic, topological descriptors and many others. We also omitted the descriptors whose values were strictly correlated with the value of another descriptor (i.e., with the correlation coefficient between them greater than or equal to 0.95) to avoid useless redundancy.

2.2. Machine learning techniques (MLTs)

2.2.1. Random forest

Random forest is an ensemble classifier consisting of a number of decision trees. The method is specially devised to operate quickly over large data sets [33]. Training procedure of the random forest's algorithm (suitable for classification and regression) includes the following steps:

1. Randomly select n times with replacement a molecule out of a total number of n molecules into a bootstrap sample. As a result approximately $n/3$ samples form a so-called out-of-bag test set while the rest form a training set in which a molecule can occur several times.
2. For each bootstrap sample, grow a tree with the following modification: for each node of the tree, randomly choose m_{try} variables on which to base the decision at that node. Calculate the best split based on these m_{try} variables in the training set. m_{try} is essentially the only tuning parameter. The tree is grown to the maximum size (which does not allow any further split) and not pruned. Test set is used to estimate the error of the tree. The model performance is internally assessed by the prediction error for the objects left out in the bootstrap procedure (out-of-bag estimation, OOB).
3. The above steps are repeated until a sufficiently large number B of trees are grown.
4. To predict new data the predictions of n_{tree} trees are aggregated (i.e., majority votes is used for classification, average value is used for regression).

An estimate of the error rate can be obtained, based on the training data, by the following:

1. At each bootstrap iteration, data left out of bootstrap sample (what Breiman calls "out-of-bag", or OOB, data) are predicted using the tree grown at this iteration.
2. OOB predictions are aggregated. Calculated error rate is called the OOB estimate of error rate.

Our experience shows that if the number of trees is big enough the OOB estimate of error rate is quite accurate.

The algorithm has a number of attractive features such as an internal procedure for descriptor selection [33]. It is not affected by correlated descriptors because random samples are used to build each tree in the forest. Furthermore, there is a measure of importance for each descriptor in the model, which can be later used in other modeling approaches to operate on smaller sets of descriptors (see Section 2.3) [33]. The importance of descriptors is calculated according to the following procedure. Use tree's OOB data to make predictions for each grown tree. All descriptors in the OOB data are randomly permuted one at a time and modified data sets are used for prediction by the tree. A margin is calculated for each molecule in the end of the model training process. The margins are based on the OOB prediction as well as the OOB predictions with each descriptor permuted. Let M is the average margin based on the OOB prediction and M_j is the average margin based on the OOB prediction with the jth descriptor permuted. If M is the average margin based on the OOB prediction and M_j is the average margin based on the OOB prediction with the jth descriptor permuted, the measure of importance for the jth descriptor is simply $M - M_j$. In the case of regression, the margins are simply replaced by squared prediction errors.

RFs were grown with the R suite of statistical software [34] using the random forest library. There are number of user-definable parameters. The influence of the number of trees and number of descriptors selected at each split point was investigated. There was no significant change on the model's predictive ability when the number of trees was increased beyond $n_{tree} = 1000$ and the number of descriptors beyond $m_{try} = 30$ in all cases. This is not surprising, because random forests avoid overfitting, as has been shown by Breiman et al. [33].

2.2.2. Associative neural network

An associative neural network (ASNN) combines an ensemble of feed-forward neural networks (FFNNs) and the method of k-nearest neighbours (kNNs) [35]. The traditional FFNN represents a memory less approach. This means that after training the initial data are no longer needed, all information necessary for predictions is stored in the neural network weights. There is no explicit storage of any presented example. On the contrary, kNN represents a memory-based

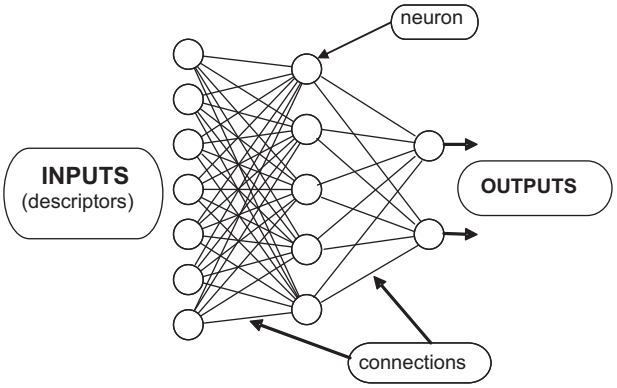

Fig. 1. Typical feed forward neural network with two outputs.

approach. This approach keeps in memory all training samples and predictions are based on some local approximation of the stored data. ASNN performs kNN in the space of ensemble residuals. Each molecule is represented as a vector of neural network predictions by the neural network ensemble. Correlation between such vectors is used by the nearest neighbour method as a measure of distance between the analyzed cases. This approach improves prediction by correcting the bias of the neural network ensemble [35].

FFNN is a supervised regression method that is trained using a data set with a known property to be modeled. Fig. 1 presents a typical topology of a neural network. The trained model can be used later to predict the required property for molecules outside the training set.

Data for each sample of the training set are presented in turn to the input neurons and propagated via the connections to neurons of the hidden layer. Each neuron of the hidden layer processes the weighted sum of its ingoing connections by a non-linear transfer function and passes the result on to the output neurons. The network is trained by comparing the response from the output neuron with the known response, propagating the error back through the network and adjusting the weights to minimize the error.

The ASNN was trained by SuperSAB algorithm [36,37]. The neural networks had the number of inputs equal to the number of descriptors and five neurons in one hidden layer. Random initial weights were used. There was also a bias neuron both of the input and hidden layers. ASNN used two output neurons for classification tasks – the target values were assigned to (0.1 and 0.9) for active and (0.9 and 0.1) for inactive compounds (Fig. 1). There was only one output neuron for regression tasks, the output values were linearly scaled between 0.1 and 0.9 [38]. The architecture of the networks was the same. Cross-validation techniques were used to rigorously control the possibility of overfitting the data [38]. Application of artificial neural network ensemble (ANNE) allowed us to avoid chance correlation and obtain satisfactory predictions of new data for a wide range of numbers of neurons in the hidden layer. Let us outline some general principles of the method. Each ANNE included $M = 200$ networks. Initial training set was divided into two equal learning/validation subsets. The first set was used for training the neural network while the second one was used to control the training procedure. Network performance was measured by root mean square error (RMSE) in two stopping points. The first point (early stopping) was determined as the point of the best performance of the network for the validation set while the second point corresponded to the minimum of error for the learning set, which usually coincided with the end of the network training. Leave-one-out cross-validation (LOO) was used to access the quality of each final model. According to the method, each molecule was removed from the training set, the remaining molecules were used to build a model, the activity of the excluded molecule predicted to evaluate the quality of the model. Further details of are provided in the earlier article [39].

2.3. Evaluation of descriptor importance

Sensitivity analysis methods estimate the level of change in the model's output resulting from the changes of model's inputs. They are primarily used to find the input descriptors which are most important to achieve accurate output values. In this work, "Mean Decrease Accuracy" value from RF models is used as a measure of descriptor's importance [34].

"Pruning methods" implemented in ASSN Software were used as an additional selection tool. These methods proved to be very efficient in QSAR studies [40,41]. Pruning algorithms introduce some measure of importance of the ASNN matrix weights by the so-called "sensitivities". These algorithms work similarly to stepwise multiple regression analysis excluding on each step one input parameter, which is considered non-significant. At each step, the model sensitivities to all weights and input nodes are estimated and the descriptor corresponding to the input neuron with the smallest sensitivities is deleted [42,43]. 300–400 most important descriptors obtained from RF models were subjected to this kind of analysis. The influence of the number of descriptors selected by RF on the ASNN model quality was also estimated (on the basis of the leave-one-out results for the training sets). It appeared that preselecting less than 300–400 descriptors by RF decreased the ASNN models predictive ability.

2.4. Statistical coefficients

Regression models were evaluated by the leave-one-out cross-validation coefficient (Q^2) [44]. Q^2 is calculated by the formula:

$$Q^2 = \frac{\text{SDY-PRESS}}{\text{SDY}} \tag{1}$$

where SDY is the sum of squared deviation of the target variable values from their mean, and PRESS is the prediction error sum of squares obtained from the leave-one-out-cross-validation procedure. Application of this coefficient makes the analysis of residuals by means of standard deviation, which is interrelated and can be obtained from Q^2, redundant. Besides that, prediction performance of the methods used in this study was compared using the root mean squared error, RMSE.

Other parameters, such as sensitivity (Sn), specificity (Sp), and overall accuracy (Ac) were calculated to assess the classification ability and to monitor the classification performance of the two classes separately. Sensitivity is also known as positive rate or positive class accuracy, while specificity is also known as true negative rate or negative class accuracy. A theoretical, optimal prediction can achieve 100% sensitivity (i.e., predict all active molecules from the active group as active) and 100% specificity (i.e., not predict any molecule from the inactive group as active).

$$\text{Sn} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{Sp} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3}$$

$$\text{Ac} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \tag{4}$$

where TP, FP, TN and FN denote true positives, false positives, true negatives and false negatives, respectively. In general, the overall accuracy, Ac, is a good measure of predictive power of models if the number of active and inactive compounds is approximately equal.

Table 1
Q^2 coefficient calculated using the two MLT for different set of descriptors.

No.	Set	Molec.	Random forest		ASNN		Consensus prediction
Descriptors			777	122	777	122	
1	Training 1	812	0.66	0.67	0.77	0.77	0.75
	Test 1	203	0.70	0.70	0.72	0.73	0.75
2	Training 2	812	0.66	0.68	0.77	0.78	0.76
	Test 2	203	0.68	0.69	0.71	0.70	0.73
3	Training 3	812	0.67	0.69	0.77	0.77	0.76
	Test 3	203	0.62	0.65	0.68	0.71	0.70
4	Training 4	812	0.66	0.68	0.77	0.76	0.76
	Test 4	203	0.69	0.69	0.73	0.71	0.74
5	Training 5	812	0.66	0.66	0.75	0.76	0.74
	Test 5	203	0.69	0.70	0.75	0.76	0.77
5-Fold validation							
6	Total set	1015	0.68	0.69	0.72	0.72	0.74

2.5. External validation experiments

The validation of a QSAR model is important if a particular QSAR model is to be used for predictive purposes. In external validation, the predictive performance of a QSAR model is evaluated using validation test compounds that were not used to generate a QSAR model. Ideally external validation compounds should be a completely independent data set. In this study, external validation experiments were randomly performed using an alternative "leave 20% out" cross-validation external validation. This procedure was sequentially repeated five times producing five different external validation data sets and corresponding training set molecules. Then the average statistical coefficients for all 5-*test sets* were computed. Therefore we developed five of predictive models by each MLT. The prediction statistics for the RF and ASNN models are given in Tables 1 and 2, respectively (see Section 3).

3. Results and discussion

The main aim of this study was to compare the accuracy of the PDE-4 models built by different machine learning methods. In these experiments the influence of the chosen descriptor set on the model accuracy was also studied. The descriptors were filtered by "pruning methods" implemented in ASSN Software. The previously mentioned methods of descriptor selection are very efficient in QSAR studies. Experiments were performed both with full set of descriptors and with the use of descriptor selection procedure. Results of both approaches were compared (see Tables 1 and 2). For the development of QSAR models, two methods, i.e., ASNN and Random Forest were used, each with descriptors calculated by Dragon programs [30]. Also for the modeling purpose, about 20% of compounds were selected from initial data set to form the external test set and the remaining molecules were used as a training set. Using 5-fold splitting procedure we received five data sets. Consequently, four different QSAR models were developed for all five modeling sets. The summary Table 1 contains the detailed information for all types of models.

3.1. Regression models

3.1.1. ASNN results

The overall best performance for the training sets was shown by the ASNN method. In the first stage, ASNN models were developed using total set of descriptors. The initial number of descriptors was submitted to the following additional reduction procedure: descriptors with constant values were removed, after which a pair wise correlation analysis was performed where a given descriptor was eliminated if the correlation coefficient with another descriptor was equal or higher than 0.95. As a result, the 777 most important descriptors were selected. The Q^2 coefficients for the training set were 0.75–0.78. The compounds in the external test sets were predicted with the accuracy, $Q^2 = 0.68–76$ (see Table 1).

In the second stage the importance of the descriptors for the observed activity was evaluated as the result of using ASNN gradually pruning methods. Since the Q^2 calculated were non-significantly different between all training sets for all analyzed methods (Table 1) we used first model for variable selection. Thus, we selected variables using the first training set, developed the model and then applied it to predict molecules in the external test set. The 122 most important descriptors were selected. After that the same set of variables was used for building the rest model. However, the application of descriptor selection methods to the

Table 2
Comparison of classification models built with different MLT.

MLT			Random forest						ASNN						Consensus prediction
Amount of descriptor			777			257			777			257			
No.	Set	Mol.	Sn	Sp	Ac	Sn	Sp	Ac	Sn	Sp	Ac	Sn	Sp	Ac	Ac
1	Training 1	812	0.85	0.88	0.87	0.87	0.88	0.87	0.88	0.93	0.91	0.88	0.92	0.91	0.87
	Test 1	203	0.83	0.92	0.88	0.81	0.92	0.87	0.80	0.94	0.88	0.80	0.93	0.87	0.86
2	Training 2	812	0.85	0.87	0.86	0.86	0.87	0.86	0.88	0.91	0.89	0.88	0.90	0.89	0.86
	Test 2	203	0.84	0.89	0.87	0.85	0.90	0.88	0.83	0.88	0.86	0.82	0.88	0.86	0.85
3	Training 3	812	0.85	0.85	0.85	0.84	0.87	0.86	0.88	0.92	0.89	0.87	0.91	0.89	0.86
	Test 3	203	0.85	0.89	0.87	0.87	0.88	0.88	0.81	0.88	0.85	0.84	0.87	0.86	0.85
4	Training 4	812	0.84	0.88	0.86	0.84	0.88	0.86	0.88	0.92	0.90	0.87	0.92	0.90	0.86
	Test 4	203	0.85	0.87	0.86	0.86	0.87	0.87	0.82	0.88	0.85	0.84	0.88	0.86	0.85
5	Training 5	812	0.85	0.88	0.87	0.85	0.88	0.87	0.89	0.93	0.90	0.82	0.92	0.90	0.87
	Test 5	203	0.87	0.82	0.84	0.90	0.82	0.85	0.84	0.86	0.85	0.82	0.84	0.83	0.82
5-Fold validation															
6	Total set	1015	0.85	0.88	0.87	0.86	0.88	0.87	0.82	0.89	0.86	0.82	0.88	0.86	0.84

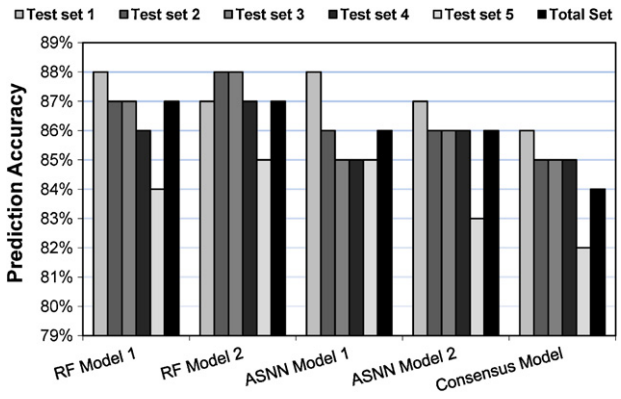

Fig. 2. Prediction accuracy (%) for the external evaluation sets.

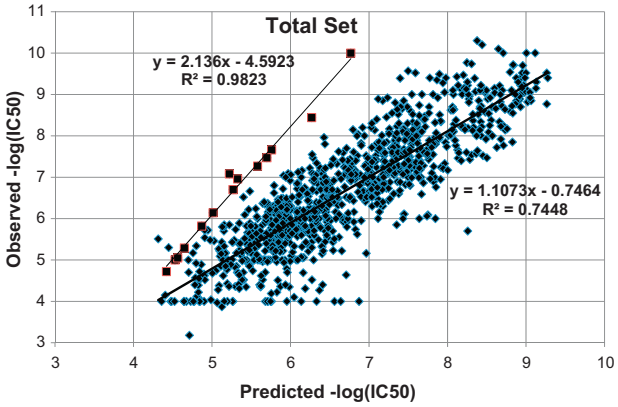

Fig. 3. Consensus observed versus predicted −log (IC_{50}) values for total set using 5-fold validation, IC_{50} = inhibitory concentration. Red line is a line of best fit for a total set. A group of underestimated outliers which can be fitted well by a separate line is also shown.

descriptor sets did not lead to better performance of the received models.

The Q^2 calculated using the 5-fold validation were 0.72 for models based on total and pruned set of descriptors respectively (Table 1).

3.1.2. RF results

The results of this method for all data sets were statistically similar to the ASNN results. However, the prediction ability was slightly lower than that for ASNNs models (see Table 1).

3.2. Classification models

Results are summarized in Table 2. Five QSAR models were developed similarly to the regression studies. Statistical coefficients for the training sets were about Sn = 0.80–0.90, Sp = 0.82–0.93, Ac = 0.83–0.88 (see Table 2).

In the second stage, the number of descriptors was reduced to 257 by ASNN pruning methods, keeping basically the same accuracy for training and test sets. The compounds in the external test sets were predicted with good accuracy, Ac = 0.82–0.88. (Fig. 2) The total accuracy Ac calculated using the 5-fold validation was about 0.86–0.87 for all analyzed methods (Table 2 and Fig. 2).

3.3. Consensus prediction

As mentioned above, application of different MLT and variable selection procedures provides a variety of possible models with nearly equivalent predictive performance (Tables 1 and 2). The models use different descriptors, reflecting different aspects of molecular structure. If we evaluate QSAR models by Q^2, the top-ranking model is ASNN model for data set 2 with $Q^2 = 0.78$ (see Table 1). However, if we consider the best prediction for the external evaluation sets, the ASNN models for data set 5 are the best with $Q^2 = 0.76$ (Table 1). It means that we should not rely on the results received by only one learner, which could obfuscate the choice of the best modeling method to achieve the most accurate external prediction of activity. It has been shown that consensus prediction based on the results obtained by all predictive models always provides the most stable decision [45,46]. Thus, it seems reasonable that a consensus QSAR model, derived by averaging the predictions made by individual models for regression QSAR models, or by majority voting for classification QSARs, should be better than any individual model. The final consensus prediction was made by integrating the PDE-4 inhibitor activity predicted by all four types of models, i.e., by two ASNN and two RF models – Tables 1 and 2. To ensure the reliability of the consensus prediction, we used the next rule: a compound was predicted correctly if the majority of models (in our case at least three) gave correct prediction. If compound was predicted by two models as active and by other two models as inactive we supposed that activity was predicted incorrectly.

The performance of the consensus prediction is always very close to the best model and often even better than the best model. Thus, the obtained results indicate that application of consensus model provides higher prediction ability compared to the individual model. As it is seen, the approach presented in this study has shown quite satisfactory results (Tables 1 and 2).

Fig. 3 shows the regression line plotting the values predicted by the consensus model and the experimental data using 5-fold validation method. Most predictions, namely 869 out of 1015 do not differ from the experimental values by more than 1 log unit. Among the underestimated outliers (prediction error < −1) there seem to be two separate groups of compounds. Each of the group forms a separate model different from the model obtained for the whole set of compounds. The group of 8 highly active and 7 other compounds (see Supplementary Materials) probably requires a special consideration. Only 12 chemicals have residuals between the experimental and predicted log IC_{50} higher than 2 log units (see Supplementary Materials). Some inactive compounds ($IC_{50} = 4.0$) have been predicted as active. Fortunately there are only 8 such compounds, which seem to be beyond chemical domain of the consensus model.

As we can see the predictive ability of classification models are better than ability of regression models (Tables 1 and 2). Therefore it will be better to use classification models on the first stage of virtual screening. Then regression models can be used as a reliability measure for classification purposes. This possibility has obvious interest for ranking compounds in virtual screening experiments.

In summary, the proposed QSAR methodology exhibited both “advantages” and some “limitations”: The advantages include:

(1) Calculation of Dragon descriptors does not require extensive conformational analysis or spatial alignment of molecules.
(2) The method is fast and efficient is typically characterized by the same or better statistics compared to known 3D-QSAR methods.
(3) The method uses an efficient criterion for the optimal stopping of network training and a procedure to avoid overtraining of ASNNs and RFs.
(4) The method’s accuracy is based on statistical averaging of predictions of ensembles of ASNN and of trees of RF. The consensus prediction for models using all statistically significant training set models was found to provide a better balance between prediction accuracy and chemical space coverage than individual models (Tables 1 and 2 and Figs. 2 and 3).

(5) The predictive models can be applied as virtual screening tools to give the best property evaluation for a molecule not yet synthesized and to help in prioritizing experimental screening for lead discovery.

On the other hand there are some limitations:

(1) The models work well for the compound classes represented in the training and validation sets, but may fail catastrophically for other classes.
(2) Additional errors may arise because biological data used as a training set are obtained from different sources and may contain considerable experimental errors (noisy data).
(3) The models use large training sets of diverse quality and large number of descriptors thus cannot be used for detailed interpretation.

4. Conclusion

We have presented a series of new predictive QSAR models using different machine learning techniques and a broad range of atomic and molecular properties. All derived QSAR models were of good predictive quality with $Q^2 > 0.65$ despite the fact that all subsets contained very different classes of compounds. It means that presented QSARs can be used to predict the PDE-4 inhibitor activity of a wide range of chemicals. Finally, this approach codifies knowledge about chemical structure and classifies compounds into univocal chemical classes. An application of RF measure of descriptor's importance and ASNN pruning algorithms was able to detect subsets of relevant input descriptors determining the molecular activity. Integration of random forest and artificial neural network methods with pruning algorithms and consensus prediction allowed us to build models with higher predictive ability. They can be applied for virtual screening of molecules not yet synthesized. Besides lead discovery, the models can be used for lead optimization.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2011.10.001.

References

[1] M.D. Cullen, Y.F. Cheung, M.D. Houslay, T.L. Hartman, K.M. Watson, R.W. Buckheit Jr, C. Pannecouque, E. De Clercq, M. Cushman, Investigation of the alkenyldiarylmethane non-nucleoside reverse transcriptase inhibitors as potential cAMP phosphodiesterase-4B2 inhibitors, Bioorg. Med. Chem. Lett. 18 (2008) 1530–1533.
[2] T. Akama, S.J. Baker, Y.K. Zhang, V. Hernandez, H. Zhou, V. Sanders, Y. Freund, R. Kimura, K.R. Maples, J.J. Plattner, Discovery and structure-activity study of a novel benzoxaborole anti-inflammatory agent (AN2728) for the potential topical treatment of psoriasis and atopic dermatitis, Bioorg. Med. Chem. Lett. 19 (2009) 2129–2132.
[3] R.P. Alexander, G.J. Warrellow, M.A. Eaton, E.C. Boyd, J.C. Head, J.R. Porter, J.A. Brown, J.T. Reuberson, B. Hutchinson, P. Turner, B. Boyce, D. Barnes, B. Mason, A. Cannell, R.J. Taylor, A. Zomaya, A. Millican, J. Leonard, R. Morphy, M. Wales, M. Perry, R.A. Allen, N. Gozzard, B. Hughes, G. Higgs, CDP840. A prototype of a novel class of orally active anti-inflammatory phosphodiesterase 4 inhibitors, Bioorg. Med. Chem. Lett. 12 (2002) 1451–1456.
[4] J.I. Andrés, J.M. Alonso, A. Díaz, J. Fernández, L. Iturrino, P. Martínez, E. Matesanz, E.J. Freyne, F. Deroose, G. Boeckx, D. Petit, G. Diels, A. Megens, M. Somers, J. Van Wauwe, P. Stoppie, M. Cools, F. De Clerck, D. Peeters, D. de Chaffoy, Synthesis and biological evaluation of imidazol-2-one and 2-cyanoiminoimidazole derivatives: novel series of PDE4 inhibitors, Bioorg. Med. Chem. Lett. 12 (2002) 653–658.
[5] O. Bruno, A. Romussi, A. Spallarossa, C. Brullo, S. Schenone, F. Bondavalli, N. Vanthuyne, C. Roussel, New selective phosphodiesterase 4D inhibitors differently acting on long, short, and supershort isoforms, J. Med. Chem. 52 (2009) 6546–6557.
[6] G.M. Buckley, N. Cooper, H.J. Dyke, F.P. Galleway, L. Gowers, A.F. Haughan, H.J. Kendall, C. Lowe, R. Maxey, J.G. Montana, R. Naylor, J. Oxford, J.C. Peake, C.L. Picken, K.A. Runcie, V. Sabin, A. Sharpe, J.B. Warneck, 8-Methoxyquinoline-5-carboxamides as PDE4 inhibitors: a potential treatment for asthma, Bioorg. Med. Chem. Lett. 12 (2002) 1613–1615.
[7] E. Boichot, J.L. Wallace, N. Germain, M. Corbel, C. Lugnier, V. Lagente, J.J. Bourguignon, Anti-inflammatory activities of a new series of selective phosphodiesterase 4 inhibitors derived from 9-benzyladenine, J. Pharmacol. Exp. Ther. 292 (2000) 647–653.
[8] J.R. Cashman, S. Ghirmai, Inhibition of serotonin and norepinephrine reuptake and inhibition of phosphodiesterase by multi-target inhibitors as potential agents for depression, Bioorg. Med. Chem. 17 (2009) 6890–6897.
[9] H. Ochiai, T. Ohtani, A. Ishida, K. Kusumi, M. Kato, H. Kohno, K. Kishikawa, T. Obata, H. Nakai, M. Toda, Highly potent PDE4 inhibitors with therapeutic potential, Bioorg. Med. Chem. Lett. 14 (2004) 207–210.
[10] Y.K. Zhang, J.J. Plattner, T. Akama, S.J. Baker, V.S. Hernandez, V. Sanders, Y. Freund, R. Kimura, W. Bu, K.M. Hold, X.S. Lu, Design and synthesis of boron-containing PDE4 inhibitors using soft-drug strategy for potential dermatologic anti-inflammatory application, Bioorg. Med. Chem. Lett. 20 (2010) 2270–2274.
[11] A. Tropsha, Best practices for QSAR model development, validation, and exploitation, Mol. Inf. 29 (2010) 476–488.
[12] N.-S. Kang, S.-J. Hong, C.-H. Chae, S.-E. Yoo, Comparative molecular field analysis (CoMFA) for phosphodiesterase (PDE) IV inhibitors, J. Mol. Struct. THEOCHEM 820 (2007) 58–64.
[13] R. Sabet, A. Fassihi, B. Moeinifard, QSAR study of PETT derivatives as potent HIV-1 reverse transcriptase inhibitors, J. Mol. Graph. Model. 28 (2009) 146–155.
[14] Q. Li, Y. Wang, S.H. Bryant, A novel method for mining highly imbalanced high-throughput screening data in PubChem, Bioinformatics 25 (2009) 3310–3316.
[15] D.A.R.S. Latino, J.A. de Sousa, Assignment of EC numbers to enzymatic reactions with MOLMAP reaction descriptors and random forests, Chem. Inf. Model. 49 (2009) 1839–1846.
[16] N. Karalı, A. Gürsoy, F. Kandemirli, N. Shvets, F.B. Kaynak, S. Özbey, V. Kovalishyn, A. Dimoglo, Synthesis and structure-antituberculosis activity relationship of 1H-indole-2,3-dione derivatives, Bioorgan. Med. Chem. 15 (2007) 5888–5904.
[17] A.K. Chakraborti, B. Gopalakrishnan, M.E. Sobhia, A. Malde, 3D-QSAR studies on thieno[3,2-d]pyrimidines as phosphodiesterase IV inhibitors, Bioorg. Med. Chem. Lett. 13 (2003) 1403–1408.
[18] A.K. Chakraborti, B. Gopalakrishnan, M.E. Sobhia, A. A.Malde, Comparative molecular field analysis (CoMFA) of phthalazine derivatives as phosphodiesterase IV inhibitors, Bioorg. Med. Chem. Lett. 13 (2003) 2473–2479.
[19] P. Ducrot, C.R. Andrianjara, R.J. Wrigglesworth, CoMFA and CoMSIA 3D-quantitative structure-activity relationship model on benzodiazepine derivatives, inhibitors of phosphodiesterase IV, Comput. Aided Mol. Des. 15 (2001) 767–785.
[20] X. Dong, W. Zheng, A new structure-based QSAR method affords both descriptive and predictive models for phosphodiesterase-4 inhibitors, Curr. Chem. Genomics 2 (2008) 29–39.
[21] A. Adekoya, X. Dong, J. Ebalunode, W. Zheng, Development of improved models for phosphodiesterase-4 inhibitors with a multi-conformational structure-based QSAR method, Curr. Chem. Genomics 3 (2009) 54–61.
[22] M. Saracoglu, F. Kandemirli, V. Kovalishyn, T. Arslan, E.E. Ebenso, ETM-ANN approach application for thiobenzamide and quinolizidine derivatives, J. Biomed. Biotechnol. (2010) (Epub September 7, 2010).
[23] 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.
[24] B. Charpiot, J. Brun, I. Donze, R. Naef, M. Stefani, T. Mueller, Quinazolines: combined type 3 and 4 phosphodiesterase inhibitors, Bioorg. Med. Chem. Lett. 8 (1998) 2891–2896.
[25] C. Deleuze-Masquéfa, G. Gerebtzoff, G. Subra, J.R. Fabreguettes, A. Ovens, M. Carraz, M.P. Strub, J. Bompart, P. George, P.A. Bonnet, Design and synthesis of novel imidazo[1,2-a]quinoxalines as PDE4 inhibitors, Bioorg. Med. Chem. 12 (2004) 1129–1139.
[26] Y. Ducharme, R.W. Friesen, M. Blouin, B. Côté, D. Dubé, D. Ethier, R. Frenette, F. Laliberté, J.A. Mancini, P. Masson, A. Styhler, R.N. Young, Y. Girard, Substituted 2-pyridinemethanol derivatives as potent and selective phosphodiesterase-4 inhibitors, Bioorg. Med. Chem. Lett. 13 (2003) 1923–1926.
[27] Z. Huang, S. Liu, L. Zhang, M. Salem, G.M. Greig, C.C. Chan, Y. Natsumeda, K. Noguchi, Preferential inhibition of human phosphodiesterase 4 by ibudilast, Life Sci. 78 (2006) 2663–2668.
[28] E. Kim, H.O. Chun, S.H. Jung, J.H. Kim, J.M. Lee, B.C. Suh, M.X. Xiang, C.K. Rhee, Improvement of therapeutic index of phosphodiesterase type IV inhibitors as anti-Asthmatics, Bioorg. Med. Chem. Lett. 13 (2003) 2355–2358.
[29] T.H. Keller, K. Bray-French, F.W. Demnitz, T. Müller, E. Pombo-Villar, C. Walker, Synthesis and structure-activity relationship of N-arylrolipram derivatives as inhibitors of PDE4 isozymes, Chem. Pharm. Bull. 49 (2001) 1009–1017.
[30] A.P. Skoumbourdis, C.A. Leclair, E. Stefan, A.G. Turjanski, W. Maguire, S.A. Titus, R. Huang, D.S. Auld, J. Inglese, C.P. Austin, S.W. Michnick, M. Xia, C.J. Thomas, Exploration and optimization of substituted triazolothiadiazines and triazolopyridazines as PDE4 inhibitors, Bioorg. Med. Chem. Lett. 19 (2009) 3686–3692.
[31] https://www.chemaxon.com/ (accessed February 2011).
[32] http://www.talete.mi.it/products/dragon_description.htm (accessed March 2011).
[33] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Chapman & Hall/CRC, Boca Raton, FL, 1984.

[34] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2004, ISBN 3-900051-07-0 (last accessed in October 2010). Fortran original by Leo Breiman, Adele Cutler, R port by Andy Liaw and Matthew Wiener http://www.R-project.org.
[35] I.V. Tetko, Neural network studies. 4. Introduction to associative neural network, J. Chem. Inf. Comput. Sci. 42 (2002) 717–728.
[36] I.V. Tetko, V.Yu. Tanchuk, N.P. Chentsova, S.V. Antonenko, G.I. Poda, V.P. Kukhar, A.I. Luik, HIV-1 reverse transcriptase inhibitor design using artificial neural networks, J. Med. Chem. 37 (1994) 2520–2526.
[37] T. Tollenaere, S.A.B. Super, Fast adaptive back propagation with good scaling properties, Neural Networks 3 (1990) 561–573.
[38] I.V. Tetko, D.J. Livingstone, A.I. Luik, Neural network studies. 1. Comparison of overfitting and overtraining, J. Chem. Inf. Comput. Sci. 35 (1995) 826–833.
[39] I.V. Tetko, A.E.P. Villa, Efficient partition of learning data sets for neural network training, Neural Networks 10 (1997) 1361–1374.
[40] Y. LeCun, J.S. Dencer, S.A. Solla, Optimal brain damage, in: D.S. Touretzky (Ed.), NIPS*2, Morgan Kaufmann, San Mateo, CA, 1990, pp. 598–605.
[41] Y.A. Chauvin, Back-propagation algorithm with optimal use of hidden units, in: D.S. Touretzky (Ed.), NIPS*1, Morgan Kaufmann, San Mateo, CA, 1989, pp. 519–526.
[42] I.V. Tetko, A.E.P. Villa, D.J. Livingstone, Neural network studies. 2. Variable selection, J. Chem. Inf. Comput. Sci. 36 (1996) 794–803.
[43] V.V. Kovalishyn, I.V. Tetko, A.I. Luik, V.V. Kholodovych, A.E.P. Villa, D.J. Livingstone, Neural network studies. 3. Variable selection in the cascade-correlation learning architecture, J. Chem. Inf. Comput. Sci. 38 (1998) 651–659.
[44] R.D. Cramer III, D.E. Patterson, J.D. Bunce, Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins, J. Am. Chem. Soc. 110 (1988) 5959–5967.
[45] R. Guha, S.C. Schürer, Utilizing high throughput screening data for predictive toxicology models: protocols and application to MLSCN assays, J. Comput. Aided Mol. Des. 22 (2008) 367–384.
[46] L. Zhang, H. Zhu, T.I. Oprea, A. Golbraikh, A. Tropsha, QSAR modeling of the blood-brain barrier permeability for diverse organic compounds, Pharm. Res. 25 (2008) 1902–1914.