



Kent Academic Repository

da Silva, Aldineia P., Chiari, Laise P.A., Guimaraes, Amanda R., Honorio, Kathia M. and da Silva, Albérico B.F. (2021) *Drug design of new 5-HT₆R antagonists aided by artificial neural networks*. *Journal of Molecular Graphics and Modelling*, 104 . p. 107844. ISSN 1093-3263.

Downloaded from

<https://kar.kent.ac.uk/115838/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1016/j.jmgm.2021.107844>

This document version

Publisher pdf

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

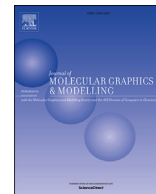
If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).



Drug design of new 5-HT₆R antagonists aided by artificial neural networks

Aldineia P. da Silva ^a, Laise P.A. Chiari ^a, Amanda R. Guimaraes ^a, Kathia M. Honorio ^b,
Albérico B.F. da Silva ^{a,*}

^a Departamento de Química e Física Molecular, Instituto de Química de São Carlos, Universidade de São Paulo, C.P 780, 13560-970, São Carlos, SP, Brazil

^b Escola de Artes, Ciência e Humanidades, Universidade de São Paulo, São Paulo, 03828-000, Brazil

ARTICLE INFO

Article history:

Received 13 August 2020

Received in revised form

6 January 2021

Accepted 8 January 2021

Available online 18 January 2021

Keywords:

Alzheimer's disease

5-HT₆ receptor

Drug-design

Molecular modeling

ANN

ABSTRACT

Alzheimer's Disease (AD) is the most frequent illness and cause of death amongst the age related-neurodegenerative disorders. The Alzheimer's Disease International (ADI) reported in 2019 that over 50 million people were living with dementia in the world and this number could potentially be around 152 million by 2050. 5-hydroxytryptamine subtype 6 receptor (5-HT₆R) has been identified as a potential anti-amnesic drug target and therefore, the administration of 5-HT₆R antagonists can likely mitigate the memory loss and intellectual deterioration associated with AD. Herein, computational tools were applied to design new 5-HT₆ antagonists and their biological activity values were predicted by our QSAR model obtained from Artificial Neural Networks (ANN). The proposed compounds here from the QSAR-ANN model presented significant biological activity values and some of them have achieved pK_i above 9.00. Furthermore, our results suggest that the presence of halogen atoms (especially bromine) linked to the aromatic ring at *para*-position (HYD) contribute considerably to the increase of the biological activity values while bulky groups in the PI position do not culminate with the increase antagonist activity of compounds here analyzed. Finally, the ADME/Tox profile as well as the synthetic accessibility of new proposed compounds qualify them to go on further with experimental procedures and thenceforward their antagonist effects can be confirmed.

© 2021 Published by Elsevier Inc.

1. Introduction

As a consequence of the increased life expectancy and aging population worldwide, the number of elderly individuals with neurodegenerative disorders grown-up exponentially [1]. According to the organization "Alzheimer's Disease International", (ADI) [2], it has been estimated that more than 50 million people worldwide are suffering from Alzheimer's disease, and this number could potentially be around 152 million by 2050.

The clinical and socioeconomic impacts are colossal, as dementia costs are more than US\$ 1 trillion, and until 2030 this number will likely double [3]. Worldwide, between 2000 and 2016, deaths from some type of dementia had increased more than twice, rising from the 14th to the 5th cause of death compared with other diseases such as ischemic heart disease, chronic obstructive pulmonary disease, intracerebral hemorrhage, and ischemic stroke [4].

Among the age related-neurodegenerative disorders, Alzheimer's disease (AD) is not only the most frequent as also the leading cause of death among elderlies [5–7]. The losses from AD are associated with progressive and chronic deterioration of cognitive functions, i.e., memory, speech disruption, space-time disorientation, forgetfulness, logic reasoning, focus, learning, change in mood and personality as well diffculted with everyday activities [8].

A healthy human brain has about 100 billion neurons, 100 trillion synapses, and several neurotransmitters. Brain's neuronal circuits interact by synapses leading to release neurotransmitters, which are in charge of carrying chemical information to other nerve cells [9]. The way neurons connect to each other leads us to understand how activity patterns are formed at the cellular level, such as learning one new language, learning math, chemistry, physics, and other ordinary activities like eating, walking, talking, dress up, among others. However, for people stricken by AD, the number of synapses fall gradually as the disease progresses, impairing the practice of the activities above mentioned [10–12].

* Corresponding author.

E-mail address: alberico@iqsc.usp.br (A.B.F. da Silva).

Hitherto, acetylcholinesterase (AChE) inhibitors (donepezil, rivastigmine, and galantamine) [13,14] are the therapeutic approaches administrated to people affected by AD. These pharmacological therapies are applied to decrease the degradation rate of acetylcholine, leading to an increase in the concentration of acetylcholine in the brain resisting the cholinergic deficit associated with cognitive impairment from AD. However, there are several limitations linked to these acetylcholinesterase inhibitors, namely peripheral side effects (nausea, vomiting, diarrhea and tremors), and the effectiveness to a small amount of diseased people that are mostly in mild to moderate stages of AD, and reasonable rewards in most people that are in advanced phases of the disease [10]. To overcome this drawback, it is necessary new alternative therapies for the treatment of Alzheimer's disease.

Addressing this problem, the 5-hydroxytryptamine subtype 6 (5-HT₆) receptor has been called attention to the scientific community as a potential anti-amnesic drug target [15]. Since it is mainly expressed in the central nervous system (CNS), assorted studies [9–12] singled out that this receptor is involved in the learning and memory processes, and the administration of 5-HT₆R antagonists improve ones.

In the field of the medicinal chemistry, computational and experimental strategies generate important information in order to identify new potential drug candidates. In modern pharmaceutical industries and universities, one of the computational tools employed in drug design is the Quantitative Structure-Activity Relationship (QSAR) [16–23]. This method, along with other methodologies such as pharmacophore-based [23–26] and homology modeling [27], allowed to identify the leading characteristic of candidates to 5-HT₆R antagonists (Fig. 1), namely:

- > a positive ionizable atom (PI);
- > a hydrogen bond acceptor group (HBA);
- > a hydrophobic site (HYD);
- > an aromatic ring-hydrophobic site (AR)

Usually, the PI group is either a piperazine ring or similar groups, HBA is represented by a sulfonamide or sulfone moiety, HYD is figured by fused hydrophobic aromatic rings (naphthalene or halogen-substituted benzene, benzothiophene, or indole). Whereas the AR group is represented by aromatic heterocyclic rings (pyridine, benzimidazole, or quinoline). These are the key structural characteristics desirable for the inactivation of the 5-HT₆ receptor as set out by López-Rodríguez and co-workers [24].

Some researches [10,27] have singled out that the interaction of the AR moiety with 5-HT₆ receptor eschew the inward movement of the transmembrane domain (TM)6 in direction to TM3, inactivating it. Likewise, the findings [10,27] have indicated that the AR moiety also interacts with the extracellular loop 2 (EL), resulting in blocking both EL2 and TM6 in the inactive conformation. The HYD moiety prevents the movements of TM3 and TM5, helices that activate the 5-HT₆ receptor. The PI (usually piperazine) group interacts with aspartate of TM3, while the HBA moiety prevents the conformational transition of TM3 through of the cysteine toward TM7 throughout the process of 5-HT₆ receptor activation (Fig. 1).

Using computational tools, several groups of researches [28–30] had dedicated to studies on the 5-HT₆ receptor. These groups were limited to construct QSAR models and to analyze the interaction of this receptor with their antagonists. We took a step further and new compounds were designed, for which biological activities were predicted by our QSAR model using the partial least squares (PLS)

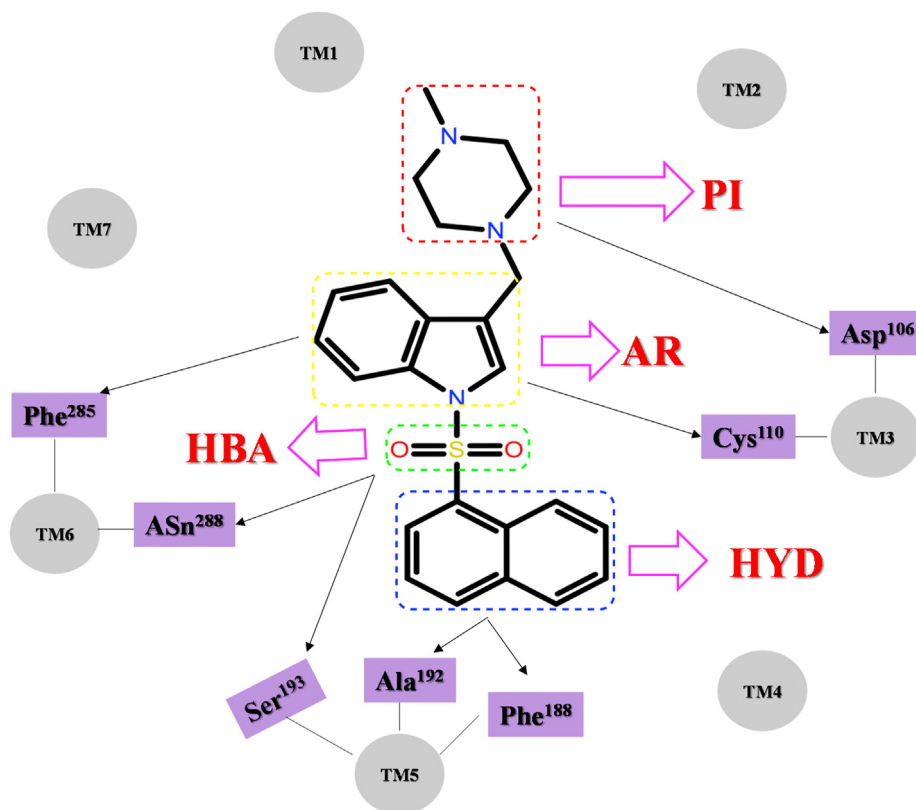


Fig. 1. Leading characteristic of candidates as 5-HT₆R antagonists - Red: positive ionizable atom (PI), Yellow: aromatic ring-hydrophobic site (AR), Green: hydrogen bond acceptor group (HBA), and Blue: hydrophobic site (HYD). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

method, as reported recently [31]. Herein, in addition to our last study [31], and to large the number of potential antagonists of this receptor, we used Artificial Neural Networks (ANN) to understand the structure-activity of a series of compounds with activity against the 5-HT₆ receptor taking into account some structural characteristics presented by López-Rodríguez and co-workers [24]. Next, we designed new candidates as anti-amnesic compounds, whose biological activity values were predicted by the ANN model.

2. Materials and methods

2.1. Data set

81 compounds were selected from the literature [32] taking into account the characteristics described in Fig. 1. Also, the compounds set presents structural diversity and biological activity over a range of values represented by their inhibitory constants (pK_i), as illustrated in Table 1.

The structures of the selected compounds (Table 1) were optimized at B3LYP/631-G (d,p) [33,34] using the Gaussian 09 software package [35]. Then, from the optimized structures, the electronic descriptors were extracted and molecular descriptors were obtained using the E-dragon platform [36]. Next, in the first screening, the descriptors were selected by the correlation coefficients of Pearson-R (Eq. (1)) between each descriptor and the pK_i values. The correlation coefficient value indicates the degree of linearity between the descriptors and the pK_i values for each compound. The descriptors, whose coefficient values were less than |0.3| in modulus, were excluded.

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

where x_i represents the descriptor values, y_i refers to the biological activity values, and $\bar{x}$ and $\bar{y}$ are arithmetic means of both x and y variables.

Even on, a plethora of descriptors remained, for which we employed the Genetic Algorithm (GA) method [37] as implemented in BuildQSAR [38] to narrow down the descriptors most correlated with the biological activity. Finally, we constructed our Artificial Neural Networks (ANN) model using MATLAB R2015a [39].

3. Theory

3.1. Genetic algorithm

One of the precautions in the construction of multivariate models is the stage of descriptor selection. It is necessary to ensure that the descriptors are the most relevant to explain the structural variation of compounds before their biological activity. Genetic Algorithm (GA) is a computational search method that brings light to this task, besides being one of the most useful in the drug design field [40].

The GA method is based on evolutionary concepts such as selection, crossover and mutation in which copies of the fittest individuals are reproduced in the next generation [37]. In the drug design field, these concepts are employed (amongst other applications) to select molecular and electronic descriptors, which will later be used in the construction of statistical models, such as QSAR-ANN ones.

Along the time, the population develops itself in nature to assure their survival that briefly occurs through three phases: a crossover to continue and improve the quality of the individuals from the same species; natural selection to choose the most

adapted individuals; genetic mutation to produce new species. Such processes are based on Darwin's evolution theory, which inspired the development of the GA method in 1960 by John Holland [37].

The GA technique generates one population with possible solutions for a given problem (initial population) and, subsequently, this population is submitted to an evolution process, as illustrated in Fig. 2, where the results are evaluated by the analysis of the statistical parameters in which the fittest individuals are selected.

In the stage of descriptors selection to the model construction, the search procedure operated by the genetic algorithm method works as follows (Fig. 3): (1) it is needed one N set of regression equations named as first-generation (**G1**) with N variables; (2) each equation (**E**) represents one individual (For example, **E**_{1,1} means individual 1 from the G1); (3) each individual has two genes (variables); (4) the sum these variables represent one chromosome.

The assessment consists in analyzing each R-value ($R_{1,1}$ to $R_{1,4}$ - Fig. 3) for each regression equation. The best R-value is equals to 0.89, so this value is considered the most promising solution so far. It is seen that the most adapted individuals (first three individuals, $R_{1,2} = 0.88$, $R_{1,2} = 0.89$ and $R_{1,3} = 0.84$, respectively) move on to the next stage and then the fourth individual is randomly selected [41].

In the next phase, there is the crossing among genes form of one or more individual's pairs to that the new ones inherit from their parent's genes (the crossing has performed between individuals with the worst R values, $R_{1,3'}$ and $R_{1,4'}$). In the stage of mutation, one or more variables of any individual presented in the G1'' generation is replaced by another that does not exist among these that remained. To ensure that there has been an evolution, the sum $R_{2,1}$, $R_{2,2}$, $R_{2,3}$, and $R_{2,4}$ values ($\sum R_2 = 3.64$) from the G2 generation must be greater than $R_{1,1''}$, $R_{1,2''}$, $R_{1,3''}$, and $R_{1,4''}$ values ($\sum R_1 = 3.14$) from the G1'' generation [41]. Finally, after the descriptor selection, the multivariate model is constructed.

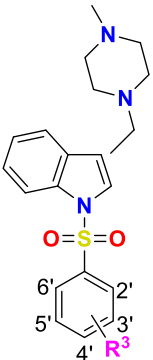
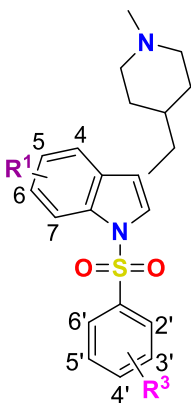
3.2. Artificial neural networks

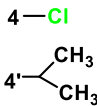
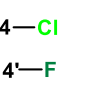
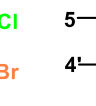
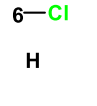
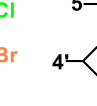
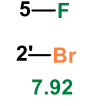
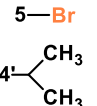
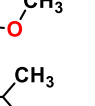
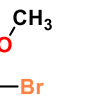
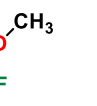
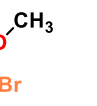
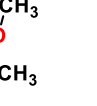
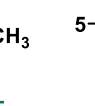
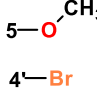
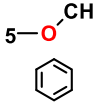
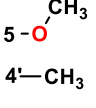
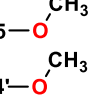
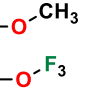
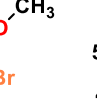
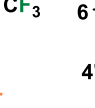
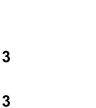
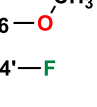
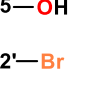
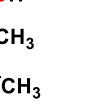
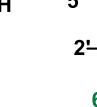
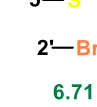
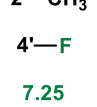
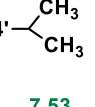
Machine learning is one of the computational approaches that provide results most rapid and more efficient with low computational resources in the field of computer-aided drug discovery [42,43]. The relationship between chemical structure and biological activity of the compounds is processed using pattern recognition algorithms and so extrapolation to predict the pK_i value of new compounds, as illustrated in this study. The conventional QSAR models, namely Hansch analysis, use basic linear regression models to correlate dependent variable (experimental property, for example, pK_i) with independent variables (chemical descriptor values) although groundbreaking and efficient, these models are limited since that the relationship between these variables is not always linear. Machine learning such as artificial neural networks (ANN) addresses this drawback, modeling nonlinear datasets [44].

The artificial neural networks method has been broadly employed to explain properties of small molecules [45] and it is sorted as supervised learning [43], in which the labels (pK_i) are attributed to the training data and, once trained, the constructed model is employed to predict labels for new data inputs. ANN method is a mathematical model inspired by operations of the biological neuron [44,46,47] where the propagation signals are activated by neurotransmission. Basically, each neuron receives several inputs signals (similar to dendrites) and executes them using the weighted sum that passes through a transfer function (alike to cell body) to produce the output signals to the next neurons layer (similar to axons), as we see in Fig. 4.

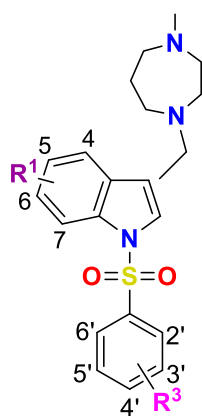
ANNs learn the relationship between input (descriptor) and output (biological property, pK_i) data by experience using a given learning algorithm. Summarily, the network learns as follows:

Table 1Structure and biological activity values of the 5-HT₆ antagonists selected to construct the ANN model (the test set is underlined and the validation set is in blue).

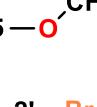
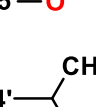
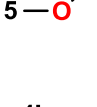
Compounds 1 to 23							
	1	2	3	4	5	6	7
	H	2'-Br	3'-Br	4'-Br	3'-Cl	2'-F	3'-F
	8.22	7.99	7.72	8.67	7.79	8.11	7.24
	8	9	10	11	12	13	
	4'-F	4'-CH ₃	4'-CH ₃	3'-O-CH ₃	4'-O-CH ₃	4'-O-CF ₃	
	8.06	8.22	8.02	7.53	7.60	6.82	
	14	15	16	17	18	19	20
	4'-C≡N	2'-F	3'-F	2'-F	2'-Cl	2'-Cl	2'-Cl
	7.04	7.52	6.92	7.19	7.55	7.00	7.54
Activity range=6.82 to 8.67		21	22	23			
		2', 5'-O-CH ₃	2', 5'-CH ₃				
		7.04	7.19	7.94			
Compounds 24 to 57							
							
Activity range=6.62 to 8.79							

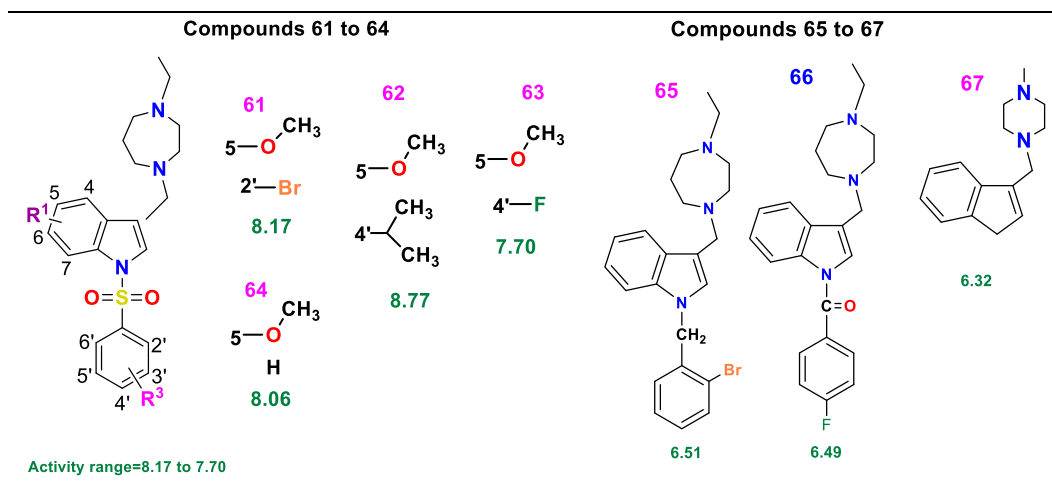
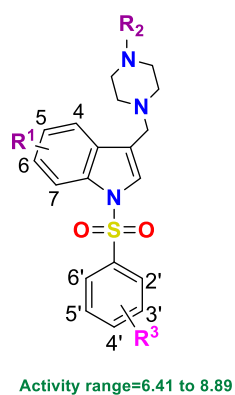
24  8.04	25  7.51	26  7.63	27  7.76	28  7.45	29  7.22	30  8.40	31  7.92	32  8.04	33 7.77
34  7.56	35  7.94	36  7.84	37  7.74	38  8.70	39  8.79	40  8.30	41  8.37		
42  7.99	43  8.11	44  7.81	45  7.74	46  7.42	47  7.44	48  7.02	49  7.44		
50  6.62	51  7.73	52  7.30	53  7.46	54  6.70	55  6.71	56  7.25	57  7.53		

Compounds 58 to 60



Activity range=8.11 to 8.70

58  8.24	59  8.70	60  8.11
--	---	--

**Compounds 65 to 67**

	R ₁	R ₂	R ₃	pK _i
68	H	H	H	8.12
69	H	2'-Br	H	8.35
70	5-O-CH ₃	H	2'-Br	8.89
71	5-O-CH ₃	H	4'-CH ₃ CH ₃	8.24
72	5-O-CH ₃	H	4'-F	8.16
73	H		2'-Br	6.44
74	H		2'-Br	6.41
75	H		2'-Br	6.88
76	5-O-CH ₃	H ₂ C-F	-CH ₂	6.46

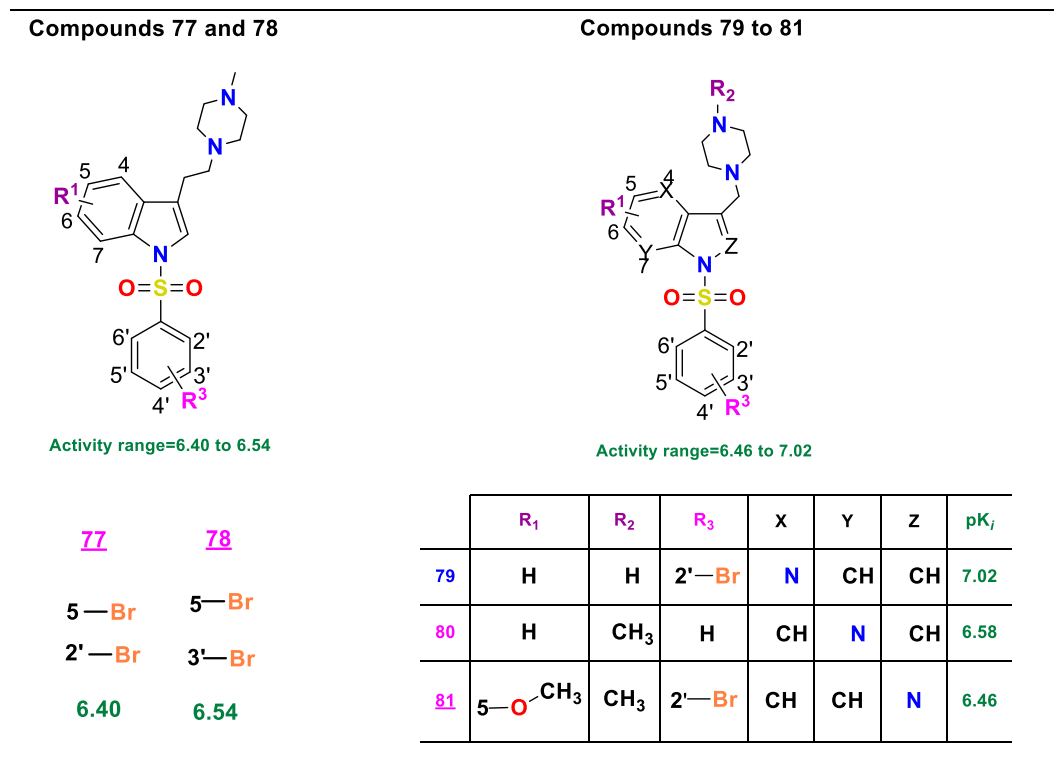


Fig. 2. General representation of the GA method.

neural activity is established by synaptic weights that in turn determine the expected answer (pK_i), and finally, through of the obtained answer we can determine the network's error. If errors are not suitable, we can reduce them by executing a few minor alterations in the synaptic weights and then, we can train the neural network again to obtain the predictive power model [48,49]. The weight update is performed making use of Eq. (2).

$$\Delta W_{ij,n} = F_n + \alpha W_{ij,n-1} \quad (2)$$

wherein ΔW_{ij} is the modification in the weight factor for each network node, α is the momentum factor, and F is a weight update function, which the role is to inform how weights are altered throughout the apprenticeship [50]. In this work, the Levenberg-Marquardt (L-M) algorithm was used as a weight update function, which stops automatically the network training when a significant increase in the error is observed [51].

3.3. Data normalization

To ensure that the ANN-QSAR model to be generated has satisfactory accuracy, it is necessary that the data set (dependent and independent variables) were normalized. This procedure prevents the algorithm used in the model generation provides biased results towards samples with higher magnitude order. So, to that data be comparable, we normalized them between 0 and 1 (Eq. (3)).

$$X'_i = \frac{(X_i - X_{min})}{(X_{max} - X_{min})} \quad (3)$$

where X'_i is the normalized value, X_{max} and X_{min} are maximum and minimum input values of the dataset.

3.4. Model validation

The most important step to construct a multivariate model is its validation. In this work, the leave-N-out cross-validation procedure was employed [52,53], in which N samples were left out of the model construction, and then the biological activity of these samples was predicted by our model (using the ANN technique, this procedure is performed through the validation set). Another

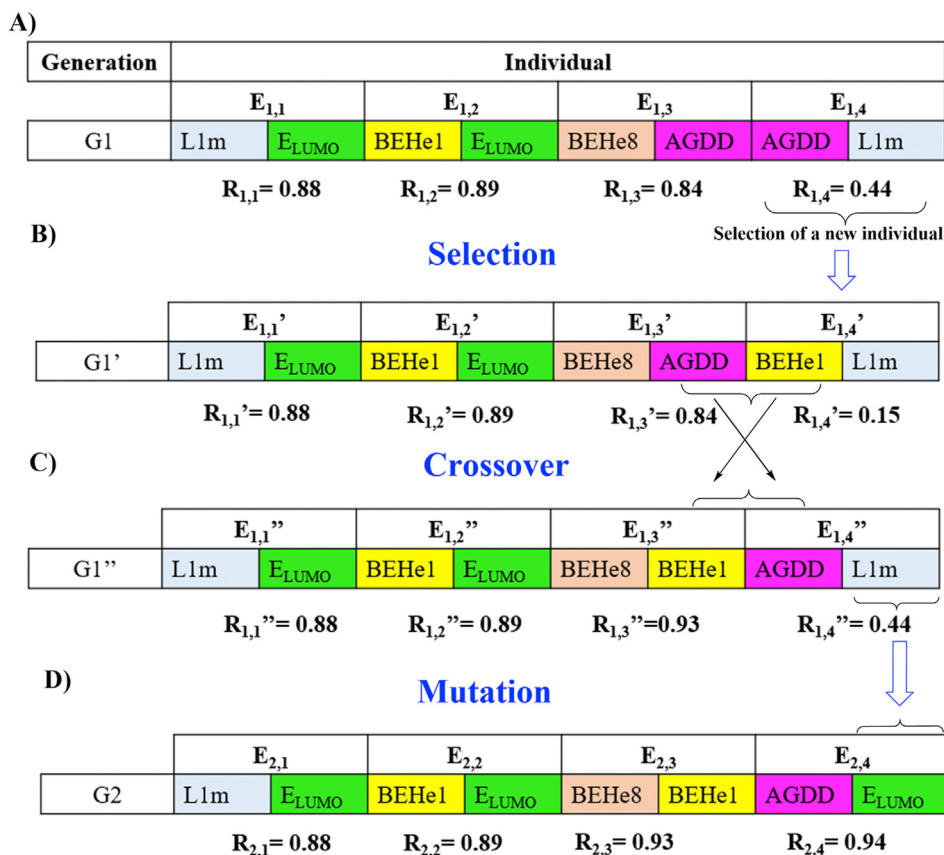


Fig. 3. Example of the search procedure operated by the GA method: (A) initial population; (B) natural selection; (C) crossover; and (D) mutation. Figure adapted from Ref. [41].

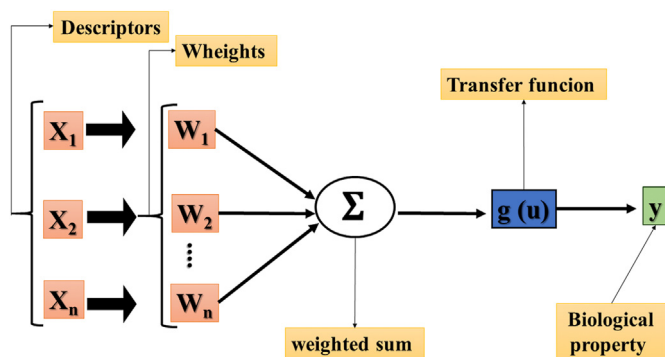


Fig. 4. Simplified diagram for the operation involving the ANN technique.

approach applied was the Y-randomization test [54] to ensure that was no haphazard correlation between the biological activity and the descriptor values. This test consists of scrambling the activity values for being assigned to other samples resulting in poor statistical models. Finally, we used a test set as the ultimate proof to attest the predictive power of our model.

4. Results and discussion

4.1. Descriptor selection

1250 descriptors were calculated (for example, geometrics, topological, 2D autocorrelation, and electronic descriptors) which were submitted to the Pearson correlation matrix to a pre-

screening and thus to cut off the less correlated descriptors with the biological activity. From this screening, 181 descriptors remained, and, for these, the genetic algorithm method was utilized to select those ones that most contributed to the biological activity. Next, we constructed 200 models varying the main statistical parameters, such as number of generations (600–1000), number of descriptors (5–9), multiple determination coefficient (r), standard deviation (S), F test (F), and cross-validation correlation coefficient (q^2). From these analyses, we selected the following descriptors (Table 2):

The generation (G_i) equal to 600 means the number of times that GA searched the best descriptors. The 0.9 criterion indicates that the descriptors more correlated with one another were excluded. Table 3 presents the categorization for each descriptor and their definitions (the values of the descriptors are displayed in Supplementary Material).

4.2. Construction of the QSAR-ANN model and validation

To construct the ANN-QSAR model, we used the network type Multilayer Perceptron (MLP) with the architecture 8-20-1 (8 neurons in the input layer, 20 neurons in the hidden layer, and 1 neuron in the output layer), *logsin-tansig* as transfer functions, and the Levenberg-Marquardt as the training algorithm. Next, the network was divided into three sets: training (70%), test (20%) and validation (10%). Besides the three sets, one model including all compounds was also generated and its R^2 value was compared to the R^2 value from the training set. As we can see (Fig. 5) both values are close, meeting the requirement proposed by the good practice in QSAR studies [53].

Table 2
Statistical parameters analyzed to select the main descriptors by GA.

G_i	Criterion	Selected descriptors	r	S	F	q^2
600	0.9	JhetZ, BEHm2, GATS3p, GATS1p, RDF025u, Au, H3v, $E_{HOMO(-2)}$	0.77	0.45	13.13	0.52

Table 3
Description of the selected descriptors.

Descriptor	Class	Definition
JhetZ	Topological	Balaban-type index from Z/weighted distance matrix
BEHm2	BCUT (Burden- CAS University of Texas Eigen values)	Highest eigenvalue n-2 of Burden matrix/weighted by atomic masses.
GATS3p	2D Autocorrelation	Geary autocorrelations- lag 3/weighted by atomic polarizabilities.
GATS1p		Geary autocorrelations- lag 1/weighted by atomic polarizabilities.
RDF025u	RDF (Radial Function Distribution)	Radial distribution function –25/unweighted
Au	WHIN (Weighted Holistic Invariant Molecular)	A total size index/unweighted
H3v	GETAWAY (Geometric Topology and Atom Weights Assembly)	H autocorrelation of lag3/weighted by atomic van der Waals volumes
$E_{HOMO(-2)}$	Quantum	Energy of the $HOMO_{(-2)}$ orbital

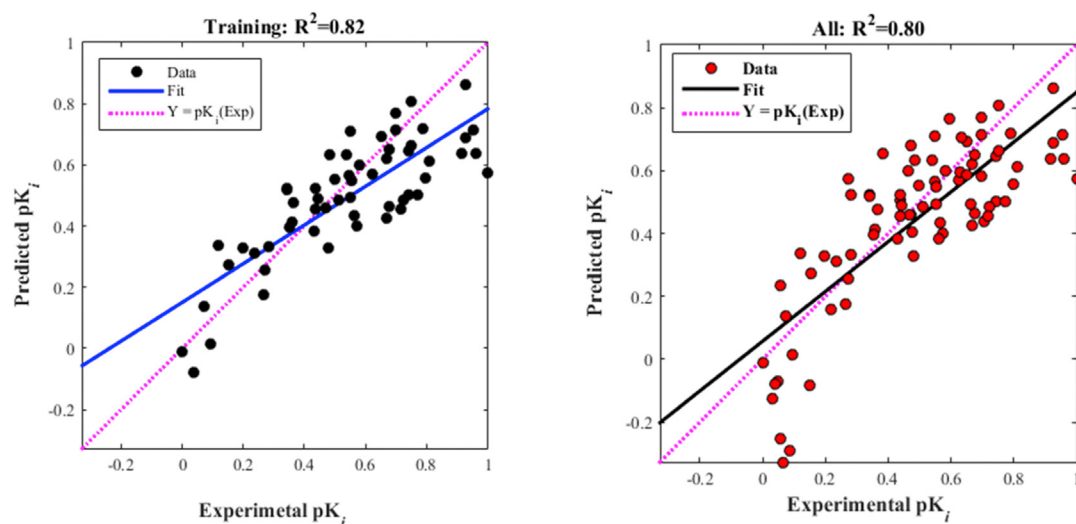


Fig. 5. R^2 values for the training and full sets.

From Fig. 5 we can see that the correlation coefficient (R^2) values for both models explain 82 (training set) and 80% (full set) information of the original data. The dashed line (in magenta color) represents the ideal results, in which the predicted pK_i values are equal to the experimental pK_i values, while the blue (training set) and black (full set) lines illustrate the best fit found by our neural network.

To ensure the robustness, the predictive ability, and the goodness-of-fit of our model, two other sets were generated (test and validation sets- Fig. 6), which presented the $R^2_{\text{Test}} = 0.82$ and $R^2_{\text{Validation}} = 0.81$ values. It is important to observe that both R^2 values (full compounds, training, test and validation sets - Figs. 5 and 6) remained close to each other, indicating the model significance. The test set is used as the ultimate proof of the model predictive power, and the validation set is necessary for certifying that the neural network has no overfitting. We can argue that this set (validation) can also act as a cross-validation test since the errors obtained from this procedure are used to estimate the model predictive capacity as well as its robustness. Moreover, from the error analyses, the network can be trained again to obtain more promising results.

The neural network performance was assessed through Mean Squared Error (MSE) for the training, test, and validation sets (Fig. 7). We emphasize that having an exacerbated MSE for the

validation set, the Levenberg-Marquardt algorithm automatically halts the network training, indicating that the network must be retrained.

The green circle shows the best neural network performance, where the minimum MSE was reached. As expected, for an ANN model to be considered robust, the error regarding the training set must be below the error from other subsets. Also, if the error from the test set increases significantly when compared to the validation set error, it is possible that overfitting in the model occurred. As illustrated in Fig. 7, this assumption was not observed, indicating the obtained model is trustworthy.

Table 4 contains the predicted pK_i values for the test set and the associated errors. It is worth mentioning that the vast majority of the compounds exhibits errors of less than 5%, except the compound **78** (5.81%). In the face of the small errors, we confirmed again the model robustness.

It was also analyzed if there are outlier samples, and the results are displayed in Fig. 8. As we can see in Fig. 8 none sample with anomaly behavior was detected (all results for this analysis are presented in Supplementary Material). So, this is another test that certifies the model predictive ability.

We performed the Y-randomization test from shuffling the pK_i values by 80-fold and excluding one sample (leave-one-out) for each randomization. This test shows the randomized model with

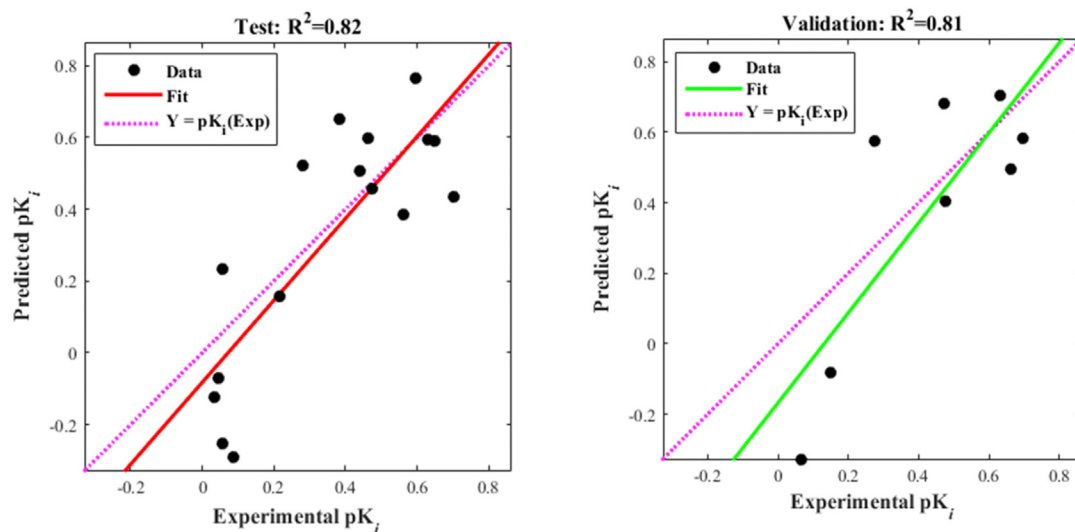


Fig. 6. R^2 values for the test and validation sets.

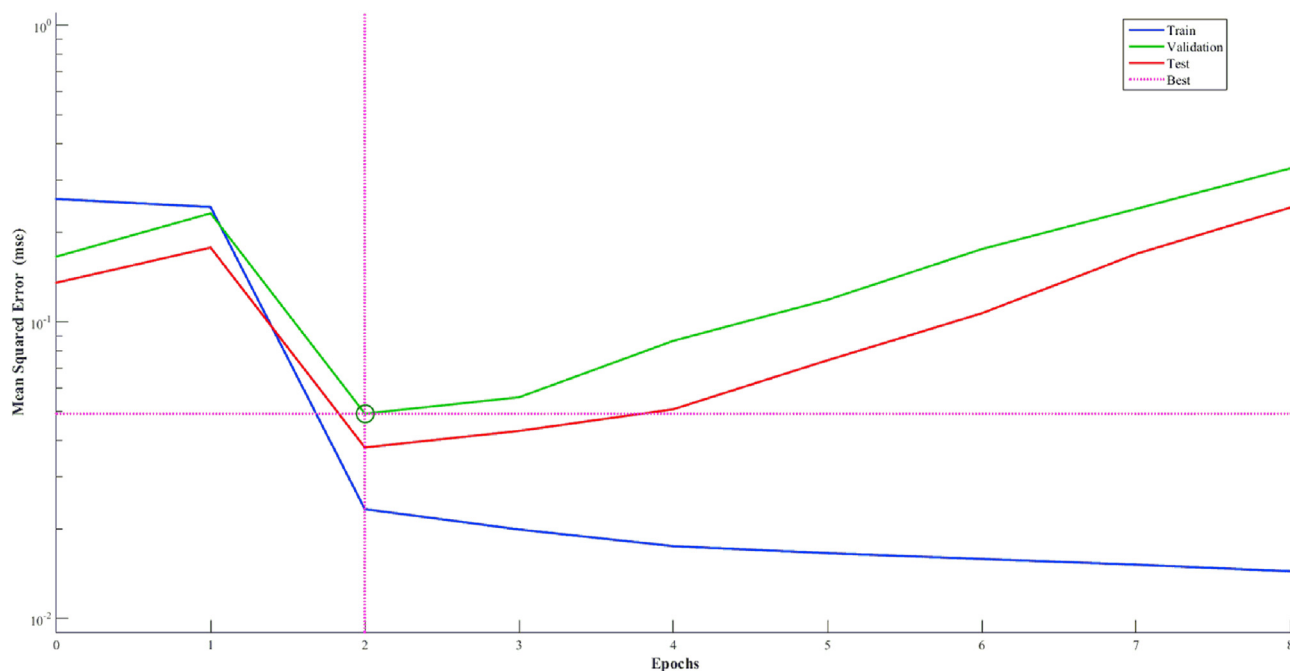


Fig. 7. Mean Squared Error for the training, test, and validation sets versus the number of epochs.

inconsiderable q^2 and r^2 values when compared to the real model value (located in the upper right quadrant of the graph-[Fig. 9](#)), proving that our QSAR-ANN model there has no obtained randomly.

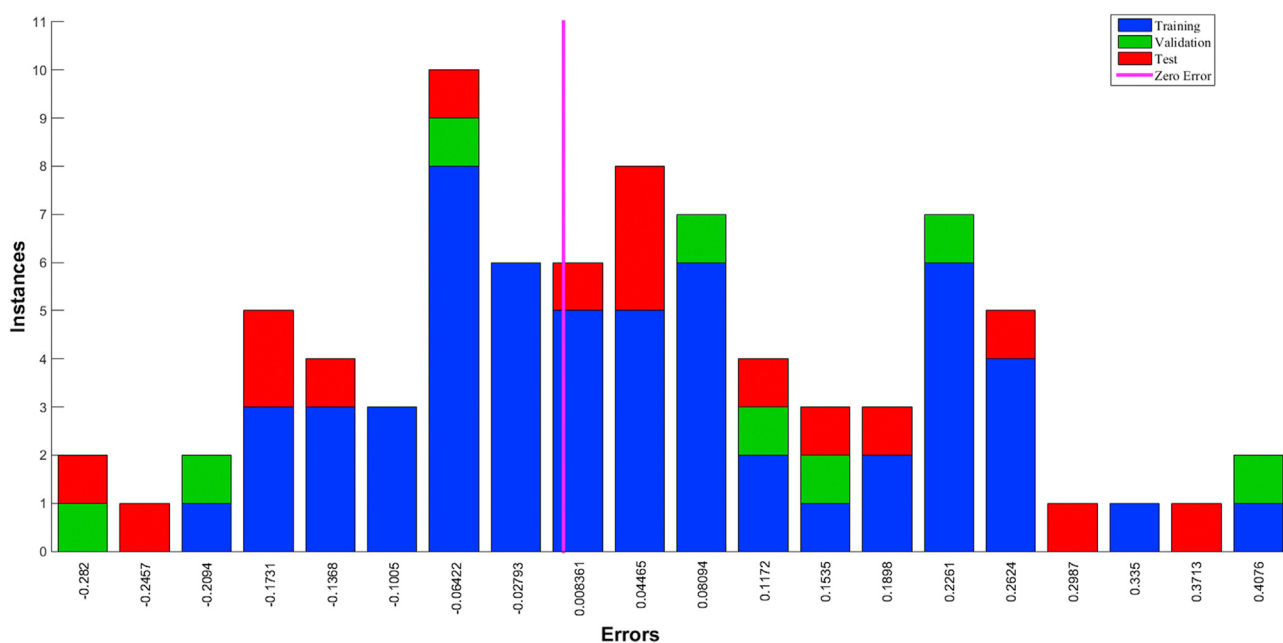
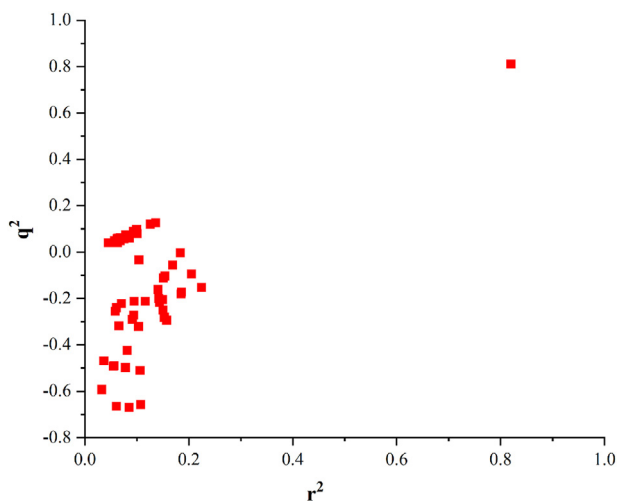
Since our model went through several statistical tests, we selected from the ChEMBL [55] database a class of 5-HT₆ antagonists with different molecular structures (molecular structures in Supplementary Material) so our model predicts the respective biological activity value (the chemical structures of the antagonists were optimized using the same protocol described in section 2.1 as well as the descriptors values). Afterward, the experimental and calculated pK_i values were compared and as it is seen in [Fig. 10](#), the pK_i values predicted by our model are very close to the experimental ones.

It is worth to mentioning that compound **215945** (ChEMBL ID) had a predicted pK_i value, by our ANN model, equal to the experimental result ([Table 5](#), compound in bold). Therefore, the results ([Table 5](#)) are a robust proof that certifies the forecasting ability of our model to predict the biological affinity of new molecules as putative 5-HT₆ antagonists.

We also emphasize that the compounds (molecules **411**, **808**, **402143**, **2159463**, **2159456** and **3260806**), whose structural characteristics differ greatly from the ones mentioned by López-Rodríguez and co-workers [24] ([Fig. 1](#)), have a low biological affinity and our ANN model was able to predict the biological activity of these compounds with a small margin of error, as we see in [Table 5](#).

Table 4
Residues calculated for the test set.

Compounds	Experimental pK_i	Predicted pK_i	Absolute residual	Relative residual %
11	7.53	7.52	0.01	0.13
21	7.04	7.28	-0.24	3.41
25	7.51	7.64	-0.13	1.73
27	7.76	7.58	0.18	2.32
28	7.45	7.51	-0.06	0.81
35	7.94	7.90	0.04	0.50
36	7.84	8.01	-0.17	2.17
42	7.99	7.93	0.06	0.75
52	7.30	7.57	-0.27	3.70
68	8.12	7.86	0.26	3.20
73	6.44	6.32	0.12	1.86
75	6.88	6.82	0.06	0.87
76	6.46	6.64	-0.18	2.79
77	6.40	6.24	0.16	2.50
78	6.54	6.16	0.38	5.81
81	6.46	6.15	0.31	4.80

**Fig. 8.** Detection of outliers in function of error.**Fig. 9.** Y-randomization test applied to QSAR-ANN model.

4.3. Analyzing the meaning of the selected descriptors

The classification of the selected descriptors is based on the dimension of its structural representation, for example, 2D and 3D descriptors (Fig. 11). The 2D descriptors (JhetZ, BEHm2, GATS3p, GATS1p) are ones of unique values extracted from 2D representations of compounds, as illustrated in Fig. 11a. From these descriptors, it is possible to extract important information, for instance, size, degree of branching, flexibility, the neighborhood of atoms as nitrogen, oxygen, and bromine atoms (of which we can obtain information about electrostatic and steric effects, and overall 2D shape). On the other hand, the 3D descriptors (Au, RDF025u, H3v, and HOMO₍₋₂₎) have substantial importance in the molecular recognition process between compounds and molecular targets.

Topological descriptor (JhetZ): this descriptor is allocated among the molecular graph descriptors, in which the chemical structure of the compound is converted into a molecular graph. Since the physical properties of the molecules are intrinsically related to their molecular geometry, the topological indices describe the molecular shape and model it from physical properties

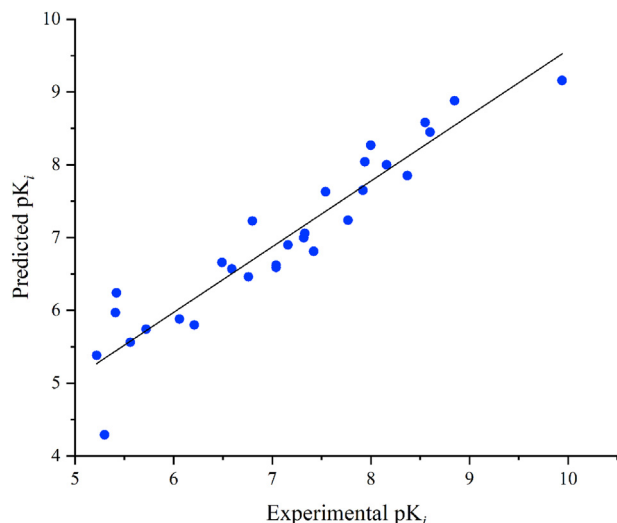


Fig. 10. Experimental versus predicted pK_i for the selected compounds from the ChEMBL database.

Table 5
Predicted pK_i values by our ANN model.

CHEMBL ID	Experimental pK_i	Predicted pK_i	Absolute residual
411	5.22	5.38	-0.16
808	5.30	4.29	1.01
62193	6.21	5.80	0.41
24474	9.94	9.16	0.78
402143	5.41	5.97	-0.56
484164	8.55	8.58	-0.03
482990	8.37	7.85	0.52
506736	8.16	8.00	0.16
483559	7.94	8.04	-0.10
482748	7.54	7.63	-0.09
347335	6.80	7.23	-0.43
2159470	6.59	6.57	0.02
2159463	5.42	6.24	-0.81
2159464	6.06	5.88	0.19
2159481	6.49	6.66	-0.17
2159456	5.56	5.56	0.00
154641	8.60	8.45	0.15
46071	7.92	7.65	0.27
3260792	7.32	7.00	0.32
3260806	5.72	5.74	-0.02
1834222	7.04	6.62	0.42
1834351	7.04	6.59	0.45
3260790	7.16	6.90	0.26
60669	8.00	8.27	-0.27
95136	8.85	8.88	-0.03
434365	7.42	6.81	0.61
378656	7.77	7.24	0.53
413049	6.76	6.46	0.30
433461	7.33	7.06	0.27

with good precision [56]. The descriptor JhetZ is a Balaban-type descriptor (Table 3) coded as a distance matrix, being a highly discriminated descriptor, which single out the existence of heteroatoms and multiple bonds in the compounds [57]. For the series of compounds under study (Table 1), most molecules presented value equal to 2 for JhetZ descriptor, except for the compounds **65**, **67**, **73**, and **75** (Fig. 12), which exhibited value equal to 1.

From JetZ descriptor values, we observed the following characteristics:

- all compounds that presented more than 3 rings (number cyclomatic) have Balaban index equal to 1;

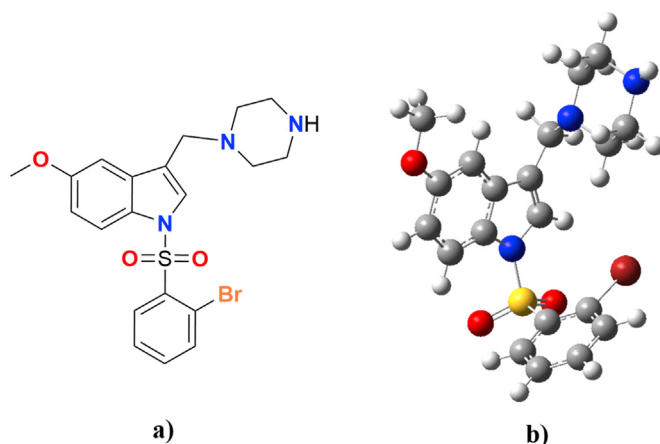


Fig. 11. Dimensionality of the structural representations-(a) 2D and (b) 3D descriptors.

- compounds that have no sulfone moiety or any group with similar characteristics (HBA moiety) had Balaban index equal to 1 (compounds **65** and **67**). Thus, this descriptor singles out the importance of the presence of these groups for the molecular recognition process;
- molecules that have aromatic rings (acetophenone and toluene, compounds **73** and **75**, respectively) attached to PI moiety exhibited the Balaban index equal to 1 and the biological activity value below 7, indicating that these aromatic rings are not important to the increase of the 5-HT₆ antagonist activity of the compounds.

BCTU descriptor (BEHm2): this type of descriptor describes the molecule geometry and combines several atomic properties in unique value using a matrix whose diagonal elements include atomic number, polarizability and information about H-bond [57]. BEHm2 is weighted by the atomic masses and we observed that, in general, the highest values of this descriptor are attributed to the compounds with the highest biological activity. However, the compounds **77** and **78** presented values equal to 6 (all values are presented in Supplementary Material) but have low pK_i values (Fig. 13).

Some characteristics from the analysis of the descriptor BEHm2 include: (a) bulky groups at the PI position do not contribute to the increase of the pK_i values of the compounds; (b) halogens attached to the AR moiety are not tolerated, thus they do not contribute to the increase the antagonist activity of the compounds and (c) in general, *o*-bromobenzene or *p*-bromobenzene are important for the molecular interactions (Receptor-Ligand).

2D Autocorrelation (GATS3p and GATS1p): these descriptors represent the relation between values of a unique variable (for instance, polarizability) for atoms in a molecule with 2D structural representation. Mathematically, they are defined as [57]:

$$AC_k = \int_a^b f(x) \cdot f(x+k) \cdot dx \quad (4)$$

wherein $f(x)$ represents some functions of the variable x (for this study, $f(x)$ is related to polarizability); k is the lag (it represents the topological distance for the selected descriptors in this study) that illustrates an interval of x (a and b). The function AC_k is obtained from the sum of the products of the function $f(x)$ values computed at x and $x+k$ range [57].

In terms of molecular graph, the descriptor (GATS) can be defined as [55]:

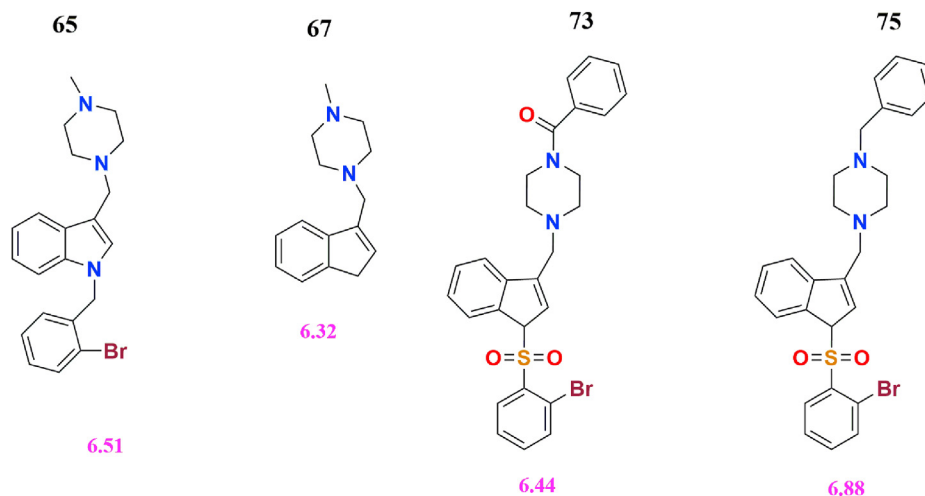
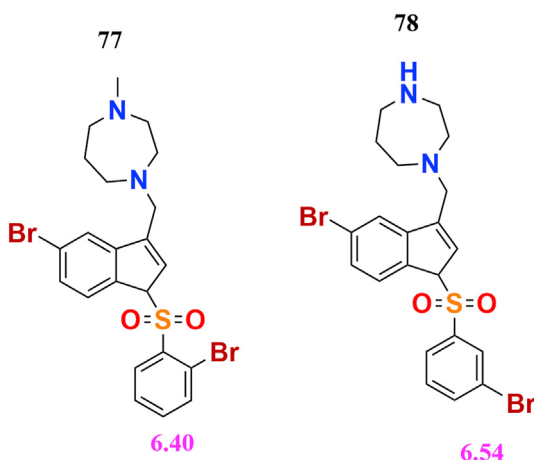


Fig. 12. Compounds with JetZ value equal to 1.

Fig. 13. Compounds with BEHm2 value equal to 6 but low pK_i values.

$$c_k = \frac{\frac{1}{2\Delta_k} \sum_{i=1}^A \sum_{j=1}^A (w_i - w_j)^2 \cdot \delta(d_{ij}; k)}{\frac{1}{(A-1)} \sum_{i=1}^A (w_i - \bar{w})^2} \quad (5)$$

in which w_i is some atomic property (polarizability for this study), $\bar{w}$ represents the mean value of polarizability, A refers to the atomic number, d_{ij} represents the topological distance between i and j atoms, and $\delta(d_{ij}; k)$ expresses the Kronecker delta $\left\{ \begin{array}{l} \delta_{ij} = 1 \text{ if } d_{ij} = k \\ \delta_{ij} = 0 \text{ if } d_{ij} \neq k \end{array} \right\}$, Δ_k symbolizes the number of atom pairs when the distance is equal to k [54]. The range of values between 0 and 1 shows positive autocorrelation, whereas values larger than 1 indicate an inverse correlation. Putting in another way, the atoms of the molecules represent a set of discrete points in a 3D space in which the function related to the atomic property (polarizability) is computed in these points [57]. Thus, both GATS3p and GATS1p descriptors refer to the autocorrelation of polarizability in intervals of topological distance equal to the lag value.

All compounds here studied presented values equal or larger than 1 for both descriptors (GATS3p and GATS1p). For GATS3p (see values in Supplementary Material), there is no autocorrelation at intervals of topological distance equal to 3, while for GATS1p with lag equals to 1 there is an inverse autocorrelation. Indeed, the

compounds with the highest biological activity have halogen atom (Br-halogen with large polarizing power) in the hydrophobic region (HYD - Figs. 1 and 12). Therefore, the polarizability in this region is equal for the compounds **58**, **61**, **69**, and **70** (blue dotted circle in Fig. 14). Analyzing the PI and AR moieties of these compounds, there are different polarization moieties in these regions (yellow and red dotted circles in Fig. 14).

RDF025u: this descriptor is defined as Radial Distribution Function – 2.5/unweighted, which belongs to the Radial Distribution Function (RDF) class. RDF descriptors are based on the geometrical representation of the distance distribution for a molecule and a radial distribution function code (RDF code) is established. This code provides important information about the three-dimensional arrangement of the atoms and can represent specific information to specific atom types or distance ranges, such as structure-activity properties of a molecule [57].

For the compounds studied in this work, RDF025 presented values of 28, 35 to 50, 52 and 54. The lowest value of these descriptor (28) was attributed to the compound **67**, which has the lowest biological activity.

All compounds with the RDF025 values in the range 35–38 exhibited at least one halogen atom in their molecular structure (with exception of the compounds **68** and **80**). We can highlight the compounds **2**, **4**, **32**, **69** and **70**, which presented pK_i greater than 7.99 and a bromine atom attached at the 2' or 4' position in the HYD moiety, evidencing the importance of the bromine atom fixed at these positions to increase the antagonist effect of the compounds (Fig. 15).

All molecules that presented RDF025 above 49 (compounds **35**, **39**, **49**, **59**, and **62**), have both the methoxy substituent attached at 4', 5', or 6' positions and the isopropyl substituent at the 4' position. These compounds are among those that have the highest biological activity of the dataset (the exception is compound **49**, which presented a moderate pK_i value). Thus, the methoxy substituent in 4', 5', or 6' positions and isopropyl in 4' position contribute enormously to the increase of the biological activity (Fig. 16).

Au: the descriptor Au belongs to Weighted Holistic Invariant Molecular (WHIN) class, which provides important 3D information on the molecule, such as molecular size, shape, symmetry, and atom distribution concerning invariant reference frames. WHIN descriptors are geometrical descriptors based on a statistical index calculated from the atoms along the principal axes, being classified as directional and global descriptors [57]. Au descriptor is defined

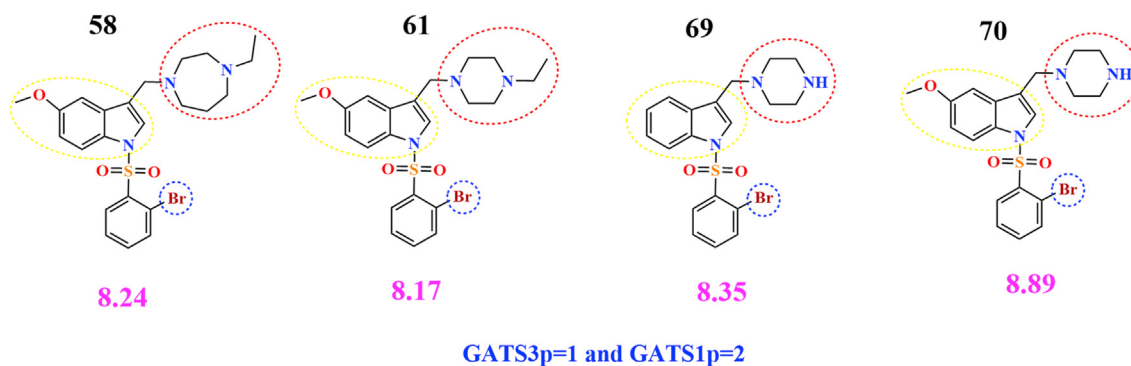


Fig. 14. Analysis of GATS3p and GATS1p for the compounds with the highest values of biological activity.

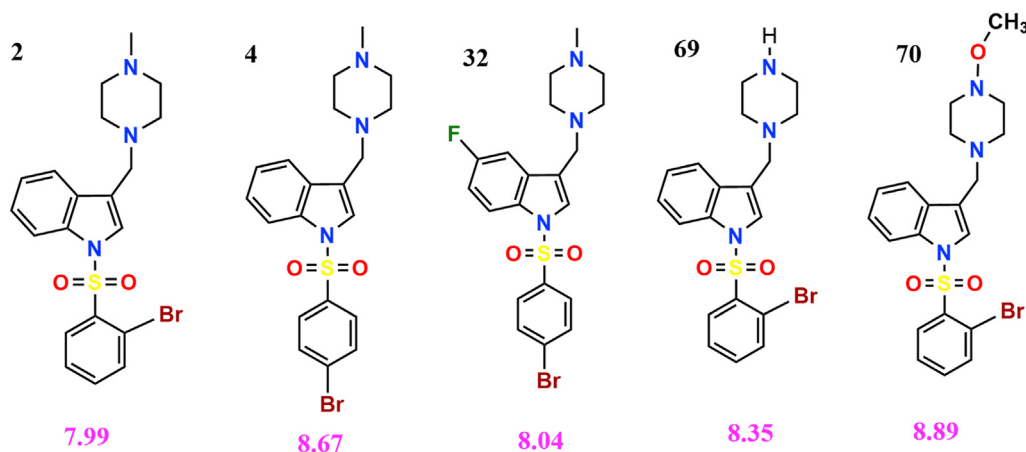


Fig. 15. Compounds that presented RDF025 values varying from 35 to 38.

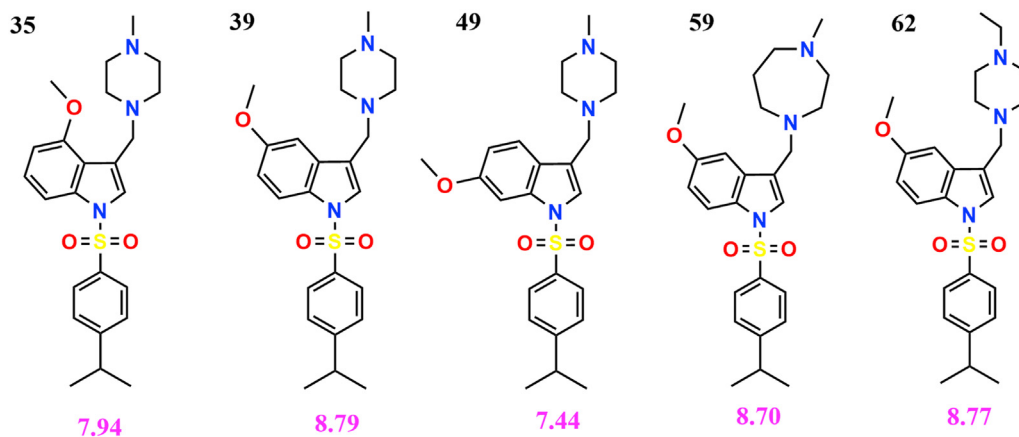


Fig. 16. Compounds that presented RDF025 values above 49.

as a total size index/unweighted and is classified as a global WHIN descriptor [57]. Au descriptor gives a global dimension on the molecular feature and the eigenvalue of Au is related to linear and quadratic contributions to the molecular size [57]. In this study, the lowest value of Au ($Au = 33$) is related to compound **67**, which also presented the lowest biological activity value ($pK_i = 6.32$) when compared to the set of compounds studied in this work (Table 1).

H3v: This descriptor is defined as H autocorrelation of lag 3/ weighted by atomic van der Waals volumes. H3v belongs to GETAWAY (GEometry, Topology, and Atom-Weights Assembly)

class, which is calculated from the leverage matrix obtained from the centered atomic coordinates (molecular influence matrix/MIM) [58].

The H3v descriptor has single values (1 or 0) for the compounds studied. Most compounds have H3v values equal to 1, except compounds **67** and **80** ($H3v = 0$). Compounds **67** ($pK_i = 6.32$) and **80** ($pK_i = 6.58$) are among the molecules of dataset with the lowest biological activity values. Moreover, both compounds **67** and **80** do not have halogen atoms in their structures while the most active compounds have them. Therefore, the H3v descriptor shows that

halogen atoms contribute to increase the antagonist activity of the compounds against the 5-HT₆ receptor.

E_{HOMO(-2)}: This electronic descriptor represents the energy of HOMO (High Occupied Molecular Orbital) for a molecule and indicates its capacity to donate electrons. The higher E_{HOMO(-2)} values, the more likely the molecule is capable of donating electrons [59–61]. This descriptor expresses the energy value of the third-highest occupied molecular orbital and it evidenced the importance of the presence of PI moieties, as the electronic transfer process is likely to occur. Indeed, the presence of nitrogen atoms proves this premise, as we can see in Fig. 17, in which the orbital

representation of HOMO₍₋₂₎ is exhibited for the most active molecules (compounds **62** and **70**), as well as the moderate ones (compounds **63** and **48**) and the least active ones (compounds **67** and **81**).

4.4. Design of new 5-HT₆ antagonists

To design new 5-HT₆ antagonists, we modified the AR, HYD and PI moieties based on the analysis of the descriptors and then we examined if there was an increase of the biological activity values (Table 6):

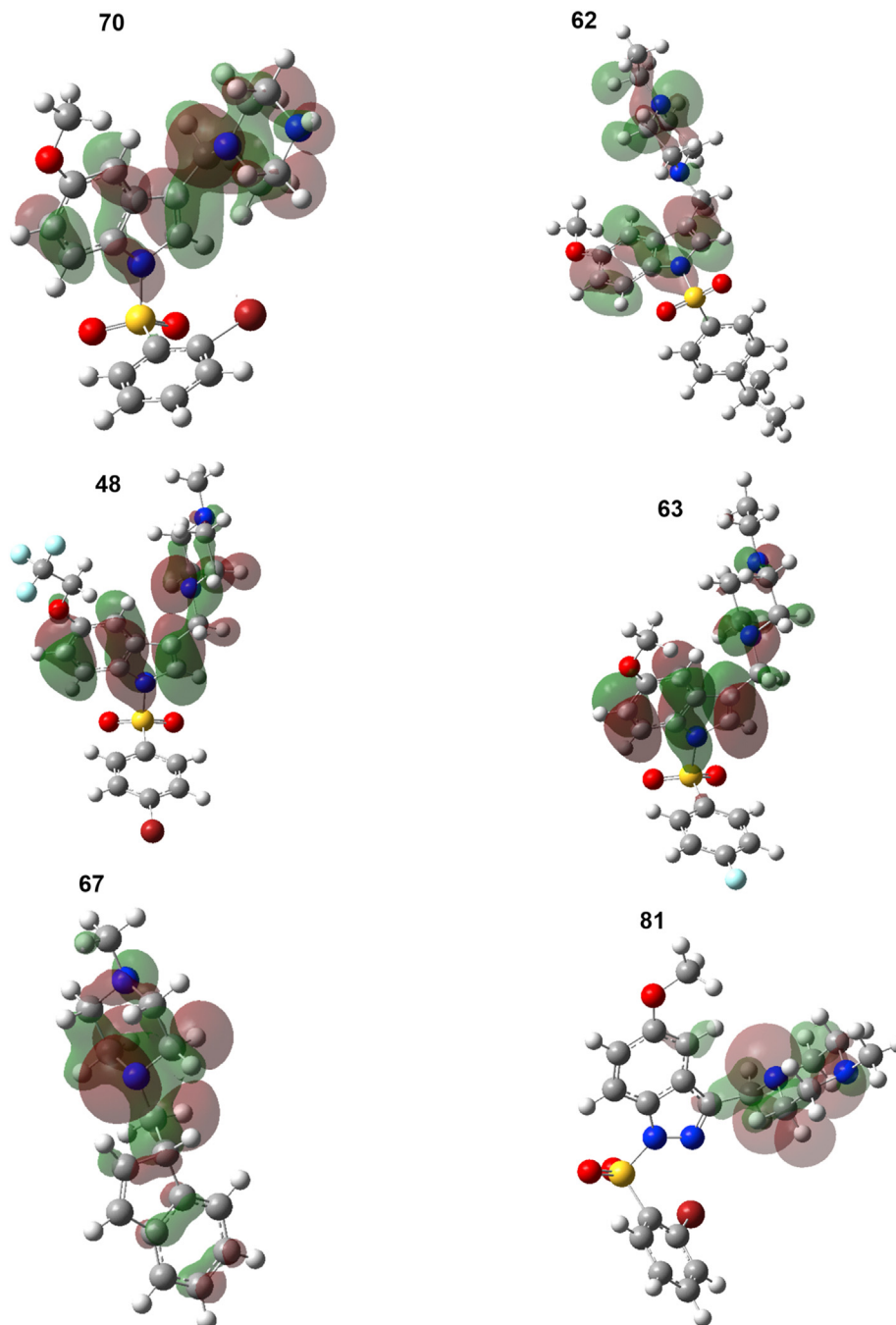
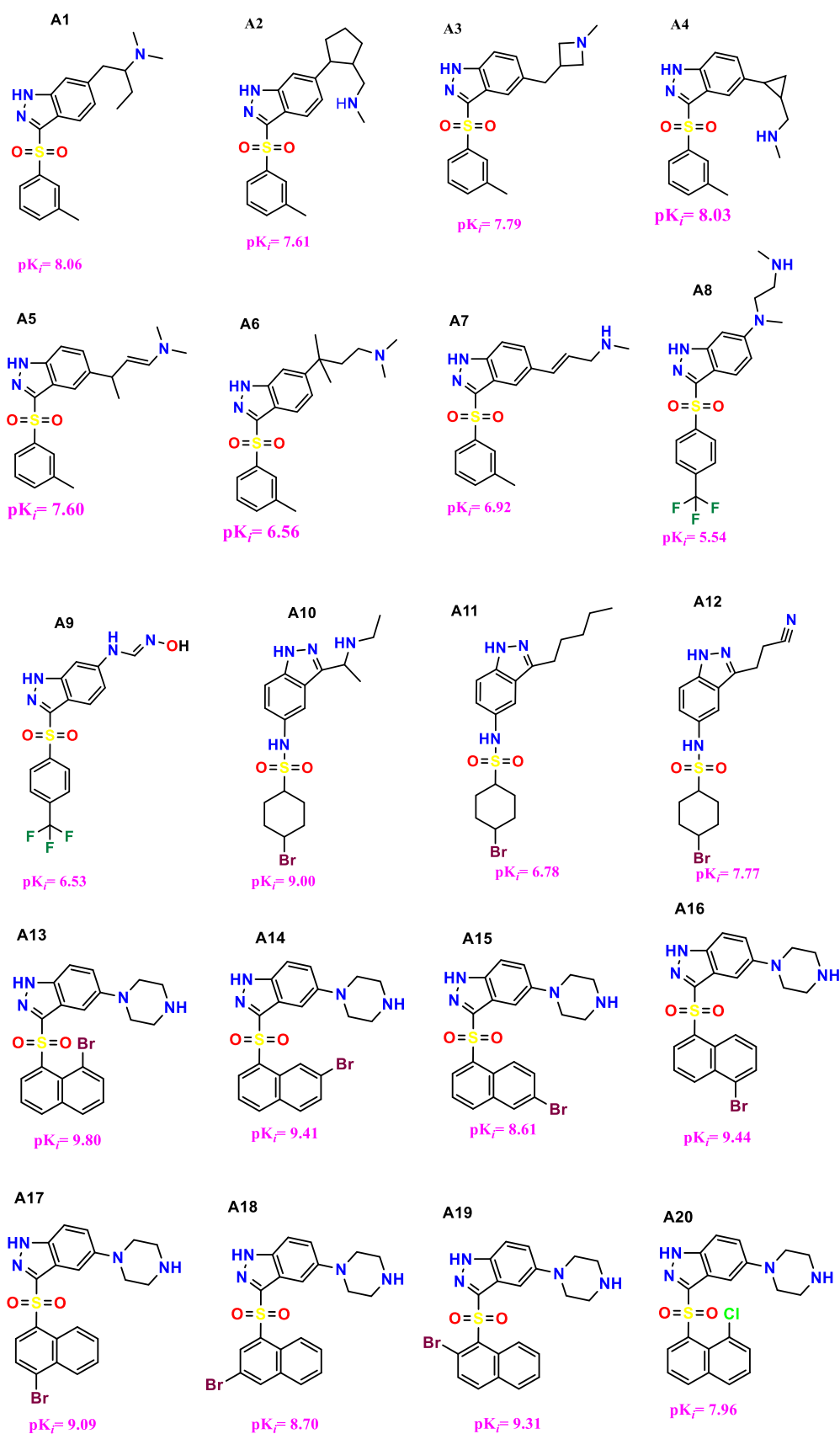


Fig. 17. Orbital representation of HOMO₍₋₂₎ for the most active molecules (compounds **62** and **70**), as well as the moderate (compounds **63** and **48**), and the least active (compounds **67** and **81**).

Table 6

Proposed compounds from the QSAR-ANN model analyses.



- in the AR moiety we replaced indole by an of the indazole group (this alteration remained fixed for all design compounds);
- we made several modifications in the PI moiety;
- other changes were also performed in the HYD group;
- several times, the modifications were made in the PI and HYD moieties simultaneously.

Finally, the proposed compounds were submitted to our ANN model so that their biological activity values could be predicted, as presented in Table 6.

For compounds A1, A2, A3, A4, A5, A6, and A7, the AR and HYD moieties remained unchanged, while gradual changes were performed in the PI moiety. From compounds A1 through A7, we highlight the significant biological activity values for A1 ($pK_i = 8.06$), A3 ($pK_i = 7.79$) and A4 ($pK_i = 8.03$) predicted by our QSAR-ANN model. As aforementioned (Section 4.3), bulky groups in the position PI are not profitable, indeed all compounds that presented this characteristic did not exhibit reasonable pK_i values (compounds A8, A9, and A11). We also insert halogen atoms (Br, Cl, and F) in several positions of the HYD moiety, and these changes resulted in higher pK_i values for some of the new compounds and reasonable pK_i values for others. Indeed, several halogen atoms linked to the aromatic ring do not contribute to the increase of the antagonist activity (compounds A8 and A9), whereas the presence of bromine in several positions at naphthalene and cyclohexane resulted in excellent pK_i values (compounds A10, A13, A14, A15, A16, A17, A18, and A19).

4.4.1. Estimation of the ADME/Tox profile

Finally, we analyze the drug-likeness properties (physico-chemical and pharmacokinetic properties) of the new designed compounds (Table 7) using QikProp module from the Schrödinger

package [62]. Drug-likeness properties of these compounds indicated that they have desired ADME/Tox profile (Table 7) and hence they can go through the experimental approaches.

Regarding synthetic accessibility of new molecules, we used the SwissADME [64] tool to evaluate them and the outcomes showed that none of the proposed compounds has a high difficulty degree (on a score from 1 to 10, where 1 is very easy and 10 is very difficult) as we can see in Table 8.

5. Conclusions

Artificial intelligence techniques have long been applied in several branches of science to solve various problems. In the field of drug discovery, the application of these techniques is growing with the increase of the computational power. In this work, we constructed a QSAR model using Artificial Neural Networks and for this purpose, the Genetic Algorithm method was essential to select the descriptors that more contributed to understand the structure-activity relationship of the 5-HT₆ antagonists here studied.

Table 8
Estimation of the synthetic accessibility of new compounds.

Compound	Score	Compound	Score
A1	5.20	A11	4.17
A2	3.87	A12	3.92
A3	2.88	A13	3.27
A4	3.60	A14	3.18
A5	3.78	A15	3.15
A6	3.22	A16	3.18
A7	3.13	A17	5.16
A8	3.00	A18	3.27
A9	2.78	A19	3.33
A10	4.21	A20	3.24

Table 7
Estimation of the ADME/Tox properties of the new proposed compounds.

Compound	MW	HBdon	HB _{accp}	QLogPo/w	QLogS	QPPCaco	QLogBB	PHOA	PSA	R05
A1	371.496	0.00	6.000	2.879	-3.396	137.849	-0.715	82.095	72.673	0
A2	383.507	1.00	5.500	3.457	-4.939	115.551	-0.719	84.108	79.380	0
A3	355.454	0.000	4.000	4.234	-5.837	674.102	-0.919	100.000	71.112	0
A4	355.454	1.00	5.500	2.969	-4.324	134.002	-0.640	82.402	79.490	0
A5	369.481	0.000	6.000	3.441	-4.761	587.204	-0.988	96.651	70.923	0
A6	385.523	0.000	6.000	3.511	-4.164	187.242	-0.608	88.172	70.444	0
A7	341.427	1.00	5.500	2.761	-4.106	118.084	-0.785	80.199	79.209	0
A8	424.440	1.00	6.500	3.287	-5.098	131.158	-0.338	84.093	80.935	0
A9	384.332	2.000	7.700	1.392	-4.292	32.691	-2.028	62.204	134.109	0
A10	429.374	3.000	7.000	2.400	-3.750	151.099	-0.525	80.004	86.854	0
A11	428.386	2.000	5.500	3.862	-5.782	579.288	-1.052	100.000	79.617	0
A12	411.316	2.000	7.000	2.247	-5.345	184.071	-1.379	80.638	100.401	0
A13	471.371	1.000	6.500	3.051	-4.419	114.591	-0.440	81.667	82.595	0
A14	471.371	1.000	6.500	3.342	-4.876	131.733	-0.308	84.453	79.699	0
A15	471.371	1.000	6.500	3.327	-4.908	123.735	-0.335	83.878	79.806	0
A16	471.371	1.000	6.500	3.255	-4.761	123.703	-0.342	83.451	79.763	0
A17	471.371	1.000	6.500	3.262	-4.770	124.519	-0.340	83.549	79.787	0
A18	471.371	1.000	6.500	3.318	-4.898	121.730	-0.345	83.699	79.905	0
A19	471.371	1.000	6.500	2.986	-4.189	116.321	-0.380	81.398	79.566	0
A20	426.920	1.000	6.500	3.010	-4.363	114.824	-0.443	81.440	82.853	0

Description of each property from reference [63].

MW- Molecular weight (130.0–725.0 Da).

HB_{don} - Recommended number of hydrogen bond donor groups (0.0–6.0).

HB_{accp} - Recommended number of hydrogen bond accept groups (2.0–20.0).

QLogP_{o/w} - Predicted octanol/water partition coefficient (-2.0 to 6.5).

QLogS - Predicted aqueous solubility, log S (-6.5 to 0.5).

QPPCaco - Predicted apparent Caco-2 cell permeability in nm/sec. (<25 poor, > 500 optimum).

QLogBB - Predicted brain/blood partition coefficient. (-3.0 to 1.2).

PHOA - Predicted human oral absorption on 0–100% scale. (>80% is optimum, <25% is poor).

PSA - Van der Waals Surface area of polar N and O atoms as well as carbonyl atoms (70–200.0).

R05 - Rule of five (Mol_MW < 500, QLogP_{o/w} < 5, donor HB ≤ 5, and accept HB ≤ 10).

Once our QSAR-ANN model has been constructed, we applied several validation procedures to certify its robustness and predictive power. Among all tests, we believe that the test set is the ultimate proof that ensures the model significance however, to have no doubt about our ANN model usefulness, we verified its predictive power against different chemical classes of 5-HT₆ antagonists, and the outcomes certified its performance. Indeed, by analyzing the descriptors as well as the structure-activity relationship, we designed new 5-HT₆ antagonist compounds, which presented good biological activity values and among them we highlight compounds A10, A13, A14, A16, A17, A18 and A19, which exhibited pK_i values above 9.00.

We also emphasize that the presence of halogen atoms (mainly Br) linked to the aromatic ring in *para*-position (HYD moiety) contributes significantly to the increase of the biological activity values, whereas the bulky groups in the PI position are not profitable. Lastly, the drug-likeness properties obtained via ADME/Tox estimation, as well as the synthetic accessibility, qualify these new compounds to move on to further experimental studies and thenceforward their antagonist effects can be confirmed.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) (Finance Code 001), CNPq and FAPESP (2016/24524–7, 2017/10118–0). Our research was carried out using the computational resources of the Center for Mathematical Sciences Applied to Industry (CeMEAI) funded by FAPESP (grant 2013/07375–0).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jmgm.2021.107844>.

References

- [1] Jet M.J. Vonk, Miguel Arce Rentería, Justina F. Avila, Nicole Schupf, M. James, Noble, Richard Mayeux, Adam M. Brickman, J. Jennifer, Manly, Secular trends in cognitive trajectories of diverse older adults, *Alzheimer's Dementia* 15 (2019) 1576–1587.
- [2] Alzheimer's Disease International, World Alzheimer report 2019 Attitudes to dementia disponível em. <https://www.alz.co.uk/research/WorldAlzheimerReport2019.pdf>. Accessed 06 mar. 2020.
- [3] M. Prince, A. Wimo, M. Guerchet, G.C. Ali, Y.T. Wu, M. Prina, et al., World Alzheimer Report 2015: the Global Impact of Dementia: an Analysis of Prevalence, Incidence, Cost and Trends, *Alzheimer's Disease International*, London, 2015.
- [4] GBD 2016 Dementia Collaborators. Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016 (2019). *Lancet Neurol.* 18:88–106.
- [5] R.S. Wilson, E. Segawa, P.A. Boyle, S.E. Anagnos, L.P. Hize, D.A. Bennett, The natural history of decline cognitive in Alzheimer's disease, *Psychol. Aging* 17 (2012) 1008–1017.
- [6] W.W. Barker, C.A. Luis, A. Kashuba, M. Luis, D.G. Harwood, D. Loewenstein, C. Waters, P. Jimison, E. Shepherd, S. Sevush, N. Graff-Radford, D. Newland, M. Todd, B. Miller, M. Gold, K. Heilman, L. Doty, I. Goodman, B. Robinson, G. Pearl, D. Dickson, R. Duara, Relative frequencies of Alzheimer disease, Lewy body, vascular and frontotemporal dementia, and hippocampal sclerosis in the State of Florida Brain Bank, *Alzheimer Dis. Assoc. Disord.* (2002) 203–212.
- [7] R. Katzman, Editorial: the prevalence and malignancy of the Alzheimer disease. A major killer, *Arch. Neurol.* 4 (2017) 217–218.
- [8] Alzheimer's association, Alzheimer's disease facts and figures (2016), *Alzheimer's Dement* 2 (2016) 459–509.
- [9] The human memory- neurons & synapses. http://www.human-memory.net/brain_neurons.html. Accessed jun 11, 2019.
- [10] B. Benhamú, M.M. Fontecha, H.V. Villa, L. Pardo, M.L.L. Rodríguez, Serotonin 5-HT₆ receptor antagonists for the treatment of cognitive deficiency in Alzheimer's disease, *J. Med. Chem.* 57 (2014) 7160–7181, <https://doi.org/10.1021/jm5003952>.
- [11] N. Upton, T.T. Chuang, A.J. Hunter, D.J. Virley, 5-HT₆ antagonist as novel cognitive enhancing agent for Alzheimer's disease, *Neurotherapeutics* 5 (2008) 458–459, <https://doi.org/10.1016/j.nurt.2008.05.008>.
- [12] D. Karila, T. Freret, V. Bouet, M. Boulouard, P. Dallemagne, C. Rochais, Therapeutic potential of 5-HT₆ receptor Agonists, *J. Med. Chem.* 58 (2015) 7901–7912, <https://doi.org/10.1021/acs.jmedchem.5b00179>.
- [13] A. Corbett, J. Pickett, A. Burns, J. Corcoran, S.B. Dunnett, P. Edison, et al., Drug repositioning for Alzheimer's disease, *Nat. Rev. Drug Discov.* 11 (2012) 833–846.
- [14] P. Anand, B. Singh, A review on cholinesterase inhibitors for Alzheimer's disease, *Arch Pharm. Res. (Seoul)* 36 (2013) 375–399, <https://doi.org/10.1007/s12272-013-0036-3>.
- [15] W. Karsten, H. Andreas, B. Anton, Investigational drugs targeting 5-HT₆ receptors for the treatment of Alzheimer's disease, *Expert Opin. Invest. Drugs* 24 (2015) 1515–1528, <https://doi.org/10.1517/13543784.2015.1102884>.
- [16] A. Cherkasov, et al., QSAR modeling: where have you been? where are you going to? *J. Med. Chem.* 57 (2014) 4977–5010, <https://doi.org/10.1021/jm4004285>.
- [17] M. Keiser, V. Setola, J. et al., Predicting new molecular targets for known drugs, *Nature* 462 (2009) 175–181, <https://doi.org/10.1038/nature08506>.
- [18] E. Lounkine, M. Keiser, S. Whitebread, et al., Large-scale prediction and testing of drug activity on side-effects targets, *Nature* 486 (2012) 361–367, <https://doi.org/10.1038/nature11159>.
- [19] J. Besnard, G. Ruda, V. Setola, et al., Automated design of ligands to polypharmacological profiles, *Nature* 492 (2012) 215–220, <https://doi.org/10.1038/nature11691>.
- [20] R. Hajjo, C.M. Grulke, A. Golbraikh, V. Setola, Xi-Ping Huang, B.L. Roth, A. Tropsha, Development, validation, and use of quantitative structure-activity relationship models of 5-hydroxytryptamine (2B) receptor ligands to identify novel receptors binders and putative valvulopathic compounds among common drugs, *J. Med. Chem.* 53 (2010) 7573–7586, <https://doi.org/10.1021/jm100600y>.
- [21] L. Zhang, D. Foursches, A. Sedykh, H. Zhu, A. Golbraikh, S. Ekins, J. Clark, M.C. Connelly, M. Sigal, D. Hodges, A. Guiguem, R.K. Guy, A. Tropsha, Discovery of novel Antimalarial compounds enabled by QSAR-based virtual screening, *J. Chem. Inf. Model.* 53 (2013) 475–492, <https://doi.org/10.1021/ci300421n>.
- [22] E.N. Muratov, et al., Per Aspera Ad Astra: application of simplex QSAR approach in Antiviral research, *Future Med. Chem.* 2 (2010) 1205–1226, <https://doi.org/10.4155/fmc.10.194>.
- [23] M.G. Cardozo, T. Kawai, Y. Limura, H. Sugimoto, Y. Yamanishi, J. Hopfinger, Conformational analyses and molecular-shape comparisons of a series of indanone-benzylpiperidine inhibitors of acetylcholinesterase, *J. Med. Chem.* 35 (1992) 590–601, <https://doi.org/10.1021/jm00081a023>.
- [24] M.L. López-Rodríguez, B. Benhamú, T. de la Fuente, A. Sanz, L. Pardo, M.A. Campilo, Three-dimensional pharmacophore model for 5-hydroxytryptamine₆ (5-HT₆) receptor antagonists, *J. Med. Chem.* 48 (2005) 416–4219, <https://doi.org/10.1021/jm050247c>.
- [25] D. Sikazwe, M.L. Bondarev, M. Dukat, J.B. Rangisetty, A. Sanz, B.L. Roth, R.A. Glennon, Binding of sulfonyl-containing arylalkylamines at human 5-HT₆ serotonin receptor, *J. Med. Chem.* 49 (2006) 5217–5225, <https://doi.org/10.1021/jm060469q>.
- [26] DoddareddyMR. KimHJ, H. Choo, Y.S. Cho, K.T. No, W.K. Park, A.N. Pae, New serotonin 5-HT₆ ligands from common feature pharmacophore hypotheses, *J. Med. Chem.* 48 (2005) 197–206, <https://doi.org/10.1021/ci700160t>.
- [27] T. De la Fuente, M. Martín-Fontecha, J. Sallander, B. Benhamú, L. Campilo, R.A. Medina, L.P. Pellissier, S. Claeysen, A. Dumuis, L. Pardo, M.L. López-Rodríguez, Benzimidazole derivatives as new serotonin 5-HT₆ receptor antagonists. Molecular mechanisms of receptor inactivation, *J. Med. Chem.* 53 (2010) 1357–1369, <https://doi.org/10.1021/jm901672k>.
- [28] J. Mella, F. Villegas, C. Morales-Verdejo, C.F. Lagos, Recabarren- GajardoG, Structure-activity relationships studies on weakly basic N-arylsulfonylindoles with an antagonistic profile in the 5-HT₆ receptor, *J. Mol. Struct.* 1139 (2017) 362–370, <https://doi.org/10.1016/j.molstruc.2017.03.067>.
- [29] M.R. Doddareddy, Y.S. Cho, H.Y. Koh, A.N. Pae, CoMFA and CoMSIA 3D QSAR analysis on N1-arylsulfonylindole compounds as 5-HT₆ antagonists, *Bioorg. Med. Chem.* 12 (2004) 3977–3985, <https://doi.org/10.1016/j.bmc.2004.06.007>.
- [30] M. Hao, Y. Li, H. Li, S. Zhang, Investigation of the structure requirement for 5-HT₆ binding affinity of arylsulfonyl derivatives: a computational study, *Int. J. Mol. Sci.* 12 (2011) 5011–5030, <https://doi.org/10.3390/ijms12085011>.
- [31] A.P. da Silva, R.M. de Angelo, H. de Paula, K.M. Honório, A.B.F. da Silva, Drug design of new 5-HT₆ antagonists: a QSAR study of arylsulfonamide derivatives, *Struct. Chem.* 31 (2020) 1585–1597, <https://doi.org/10.1007/s11224-020-01513-z>.
- [32] Nirogi, et al., Discovery and development of 1-[(2-bromophenyl) sulfonyl]-5-methoxy-3-[4-[(4-methyl-1-piperazinyl) methyl]-1H-indole dimethylate Monohydrate (SUVN-502): a novel, potent, selective and orally active serotonin 6 (5-HT₆) receptor antagonist for potential treatment of Alzheimer's disease, *J. Med. Chem.* 60 (2017) 1843–1859, <https://doi.org/10.1021/acs.jmedchem.6b01662>.

- [33] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density (1980), *Phys. Rev. B* 37 (1988) 785, <https://doi.org/10.1103/physrevb.37.785>.
- [34] R. Ditchfield, W.J. Hehre, J.A. Pople, Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules, *J. Chem. Phys.* 54 (1971) 724–728, <https://doi.org/10.1063/1.1674902>.
- [35] Gaussian 09, Revision A.02 M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian, Inc., Wallingford CT, 2016.
- [36] I.V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V.A. Palyulin, E.V. Radchenko, N.S. Zefirov, A.S. Makarenko, V.Y. Tanchuk, V.V. Prokopenko, Virtual computational chemistry laboratory - design and description, *J. Comput. Aided Mol. Des.* 19 (2005) 453–463, <https://doi.org/10.1007/s10822-005-8694-y>.
- [37] P. Sutton, S. Boyden, Genetic algorithms: a general search procedure, *Am. J. Phys.* 62 (1994) 5994, <https://doi.org/10.1119/1.17516>.
- [38] D.B. De Oliveira, A.C. Gaudio, BuildQSAR: a new computer program for QSAR analysis, *Quant. Struct.-Act. Relat.* 19 (6) (2000) 599–601, <https://doi.org/10.1119/1.17516>.
- [39] MATLAB R2015a, The MathWorks, Natick, 2015.
- [40] V.A. Pulido, M. Gestal, M.C. Monteagudo, J.R. Rabuñal, J. Dorado, C.R. Munteanu, Evolutionary computation and QSAR research, *Curr. Comput. Aided Drug Des.* 9 (2013) 206–225, <https://doi.org/10.2174/1573409911309020006>.
- [41] M.M.C. Ferreira, C.A. Montanari, A.C. Gaudio, Variable selection in QSAR, *Quim. Nova* 25 (2002) 439–448, <https://doi.org/10.1590/S0100-40422002000300017>.
- [42] J. Ma, R.P. Sheridan, A. Liaw, G.E. Dahl, V. Svetnik, Deep neural nets as a method for quantitative structure-activity relationships, *J. Chem. Inf. Model.* 55 (2015) 263–274, <https://doi.org/10.1021/ci500747n>.
- [43] I.B. Campbell, J.F. Macdonald, P.A. Procopiou, Medicinal Chemistry in drug discovery in big pharma: past, present and future, *Drug Discov. Today* 23 (2018) 219–234, <https://doi.org/10.1016/j.drudis.2017.10.007>.
- [44] Yu-Chen, S.E. Rensi, W. Torng, R.B. Altman, Machine Learning in chemoinformatics and drug discovery, *Drug Discov. Today* 23 (2018) 1538–1546, <https://doi.org/10.1016/j.drudis.2018.05.010>.
- [45] K.Z. Myint, L. Wang, Q. Tong, X.Q. Xie, Molecular fingerprint-based artificial neural networks QSAR for ligand biological activity predictions, *Mol. Pharm.* 9 (2012) 293–2912, <https://doi.org/10.1021/mp300237z>.
- [46] S. Agatonovic-Kustrin, R. Beresford, Basic concepts of artificial neural networks (ANN) modeling and its application in pharmaceutical research, *J. Pharmaceut Biomed* 22 (2000) 717–727, [https://doi.org/10.1016/S0731-7085\(99\)00272-1](https://doi.org/10.1016/S0731-7085(99)00272-1).
- [47] K.M. Honório, E.F. de Lima, M.G. Quiles, R.A.F. Romero, F.A. Molfetta, A.B.F. da Silva, Artificial neural networks and the study of the psychoactivity of cannabinoid compounds, *Chem. Biol. Drug Des.* 75 (2010) 632–640.
- [48] K.G. Sheela, S.N. Deepa, Review on methods to fix number of hidden neuron in neural networks, *Math. Probl Eng.* (2013) 1–11, <https://doi.org/10.1155/2013/425740>, 2013.
- [49] T.P. Lillicrap, A. Santoro, L. Marris, et al., Backpropagation and the brain, *Nat. Rev. Neurosci.* 21 (2020) 335–346, <https://doi.org/10.1038/s41583-020-0277-3>.
- [50] M. Jalali-Heravi, M. Asadollahi-Baboli, P. Shahbazikhah, QSAR study of heparanase inhibitors activity using artificial neural networks and Levenberg–Marquardt algorithm, *Eur. J. Med. Chem.* 43 (2008) 548–556, <https://doi.org/10.1016/j.ejmech.2007.04.014>.
- [51] N. Khan, D. Gaurav, T. Kandl, Performance evaluation of levenberg-Marquardt thechinique in error reduction for diabetes condition classification, *Procedia Comput. Sci.* 18 (2013) 2629–2637, <https://doi.org/10.1016/j.procs.2013.05.455>.
- [52] A. Tropsha, P. Gramatica, V.K. Gombar, The importance of being earnest: validation ins the absolute essential for successful application and interpretation of QSPR model, *QSAR Comb. Sci.* 22 (2003) 69–77, <https://doi.org/10.1002/qsar.200390007>.
- [53] R. Kiralj, M.M.C. Ferreira, Basic validation procedures for regression models in QSAR and QSPR studies: theory and application, *J. Braz. Chem. Soc.* 20 (2009) 770–787, <https://doi.org/10.1590/S0103-50532009000400021>.
- [54] J.P.A. Martins, M.M.C. Ferreira, QSAR modeling: a new open source computational package to generate and validate QSAR models, *Quim. Nova* 36 (2013) 554–560, <https://doi.org/10.1590/S0100-40422013000400013>.
- [55] A. Gaulton, L.J. Bellis, A.P. Bento, J. Chambers, M. Davies, A. Hersey, et al., ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Res.* 40 (2012) D1100–D1107, <https://doi.org/10.1093/nar/gkr777>.
- [56] K.J. Burck, Chemical applications of graph theory, in: J.E. House, S.M. Blinder (Eds.), *Mathematical Physics in Theoretical Chemistry* (2019) 261–294, <https://doi.org/10.1016/C2016-0-04521-7>.
- [57] Todeschini, V. Consonni, *Handbook of Molecular Descriptor*, Wiley-VCH, Weinheim, 2009.
- [58] V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies, *J. Chem. Inf. Comput. Sci.* 42 (2002) 693–705, <https://doi.org/10.1021/ci0155053>.
- [59] M. Karelson, V.S. Lobnov, A.R. Katritzky, Quantum-chemical descriptors in QSAR/QSPR studies, *Chem. Rev.* 96 (1996) 44–1027, <https://doi.org/10.1021/cr950202r>.
- [60] A. Arroio, K.M. Honorio, A.B.F. da Silva, Quantum chemical properties used in Structure-Activity Relationship studies, *Quim Nov* 33 (2010) 694–699, <https://doi.org/10.1590/S0100-40422010000300037>.
- [61] K.M. Honorio, A.B.F. da Silva, Na AM1 study on the electron-donating and electron-accepting character of biomolecules, *OCT* 95 (2003) 126–132, <https://doi.org/10.1002/qua.10661>.
- [62] (2017) Schrödinger, Maestro–Schrödinger Release, 2017, p. 1.
- [63] Schrödinger Release 2020-4, QikProp, Schrödinger, LLC, New York, NY, 2020.
- [64] A. Daina, O. Michielin, V. Zoete, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules, *Sci. Rep.* 7 (2017) 42717, <https://doi.org/10.1038/srep42717>.