# QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP MODELING OF SEROTONIN TYPE-6 RECEPTOR ANTAGONISTS

By

DANIEL P. RUSSO

A thesis submitted to the Graduate School – Camden

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements for the degree of

Master of Science Graduate Program in Biology

written under the direction of

Joseph V. Martin

and approved by

_____

Joseph V. Martin

_____

Hao Zhu

_____

Simeon O. Kotchoni

_____

William Saidel

Camden, New Jersey May, 2014

ABSTRACT OF THE THESIS

QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP MODELING OF SEROTONIN TYPE-6 RECEPTOR ANTAGONISTS

By DANIEL P. RUSSO

Thesis Director:

Joseph V. Martin

The serotonin type-6 receptor (5-HT6) is a drug target for many psychotic diseases, especially cognitive disorders. The traditional method to design novel 5-HT6 binding agents (e.g. antagonists) is to experimentally screen a large chemical dataset that is randomly selected from a drug-like chemical library. This process is normally very costly and has a low success rate. Computer Aided Drug Discovery (CADD) uses computational models to virtually screen a chemical library and select promising candidates for experimental testing. Using CADD, the resources could be saved and the success rate could be increased by excluding unsuitable compounds. Quantitative Structure-Activity Relationship (QSAR) is the most frequently used method for developing various predictive models within the drug discovery process. In this work, a 5-HT6 dataset of 488 unique compounds was compiled. Among them, 225 were experimentally identified as 5-HT6 antagonists and the remaining were diverse anti-cancer compounds, which were considered to be unable to bind to 5-HT6. I applied various QSAR modeling approaches to develop several computational binary 5-HT6 models. The resulting models were validated by a five-fold cross-validation approach and the resulting predictivity, which was measured using Correct Classification Rate (CCR),

was 96%. The resulting models were used to predict an external data set and the predictivity (CCR=88%) was similar to the cross validation. Thus, the models developed in this study could be used to detect novel 5-HT6 ligands in the future drug discovery process.

# ACKNOWLEDGMENTS

I would like to first and foremost thank my mentor, Dr. Hao Zhu. The skills and encouragement I have received in your lab are vast and not underappreciated. Your door was always open and a very genuine inspiration to learn was always apparent. I feel very lucky and grateful to have had the opportunity to study under your tutelage.

I would also like to thank Dr. Joseph V. Martin and Dr. Simeon O. Kotchoni. You both have always been very kind and knowledgeable whenever I came to you with either questions or in need of advice.

Finally, I would like to thank my good friend and lab mate, Marlene Kim. Your dedication only made me want to work harder. Your help was always available. And, of course, your companionship made me look forward to the late nights in the lab.

# DEDICATION

I would like to dedicate this work to my mother, father, and brothers. Your support and encouragement has always been and will always be my strongest motivator. I would also like to dedicate this work to my dog, Charles Barkley.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

## Introduction

The neurotransmitter serotonin (5-HT) is critical in a variety of neurological and cognitive dysfunctions including anxiety, migraines, depression, and schizophrenia.[1,2] The serotonin receptors are classified into seven major families with 14 distinctive receptor subtypes.[3,4] Among them, the serotonin type-6 receptor (5-HT6) contains a 440 amino acid chain with seven hydrophobic, membrane-spanning regions placing it in the G-protein-coupled receptor (GPCR) family.[5,6] The 5-HT6 receptor has attracted great attention since it is almost exclusively localized to the central nervous system and is primarily expressed in regions of the brain known to be associated with learning and memory.[7] It is also relevant to adenylate cyclase stimulation.[6] The inhibition of the 5-HT6 receptor will, specifically, cause an increase of neurotransmission in glutamatergic and cholinergic pathways, and ultimately increases cognition in rats.[8–13] All these characteristics make the 5-HT6 receptor a potential therapeutic target for many CNS diseases, such as Alzheimer's disease.[13,14] Additionally, 5-HT6 shows high-affinity to many non-selective anti-depressant and anxiolytic compounds.[15] Thus, there has been lots of research on the development of novel and selective 5-HT6 antagonists.

The Computer Aided Drug Discovery (CADD) techniques, such as Quantitative Structure-Activity Relationship (QSAR) modeling, have been applied to reduce the experimental costs of developing new drugs. Several QSAR studies have been published to discover novel 5-HT6 antagonists. In 2004, Doodareddy *et al*. investigated a three-dimensional QSAR study consisting of 33 compounds with dissociation constant (*K*i) values ranging from 1.3 to 1700 nM.[16] In another study, they developed hologram-QSAR models for two data sets; one consisting of 48 congeneric 5-HT6 antagonists and

another one with 30 structurally diverse 5-HT6 antagonists.[17] In 2005, López-Rodríguez *et al*. developed a pharmacophore model based on 45 diverse 5-HT6 antagonists.[18] In 2010, Goodarzi *et al.* modeled 52 antagonists and agonists.[19] Sharma *et al.* used 50 compounds, representing 5 different classes of chemicals, to develop the predictive models.[20] In a recent study, a comprehensive data set consisting of 223 compounds was compiled by Hao *et al.* and used for modeling purposes.[21] These compounds were either 5-HT6 receptor agonists or antagonists. However, the applicability of these models is limited, mostly due to the small number of compounds used to develop the models. Furthermore, most of these modeling set compounds are congeneric and not very useful for identifying novel agents with new chemical scaffolds.

In this study, I compiled a modeling set consisting of 488 structurally diverse active and inactive 5-HT6 receptor antagonists and developed several QSAR models based on this set. The resulting models were validated by a five-fold cross validation approach. Furthermore, all the models were used to predict a data set which was compiled after the models were developed. The models developed in this study could be used to prioritize compounds with novel chemical scaffolds for future experimental testing**.**

## Materials and Methods

### 5-HT6 Antagonist Data Set

The active compounds for the modeling set were compiled from multiple resources. All of the assays compiled used the same radioligand, radiolabeled lysergic acid ([3H]-LSD).[5,22–29] The activities of these compounds are expressed in *K*i values,

which are defined as the concentration at which the inhibitor can displace 50% of the radiolabeled ligand. In this study, I classified all the compounds with an experimental *K*i less than 10,000 nM as 5-HT6 antagonists (or "actives"). Afterwards, I compiled around 2,000 anti-cancer drug-like compounds. Since these anti-cancer drugs are cytotoxic compounds, it is reasonable to consider them as 5-HT6 non-binders (or inactives). After removing duplicates, inorganic compounds, and mixtures, 225 unique 5-HT6 actives and 1,558 inactives were used for model development. In order to develop a more manageable dataset with better predictability, I used a fragment-based similarity search. Using our actives as probe molecules, I used a Tanimoto coefficient of 0.6 as a threshold.[30] This generated 296 inactive compounds that were used for modeling.

After the models were developed, I found an external validation set consisting of 82 compounds from other resouces.[31–33] There are 38 5-HT6 antagonists and the remaining compounds are non-binders based on our above definitions.

## Chemical Descriptors

The chemical descriptors used in this study were obtained from Dragon version 6.0 (Talete SRL, Milano, Italy) and Molecular Operating Environment (MOE) version 2011. The Dragon descriptors include E-state values and E-state counts, constitutional descriptors, topological descriptors, walk and path counts, connectivity and information indices, 2D autocorrelations, Burden eigenvalues, molecular properties (i.e., the octanol-water partition coefficient), Kappa, hydrogen bond acceptor/donor counts, molecular distance edge, and molecular fragment counts. The MOE descriptors include topological indices, structural keys, E-state indices, physical properties (i.e. LogP, molecular weight,

and molar refractivity), and topological polar surface area. There were over 4,000 Dragon descriptors (DRGN) initially generated but most of them were redundant. I removed redundant Dragon descriptors by using pairwise comparisons between each pair of descriptors. Eventually there were 1,052 Dragon descriptors left for the modeling set in this study. MOE generated 186 descriptors, all of which were used in the modeling process.

## Modeling Approaches

The generated descriptors were used in the machine learning algorithms Random Forest (RF),[34] Support Vector Machine (SVM),[35–38] and *k* nearest neighbor algorithm (*k*NN). [39] RF and SVM algorithms available in R.2.15.1 were used.[40] The *k*NN models were built using Chembench (chembench.mml.unc.edu).

RF is a machine learning algorithm that can be used for either classification or regression models first described by Leo Breiman.[34] Simply put, it involves the generation of decision trees based on input variables of a data set. The trees are grown by selecting with replacement N samples from a training set of N. With M variables for each data point, *m* variables are selected at random from M and the best split on these *m* are used. At prediction time, each new data point is pushed down each tree and the predicted output for each tree is averaged.

Support Vector Machine is a machine learning algorithm constructed by Vapnik, et al and is a popular method for classifying data.[35] It works by mapping the training set in *n*-dimensional space. A hyperplane is constructed that represents the largest separation between classes. Compounds are assigned activity, active (1) or inactive (-1). Using

optimization, a hyperplane is sought to find the separation of data according to classification with the largest marginalization possible.

The *k* Nearest Neighbor algorithm employs a classification and variable selection procedure.[41] During the building of the model, a random subset of descriptors are selected. This subset is set to different values and the models are developed with leave-one-out-cross validation. Here, each compound is eliminated from the training set and its activity on 5-HT6 is predicted as the average activity of *k* most similar molecules where the value of *k* is optimized as well (*k*=1-5). Similarity is characterized by Euclidean distance between compounds in multidimensional descriptor space.

## Modeling Workflow

Individual models were developed using Dragon or MOE descriptors and one of the modeling methods (RF, SVM, or *k*NN). The combination of descriptors and modeling methods resulted in seven individual models: RF_MOE, SVM_MOE, *k*NN_MOE, RF_DRGN, SVM_DRGN, *k*NN_DRGN. The predicted values for each compound obtained from all several individual models were averaged to generate a consensus prediction.

## Universal Statistical Figures of Merit for All Models

Since various modeling approaches and different descriptors were used in the modeling process, universal statistical metrics were needed to evaluate the performance of the models developed individually. The results were harmonized by 1) using sensitivity (percentage of active compounds predicted correctly), specificity (percentage

of inactive compounds predicted correctly), and CCR (correct classification rate or balanced accuracy). These parameters are defined as follows:

$$\%\ Sensitivity = \left(\frac{true\ positives}{true\ positives + false\ negatives}\right) 100 \qquad (1)$$

$$\%\ Specificity = \left(\frac{true\ negatives}{true\ negatives + false\ positives}\right) 100 \qquad (2)$$

$$\%\ CCR = \left(\frac{sensitivity + specificity}{2}\right) 100 \qquad (3)$$

**Results**

Overview of Dataset

The chemical space of all the compounds was analyzed by performing a Principle Component Analysis (PCA) of the chemical descriptor values used in this study. After the PCA with the 186 MOE descriptors for the compounds in both modeling and external validation sets, I selected the three most important components to generate a three-dimensional plot (**Figure 1**) for these 570 (488 modeling and 82 external set compounds) compounds. These three principal components represented around 56% of the variance in the database. This plot could be viewed as the chemical space covered by all the compounds used in this study.

Modeling Results

Six individual models and one consensus model was developed for 225 5-HT6 antagonists (or "actives") and 263 non-binders (or "inactives") (see experimental section). The consensus model was generated by averaging all individual model

predictions. The five-fold external cross validation results for all models are shown in **Figure 2a**. The sensitivity, specificity, and correct classification rate (CCR) values ranged from 89-99%, 94-98%, and 94-97%, respectively. All the individual models have similarly good performance when considering the cross validation results. The best models overall, by ranking of the CCR values, were those two models using *k*NN approach (*k*NN_MOE and *k*NN_DRGN models). The consensus model showed comparable or superior results to individual models with a CCR of 97% (**Figure 2a**).

The developed models were used to predict an external validation set of 82 compounds (38 actives, 44 inactives). These 82 compounds were collected after all the models were developed, so they are truly "unknown" compounds to the models. The prediction results are shown in **Figure 2b**. The sensitivity, specificity, and CCR values of individual models ranged from 66-97%, 71-91%, and 68-91%, respectively. It is noticeable that for most models the predictivity of the new compounds is lower than the cross-validation results. For example, the CCR of RF_MOE (the mode using random forest approach and MOE descriptors) has the CCR as 97% for the 5-fold cross validation, but the relevant CCR of external 82 compounds decreased to 78%, because of the poor prediction of actives. The major reasons for this difference are due to 1) substantial different chemical structures of new compounds compared to the modeling set, as shown in the different chemical space locations of the modeling set and new compounds in **Figure 1**; 2) the hypothesis of using anticancer drugs as non-binders may be incorrect for some compounds and it may cause false negative predictions. The first issue could be addressed when an applicability domain is correctly implemented with more experimental data available. Similarly, if experimental non-binders could be

obtained from 5-HT6 binding assays, the second issue could also be addressed. Compared to individual models, the consensus prediction still showed superior results of new 82 compounds (CCR=88%), which is also similar to the cross validation predictivity. Although the consensus model requires the development of various QSAR models using different approaches, the clear advantage of predictivity without selecting individual models makes consensus prediction more suitable for virtual screening purpose compared to using a single model.

**Discussion**

Interpretation of QSAR Models

I evaluated chemical features potentially related to 5-HT6 binding by analyzing the important descriptors that may be necessary for 5-HT6 binding. The Dragon descriptors for *k*NN models are chosen for this analysis because: 1) as a result of the stochastic variable selection procedure, the *k*NN approach maximizes the correlation between descriptors and the target bioactivity; and 2) the Dragon descriptors contain more detailed structural information than MOE descriptors. I reasoned that the analysis of the occurrence of various Dragon descriptors may be able to interpret their relative information contents with respect to the 5-HT6 binding potential. I ranked all the Dragon descriptors by their occurrences and analyzed the top 20, with frequencies ranging from 5% to 79% among all *k*NN models (**Figure 3**).

In order to gain insight into the potential mechanisms, I selected eight fragmental type descriptors out of these twenty Dragon descriptors for further analysis. The difference between the selected descriptor values in active and inactive compounds were

compared (**Table 1**). Two descriptors (F04[C-S] and MAXDN) had much higher values among active compounds than inactives, suggesting they contribute positively to 5-HT6 binding (**Table 1**). Specifically, the F04[C-S] descriptor represents the number of fragments in which a carbon is separated by 4 bonds from a sulfur atom. Besides the great difference between the average values of this descriptor in 5-HT6 antagonists and non-binders as shown in Table 1, I found that 91% of the actives containing this fragment, in comparison to 33% of the inactives. Sulfones and sulfonamides are two types of molecules containing this structural feature in the dataset. They are well-known 5HT6 antagonists in previous reports.[22,23] The other descriptor MAXDN is the maximal electrotopological negative variation and it represents the nucleophilicity of the molecules. Although it does not specifically refer to any substructures, it is relevant to the hydrogen bond formation, which was considered to be critical for 5-HT6 binding.[18] Another relevant descriptor is O-058, which refers to the number of double-bonded oxygen atoms (**Table 1**). This descriptor may still be relevant to the $SO_2$ functional group in sulfonamides and sulfones. Since this descriptor is defined very broadly and covers many other functional groups, the difference of descriptor values within actives and inactives is not significant (**Table 1**).

Another descriptor that contributes potentially to 5-HT6 binding is NRS, which represents the number of rings in the chemical structures. An aromatic ring is another validated substructural feature that is responsible for 5-HT6 binding, which is believed to interact with a phenylalanine group in the $6^{th}$ transmembrane segment of the 5-HT6 receptor.[18,42] The descriptor NRS was ranked high in the models and supported this

finding. However, it is clear that it is another broadly defined descriptor which is important in the model but does not significantly contribute to the 5-HT6 binding.

The other four descriptors, F03[N-N], nDB, TPSA and ndssC, all contribute to the inactives more than actives. They represent some general features of chemical cytotoxicity, such as polyaromatic cyclic hydrocarbons.[43–46] Due to the limitation of the current database (the use of cytotoxic compounds to define inactives), it is not that meaningful to discuss the effect of these features on 5-HT6 binding. However, it is interesting to notice that a di-amine system, represented by the descriptor F03[N-N], has almost equivalent impact on the actives and inactives. The positively ionizable group, such as amines, is considered to interact with the aspartic acid in the third transmembrane segment of the 5-HT6 receptor.[18] The contribution of this descriptor on 5-HT6 binding becomes obscure due to the cytotoxicity of compounds with amine groups.[47,48]

**Conclusions**

CADD is a burgeoning field in pharmaceutics with promising and far-reaching applications. The early stages of drug discovery have been shown to benefit greatly from the use of computer science and statistical methods. The quick and logical optimization of lead compounds by computer models has cut down drastically on this costly stage. The interest sparked by the most recently discovered serotonin receptor and its role in cognition has generated much attention in the development of novel binding agents and their possible exploitation for use in therapy of various behavioral disorders.

In this study, I compiled a modeling set containing 225 antagonists and 263 non-binders for the 5-HT6 receptor. This dataset was used to develop several various QSAR

models. The developed models were validated by using 5-fold cross validation approach and by predicting an external dataset of 82 compounds. The consensus model, which was the average of all seven individual models, shows advantages of predictivity, especially for new compounds. The models developed in this study will be useful to prioritize new compounds which are potential 5-HT6 antagonists for future experimental testing.

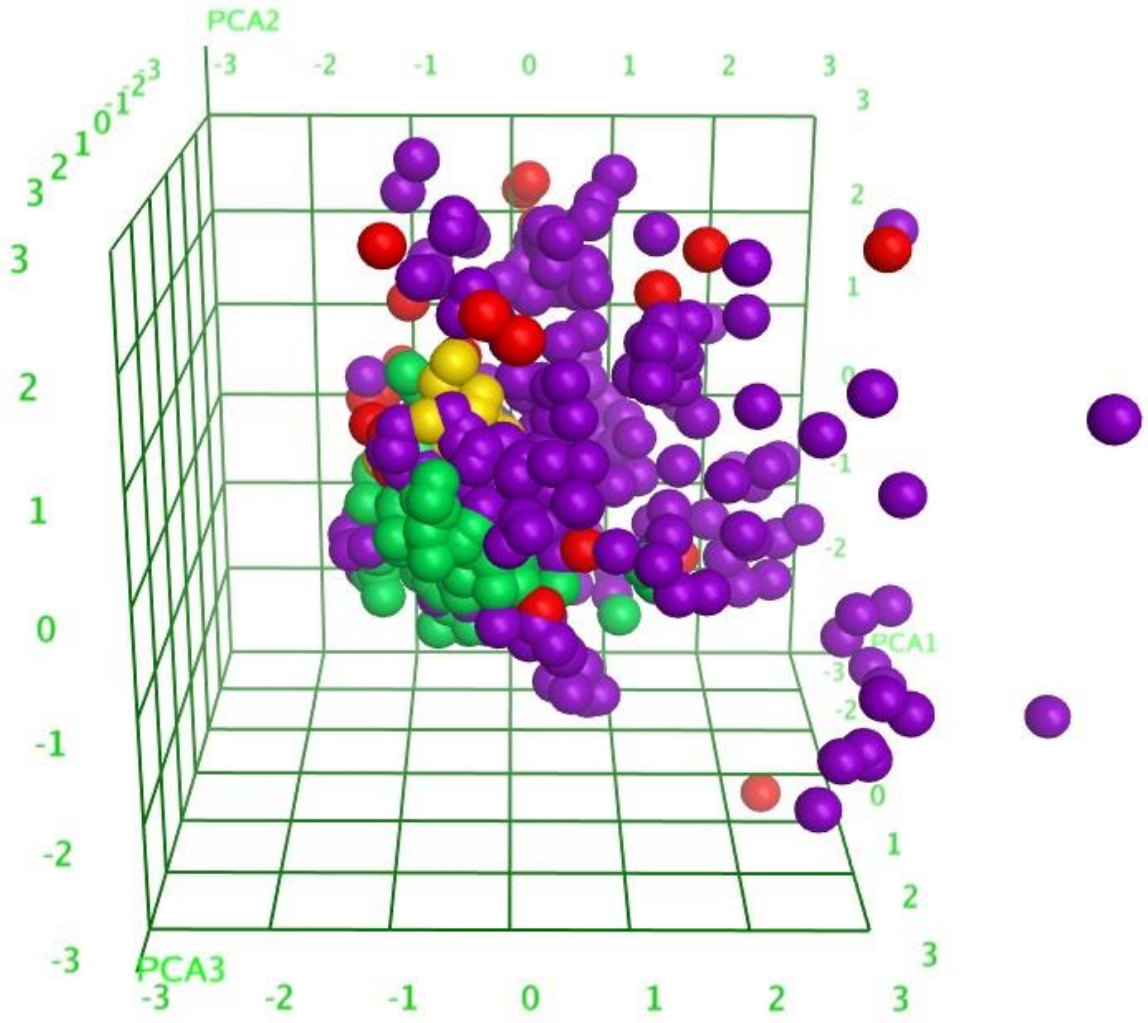
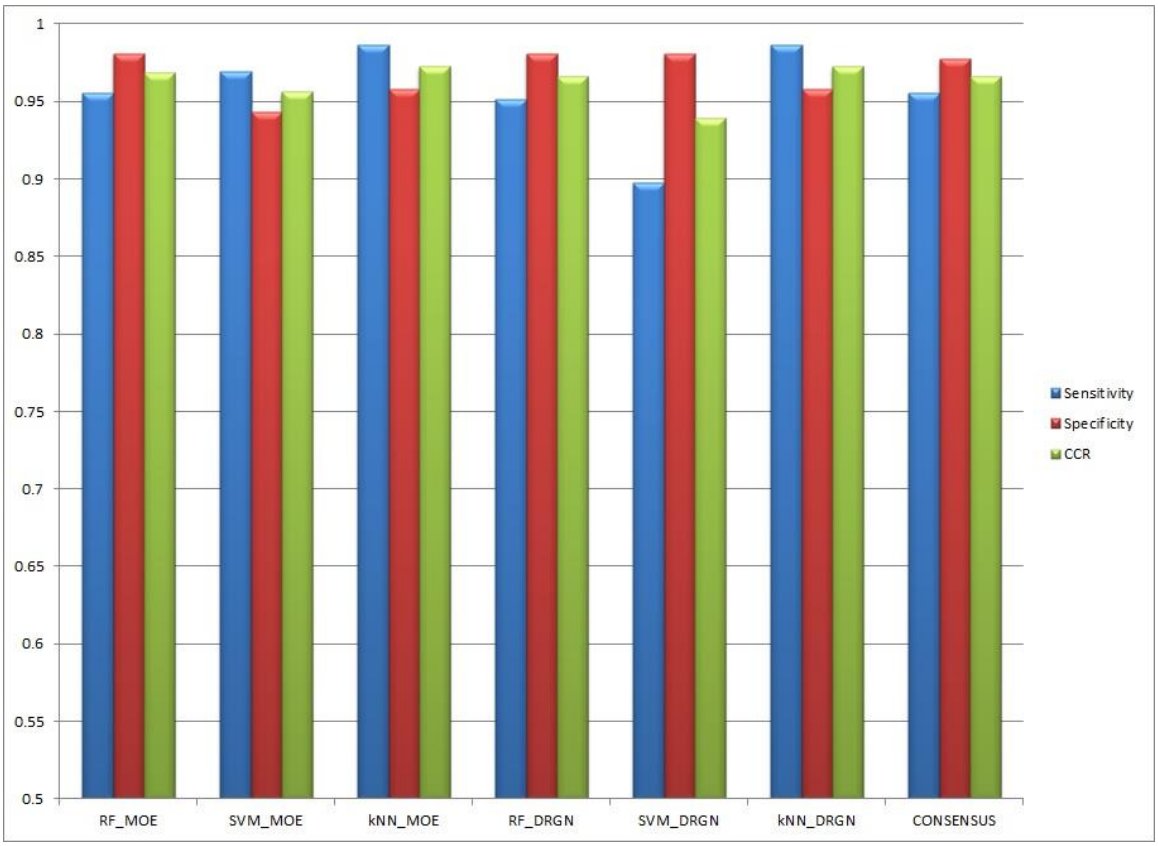
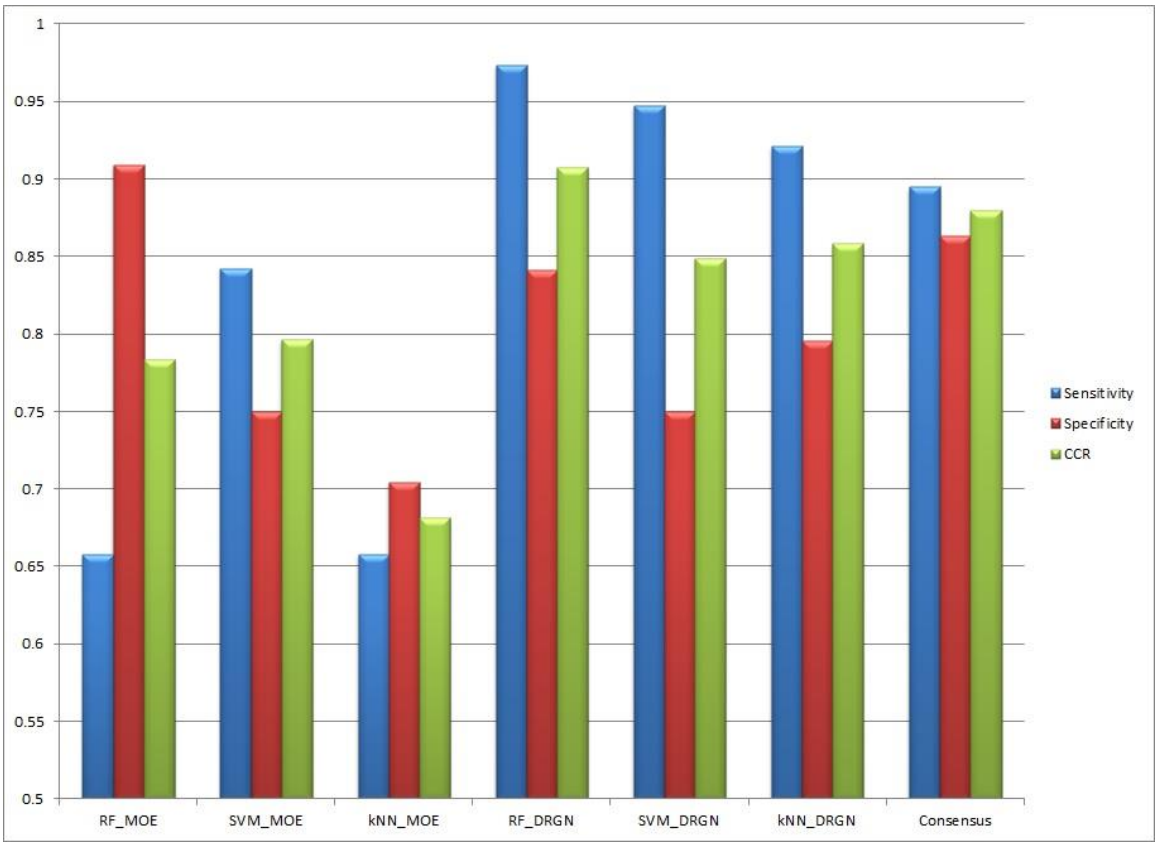



**Figure 1.** The chemical structure space of the modeling set (225 5-HT6 antagonists (green) and 263 non-binders (purple)) and the 82 compound external validation set (38 5-HT6 antagonists (yellow) and 42 non-binders (red)) using top 3 principal components of MOE descriptors.

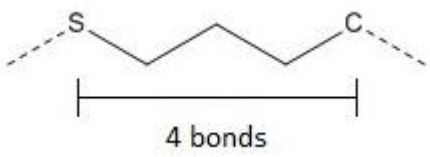
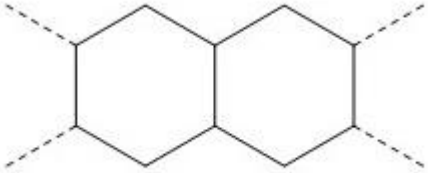
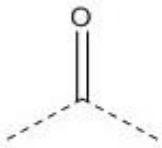
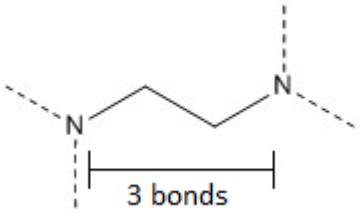

1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
RF_MOE
SVM_MOE
kNN_MOE
RF_DRGN
SVM_DRGN
kNN_DRGN
CONSENSUS
Sensitivity
Specificity
CCR

(a)




(b)

**Figure 2.** Statistics of seven individual and consensus models: (a) the 5-fold cross validation results; (b) the prediction results of 82 external validation set.

| Descriptor Name | Description | Illustration | Normalized average values | |
|---|---|---|---|---|
| | | | Actives | Inactives |
| F04[C-S] | Frequency of C-S at a topological distance of 4 |   | 0.42 | 0.08 |
| MAXDN | Maximal electrotopological negative variation | N/A | 0.70 | 0.43 |
| NRS | Number of ring systems |  | 0.39 | 0.31 |
| O-058 | Number of doubly bonded oxygen atoms |   | 0.31 | 0.29 |
| F03[N-N] | Frequency of N-N at a topological distance of 3 |   | 0.10 | 0.15 |

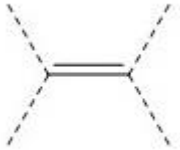
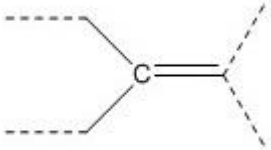
| | | | | |
|---|---|---|---|---|
| nDB | Number of double bonds |  | 0.18 | 0.25 |
| TPSA(Tot) | Topological polar surface area (TPSA) using Nitrogen, Oxygen, Phosphate, and Sulfur as polar contributors | N/A | 0.21 | 0.32 |
| ndssC | Number of atoms of type dssc |   | 0.01 | 0.20 |

**Table 1.** A few of most important Dragon descriptors used to develop 5-HT6 antagonist *k*NN QSAR models.

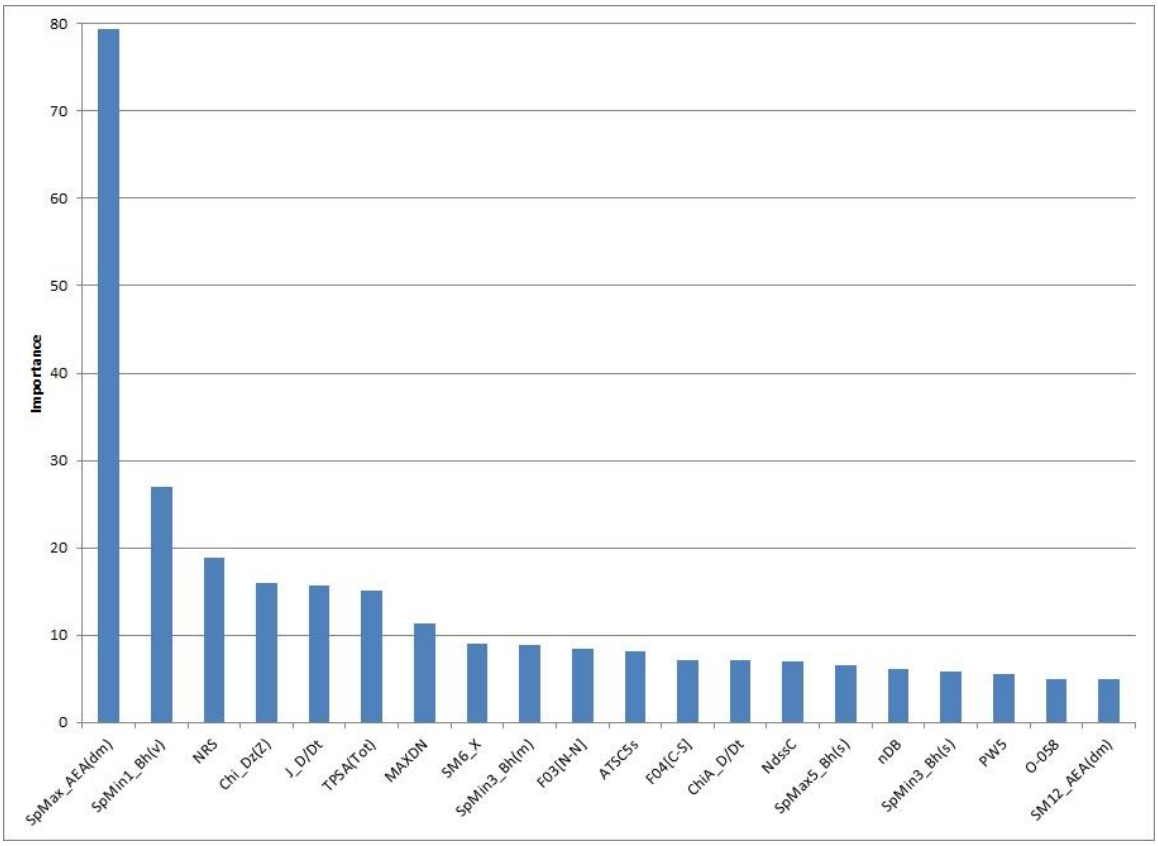



**Figure 3.** Top 20 most frequently used descriptors in *k*NN_DRGN model. Importance is the frequency of occurrence in acceptable models.

## References


1. Hoyer, D.; Hannon, J. P.; Martin, G. R. *Pharmacol. Biochem. Behav.* **2002**, *71*, 533–554.

2. Jones, B. J.; Blackburn, T. P. *Pharmacol. Biochem. Behav.* **2002**, *71*, 555–568.

3. Uphouse, L. *Neurosci. Biobehav. Rev.* **1997**, *21*, 679–698.

4. Murphy, P. M. *PharmacolRev* **2002**, *54*, 227–229.

5. Kohen, R.; Metcalf, M. A.; Khan, N.; Druck, T.; Huebner, K.; Lachowicz, J. E.; Meltzer, H. Y.; Sibley, D. R.; Roth, B. L.; Hamblin, M. W. *J. Neurochem.* **1996**, *66*, 47–56.

6. Monsma, Jr., F. J.; Shen, Y.; Ward, R. P.; Hablin, M. W.; Sibley, D. R. *Mol. Pharmacol.* **1992**, *42*, 320–327.

7. Hamon, M.; Ph, D.; Doucet, E.; Lefèvre, K.; Miquel, M.; Rio, J. Del; Vergé, D. *Neuropsychopharmacology* **1999**, *21*, 68S–76S.

8. Austin, H.; Sleight, J. *J. Pharmacol. Exp. Ther.* **1995**, *274*, 173–180.

9. Bentley, J.; Bourson, A. *Br. J. Pharmacol.* **1999**, *6*, 1537–1542.

10. Lacroix, L. P.; Dawson, L. A.; Hagan, J. J.; Heidbreder, C. A. *Synap. New York Ny* **2004**, *51*, 158–164.

11. Dawson, L. A.; Nguyen, H. Q.; Li, P. *Neuropsychopharmacology* **2001**, *25*, 662–668.

12. Rogers, D. C.; Hagan, J. J. *Psychopharmacology (Berl).* **2001**, *158*, 114–9.

13. Upton, N.; Chuang, T. T.; Hunter, A. J.; Virley, D. J. *Neurother. J. Am. Soc. Exp. Neurother.* **2008**, *5*, 458–469.

14. Branchek, T. a; Blackburn, T. P. *Annu. Rev. Rharmacology Roxicology* **2000**, *40*, 319–34.

15. Wesoowska, A. *Pharmacol. Reports* **2010**, *62*, 564–577.

16. Doddareddy, M. R.; Cho, Y. S.; Koh, H. Y.; Pae, A. N. *Bioorg. Med. Chem.* **2004**, *12*, 3977–85.

17. Doddareddy, M. R.; Lee, Y. J.; Cho, Y. S.; Choi, K. Il; Koh, H. Y.; Pae, A. N. *Bioorg. Med. Chem.* **2004**, *12*, 3815–24.

18. López-Rodríguez, M. L.; Benhamú, B.; De La Fuente, T.; Sanz, A.; Pardo, L.; Campillo, M. *J. Med. Chem.* **2005**, *48*, 4216–4219.

19. Goodarzi, M.; Freitas, M. P.; Ghasemi, N. *Eur. J. Med. Chem.* **2010**, *45*, 3911–5.

20. Sharma, B. K.; Singh, P.; Sarbhai, K.; Prabhakar, Y. S. *SAR QSAR Environ. Res.* **2010**, *3-4*, 369–388.

21. Hao, M.; Li, Y.; Li, H.; Zhang, S. *Int. J. Mol. Sci.* **2011**, *12*, 5011–30.

22. Ahmed, M.; Briggs, M. a; Bromidge, S. M.; Buck, T.; Campbell, L.; Deeks, N. J.; Garner, A.; Gordon, L.; Hamprecht, D. W.; Holland, V.; Johnson, C. N.; Medhurst, A. D.; Mitchell, D. J.; Moss, S. F.; Powles, J.; Seal, J. T.; Stean, T. O.; Stemp, G.; Thompson, M.; Trail, B.; Upton, N.; Winborn, K.; Witty, D. R. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4867–71.

23. Bard, J. a; Zgombick, J.; Adham, N.; Vaysse, P.; Branchek, T. a; Weinshank, R. L. *J. Biol. Chem.* **1993**, *268*, 23422–6.

24. Bromidge, S. M.; Griffith, K.; Heightman, T. D.; Jennings, a; King, F. D.; Moss, S. F.; Newman, H.; Riley, G.; Routledge, C.; Serafinowska, H. T.; Thomas, D. R. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2843–6.

25. Bromidge, S. M.; Clarke, S. E.; Gager, T.; Je, P.; Jennings, A. J.; Joiner, G. F.; King, F. D.; Lovell, P. J.; Moss, S. F.; Newman, H.; Riley, G.; Rogers, D.; Routledge, C.; Smith, D. R. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 55–58.

26. Glennon, R. a; Lee, M.; Rangisetty, J. B.; Dukat, M.; Roth, B. L.; Savage, J. E.; McBride, a; Rauser, L.; Hufeisen, S.; Lee, D. K. *J. Med. Chem.* **2000**, *43*, 1011–8.

27. Tsai, Y.; Dukat, M.; Slassi, A.; Maclean, N.; Demchyshyn, L.; Savage, J. E.; Roth, B. L.; Hufesein, S.; Lee, M.; Glennon, R. A. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2295–2299.

28. Ivachtchenko, A. V; Golovina, E. S.; Kadieva, M. G.; Kysil, V. M.; Mitkin, O. D.; Okun, I. M. *Pharm. Chem. J.* **2012**, *46*, 337–345.

29. Kalkman, H. O.; Subramanian, N.; Hoyer, D. *Neuropsychopharmacology* **2001**, *25*, 904–14.

30. Willett, P.; Barnard, J. M.; Downs, G. M.; Information, B. C.; Road, U.; Sheffield, S. **1998**, 983–996.

31. Liu, K. G.; Lo, J. R.; Comery, T. a; Zhang, G. M.; Zhang, J. Y.; Kowal, D. M.; Smith, D. L.; Di, L.; Kerns, E. H.; Schechter, L. E.; Robichaud, A. J. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 2413–5.

32. Mesquida, N.; López-Pérez, S.; Dinarès, I.; Frigola, J.; Mercè, R.; Holenz, J.; Pérez, R.; Burgueño, J.; Alcalde, E. *J. Med. Chem.* **2009**, *52*, 6153–7.

33. Sundar, B. G.; Bailey, T. R.; Dunn, D. D.; Bacon, E. R.; Salvino, J. M.; Morton, G. C.; Aimone, L. D.; Zeqi, H.; Mathiasen, J. R.; Dicamillo, A.; Huffman, M. J.; McKenna, B. a; Kopec, K.; Lu, L. D.; Brown, R.; Qian, J.; Angeles, T.; Connors, T.; Spais, C.; Holskin, B.; Galinis, D.; Duzic, E.; Schaffhauser, H.; Rosse, G. C. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 120–3.

34. Breiman, L. *Mach. Learn.* **2001**, *45*, 5–32.

35. Mukherjee, S.; Vapnik, V. *MIT* **1999**, *1653*.

36. Kovatcheva, A.; Golbraikh, A.; Oloff, S.; Xiao, Y.-D.; Zheng, W.; Wolschann, P.; Buchbauer, G.; Tropsha, A. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 582–95.

37. Kovatcheva, A.; Golbraikh, A.; Oloff, S.; Feng, J.; Zheng, W.; Tropsha, A. *SAR QSAR Environ. Res.* **2005**, *16*, 93–102.

38. Votano, J. R.; Parham, M.; Hall, L. M.; Hall, L. H.; Kier, L. B.; Oloff, S.; Tropsha, A. *J. Med. Chem.* **2006**, *49*, 7169–81.

39. Zheng, W.; Tropsha, A. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 185–194.

40. Dalgaard, P. *Introductory Statistics with R*; Springer: New York, **2008**.

41. Altman, N. S. *Am. Stat.* **1992**, *46*, 175–185.

42. Hayat, F.; Cho, S.; Rhim, H.; Indu Viswanath, A. N.; Pae, A. N.; Lee, J. Y.; Choo, D. J.; Choo, H.-Y. P. *Bioorg. Med. Chem.* **2013**, *21*, 5573–82.

43. Samanta, S. K.; Singh, O. V.; Jain, R. K. Polycyclic aromatic hydrocarbons: Environmental pollution and bioremediation. *Trends Biotechnol.* **2002**, *20*, 243–248.

44. Ramesh, A.; Walker, S. A.; Hood, D. B.; Guillén, M. D.; Schneider, K.; Weyand, E. H. *Int. J. Toxicol.* **2004**, *23*, 301–333.

45. Sikkema, J.; de Bont, J. A.; Poolman, B. *Microbiol. Rev.* **1995**, *59*, 201–222.

46. Flowers-Geary, L.; Bleczinski, W.; Harvey, R. G.; Penning, T. M. *Chem. Biol. Interact.* **1996**, *99*, 55–72.

47. Ambs, S.; Neumann, H. G. *Toxicol Appl Pharmacol* **1996**, *139*, 186–194.

48. Benigni, R.; Passerini, L. Carcinogenicity of the aromatic amines: From structure-activity relationships to mechanisms of action and risk assessment. *Mutat. Res. - Rev. Mutat. Res.* **2002**, *511*, 191–206.