Virtual Screening Tailored Ensembles of QSAR Models for the Discovery of Dual A₂A Adenosine Receptor Antagonists / Monoamine Oxidase B Inhibitors

Aliuska Morales Helguera[1], Yunierkis Perez-Castillo[1,2], M. Natália D. S. Cordeiro[3], Eduardo Tejera[4], Cesar Paz-y-Miño[4], Aminael Sánchez-Rodríguez[5], Marta Teijeira Bautista[6,7], Evys Ancede-Gallardo[1], Fernando Cagide[8], Fernanda Borges*,[8], Maykel Cruz-Monteagudo*,[4,8]

[1] Molecular Simulation and Drug Design Group, Centro de Bioactivos Químicos (CBQ), Central University of Las Villas, Santa Clara, 54830, Cuba;

[2] Sección Físico Química y Matemáticas, Departamento de Química, Universidad Técnica Particular de Loja, San Cayetano Alto S/N, EC1101608 Loja, Ecuador;

[3] REQUIMTE, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal;

[4] Instituto de Investigaciones Biomédicas (IIB), Universidad de Las Américas, 170513 Quito, Ecuador;

[5] Departamento de Ciencias Naturales, Universidad Técnica Particular de Loja, Calle París S/N, EC1101608 Loja, Ecuador;

[6] Departamento de Química Orgánica. Facultade de Química, Universidade de Vigo, 36310 Vigo, Spain;

[7] Instituto de Investigación Biomédica (IVIB), Universidade de Vigo, 36310 Vigo, Spain;

[8] CIQUP/Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Porto 4169-007, Portugal.

* Author to whom correspondence should be addressed; E-Mail: gmailkelcm@yahoo.es (MCM); fborges@fc.up.pt (FB); Tel.: +351 220402502; Fax: +351 220402659.

Abstract: Virtual Screening methodologies have emerged as efficient alternatives for the discovery of new drug candidates. At the same time, ensemble methods are nowadays frequently used to overcome the limitations of employing a single model in ligand-based drug design. However, many applications of ensemble methods to this area do not consider important aspects related to both virtual screening and the modeling process. During the application of ensemble methods to virtual screening the proper validation of the models in virtual screening conditions is often neglected. Frequently no analysis is performed of the diversity of the ensemble members or no considerations regarding the applicability domain of the base model are made. In this research we propose a method employing genetic algorithms optimization for the generation of virtual

screening tailored ensembles that address problems in the current applications of ensemble methods to virtual screening. The proposed methodology is successfully applied to the generation of ensemble models for the ligand-based virtual screening of dual target A2A adenosine receptor antagonists and MAO-B inhibitors as potential Parkinson's disease therapeutics.

.

1. Introduction

During the last decades, Virtual Screening (VS) methodologies have emerged as efficient alternatives to the expensive, in terms of time and money, High Throughput Screening (HTS) approaches for the discovery of new drug candidates [1]. In terms of efficiency, the hit rates obtained when VS tools are employed to filter large databases of chemical compounds are considerably higher than those obtained with HTS techniques [2]. Literature reports where VS experiments conducted to the identification of hit molecules in a wide range of application can be found elsewhere [3,4].

VS techniques can be divided into two main categories: Structure-Based VS (SBVS) and Ligand-Based VS (LBVS) [5]. The first one includes all the modeling approaches such as Molecular Docking and Molecular Dynamics that depend on the structure of a molecular receptor. In some cases the structure of the molecular target is not available or the research process focuses in a mechanism where a molecular target cannot be defined. It can also be the case that the amount of data to process is too large to complete a SBVS campaign in a reasonable amount of time. In these cases LBVS tools can aid in the drug discovery process. These types of techniques include Similarity methods, Shape-based methods, Pharmacophore modeling and Quantitative Structure-Activity

Relationships (QSAR) studies [5,6]. Specifically, QSAR approaches have been used for VS in the early stages of drug discovery [7].

To exploit the full potential of QSAR modeling some general guidelines have to be followed. These guidelines have been summarized as best practices for QSAR [8]. In addition to these guidelines, in our previous research we pointed out the importance of the proper validation of QSAR models in VS conditions [9].

Ensemble methods (EM) have gained popularity among QSAR practitioners, providing a set of tools for combining the predictions of single models in a more robust and generalizable model [9-11]. Although the success of these methods in Ligand-Based Drug Design (LBDD) and LBVS, the currently published applications of ensemble methods to LBVS suffer from a limitation common to most LBDD workflows: their VS performance is not retrospectively evaluated prior to their use in VS campaigns. In addition, in most reports of applications of ensemble methods to LBVS no considerations are made regarding the diversity of the base classifiers. The optimal size of the developed ensembles is neither considered and all available base classifiers are combined instead. These two last factors can drastically reduce the performance of ensemble methods [12].

On the other side, Parkinson Disease (PD) causes chronic disability and it is the second commonest degenerative condition of the nervous system. The standard treatment for PD is levodopa, which helps to increase the dopamine levels in the brain [13]. However there is a need of finding alternative therapies since levodopa has many side effects and can become ineffective over time. To this end, multicomponent therapies (combination of different drugs) have been used. However the discovery of new multi-target drugs (a single molecule that acts on multiple targets) is attracting more and more attention [14]. Multi-target drugs, compared with the use of combinations of different drugs, have more predictable pharmacokinetic and pharmacodynamic relationships as a consequence of the administration of a single drug [15].

Here, are summarized the results obtained in the development of a GA-based ensemble selection method and its application to the search of ensembles suitable for the VS identification of dual target ligands for PD. It should be mentioned that the application of LBVS tools for the planned design of ligands with a predefined dual-target profile is a recent development and represents an important challenge to medicinal chemists [16].

.

2. Results and Discussion

The collection, curation, class assignment, representation and dataset splitting processes were performed following the procedures described in the Materials and Methods section. Both modeling sets were split into training and test subsets using three sphere exclusion algorithms 1M, 2M and 3M. This partition scheme lead to the selection of 20-23 % of the modeling data for the test set and guaranties the representativeness of each class in the test set.

Fragment descriptors were calculated with ISIDA Fragmentor software [17,18] and they were filtered to discard those with zero or close to zero variance. Afterward, the mRMR algorithm [19] was employed to only keep the 250 more relevant descriptors. ISIDA's header files listing the 250 most relevant fragments for each dataset and their partitions are provided as Supporting Information.

The main goal of this paper is to find models that can be effective in the identification of dual target ligands for PD through VS. To this end, two sets of QSAR models were independently developed for $A_{2A}AR$ binding and MAO-B inhibition following the procedure summarized in the Materials and Methods section. For each endpoint, three feature selection methods and three different classification algorithms were combined to generate nine different classifier types, which were trained and validated using three modeling set partitions obtained with the sphere exclusion algorithms 1M, 2M and 3M. In consequence, 27 different classification experiments per endpoint were conducted. At this point, it is important to remark that the external set was only used to evaluate the real predictive power of the models. In Table 1 are presented the statistics of the optimal models for each partition of the dataset. These results show that the employed QSAR modeling framework provides accurate, robust and generalizable models.

All 16 known dual ligands, MAO-B inhibitors and $A_{2A}AR$ antagonists, were compiled from the literature [20,21]. For each known dual-target compound 1608 decoys were selected based on desirability-based home-developed algorithm that has been previously employed in the

selection of tailored decoy sets for the validation of virtual screening strategies [9]. These ligands and decoys were prepared for virtual screening following the same protocol described for the datasets.

To evaluate the performance of the models in virtual screening the following metrics were employed: Area Under the Accumulation Curve (AUAC), Area under the Receiver Operating Characteristic Curve (ROC), Enrichment Factor (EF) and Boltzmann-Enhanced Discrimination of ROC (BEDROC). These metrics were computed as proposed by Truchon *et. al.*[22]

In our previous research [9] the VS-tailored ensembles were obtained using only a very limited set of predictive models that were selected as the best ones derived from each modeling approach. However, during the modeling process a larger number of high quality models are obtained that we did not considered for ensemble modeling in that investigation.

To discard low performing models, the whole pool of classification models was first filtered and a model was considered to be accurate, robust and predictive if it had values of accuracy higher than 0.75 in predicting the train, selection and external sets as well as in the cross validation experiments. The models fulfilling the above criteria are considered as valid models from here on.

For the three modeling set partitions obtained after applying the sphere exclusion algorithms, there were a total of 1526 and 1347 valid models for the A2A adenosine receptor antagonists and MAO-B inhibitors datasets respectively. Since one of the factors negatively influencing the performance of ensemble models is the redundancy of the base classifiers, these valid models were clustered following the protocol above to obtain a representative set of them. This procedure yielded 33 and 48 representative

models for the A2A adenosine receptor antagonists and MAO-B inhibitors datasets respectively. The representative model of each cluster was selected as the one having the highest value of BEDROC for $\alpha=160.9$. The best performance among the representative models corresponded to a value of BEDROC=0.20 for $\alpha=160.9$.

We then evaluated the proposed GA-guided algorithm for the search of the ensemble maximizing BEDROC at a given fraction of screened data. The algorithm was run five times with different initial populations of 100 individuals each one, yielding 500 solutions. That is, 500 ensembles are obtained. The GA fitness function maximizes BEDROC for $\alpha=160.9$.

In Table 2 are presented the enrichment metrics for the best individual found after running the GA-guided ensemble generation algorithm presented in this communication, when we search for the ensemble maximizing the initial enrichment of dual ligands. The accumulative curves for this ensemble as well as for its members are shown in Figure 1.

The analysis of these results show that the obtained ensemble is composed by 14 models, five of them related to the prediction of the antagonist activity of the A2A adenosine receptor and nine related to the inhibition of the MAO-B enzyme. This ensemble outperforms the VS metrics of all its members. Specifically, the improvement in the value of BEDROC for $\alpha=160.9$ for the obtained ensemble relative to the best single model it is composed of is 0.05. The model with the highest value of BEDROC ($\alpha=160.9$) among all single models within the ensemble is also the one with the best value of this metric among the set of representative valid models. This improvement might seem meaningless, however it can be interpreted as 5%

more probability of retrieving a dual target ligand in the first 1% of the ranked list using the obtained ensemble compared to the best performing individual model [22]. From Table 2 it can also be seen that the values of BEDROC for α=32.2 and α=20 are also higher when the ensemble model is compared with its members. In addition, improvements for the EF metric at the three analyzed fractions of screened data were obtained. The advantages of using the obtained ensemble for VS experiments over the use of single models can be visualized in Figure 1 where the accumulative curves of the ensemble and the models it is composed of are plotted. From this figure it is clear that the ensemble model is able to retrieve more known dual ligands and at lower positions in the ranked list than any of the single models it is composed of.

Another advantage of employing the proposed ensemble for VS tasks is that its applicability domain covers 93% of the whole virtual screening validation set, representing an improvement of 17% of coverage relative to the single model with the highest applicability domain coverage. If the applicability domain coverage of the ensemble is compared to that of the model with the highest BEDROC value for α=160.9, then this improvement increases to 33%. In contrast, the value of ROC of the ensemble is lower than the mean value of ROC across the ensemble members. This last result is a consequence of designing the GA search to find ensembles maximizing the initial enrichment of known dual ligands. In other words, the calculations here performed focus in retrieving the more dual ligands at the very first part of the ranked list and neglect the position of the ligands in the remaining of the list.

.

Table 1. Statistics for the best performing models for each dataset partition.

Target	Method[a]	Size[b]	Train [c]	Test[d]	LOO[e]	Boot[f]	5-Fold[g]	Ext [h]
A2A 1M	GA-AB	10	90 (93/86)	80 (90/72)	87.13	85.42	88.12	81 (81/81)
A2A 2M	GA-LSSVM	11	94 (90/97)	82 (77/88)	88.56	86.32	89.55	84 (84/84)
A2A 3M	GA-LSSVM	10	91 (87/93)	80 (78/83)	88.56	84.67	88.56	86 (84/88)
MAO 1M	GA-LDA	10	85 (90/81)	85 (83/86)	84.26	81.23	83.25	75 (79/67)
MAO 2M	BT-LSSVM	18	95 (93/97)	93 (92/94)	77.50	73.81	77.50	72 (63/88)
MAO 3M	GA-LSSVM	14	93 (94/92)	84 (79/89)	86.07	82.46	86.57	76 (78/73)

[a] Modeling method the classifier is based on, GA stands for Genetic Algorithm, BT for Bagged Trees, AB for Adaboost, LSSVM for Least Squares Support Vector Machines and LDA for Linear Discriminant Analysis.

[b] Number of features in the model.

[c,d,h] Accuracy in predicting the training, test and external sets respectively

[e,f,g] Accuracy of the Leave One Out, Bootstraping and 5-fold cross-validation experiments respectively

Table 2. Enrichment metrics for the ensemble maximizing the initial enrichment of known dual ligands.

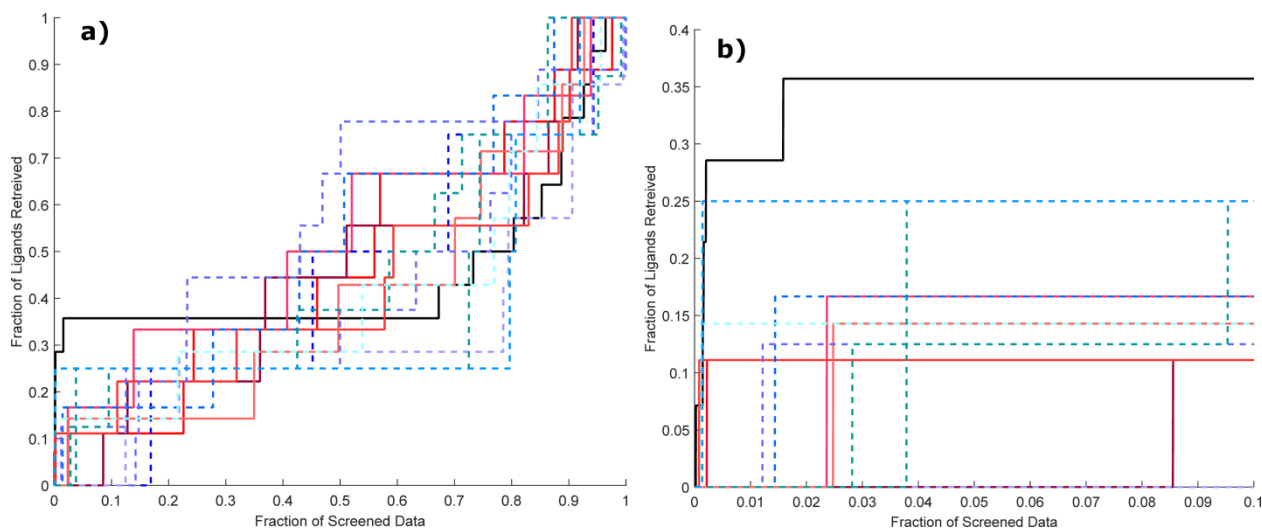| Model	BEDROC Alpha			EF %			ROC	Cov. Domain
	160.9	32.2	20	1	5	8		
Ensemble	0.25	0.32	0.33	28.47	7.14	4.46	0.45	0.93
Mean Indiv. Models	0.04	0.08	0.10	4.39	2.27	1.42	0.44	0.69

Figure 1. Accumulation curves for the best ensemble as well as for its base classifiers: a) For the whole virtual screening validation datasset and b) For the first 10% of screened data. The black line corresponds to the obtained ensemble, red continuous lines to A2A models and blue discontinuous lines to the MAO-B models

3. Materials and Methods

Data sets

Two data sets were used in this paper; A2AAR antagonists and human MAO-B (hMAO-B) inhibitors. The A2AAR antagonist data set was retrieved from 18 different literature sources. The compounds were divided into two classes according to their Ki values. The first class, designated as potent antagonists, included all chemicals with a Ki ≤ 1000 nM (pKi ≥ 6). The second class, named as weak antagonists, was formed by compounds with Ki >1000 nM (pKi < 6). As result of this categorization a balanced dataset was obtained that included 161 potent antagonists and 166 weak antagonists.

The hMAO-B inhibitors data set was compiled from [23] and it contains 474 compounds. the compounds were classified in two groups according with theirs IC_{50} values. The first group, named hMAO-B inhibitors, included those chemicals with $IC_{50} \leq 20$ µM ($pIC_{50} \geq 4.70$), while the second one, designated as hMAO-B non-inhibitors comprises those

compounds with $IC_{50} > 20$ µM ($pIC_{50} < 4.70$). Thus, the 474 ligands were split in 313 inhibitors and 161 non-inhibitors of hMAO-B.

The chemical structures were represented in smiles format and then converted to a SD file (SDF) using the ChemAxon's JChem for Excel (6.3.1.1807) program [24]. Each data set, as well as the decoy compound candidates were curated following the guidelines proposed in the literature [25] using ChemAxon's Standardizer [26]

Datasets splitting

The external sets were randomly selected as 20% of the entire initial datasets. The modeling sets were subsequently partitioned into training and test sets using the three sphere exclusion (SE) algorithms proposed by Golbraikh [27] and implemented in our laboratory that ensure the closeness in chemical spaces of the train and test sets.

The three variants 1M, 2M, and 3M of the sphere exclusion algorithm proposed by the

developers and used here to divide the balanced modeling sets were implemented in MATLAB [28]. Unlike the original algorithms, for the SE based partitioning of the data the structure of the compounds was encoded as 1024 bits Chemaxon's Topological Fingerprints from GenerateMD program [26]; and the Tanimoto distance was selected as the distance metric. The radius of the spheres was varied between 0.05 and 1.0 with a step of 0.05

Molecular Descriptors

The ISIDA Fragmentor software (freely available at http://infochim.u-strasbg.fr/spip.php?rubrique49) was used to calculate 2D fragment descriptors [17,18].

Descriptors were calculated for the training dataset. Afterward, fragments with the same value for 99% of the samples were removed. The minimal Redundancy Maximal Relevance (mRMR) algorithm [19] was applied to the reduced data set to keep only the top 500 fragments according to the MIQ score. The same subset of 500 fragments was then computed for the selection and external sets.

Classification-based modeling methodology

QSAR modeling for each dataset was performed using the previously proposed QSAR modeling framework. Here the main steps involved in the modeling process are summarized and the detailed description of the QSAR modeling framework is available in [9]

Methodology of models ensemble for virtual screening

For both targets QSAR models were first filtered, regardless the dataset partition or modeling approach they were obtained with, to ensure that only accurate, robust and predictive models remain eligible as ensemble members. For this step we set the same cutoff value of 0.75 for the accuracy in predicting the training, selection and external datasets as well as in the

cross-validation experiments. The models fulfilling these conditions were then clustered using the Hamming distance between the predictions they made on the external dataset to select a representative set of models to ensure diversity in the pool of base models. Clustering was carried out using the K-means algorithm implemented in MATLAB [28]. To obtain the optimal number of clusters we examined the silhouette plot [29] when the number of clusters varied between 3 and 50 and selected the number of clusters corresponding to the maximum of this plot. This procedure leads to the selection of N_{A2A} and N_{MAOB} representative models which form the final pool of diverse models candidates for ensembles.

To build the ensembles for VS we followed the same protocol described in our previous publication [9]. In brief, for a given subset of QSAR models forming the ensemble and a dataset to be evaluated we first search for the samples included within each model applicability domain. Next, the scores produced by each model for the compounds inside its applicability domain are used to obtain a ranking for them and the relative ranking of each sample in each model is computed. Once the relative rank of every sample in each model considering the model's applicability domain is determined, these relative rank values are averaged over the models the compound is inside their applicability domains to obtain the final aggregated score. Finally, the compounds are sorted according to this aggregated score in ascending order to obtain the final ensemble ranking.

Given that to evaluate the virtual screening performance of all ensembles formed by all possible combination of size two to $N_{A2A}+N_{MAOB}$ of the selected models is computationally unfeasible, we implemented a novel GA search strategy which could find combinations of

models maximizing the EF at a given fraction of screened data. Each individual for the GA search represents an ensemble and they are encoded as binary vectors of length $N_{A2A}+N_{MAOB}$. In an individual the "on" bits encode the set of models considered for the ensemble while "off" bits represent models excluded from the ensemble. The initial population was set to 100 randomly generated individuals and the population evolved for 100 generations. The crossover and mutation rates were set to 0.7 and 0.3 respectively while the best two individuals survived to the next generation. The selection operator was set to a tournament of size 2. For the crossover operator, the offspring chromosomes were randomly selected position by position from the two selected parents. The mutation operator changed a randomly selected "on" bit to "off" and one randomly selected "off" bit to "on" in the individual. In each case study the GA was run five times using different initial populations. The objective function for the GA was selected as BEDROC for $\alpha=160$.

..

4. Conclusions

We designed a methodology for the generation of virtual screening tailored ensembles capable of overcoming the previously identified problems. This methodology considered the diversity of the base models for ensemble generation and their applicability domains. The main advantage of the proposed algorithm is that it is able of finding the combination of models providing the best VS performance for a specific problem.

The proposed algorithm was applied to the VS simulation of dual target A2A adenosine receptor antagonists and MAO-B inhibitors. The obtained results showed that the obtained ensemble outperformed the best individual model according to the evaluated enrichment metrics. Thus, confirming the expected improved performance of ensemble models over single ones. In the specific problem being addressed, the results of the ensemble modeling process highlighted the importance of considering information from both targets for the discovery of dual target ligands.

Conflicts of Interest
The authors declare no conflict of interests.

References and Notes
1. Macalino, S.J.; Gosu, V.; Hong, S.; Choi, S., Role of computer-aided drug design in modern drug discovery. *Archives of pharmacal research* **2015**, *38*, 1686-1701.
2. Zhu, T.; Cao, S.; Su, P.C.; Patel, R.; Shah, D.; Chokshi, H.B.; Szukala, R.; Johnson, M.E.; Hevener, K.E., Hit identification and optimization in virtual screening: Practical recommendations based on a critical literature analysis. *J Med Chem* **2013**, *56*, 6560-6572.
3. Castillo-Gonzalez, D.; Mergny, J.L.; De Rache, A.; Perez-Machado, G.; Cabrera-Perez, M.A.; Nicolotti, O.; Introcaso, A.; Mangiatordi, G.F.; Guedin, A.; Bourdoncle, A*., et al.*, Harmonization of qsar best practices and molecular docking provides an efficient virtual screening tool for discovering new g-quadruplex ligands. *Journal of Chemical Information and Modeling* **2015**, *55*, 2094-2110.

4. Miller, Z.; Kim, K.S.; Lee, D.M.; Kasam, V.; Baek, S.E.; Lee, K.H.; Zhang, Y.Y.; Ao, L.; Carmony, K.; Lee, N.R.*, et al.*, Proteasome inhibitors with pyrazole scaffolds from structure-based virtual screening. *Journal of Medicinal Chemistry* **2015**, *58*, 2036-2041.

5. Lill, M., Virtual screening in drug design. *Methods in molecular biology (Clifton, N.J.)* **2013**, *993*, 1-12.

6. Lavecchia, A.; Di Giovanni, C., Virtual screening strategies in drug discovery: A critical review. *Current medicinal chemistry* **2013**, *20*, 2839-2860.

7. Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R., Qsar modeling: Where have you been? Where are you going to? *Journal of medicinal chemistry* **2014**.

8. Tropsha, A., Best practices for qsar model development, validation, and exploitation. *Mol Inf* **2010**, *29*, 476-488.

9. Perez-Castillo, Y.; Cruz-Monteagudo, M.; Lazar, C.; Taminau, J.; Froeyen, M.; Cabrera-Perez, M.A.; Nowe, A., Toward the computer-aided discovery of fabh inhibitors. Do predictive qsar models ensure high quality virtual screening performance? *Mol Divers* **2014**, *18*, 637-654.

10. Bonet, I.; Franco-Montero, P.; Rivero, V.; Teijeira, M.; Borges, F.; Uriarte, E.; Morales Helguera, A., Classifier ensemble based on feature selection and diversity measures for predicting the affinity of a(2b) adenosine receptor antagonists. *Journal of Chemical Information and Modeling* **2013**, *53*, 3140-3155.

11. Zhang, L.; Fourches, D.; Sedykh, A.; Zhu, H.; Golbraikh, A.; Ekins, S.; Clark, J.; Connelly, M.C.; Sigal, M.; Hodges, D.*, et al.*, Discovery of novel antimalarial compounds enabled by qsar-based virtual screening. *Journal of Chemical Information and Modeling* **2013**, *53*, 475-492.

12. Polikar, R., Ensemble based systems in decision making. *Circuits and Systems Magazine, IEEE* **2006**, *6*, 21-45.

13. Brunton, L.L., *Goodman & gilman's the pharmacological basis of therapeutics*. 11th ed.; The McGraw-Hill: New York, NY, 2007.

14. Youdim, M.B.; Geldenhuys, W.J.; Van der Schyf, C.J., Why should we use multifunctional neuroprotective and neurorestorative drugs for parkinson's disease? *Parkinsonism Rel Disord* **2007**, *13*, S281-S291.

15. Morphy, R.; Kay, C.; Rankovic, Z., From magic bullets to designed multiple ligands. *Drug Discovery Today* **2004**, *9*, 641-651.

16. Wang, Y.; Ge, H.; Li, Y.; Xie, Y.; He, Y.; Xu, M.; Gu, Q.; Xu, J., Predicting dual-targeting anti-influenza agents using multi-models. *Molecular diversity* **2014**, *19*, 123-134.

17. Varnek, A.; Fourches, D.; Hoonakker, F.; Solov'ev, V.P., Substructural fragments: An universal language to encode reactions, molecular and supramolecular structures. *Journal of computer-aided molecular design* **2005**, *19*, 693-703.

18. Varnek, A., Isida-platform for virtual screening based on fragment and pharmacophoric descriptors. *Current Computer Aided-Drug Design* **2008**, *4*.

19. Peng, H.; Long, F.; Ding, C., Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **2005**, *27*, 1226-1238.

20. Stossel, A.; Schlenk, M.; Hinz, S.; Kuppers, P.; Heer, J.; Gutschow, M.; Muller, C.E., Dual targeting of adenosine a(2a) receptors and monoamine oxidase b by 4h-3,1-benzothiazin-4-ones. *J Med Chem* **2013**, *56*, 4580-4596.

21. S., R.; Piersanti, G.; Bartoccini, F.; Diamantini, G.; Pala, D.; Riccioni, T.; Stasi, M.A.; Cabri, W.; Borsini, F.; Mor, M*., et al.*, Synthesis of (e)‑8-(3-chlorostyryl)caffeine analogues leading to 9-deazaxanthine derivatives as dual a2a antagonists/mao-b inhibitors. *J Med Chem* **2013**, *56*, 1247−1261.

22. Truchon, J.F.; Bayly, C.I., Evaluating virtual screening methods: Good and bad metrics for the "early recognition" problem. *Journal of Chemical Information and Modeling* **2007**, *47*, 488-508.

23. Helguera, A.M.; Pérez-Garrido, A.; Gaspar, A.; Reis, J.; Cagide, F.; Vina, D.; Cordeiro, M.N.D.S.; Borges, F., Combining qsar classification models for predictive modeling of human monoamine oxidase inhibitors. *European Journal of Medicinal Chemistry* **2013**, *59*, 75-90.

24. ChemAxon (http://www.chemaxon.com) *Jchem for excel, version 6.3.1.1807*, Budapest, Hungary, 2013.

25. Fourches, D.; Muratov, E.; Tropsha, A., Trust, but verify: On the importance of chemical structure curation in cheminformatics and qsar modeling research. *Journal of Chemical Information and Modeling* **2010**, *50*, 1189-1204.

26. ChemAxon (http://www.chemaxon.com) *Jchem, version 6.3.1*, Budapest, Hungary, 2013.

27. Golbraikh, A.; Tropsha, A., Predictive qsar modeling based on diversity sampling of experimental datasets for the training and test set selection. *Molecular diversity* **2000**, *5*, 231-243.

28. MATLAB *Version 8.1.0.604 (r2013a)*, The MathWorks Inc.: Natick, Massachusetts, 2009.

29. Rousseeuw, P.J., Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* **1987**, *20*, 53-65.