Full Paper

www.molinf.com

molecular informatics

DOI: 10.1002/minf.201900024

Molecular Modelling of Potential Candidates for the Treatment of Depression

Daniela Rodrigues Silva,[a] Stephen J. Barigye,[b] Letícia Santos-Garcia,[a] and Elaine Fontes Ferreira da Cunha*[a]

Abstract: A lot of research initiatives in the last decades have been focused on the search of new strategies to treat depression. However, despite the availability of various antidepressants, current treatment is still far from ideal. Unwanted side effects, modest response rates and the slow onset of action are the main shortcomings. As a strategy to improve symptomatic relief and response rates, the dual modulation of the serotonin transporter and the histamine H_3 receptor by a single chemical entity has been proposed in the literature. Accordingly, this work aims to elucidate key structural features responsible for the dual inhibitory activity of the hexahydro-pyrrolo-isoquinoline derivatives. For this purpose, two approaches were employed, four-dimensional quantitative structure-activity relationship (4D-QSAR) and molecular docking. The 4D-QSAR models for both receptors allowed the identification of the pharmacophore groups critical for the modelled biological activity, whereas the binding mode of this class of compounds to the human serotonin transporter was assessed by molecular docking. The findings can be applicable to design new antidepressants.

Keywords: antidepressants · structure-activity relationships · molecular docking · drug design · hexahydro-pyrrolo-isoquinoline derivatives

1 Introduction

Depression is a disabling psychological disorder affecting millions of people worldwide.[1] Low self-esteem, feelings of worthlessness, sadness, loss of interest or pleasure and suicidal thoughts are common depressive symptoms. However, despite the availability of various antidepressants and therapies, the biological mechanism underpinning depression remains unclear and the efficacy of the current treatment is still a concern.[2,3]

Drug development efforts have mostly been based on the monoamine hypothesis of depression, which states that depressive symptoms are attributed to low levels of monoamine neurotransmitters, specially serotonin and noradrenaline.[4] Currently, the first line antidepressants employed in the treatment of depression are the selective serotonin reuptake inhibitors (SSRIs). The SSRIs increase synaptic levels of serotonin by inhibiting its reuptake into the presynaptic cell. These compounds selectively target the serotonin transporter (SERT), and have an attenuated side-effect profile and similar efficacy in comparison with older antidepressants.[3,4] Nevertheless, these agents are not without shortcomings. Available antidepressants have modest response rates and have also been associated with a slow onset of action, *i.e.*, it requires several weeks for clinical improvement be observed. Furthermore, they have still been related to some unwanted side effects. All these factors underscore the need for the design of novel antidepressants and strategies for the treatment of depression.[3–5]

One approach used to improve the relief of some depressive symptoms, such as fatigue and cognitive impairment, in the initial period of treatment is the co-prescription of SSRIs with stimulants. Stimulants would provide symptomatic relief before the clinical improvement provided by antidepressants is observed, they could also assist those patients that present only a partial response to SSRIs. Histamine H_3 receptor antagonists have shown to improve cognition and wakefulness in animal models, without showing nonspecific stimulant effects.[6–8] Therefore, the dual inhibition of SERT and H_3 receptor by a single chemical entity is an interesting alternative for the treatment of depression. In this sense, Apodaca and colleagues (2012) developed a series of hexahydro-pyrrolo-isoquinoline compounds (Figure 1), which are modulators of SERT/H_3 receptor, *i.e.*, they work as potential antidepressant candidates.[9]

The use of computer-aided or in silico approaches has become a common practice in the design and optimization of novel chemical entities in the last decades. Many research groups as well as pharmaceutical companies have adopted these methods to complement experimental routines, because they contribute on the minimization of

[a] *D. R. Silva, L. Santos-Garcia, E. Fontes Ferreira da Cunha*
Department of Chemistry, Universidade Federal de Lavras, P.O. Box 3037, 37200-000, Lavras, MG, Brazil
phone/fax: +553538295129
E-mail: elaine_cunha@dqi.ufla.br
[b] *S. J. Barigye*
Department of Chemistry, McGill University, 801 Sherbrooke Street W., Montréal, QC, Canada H3 A 0B8

Supporting information for this article is available on the WWW under https://doi.org/10.1002/minf.201900024

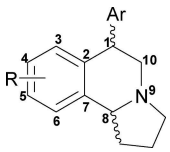

Figure 1. General structure of hexahydro-pyrrolo-isoquinoline derivatives that are modulators of SERT and H_3 receptor.[9]

time and costs required for chemical synthesis and biological evaluation.[10,11]

Quantitative Structure-Activity Relationship (QSAR) is an example of an in silico method that has been widely employed in the drug development process in the recent times.[12–14] This method is based on the premise that chemical structural differences in a set of compounds are closely associated with corresponding bioactivity differences. QSARs correlate chemical structures (encoded within molecular descriptors) to biological activities through mathematical models, which can then be applied to predict the activity of untested compounds.[12,15] There are several QSARs approaches, which include various molecular descriptors and modelling tools.[16–18] Among them, the 4D-QSAR formalism incorporates the conformational ensemble of the ligands into the model development. The molecular descriptors correspond to the grid cell occupancy measures of each pharmacophore interaction element, which allows a straightforward interpretation of how the individual parts of the compounds contribute to the observed activity. Then, based on this information, structural modifications can be proposed to further improve their performance.[19,20]

Therefore, this work aims to develop a quantitative relationship analysis between the structure of hexahydro-pyrrolo-isoquinoline derivatives and their inhibitory activity (expressed in terms of the inhibitory constant pK_I) to SERT and H_3 receptor, in order to identify important structural characteristics for the biological activity of these compounds. Finally, the most active compound was docked inside the binding pocket of human serotonin transporter (hSERT), recently deposited in the Protein Data Base (PDB),[21] to provide insight on key amino acids residues and intermolecular interactions.

2 Computational Methods

2.1 Biological Data

The dataset was comprised of hexahydro-pyrrolo-isoquinoline derivatives (Figure 1),[9] 108 of which possess H_3 receptor inhibitory activity and 43 SERT inhibitory activity, respectively. The structures and inhibitory constants, in terms of pK_I ($-\log K_I$), are given in Electronic Supporting Information (ESI, Table S1). It is worth to mention that all data were extracted from the same source in the literature.

In addition, since the biological tests were performed at physiological pH, the piperidine moiety of the compounds was considered in its protonated state. Such ionic interaction appears be relevant in the binding of compounds to SERT[21,22] as well as to H_3 receptor.[23–25]

2.2 4D-QSAR Methodology

Statistical correlation between the structure and activity for the hexahydro-pyrrolo-isoquinoline derivatives was established in light of the receptor independent 4D-QSAR formalism,[20] which involves the following steps.

2.3 Training and Test Sets

First of all, the dataset was split into two groups: the training set, used to build the QSAR models, and the test set, used to externally validate the models. There are several methods that can be employed for the test set selection.[26] In this work, three test sets were evaluated for each QSAR model, all set containing about 20% of the total number of compounds.[14] The selection was made through random splitting, Kennard-Stone sampling and cluster analysis (using the square of Euclidian distance and complete linkage) for the test set 1, 2 and 3, respectively. Test set compounds are given in ESI.

2.4 Molecular Dynamics Simulation

The molecular dynamics simulation step is carried out to generate the conformational ensemble profile (CEP) for each ligand, the fourth dimension in the 4D-QSAR formalism.[20] The simulation was performed using the Molsim 3.0 package and MM2 force field, available in the 4D-QSAR program. Temperature was set at 300 K, similar to the one used in the pharmacological essays, with a simulation sampling time of 50 ps and intervals of 0.001 ps. Therefore, a total of 500 conformations for each ligand were obtained. To modulate the solvent effects in the absence of an explicit solvent, a distance-dependent dielectric constant of 3.5 was applied.[27,28]

2.5 Alignment and Interaction Pharmacophore Elements Definition

The CEP for all compounds was then overlaid into a three-dimensional cubic box, according to the three-ordered atom alignment. This three-dimensional box is comprised of cubic grid cells that record the distribution of the spatial occupancy of each atom in ligand. In this work, a grid cell size of 1 Å was employed and ten different alignments were tested. It was assumed that the compounds bind to the

receptor in a similar mode, since they are structural analogues, and the alignment atoms were selected to span the common structural moiety of the compounds.[29,30] The three-ordered atom alignments tested are **(1)** 4-7-10, **(2)** 4-3-2, **(3)** 5-6-7, **(4)** 3-7-1, **(5)** 4-6-2, **(6)** 5-2-9, **(7)** 6-2-8, **(8)** 2-8-10, **(9)** 7-1-9 and **(10)** 5-3-7 (atoms number is in accordance to Figure 1).

Each atom in a ligand was classified according to seven types of Interaction Pharmacophore Elements (IPE), which correspond to possible interactions that may take place between the ligand and receptor. The IPEs are related to the pharmacophore groups: non-specific (any), non-polar (np), polar positive (p+), polar negative (p−), hydrogen bond acceptor (hba), hydrogen bond donor (hbd) and aromatic (ar).[20] Thus, the occupation of the grid cells by the IPEs provides the grid cell occupancy descriptors (GCOD), the dependent variables used to build the 4D-QSAR models.[20]

2.6 Construction of the 4D-QSAR Models

The 4D-QSAR analysis generates a huge number of GCODs. Therefore, in order to select the most relevant variables to explain the biological response, two variable selection criteria were employed: in the first, GCODs with variance equal to zero were excluded; and in the second, the 2000 GCODs with higher weight in the Partial Least Square (PLS) analysis were selected.

The 4D-QSAR models were built using a Genetic Function Approximation (GFA)[31] based optimization algorithm, as implemented in the 4D-QSAR program.[20] Herein, only linear terms were used in a multiple linear regression (MLR) analysis. Therefore, the selected GCODs and the inhibitory constants were optimized according to a GA-MLR optimization for model building. The optimization was initiated with 100 randomly generated models with a mutation probability of 50%. The number of crossover operations were set to a range of 30000–90000, and the smoothing factor from 0.1 to 1.0. The models were ranked with respect to the Friedman's lack-of-fit (LOF) measure,[31] the penalized least square error (LSE), and the best model from each alignment was selected for further validation.

The best models were validated using the leave-one-out cross validation, external validation (the model was used to predict the activity of the test set compounds) and y-randomization.[26,32] The following parameters were considered: the correlation coefficient for calibration (R^2), cross-validated correlation coefficient (Q^2), correlation coefficient for external validation (R^2_{pred}), modified R^2 for external validation (r^2_m), correlation coefficient for y-randomization (R^2_{rand}) and penalized correlation coefficient for y-randomization (R^2_p).[32,33] Moreover, the application domain was analysed in order to establish the model's chemical space.

2.7 Representative Conformation Selection

The last step in the 4D-QSAR analysis is the definition of the representative conformation for each ligand, *i.e.* the structure that would most resemble the active conformation. This is achieved by identifying those conformers that are within a defined ΔE relative to the energy minimum (here a $\Delta E = 5$ kcal mol^{-1} was employed) for each CEP. Then, from these low energy conformers, the one with the highest activity as predicted by the 4D-QSAR model is selected as the representative conformation.

2.8 Molecular Docking

To better understand the binding mode of hexahydro-pyrrolo-isoquinoline derivatives, the most active compound was docked inside the active site of hSERT, using the X-ray structure of hSERT co-crystallized with paroxetine at 3.14 Å resolution (PDB ID: 5I6X).[21] The hSERT is responsible for the reuptake of the serotonin neurotransmitter and its structure is comprised of 12 transmembrane helices. The molecular docking was performed using the Molegro Virtual Docker (MVD) program.[34] The MolDock score function was employed, this is an extension of the piecewise linear potential (PLP) including improved electrostatic and hydrogen bond terms.[34] Default parameters were used, except for the number of runs that was increased to 100, wherein each run returns one pose. The best pose was selected based on the lowest MolDock re-ranking score, a more accurate function which considers additional terms for the Lennard-Jones 12-6 potential and the sp2-sp2 torsion.[34]

3 Results and Discussion

Separate 4D-QSAR models were developed for the hexahydro-pyrrolo-isoquinoline derivatives' SERT and H_3 receptor inhibitory activities, respectively. The models were built using the inhibitory constant (pK_i) as response variable and the molecular descriptors from 4D-QSAR as dependent variables. Separate discussions are thus provided for each of the modelled activities.

3.1 Histamine H_3 Receptor

In this work three different training and test sets were considered, to attest each model's earnest predictivity and robustness.[14] Therefore, the 4D-QSAR model for each alignment is considered as an average of the three test and training sets. Statistical parameters for the 4D-QSAR models are shown in Table 1.

As can be observed in Table 1, all models presented adequate values for the calibration parameters R^2 and R^2_{adj}, ranging from 0.74 to 0.81 and 0.73 to 0.79, respectively.

Table 1. Mean statistical parameters of the calibration, internal and external validation for the 10 alignments evaluated in the 4D-QSAR models developed for the H_3 receptor.

Align	GCOD	R^2	R^2_{adj}	Q^2	Q^2_{adj}	R^2_{ext} [a]	r^2_m	R^2_{rand}	R^2_p
1	7	0.79	0.77	0.75	0.73	0.79	0.73	0.09	0.74
2	7	0.80	0.79	0.75	0.74	0.75	0.55	0.08	0.76
3	7	0.79	0.77	0.72	0.70	0.77	0.68	0.08	0.75
4	7	0.81	0.79	0.74	0.72	0.79	0.67	0.07	0.77
5	6	0.75	0.74	0.70	0.68	0.73	0.58	0.08	0.71
6	7	0.78	0.77	0.74	0.72	0.74	0.59	0.08	0.74
7	6	0.78	0.76	0.73	0.71	0.78	0.70	0.07	0.74
8	6	0.76	0.75	0.68	0.66	0.73	0.54	0.06	0.73
9	7	0.76	0.75	0.72	0.70	0.77	0.67	0.08	0.72
10	6	0.74	0.73	0.66	0.63	0.81	0.75	0.07	0.71

[a] The $RMSE_p$ (root mean square error of the external validation) for the 4D-QSAR models developed from alignment 1 to 10 are: 0.35, 0.37, 0.36, 0.34, 0.37, 0.36, 0.36, 0.37, 0.35 and 0.34, respectively.

These parameters reflect the goodness of fit of the models; the R^2_{adj} is used to compare models with different numbers of variables. In addition, all models also presented adequate values for the leave-one-out cross validation parameters Q^2 and Q^2_{adj}, ranging from 0.66 to 0.75 and 0.63 to 0.74, respectively. The cross validation procedure assesses the model robustness, it tests the statistical significance of each component of the model. In respect to y-randomization, the models have R^2 higher than R^2_{rand} and R^2_p higher than 0.5,[33] thereby assuring the inexistence of chance correlation. Finally, all models presented adequate values for the external validation parameter R^2_{ext} and r^2_m, which are also higher than 0.5.[33] Therefore, it can be concluded that the developed models have good performance, with suitable goodness of fit, robustness and predictivity.

To select the best model to represent the dataset, we selected the models with the best performance in all validations. The interpretation of the molecular descriptors contained in best model is performed to understand better the structural features that are essential for the H_3 receptor inhibitory activity. With this in mind, the models from alignments 1, 2 and 4 were selected, their statistical parameters are: alignment 1: $R^2=0.79$ and $Q^2=0.75$; alignment 2: $R^2=0.80$ and $Q^2=0.75$; alignment 4: $R^2=0.81$ and $Q^2=0.74$. Nevertheless, when the applicability domain is taken into account, the model from alignment 1 has more outliers (the applicability domain is discussed in more details below; see Figure S1 in the ESI and Figure 4 for the William's plot[26]). Furthermore, if we compare models 2 and 4 in the external validation, model 4 performs slightly better ($R^2_{ext}=0.75$ and 0.79, respectively). Therefore, model 4 was selected for further evaluation and it is represented below:

$$pK_i = 9.95 - 55.36\ (2,-7,-2,p-) + 28.59\ (-6,-8,-2,any) + 23.06\ (4,-9,-1,any) + 4.96\ (1,4,1,np) + 14.76\ (-6,-8,2,np) - 2.46\ (0,1,-1,any) - 50.05\ (3,-3,-5,np)$$

$$M = 86\ LOF = 0.15\ LSE = 0.11$$

The 4D-QSAR model was used to predict the activity of training and test set compounds. The experimental and predicted activities along with residual values are given in ESI (Table S2).

In the 4D-QSAR formalism, the molecular descriptors are represented by the Cartesian coordinates and the Interaction Pharmacophore Element – "(x, y, z, IPE)" – which corresponds to prospective interactions between the ligand and receptor. The model described above comprises of 7 descriptors, 4 with positive coefficients and 3 with negative coefficients. In this sense, positive coefficients represent those interactions that favor ligand-receptor binding. On the other hand, negative coefficients represent those interactions that disfavor ligand-receptor binding.

Figure 2 depicts the GCODs positions on the most active compound (1B) as well as the superposition of all compounds in the dataset. Positive and negative GCODs are represented by light and dark spheres, respectively. The occupation frequency and correlation matrix of all GCODs are given in ESI (Table S3 and S4, respectively).

As can be observed, six GCODs are located near the substituent R, which varies with regard to the composition and position on the aromatic ring (Figure 1 and Table S1). The GCOD (0,1,−1,any) has a negative coefficient, a non-specific interaction and higher occupation frequency for compounds with substituent R on positions 3 and 4 of the ring (Figure 2). Therefore, this descriptor indicates that groups on this position decrease the activity of the compounds. In fact, most compounds that possess the R group on position 3 have lower pK_i values (Table S1). For example, the difference in the activity of compounds 1B–1C ($pK_i=9.70$ and 7.17, respectively) and 4A–4C ($pK_i=9.00$ and 7.01, respectively), which only differ at the position of the R group, are higher than two logarithm units. Likewise, the same trend can be observed for compounds with R group on position 4, as exemplified by compounds 1A–45 ($pK_i=$

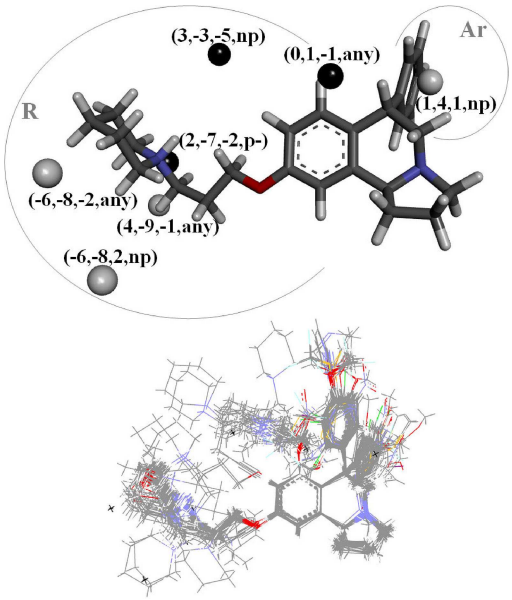

Figure 2. Top – GCODs from alignment 4 positioned on the most active compound, 1B. Light and dark spheres represent descriptors with positive and negative coefficient, respectively. Bottom – superposition of all hexahydro-pyrrolo-isoquinoline compounds studied along with the GCODs (x).

9.05 and 7.80, respectively) and 1B–44 ($pK_i = 9.70$ and 9.00, respectively). The only exception is compounds 97A and 98A, which have an alkyne group on substituent R and are almost equally active ($pK_i = 8.52$ and 8.40, respectively).

The GCOD (3,−3,−5,np) has a negative coefficient, a non-polar interaction and occupation frequency only for compounds with substituent R on the position 3 of the aromatic ring (Figure 2). Thus, it also indicates that groups at this position disfavor the biological activity of the compounds. Compounds with higher occupation frequency for each descriptor are represented in Figure 3.

The GCOD (−6,−8,−2,any) has a positive coefficient, a non-specific interaction and higher occupation frequency for compound 11A. In this compound, the substituent R is attached to the aromatic ring at position 5 (Figure 2) and this descriptor is located near the piperidine moiety of the compound (Figure 3). Therefore, the GCOD would indicate that this is the optimal position for the R group, which is in agreement with the information discussed for the previous descriptors.

The GCOD (2,−7,−2,p−) has a negative coefficient, a polar negative interaction and higher occupation frequency for compounds 11A and 50B. This descriptor is in the proximity of the nitrogen atom of the piperidine and morpholine groups, respectively (Figure 3). Nevertheless, given that all compounds possess a nitrogen atom in this position, it appears that the information provided by this descriptor is related to the orientation of the group. One of the main structural characteristics of H_3 receptor antagonists is the presence of at least one amino group.[35] An important ionic interaction seems to take place between the amino group and a negatively charged amino acid residue in the H_3 receptor active site.[36] In this sense, the GCOD (2,−7,−2,p−) would be related to the conformation adopted by the R substituent and its impact on the activity of the compounds.

The descriptors (4,−9,−1,any) and (−6,−8,2,np) are also near the R substituent at position 5. Both GCODs have positive coefficients; the former represents a non-specific interaction while the latter represents a non-polar interaction. However, these descriptors are positioned far from the structure of the compounds and are thus probably associated with the flexibility of the R group. As can be seen in Figure 2, this part of the structure varies greatly among the structures, *i.e.* this group experienced high conformation freedom in the molecular dynamic simulation.

Lastly, GCOD (1,4,1,np) are located near the aromatic substituent of the compounds. This descriptor has a negative coefficient, a non-polar interaction and higher occupation frequency for compound 4D (Figure 3). It worth mentioning that this descriptor has occupation frequency only for compounds with the aromatic ring in the *trans* orientation (like compound 4D), regardless the chemical nature and position of the group on the ring. As can be observed in the dataset, generally, compounds with the aromatic ring in the *trans* orientation are equally or more

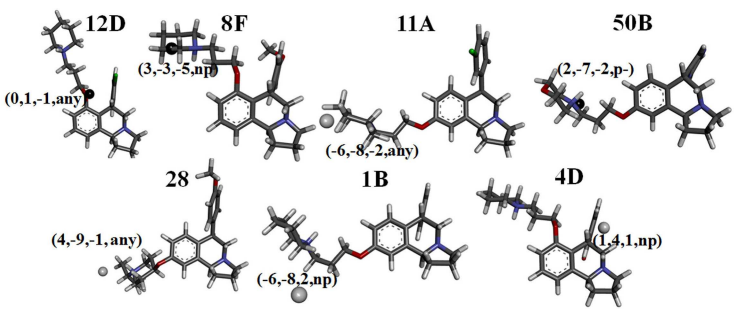

Figure 3. Descriptors from alignment 4 positioned on the compounds with higher occupation frequency.

Table 2. Mean statistical parameters of the calibration, internal and external validation for the 10 alignments evaluated in the 4D-QSAR models developed for the SERT.

Align	GCOD	R^2	R^2_{adj}	Q^2	Q^2_{adj}	R^2_{ext} [b]	r^2_m	R^2_{rand}	R^2_p
1	4	0.68	0.65	0.57	0.52	0.34	0.11	0.12	0.62
2	5	0.73	0.69	0.62	0.57	0.33	0.12	0.16	0.65
3	6	0.75	0.71	0.61	0.54	0.35	0.23	0.17	0.67
4	6	0.76	0.72	0.61	0.54	0.77	0.48	0.20	0.65
5	6	0.61	0.54	0.36	0.24	0.21	0.03	0.17	0.52
6	4	0.64	0.61	0.47	0.42	0.35	0.14	0.11	0.58
7	6	0.82	0.79	0.71	0.66	0.76	0.58	0.17	0.73
8	6	0.71	0.66	0.57	0.49	0.40	0.09	0.16	0.62
9	6	0.73	0.68	0.50	0.41	0.12	0.00	0.18	0.63
10	6	0.76	0.71	0.63	0.56	0.47	0.10	0.21	0.64

[b] The $RMSE_p$ for the 4D-QSAR models developed from alignment 1 to 10 are: 0.43, 0.47, 0.51, 0.28, 0.95, 0.46, 0.33, 0.42, 0.63 and 0.38, respectively.

potent than their respective analogues in the *cis* configuration, for example compounds 1A–1B ($pK_i = 9.05$ and 9.70, respectively), 5A–5B ($pK_i = 8.70$ and 9.00, respectively) and 7A–7B ($pK_i = 8.30$ and 8.70, respectively). Furthermore, the inhibitory activity seems to be little influenced by the substituent position on the aromatic ring. For instance, when the hydroxyl group is in the para (9A, $pK_i = 8.70$), meta (15, $pK_i = 9.15$) or orto (16, $pK_i = 8.40$) positions, the difference in their activities is small.

A pharmacophoric model developed by Lebois et al.[35] identified key characteristics conserved in the structure of H_3 receptor antagonists. In general, the structures contain an amino group bridged by an alkyl chain to an aromatic ring. This aromatic ring is linked to a variety of chemical groups, ranging from a non-polar to another amino group. This is in agreement with the information provided by GCOD (1,4,1,np), which indicates that substituents attached to the aromatic ring have little influence on the biological activity of these compounds.

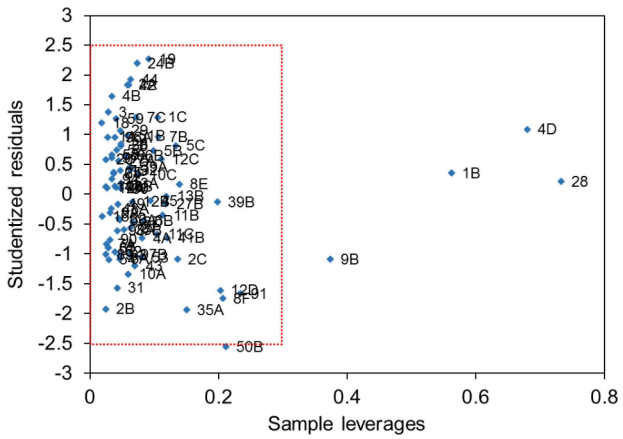

Figure 4. Applicability domain according to the plot of sample leverages against Studentized residuals for the 4D-QSAR model from alignment 4 of the H_3 receptor.

Finally, the applicability domain of the 4D-QSAR model was assessed, using the William's plot[26] (Figure 4), to identify possible atypical behavior. As can be observed in Figure 4, compound 50B is the only activity outlier[15] (the residual value is slightly below the lower limit), its inhibitory activity is overestimated by the model ($pK_i = 6.74$ and $pK_{i,pred} = 7.51$, Table S2). Examining the structures of the training set, compounds with a pyridine moiety possess high activity ($pK_i > 8$). The introduction of a second nitrogen atom, to form the pyrazine group of 50B, decreases the inhibitory constant to 6.74. Therefore, it seems that the model was not able to capture this difference. Regarding the samples leverages, four compounds (9B, 1B, 4D and 28) are above the stipulated limit, and so they are classified as structural outliers.[15] Compound 28, for example, which possesses the highest leverage, is the only compound in the training set with no carbon atom linking the aromatic ring to the piperidine moiety in the substituent R (Table S1).

3.2 Serotonin Transporter

The same procedure applied to the H_3 receptor was used to develop 4D-QSAR models for the serotonin transporter (SERT). The hexahydro-pyrrolo-isoquinoline compounds were split into training and test sets, the separation was made by three different approaches, and the 4D-QSAR model for each alignment is considered as an average of the three test and training sets. Table 2 shows the statistical parameters of the 4D-QSAR models.

The 4D-QSAR models presented calibration parameters R^2 and R^2_{adj} ranging from 0.61 to 0.82 and 0.54 to 0.79, respectively. The models from alignments 1, 5 and 6 yielded R^2 lower than 0.7. In addition, models from alignments 1, 2, 5, 6, 8 and 9 exhibited R^2_{adj} lower than 0.7. Therefore, these models were not considered in the subsequent analysis. The four remaining models presented leave-one-out cross validation parameters Q^2 and Q^2_{adj} higher than 0.5, as recommended in the literature.[37] As for y-randomization,

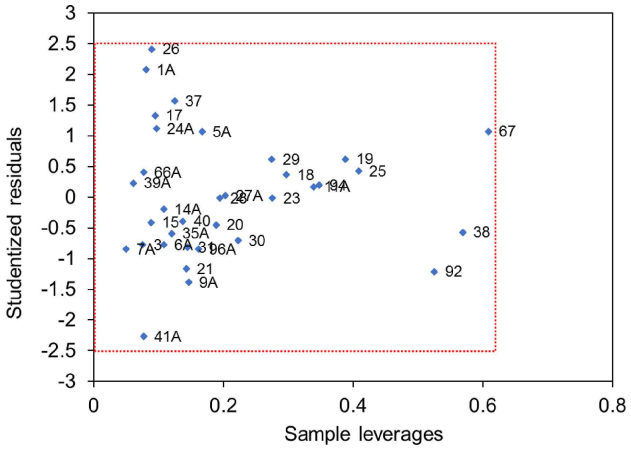

Figure 5. Applicability domain according to the plot of sample leverages against Studentized residuals for the 4D-QSAR model from alignment 7 of the SERT.

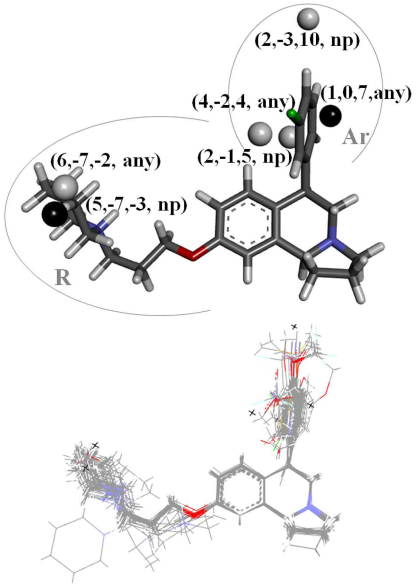

Figure 6. Top – GCODs from alignment 7 positioned on the most active compound, 11 A. Light and dark spheres correspond to positive and negative coefficient, respectively. Bottom – superposition of all hexahydro-pyrrolo-isoquinoline compounds studied along with the GCODs (x).

the models presented R^2 higher than R^2_{rand} and R^2_p higher than 0.5.[33] Regarding the parameters of the external validation, the models from alignments 4 and 7 have R^2_{pred} of 0.77 and 0.76, respectively. However, only the model from alignment 7 yielded r^2_m higher than 0.5.[33] In addition, by analyzing the applicability domain using the William's plot,[26] model 7 has no outliers (Figure 5) while model 4 (as well as other models that perform well in the internal validation, but poorly in the external validation, such as models 2, 3 and 10) has at least one (Figure S2). Therefore, taking into account the performance in all validation procedures, model 7 was selected to represent the dataset and posteriorly considered in the descriptors interpretation. The 4D-QSAR model from alignment 7 is shown below:

$$pK_i = 8.93 + 28.27\ (6,-7,-2,any) - 17.79\ (5,-7,-3,np)$$
$$-11.76\ (1,0,7,any) + 2.51\ (2,-1,5,np)$$
$$+9.72\ (2,-3,10,np) + 2.09\ (4,-2,4,any)$$

$$M = 34\ LOF = 0.11\ LSE = 0.05$$

The 4D-QSAR model for SERT was also used to predict the activity of training and test set compounds. The experimental and predicted activities along with residual values are given in the ESI (Table S5).

The model contains 6 molecular descriptors, 2 with negative coefficients, corresponding to those interactions that disfavor ligand-receptor binding, and 4 with positive coefficient, corresponding to those interactions that favor ligand-receptor binding.

Figure 6 depicts the GCODs positions on the most active compound (in this case, compound 11A) along with the superposition of all compounds in the dataset. The occupation frequency and correlation matrix for all GCODs are given in ESI (Table S6 and S7, respectively).

As can be observed in Figure 6, four GCODs are in proximity of the aromatic substituent of the compounds. The GCOD (1,0,7,any) has a negative coefficient, a non-specific interaction and higher occupation frequency for compound 18. Compounds with higher occupation frequency for each descriptor are represented in Figure 7. This descriptor has occupation frequency for all compounds, except for 92. This compound is the only one with S stereochemistry of the carbon connected to the aromatic ring. When the compounds 38 ($pK_i = 8.15$), 92 ($pK_i = 8.40$) and 94 ($pK_i = 6.95$) are compared, their structural differences consist only in the stereochemistry (Table S1), compound 92 is the most active. Therefore, it seems that the orientation of the aromatic ring is relevant for the potency of the compounds, but there is little information about this in the dataset.

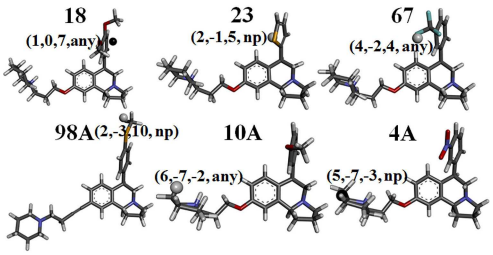

Figure 7. Descriptors from alignment 7 positioned on the compounds with higher occupation frequency.

The GCOD (2,−1,5,np) has a positive coefficient, a non-polar interaction and higher occupation frequency for compound 23 (Figure 7). The descriptor is located on the sulphur atom of thiophene, thus indicating that groups in this position favor the biological activity. It worth mentioning that this GCOD does not have any occupation frequency for compounds with groups in the orto position of the aromatic ring. In fact, the activity greatly changes if the group is in the para, meta or orto positions, as exemplified by compounds 9A ($pK_i = 7.24$), 15 ($pK_i = 8.52$) and 16 ($pK_i = 6.95$), respectively. Among them, compound 16 (orto substituted) is the least active. The same trend can be observed for compounds 66A ($pK_i = 8.22$), 11A ($pK_i = 9.22$) and 12A ($pK_i = 8.10$).

The GCOD (4,−2,4,any) has a positive coefficient, a non-specific interaction and higher occupation frequency for compound 67 (Figure 7). This descriptor is located near to the group at the meta position of the aromatic ring, thus indicating that groups in this position favor the activity of the compounds. In general, compounds with polar groups at the meta position have high pK_i values. For instance, compounds 67 ($pK_i = 8.05$) and 7A ($pK_i = 7.92$), 15 ($pK_i = 8.52$) and 9A ($pK_i = 7.24$), as well as 11A ($pK_i = 9.22$) and 66A ($pK_i = 8.22$). In all cases, structures with the group at the meta position are more active than those at the para position.

The GCOD (2,−3,10,np) has a positive coefficient, a non-polar interaction and higher occupation frequency for compound 98A. In this compound, the descriptor is located near the methyl group at the para position of the aromatic ring (Figure 7), thereby suggesting that non-polar groups in this position favor the activity of the compounds. According to the inhibitory constant, compounds with an ether or thioether group at the para position have high biological activity ($pK_i > 8$).

Lastly, the GCODs (6,−7,−2,any) and (5,−7,−3,np) are positioned in the proximity of the R substituent (Figure 6). The GCOD (6,−7,−2,any) has a positive coefficient, a non-specific interaction and higher occupation frequency for compounds 10A and 11A (Figure 7). On the other hand, the GCOD (5,−7,−3,np) has a negative coefficient, a non-polar interaction and higher occupation frequency for compound 4A (Figure 7). Both descriptors are located near the piperidine moiety, and thus seem to be related to the spatial orientation of the R group.

3.3 Molecular Docking

Recently, the first X-ray structure of the human serotonin transporter (hSERT) co-crystallized with the antidepressants escitalopram and paroxetine was reported in the literature.[21] To the best of our knowledge, there is no study on the interaction of hexahydro-pyrrolo-isoquinoline derivatives with hSERT. Therefore, in order to better understand the binding mode of this class of compounds, the most active compound, namely 11A, was docked inside the binding pocket of the hSERT using the MVD program.[34] Binding energies (Table 3) and the structure of 11A within the binding pocket (Figure 8) are given above.

Table 3. Energies from molecular docking of compounds 11 A, the most active in the dataset, inside the binding pocket of hSERT.

Compound	MolDock Score	Rerank Score	Interaction Energy
11 A	−143.60	−116.79	−157.61

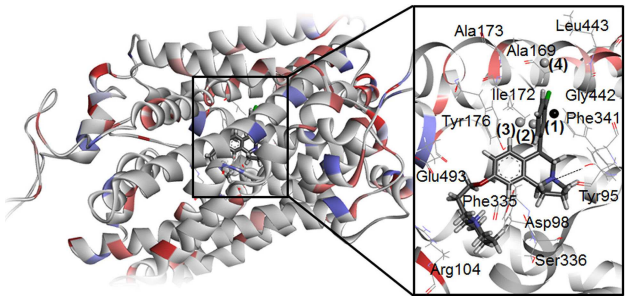

Figure 8. Left – Tridimensional structure of compounds 11 A within the binding pocket of the hSERT. Right – Binding site of hSERT with key amino acid residues highlighted along with molecular descriptors from 4D-QSAR model (1: GCOD (1,0,7,any), 2: GCOD (2,−1,5,np), 3: GCOD (4,−2,4,any) and 4: GCOD (2,−3,10,np)).

In Figure 8, the main amino acid residues of the hSERT binding pocket are highlighted. As can be observed, the amino group of the hexahydro-pyrrolo-isoquinoline portion are in the region delimited by Tyr95, Asp98 and Ser336 amino acids, similar to the amino group of paroxetine.[21] The hydroxyl group of Tyr95 is within a 3.52 Å distance of the nitrogen atom of 11A, interacting with it through a weak hydrogen bond (black dashed line in Figure 8). The R group of 11A adopted a conformation in which it interacts with Phe335, Arg104, Ser336 and Glu493 amino acid residues at the outer region of the binding site. Furthermore, the aromatic substituent is positioned in a more hydrophobic region delimited by the side chain of the Tyr176, Ile172, Gly442, Phe341 and Leu443 amino acid residues.

Moreover, we sought to validate the results obtained with the 4D-QSAR analysis using the molecular docking perspective. To this end, the GCODs were mapped into the protein-ligand interaction model obtained with the docking strategy (Figure 8). Two descriptors, *i.e.* GCODs (6,−7,−2, any) and (5,−7,−3,np) could not be mapped into the docking model due to the different orientations adopted by the R group in each method (Figure 7 and 8). In Figure 8, the GCOD (1,0,7,any)**(1)** is near the aromatic substituent of 11A, surrounded by Ile172, Phe341 and Gly442 amino acids. The GCODs (2,−1,5,np)**(2)** and (4,−2,4,any)**(3)** are positioned between the aromatic substituent of 11A and the side chain of Tyr176. The two aromatic rings are almost parallel to

each other, thus they could interact through aromatic interactions, and the nature of the substituent group may affect the strength of such an interaction. Furthermore, Tyr176 is the amino acid that most contributes to the stability of 11A ($E = -16.85$ kcal mol^{-1}) in the binding site. Finally, the GCOD (2,−3,−10,np)(**4**) is on the top of the aromatic ring of 11A, near to the Ala169, Ala173 and Leu443 amino acids. Consequently, non-polar groups on the para position of the ring would improve the interaction with the side chain of these amino acid residues.

In a nutshell, the docking of the most active compound (11A) inside the binding pocket of hSERT aided to further unravel the information from 4D-QSAR analysis, wherein the descriptors position could be related to key intermolecular interactions with some amino acid residues. Therefore, all this information together can assist the design of novel antidepressants for the treatment of depression, with improved inhibitory activity to H_3 receptor and SERT.

4 Conclusions

The biological activity of the hexahydro-pyrrolo-isoquinoline derivatives to the H_3 receptor and SERT was computationally evaluated by the 4D-QSAR formalism and molecular docking, since these compounds have the potential to act as antidepressants. The 4D-QSAR models were separately developed for each receptor, in order to characterize key structural features relevant for the activity of these compounds. The best model for each receptor presented adequate statistical parameters that ensure their goodness of fit, robustness and predictivity (histamine: $R^2 = 0.81$, $Q^2 = 0.74$, $R^2_{pred} = 0.78$ and $R^2_p = 0.77$; serotonin: $R^2 = 0.82$, $Q^2 = 0.71$, $R^2_{pred} = 0.76$ and $R^2_p = 0.73$), in addition to the important pharmacophore groups used to explain the biological activity of the compounds. Through the interpretation of the molecular descriptors, it was observed that the inhibitory activity to H_3 receptor is mainly influenced by the R group, while changes in the aromatic substituent greatly affect the activity towards SERT. The binding mode to SERT was assessed by docking the most active compound inside the active site, wherein the positions of the molecular GCOD descriptors could be related to specific intermolecular interactions with amino acid residues. Therefore, structural modifications can be proposed to increase the affinity of these compounds to both receptors and, consequently, to improve the biological activity profile and broaden the therapeutic options for the treatment of depression.

Acknowledgements

The authors are thankful to the Brazilian agencies "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" (CAPES), "Conselho Nacional de Desenvolvimento Científico e Tecnológico" (CNPq) and "Fundação de Amparo ao Ensino e Pesquisa de Minas Gerais" (FAPEMIG) for the financial support, and to A. J. Hopfinger who kindly provided the 4D-QSAR program for academic use.

References

[1] WHO, "Depression," can be found under http://www.who.int/topics/depression/en/, **2017**.
[2] B. Bondy, *Dialogues Clin. Neurosci.* **2002**, *4*, 7–20.
[3] K. Pytka, K. Podkowa, A. Rapacz, A. Podkowa, E. Zmudzka, A. Olczyk, J. Sapa, B. Filipek, *Pharmacol. Rep.* **2016**, *68*, 263–274.
[4] H. Sharma, S. Santra, A. Dutta, *Future Med. Chem.* **2015**, *7*, 2384–2405.
[5] F. Artigas, *ACS Chem. Neurosci.* **2013**, *4*, 5–8.
[6] J. M. Keith, L. a. Gomez, A. J. Barbier, S. J. Wilson, J. D. Boggs, B. Lord, C. Mazur, L. Aluisio, T. W. Lovenberg, N. I. Carruthers, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 4374–4377.
[7] J. M. Keith, L. A. Gomez, R. L. Wolin, A. J. Barbier, S. J. Wilson, J. D. Boggs, C. Mazur, I. C. Fraser, B. Lord, L. Aluisio, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2603–2607.
[8] E. M. Stocking, M. A. Letavic, P. Bonaventure, N. I. Carruthers, *Curr. Top. Med. Chem.* **2010**, *10*, 596–616.
[9] R. Apodaca, A. J. Barbier, N. I. Carruthers, L. A. Gomez, J. M. Keith, T. W. Lovenberg, R. L. Wolin, **2012**, US_20120321559_A1.
[10] I. M. Kapetanovic, *Chem.-Biol. Interact.* **2008**, *171*, 165–176.
[11] S. J. Y. Macalino, V. Gosu, S. Hong, S. Choi, *Arch. Pharmacal Res.* **2015**, *38*, 1686–1701.
[12] J. Verma, V. M. Khedkar, E. C. Coutinho, *Curr. Top. Med. Chem.* **2010**, *10*, 95–115.
[13] C. H. Andrade, K. F. M. Pasqualoto, E. I. Ferreira, A. J. Hopfinger, *Molecules* **2010**, *15*, 3281–3294.
[14] A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, *J. Med. Chem.* **2014**, *57*, 4977–5010.
[15] A. Tropsha, *Mol. Inf.* **2010**, *29*, 476–488.
[16] A. Tropsha, A. Golbraikh, *Curr. Pharm. Des.* **2007**, *13*, 3494–3504.
[17] M. A. Lill, *Drug Discovery Today* **2007**, *12*, 1013–1017.
[18] S. Pirhadi, F. Shiri, J. B. Ghasemi, *RSC Adv.* **2015**, *5*, 104635–104665.
[19] G. G. Cardoso, D. R. Silva, S. J. Barigye, E. F. F. da Cunha, *Mol. Inf.* **2017**, *36*, 1700080.
[20] A. J. Hopfinger, S. Wang, J. S. Tokarski, B. Jin, M. Albuquerque, P. J. Madhav, C. Duraiswami, *J. Am. Chem. Soc.* **1997**, *119*, 10509–10524.
[21] J. A. Coleman, E. M. Green, E. Gouaux, *Nature* **2016**, *532*, 334–339.
[22] J. Andersen, O. Taboureau, K. B. Hansen, L. Olsen, J. Egebjerg, K. Stromgaard, A. S. Kristenen, *J. Biol. Chem.* **2009**, *284*, 10276–10284.
[23] F. U. Axe, S. D. Bembenek, S. Szalma, *J. Mol. Graphics Modell.* **2006**, *24*, 456–464.
[24] N. Levoin, O. Labeeuw, S. Krief, T. Calmels, O. Poupardin-Olivier, I. Berrebi-Bertrand, J. M. Lecomte, J. C. Schwartz, M. Capet, *Bioorg. Med. Chem.* **2013**, *21*, 4526–4529.
[25] B. Schlegel, C. Laggner, R. Meier, T. Langer, D. Schnell, R. Seifert, H. Stark, H. D. Höltje, W. Sippl, *J. Comput.-Aided Mol. Des.* **2007**, *21*, 437–453.
[26] P. Gramatica, *QSAR Comb. Sci.* **2007**, *26*, 694–701.
[27] E. F. F. da Cunha, R. C. A. Martins, M. G. Albuquerque, R. B. de Alencastro, *J. Mol. Model.* **2004**, *10*, 297–304.

[28] E. F. F. da Cunha, M. G. Albuquerque, O. A. C. Antunes, R. B. de Alencastro, *QSAR Comb. Sci.* **2005**, *24*, 240–253.
[29] K. F. M. Pasqualoto, E. I. Ferreira, O. A. Santos-Filho, A. J. Hopfinger, *J. Med. Chem.* **2004**, *47*, 3755–3764.
[30] N. C. Romeiro, M. G. Albuquerque, R. B. de Alencastro, M. Ravi, A. J. Hopfinger, *J. Comput.-Aided Mol. Des.* **2005**, *19*, 385–400.
[31] D. Rogers, A. J. Hopfinger, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.
[32] A. Tropsha, P. Gramatica, V. K. Gombar, *QSAR Comb. Sci.* **2003**, *22*, 69–77.
[33] P. P. Roy, S. Paul, I. Mitra, K. Roy, *Molecules* **2009**, *14*, 1660–1701.
[34] R. Thomsen, M. H. Christensen, *J. Med. Chem.* **2006**, *49*, 3315–3321.
[35] E. P. Lebois, C. K. Jones, C. W. Lindsley, *Curr. Top. Med. Chem.* **2011**, *11*, 648–660.
[36] R. Kiss, G. M. Keserű, *Expert Opin. Drug Discovery* **2016**, *11*, 1165–1185.
[37] R. Kiralj, M. M. C. Ferreira, *J. Braz. Chem. Soc.* **2009**, *20*, 770–787.

Received: February 15, 2019
Accepted: May 7, 2019
Published online on May 27, 2019

