J Mol Model (2014) 20:2420
DOI 10.1007/s00894-014-2420-4

ORIGINAL PAPER

Application of 4D-QSAR studies to a series of benzothiophene analogs

Giovana Baptista Caldas · Teodorico C. Ramalho · Elaine F. F. da Cunha

Received: 30 April 2014 / Accepted: 6 August 2014 / Published online: 16 September 2014

Abstract Four-dimensional quantitative structure-activity relationship (4D-QSAR) analysis was applied to a series of 52 benzothiophene analogs synthesized by Hiroshi Yamashita et al. (2011, United Sates Patent no. US8,349,840) and evaluated as dopamine D2 receptor inhibitors. The QSAR equations, generated by a combined scheme of genetic algorithms (GA) and partial least squares (PLS) regression, were evaluated by leave-one-out cross-validation, using a training and test set of 42 and ten compounds, respectively. Four different alignments were tested, and model 2, generated from Eq. 10, showed the best statistical results; it was therefore chosen to represent the data set. This study allowed a quantitative prediction of compounds potency and supported the design of the new benzothiophene.

Keywords Benzothiophene · Dopamine receptor · QSAR

Introduction

Dopamine is an important neurotransmitter present in the brain. In fact, it plays a crucial role as a regulator of many physiological functions in the central nervous systems, as well as in the periphery, as a modulator of cardiovascular and renal functions, among others [1]. It is well-known that dopamine receptors are a class of G protein-coupled receptors. The D1 and D5 receptors are coupled to the Gs alpha and cause the conversion of adenosine triphosphate to cyclic adenosine monophosphate (cAMP). In contrast, the D2, D3, and D4 receptors couple to Gi alpha and inhibit adenylyl cyclase, so that the concentration of cAMP is reduced [2]. All of these receptor subtypes have been identified in specific regions of the central nervous system, where they play critical roles in the regulation of cognitive function and motor control [3, 4]. Therefore, the D2 receptor, also known as D2R, is of great interest since it is the main pharmacological target of the typical, or classic, antipsychotics (such as haloperidol, chlorpromazine, and fluphenazine). Those drugs are effective in the treatment of positive symptoms of schizophrenia (mental disorder) by blocking dopamine D2 receptors in the mesolimbic area [5]. Since the discovery of dopamine and its receptors, several research works have been conducted to aggregate knowledge about the functionality of dopamine receptors and the search for new drugs for the treatment of psychological disorders such as Schizophrenia and others [3, 4].

Recently, Yamashita et al. reported a series of benzothiophene analogs useful for inhibiting dopamine D2 receptors [6]. In fact, the benzothiophene analogs bind non-selectively to both D2 and D3 dopamine receptors. Molecular mechanics based methods involving docking calculations and also QSAR studies are suitable tools to adjust ligands at target sites and to estimate interaction energy (affinity) and associate the chemical structure to the biological activity [7]. Nowadays, it is a well-established technique applied to numerous cases [8, 9]. Among the molecular modeling techniques, QSAR studies are known to be a method used in order to connect the chemical structure and physicochemical properties to the biological activity of compounds. This method has its greatest use in medicinal chemistry, which helps explain the forces involved in the action of drugs and considers whether their biological properties are the desired. Recently, a promising computational approach for the direct study of drug features most closely associated with particular biological properties at a given receptor is four-dimensional quantitative structure-activity relationship (4D-QSAR) [10–13]. In this

This paper belongs to Topical Collection Brazilian Symposium of Theoretical Chemistry (SBQT2013)

G. B. Caldas · T. C. Ramalho · E. F. F. da Cunha (✉)
Department of Chemistry, Federal University of Lavras, Caixa Postal 3037, CEP 37200-000 Lavras, Minas Gerais, Brazil
e-mail: elaine_cunha@dqi.ufla.br

approach, the affinity correlates with the three-dimensional structure and multiple representations of ligand conformation/orientation [14]. Data sets for 4D-QSAR include all possible conformations, orientations and, in some cases, protonation states [14]. In fact, methods that can incorporate molecular flexibility proved advantageous, since they allowed the identification of the conformation that maximized the activity from 3D-QSAR models [12]. In the current study, four-dimensional quantitative structure-activity relationship (4D-QSAR) [10–14] models were developed for this series of benzothiophene analogs, in order to identify characteristics that may enhance the potency of these compounds.

Methods

Biological data

The 4D-QSAR [10–14] models were developed using 42 compounds (Table 1), the training set, and externally validated using ten compounds (Table 1), the test set, randomly selected from a series of benzothiophene analogs useful for inhibiting dopamine D_2 receptors, developed by Yamashita et al. [6]. The K_i (nM) values were converted into molar units, and then expressed in negative logarithmic units (pK_i). In addition, all pharmacological data were obtained from the same laboratory, eliminating the potential noise that might have been introduced by the pooling of data sets from different sources.

The three-dimensional (3D) structures of the 52 analogues (Table 1) were constructed using the *HyperChem 7.0* software [15]. Each structure was geometry-optimized in vacuum, without any restriction, using the MM + molecular mechanics force field, and subsequently using the semi-empirical AM1 Hamiltonian, in order to assign the partial atomic charges [13].

Molecular dynamic simulation

The details of the RI-4D-QSAR (receptor independent 4D-QSAR) formalism has been reported by Hopfinger et al. [16]. Molecular dynamics simulation (MDS) was carried out using the MOLSIM 3.0 package [17] in the 4D-QSAR program with an extended MM2 force field [18]. The temperature for the MDS was set at 300 K, close to the temperature assays, with a simulation sampling time of 100 ps, and intervals of 0.001 ps. Appling this scheme, a total sampling of 10,000 conformations of each compound was obtained. MDS calculations were carried out applying a distance-dependent dielectric function, $\varepsilon_r = D*r_{ij}$, which was set to $3*r_{ij}$ in order to try to model the solvent effect in the absence of explicit solvent.

Alignment definition

In this study, it is assumed that all molecules bind to the receptor in a similar mode, since the compounds are structural analogs. In general, the alignments are chosen to span the common framework of the molecules in the training and test sets. Alignments using atoms from the right, left, and middle of the common framework and alignments that use atoms that span the common framework should be used to ensure a complete alignment analysis [13]. Three-ordered atom trial alignments were selected in this study: (1) a-b-c, (2) b-a-c, (3) c-b-a, and (4) b-c-a, using a common framework (Fig. 1).

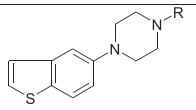
Interaction pharmacophore elements

The 4D-QSAR method currently defines seven types of interaction pharmacophore elements (IPEs), corresponding to atom types that may occupy the grid cells. These IPEs correspond to the interactions that may occur in the active site, and are related to the pharmacophore groups. In this study, the following trial set of interaction pharmacophore elements were selected: i) any type (any); ii) nonpolar (np); iii) polar-positive charge density (p+); iv) polar-negative charge density (p-); v) hydrogen bond acceptor (hba); (hbd) hydrogen bond donor; and vi) aromatic systems (ar) [11]. The use of IPEs allows each of the compounds in a training set to be partitioned into sets of structure types and/or classes with respect to possible interactions with a common receptor. The occupancy of the grid cells by each IPE type is recorded over the conformational assembly profile, and forms the set of grid cell occupancy descriptors (GCOD), to be utilized as the pool of trial descriptors in the model building and optimization process.

Grid cell

The conformational ensemble profile for each compound, obtained after the MDS step, was overlaid onto a cubic lattice of a selected grid cell size, according to each selected alignment. The cubic lattice serves to record the distribution of spatial occupancy for each atom of each ligand in the training. The grid cell occupancy measure is defined as the absolute-occupancy (AO) of a grid cell (x,y,z-coodinate) for a molecule. AO is calculated as the sum of all IPE atoms of the molecule located in a cell, summed over all conformations recorded in the MDS. The occupancy measures can be normalized by dividing the sampling values by the number of sampling steps. The grid cell size defined as 2 Å was used, which corresponds to the integral number closest to twice the hydrogen atom van der Waals radii (r_{vdW} = 1.2 Å) and thus, is large enough to encompass a hydrogen atom.

Table 1 Structures of the 53 D2R inhibitors and K_i values. Training set compound numbers are in bold and test set compounds numbers are in italic



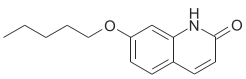
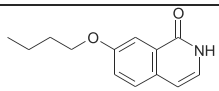
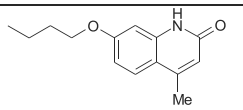
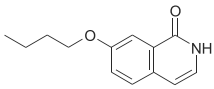
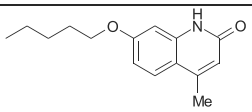
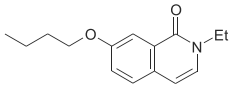
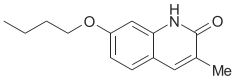
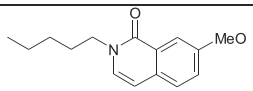
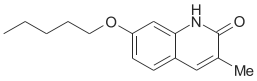
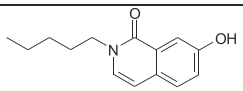
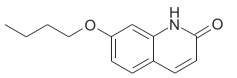
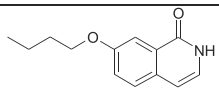
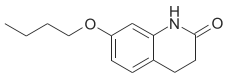
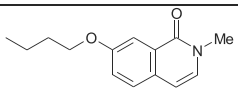
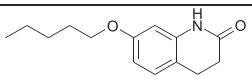
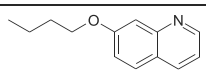
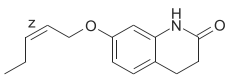
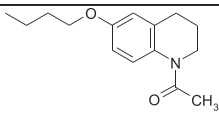
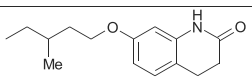
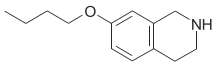
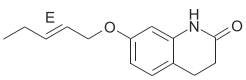
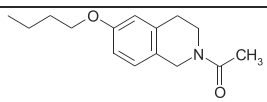
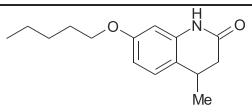
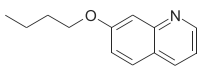
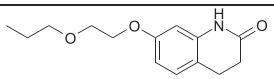
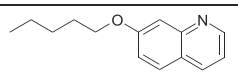
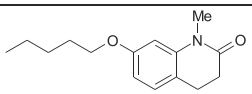
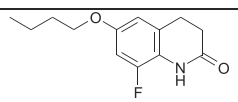
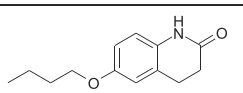
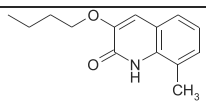
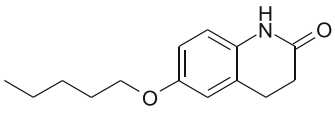
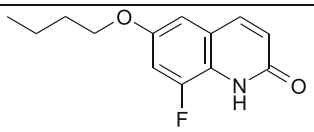
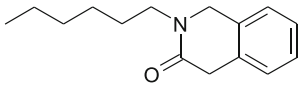
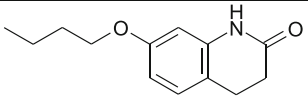
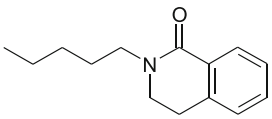
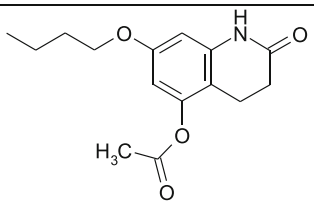
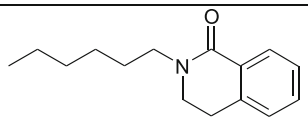
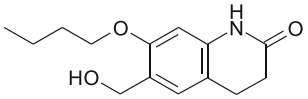
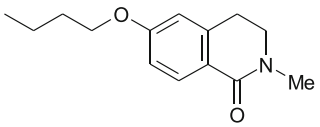
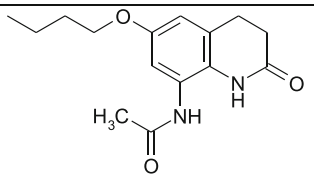
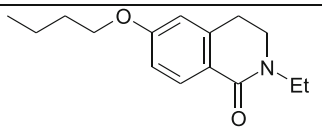
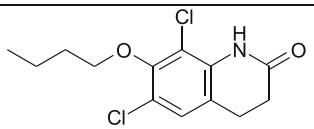
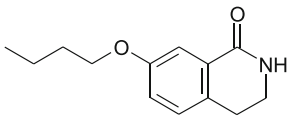
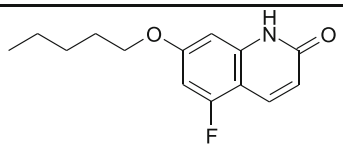
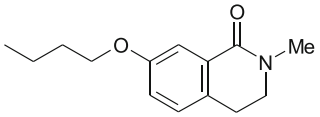
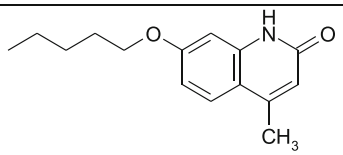
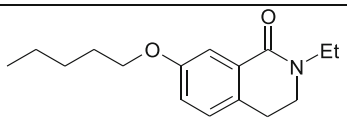
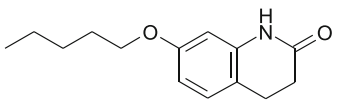
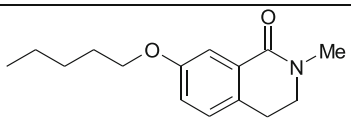
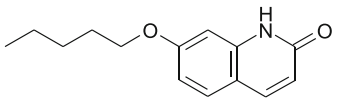
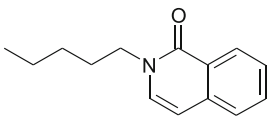
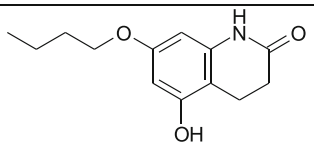
	R	K_i		R	K_i
1		9.20	**27**		8.79
2		9.30	**28**		9.00
3		9.30	**29**		8.72
4		9.22	**30**		8.92
5		9.10	*31*		8.92
6		9.30	**32**		8.32
7		9.40	**33**		8.72
8		10.00	**34**		8.40
9		8.62	**35**		8.33
10		8.49	**36**		8.82
11		9.70	**37**		9.10
12		9.15	*38*		8.62
13		8.65	**39**		8.70
14		8.58	**40**		8.35
15		8.92	*41*		8.33

Table 1 (continued)

16		8.82	**42**		8.45
17		8.40	**43**		8.48
18		9.15	*44*		8.89
19		8.30	**45**		9.70
20		8.45	**46**		8.64
21		8.31	**47**		8.55
22		8.92	**48**		8.58
23		9.15	**49**		8.57
24		8.82	**50**		8.64
25		8.96	*51*		8.80
26		8.92	**52**		9.10

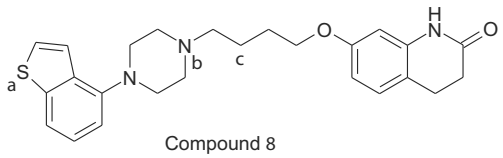


Fig. 1 Ordered atom letter codes (**a**, **b**, and **c**) used in the 4D-QSAR analysis defining the three trial alignments: (*1*) a-b-c, (*2*) b-a-c, (*3*) c-b-a, and (*4*) b-c-a

Conformational selection

In the 4D-QSAR method, the conformation of each compound can be postulated as the lowest-energy conformer state from the set sampled for each compound, which predicted the maximum activity using the optimum 4D-QSAR model [11].

Results and discussion

4D-QSAR model calculations

In order to exclude noise or useless data, databases named DB1, DB2, DB3, and DB4 were generated. Databases DB1 were constructed excluding the variables where GCODs are equal to zero for all molecules, and databases DB2-DB4 were constructed excluding the variables where GCODs have variance cutoff values of up to 0.1, 0.5, and 1.0, respectively. The GCODs from the data reduction were optimized using a combined genetic algorithm (GA) and partial least-squares (PLS) approach [11], implemented in the 4D-QSAR program [10]. Their optimizations were initiated using 10,000 randomly generated models, each having initially six variables. Mutation probability over the crossover optimization cycle was set at 100 %. The smoothing factor, the variable that specifies the number of descriptors in QSAR models, ranged in order to determine equations with no more than ten terms. Each alignment was evaluated using the procedure described above.

Validation is a crucial aspect of any QSAR modeling. High value of q^2 appears to be the necessary but not the sufficient condition for the model to have a high predictive power. Thus, the best models, resulting from the 4D-QSAR study, were based on different criteria [19]:

1) Coefficient of determination (r^2): is a measure of how well the regression line represents the data.

$$r^2 = 1 - \frac{\sum_{1}^{n} \left(Y_{Exp} - Y_{calib}\right)^2}{\sum_{1}^{n} \left(Y_{Exp} - \overline{Y}_{calib}\right)^2}, \quad (1)$$

where Y_{Exp} corresponds to the experimental pK_i values, Y_{calib} are the calculated pKi values, and $\overline{Y}_{calib}$ corresponds to the mean pKi values.

2) Leave-one-out cross-validation (LOOcv) correlation coefficient (q^2): estimating the performance of a predictive model.

$$q^2 = r^2_{CV} = 1 - \frac{\sum_{1}^{n} \left(Y_{Exp} - Y_{cv}\right)^2}{\sum_{1}^{n} \left(Y_{Exp} - \overline{Y}_{cv}\right)^2}, \quad (2)$$

where Y_{cv} corresponds to the calculated pK_i values from cross validation and $\overline{Y}_{cv}$ corresponds to the mean pKi values.

3) Adjusted r^2 (r^2_{adj}): compares the explanatory power of regression models that contain different numbers of predictors. One of the claimed benefits for adjusted R2 is that it "punishes" you for including extraneous and irrelevant variables in the model.

$$r^2_{adj} = 1 - \frac{(n-1)(1-r^2)}{(n-k-1)}, \quad (3)$$

where n is the number of samples and k is the number of independent variables.

4) Root mean square error of calibration (RMSEC), root mean square error of cross validation (RMSECV) or root mean square error of prediction (RMSEP):

$$RMSEC/RMSECV/RMSEP = \sqrt{\frac{\sum_{1}^{n} \left(Y_{Exp} - Y_{calib/cv/pred}\right)^2}{n}}, \quad (4)$$

where "n-k-1"expression corresponds to model degrees of freedom.

5) Friedman's lack-of-fit (LOF): estimates the quality of a model. This measure penalizes appropriately for the addition of terms to the equation (and consequent loss of degrees of freedom) in such a way to resist over-fitting. A "fitness function" or lack of fit (LOF) is used to estimate the quality of a model, so that best model receives the best fitness score.

$$LOF = \frac{LSE}{\left\{1 - \frac{(c+dp)}{m}\right\}^2}, \quad (5)$$

where LSE is the least-squares error (calculated from the difference between actual and calculated values for the activity index over the data set), c is the number of basis functions in the model, d is the smoothing factor, p is the

total number of variables contained in all basis functions, and *m* is the number of samples (compounds) in the training set.

6) Correlation coefficient of external validation set (r^2_{pred}): models are generated based on training set compounds, and the predictive capacity of the models is judged based on the r^2_{pred} values calculated according to the following equation:

$$R^2_{pred} = 1 - \frac{\sum_{1}^{n} \left(Y_{Exp(test)} - Y_{Pred(test)}\right)^2}{\sum_{1}^{n} \left(Y_{Exp(test)} - \overline{Y}_{training}\right)^2}, \quad (6)$$

where $Y_{pred(test)}$ and $Y_{Exp(test)}$ indicate, respectively, predicted and experimental activity values of test set compounds and $Y_{training}$ indicates the mean activity value of the training set compounds.

7) $r^2_{m\,(training)}$: it is a modified r^2 and can also be applied for the test set [20, 21].

$$r^2_{m(training)} = r^2\left(1 - \sqrt{r^2 - r_0^2}\right), \quad (7)$$

where r_0^2 represents the squared correlation coefficient between the observed and predicted values of the training set compounds when the intercept is set to 0.

8) r_p^2 [22]: penalizes the model r^2 for the difference between the squared mean correlation coefficient of randomized models and the square correlation coefficient of the non-randomized model.

$$r_P^2 = r^2 * \sqrt{r^2 - r_r^2} \quad (8)$$

QSAR models

According to the standard procedure, the compounds of Table 1 were grouped in a training set consisting of 42 compounds of the total 52 compounds, and a prediction set, which is formed by the remaining ten compounds. From the pool of equations generated by the 4D-QSAR program, the best QSAR equations for each alignment were selected. These equations were evaluated for their statistical parameters, as well as the number of outlier compounds to select the best equation for each DB.

Alignment 1

$$pIC_{50} = 8.46 + 3.93\,(-4, 2, 7, \text{ahb}) + 4.42\,(-2, 1, 9, \text{p-}) + 1.50\,(-1, -1, 8, \text{p-}) + 2.23(-1, 2, 6, \text{np}) + 2.13(0, -1, 7, \text{dhb}) + 0.83(1, 1, \text{ar}) + 0.44(-1, 1, 8, \text{np}) \quad (9)$$

Model 1

n	r^2	q^2	r adj^2	RMSEC	RMSECV	LOF	Rm^2	Y-rand	Rp^2
42	0.81	0.68	0.78	0.18	0.23	0.095	0.65	0.15	0.66

Alignment 2

$$pIC_{50} = 8.51 + 2.92(3, 2, -3, \text{ahb}) - 0.89(2, -1, -2, \text{any}) + 1.43(0, 2, -3, \text{np}) + 1.32(1, 1, -5, \text{ahb}) + 1.8246\,\text{acp}(0, -1, -3, \text{dhb}) + 1.84(0, 1, -4, \text{ar}) - 0.49(-1, 1, -5, \text{any}) + 0.52(1, 1, -4, \text{np}) \quad (10)$$

Model 2

n	r^2	q^2	r_{adj}^2	RMSEC	RMSECV	LOF	R_m^2	Y−rand	R_p^2
42	0.84	0.76	0.81	0.19	0.24	0.098	0.72	0.18	0.66

Alignment 3

$$pIC_{50} = 8.62 + 2.39(1, -2, -1, \text{dhb}) - 1.20(-1, -2, -2, \text{ahb}) + 1.54(2, -3, -4, \text{any}) + 88.20(1, 1, -1, \text{np}) - 86.20(1, 1, -1, \text{any}) + 0.43(-1, -3, -2, \text{any}) \quad (11)$$

Model 3

n	r^2	q^2	r_{adj}^2	RMSEC	RMSECV	LOF	R_m^2	Y−rand	R_p^2
42	0.68	0.61	0.62	0.25	0.28	0.13	0.48	0.11	0.51

Alignment 4

$$pIC_{50} = 8.59 - 0.57(2,-1,1,\ np) + 3.92(-1,-2,3,\ ahb) + 1.22(-1,1,2,\ any) - 0.78(-2,-2,2,\ np) - 1.13(-1,-3,4,\ np) - 2.44(-4,0,4,\ np) + 2.59(-4,0,4,\ any) + 1.54(-1,-2,4,\ np) - 0.28(-1,0,-2,\ any)$$

Model 4

n	r^2	q^2	r_{adj}^2	RMSEC	RMSECV	LOF	R_m^2	Y−rand	R_p^2
42	0.79	0.69	0.75	0.19	0.25	0.14	0.67	0.18	0.62

(12)

As model 2 showed higher r^2, q^2, R_m^2, and R_p^2 values, it was then selected as the best alignment in this 4D-QSAR study. An r-square value greater than 0.7 indicates that the model is correlated and may be used to calculate the activity of the test set [23]. The modified r^2 (r^2_m) value calculated was equal to 0.70. Values over 0.5 are acceptable. The statistical significance of the relationship between the biological response and the chemical structure descriptors was also further demonstrated by a cross-validation analysis. LOO-$CV_{(training)}$ analysis of Model 2 had a q^2 value of 0.76, with a standard error of 0.19. LOO-CV values over 0.5 reveal that the model is a useful tool for predicting affinities for new compounds in this set. The optimum number of latent variables (PLS components) used for further analyses was seven. A robustness test, based on randomization, was performed and a value of 0.18 was obtained. The r^2_P value found in this study was 0.66 (>0.5).

The calculated pK_i values for the training set were computed using model 2. The residual values were calculated and are shown in Table 2. The standard deviation (SD) of the residual values is 0.13. To establish outlier compounds, the residuals which are more than twice the SD of the residual of fit were observed. Analyses of the data showed that model 2 has no outliers.

In order to better understand the behavior of the data fitted to the models, the cross-correlation matrix among the different GCODs in model 2 was calculated. There is no correlation (r> 0.7) between pairs of GCODs.

A graphic representation of the 3D-pharmacophore embedded in the 4D-QSAR model 3 is shown in Fig. 2. Each descriptor (GCOD) is labeled as "x,y,z,IPE" which represent the cartesian coordinates position of the selected grid cell (x,y,z) and the respective atom type (IPE). They represent the possible intermolecular interactions that may occur between the ligand and the receptor. The GCODs (3,2,−3,ahb), (0,2,−3,np), (1,1,−5,ahb), (0,−1,−3,dhb), (0,1,−4,ar), and (1,1,−4,np) exhibited positive regression coefficients correspond to favorable interactions between the molecules that have this descriptor and the amino acid residues in the receptor active site. Therefore, the activity of the inhibitors should increase with increasing ligand atom occupancy. The GCODs (2,−1,−2,any) and (−1,1,−5,any) exhibited negative coefficients which correspond to unfavorable interactions between the substituent and the amino acid residues in the receptor active site. Therefore, substituents in these positions decrease the potency.

The most important GCOD should be (3,2,−3,ahb) since it is most frequently selected by the GA analysis. It is located close to the oxygen atom of the quinolinone ring and represents an acceptor hydrogen bond IPE and shows the highest frequency of occupation for compound **8** (K_i = 0.1nM). GCODs (1,1,−5,ahb) and (0,−1,−3,dhb) are also located close to oxygen and nitrogen atoms, respectively, of the quinolinone ring. These three descriptors suggest an acceptor/donor hydrogen bond region in the receptor close to this ring.

GCOD (2,−1,−2,any) is present as a non-specific IPE, with a negative regression coefficient. It shows the highest frequency of occupation for compound **9** (K_i = 2.4 nM). On the other hand, GCOD (0,2,−3,np) represents a non-polar IPE, positive regression coefficient and the highest frequency of occupation for compound **11** (K_i = 0.2 nM). Compounds **9** and **11** are isomers and the location spatial of quinolinone ring in the E isomer suggests that the occupation of this region is detrimental to the potency.

GCOD (−1,1,−5,any) is located near the methyl group and represents a non-specific IPE. Since the coefficient of this GCOD is negative, substituents such as compound **35** are detrimental to the activity.

GCOD (0,1,−4,ar) corresponds to an aromatic IPE. This grid cell suggests a hydrophobic region in the receptor close to this ring. Since the coefficient of this grid cell is positive, potential inhibitors would benefit from the exploitation of this region with other types of aromatic groups.

Table 2 Experimental and predicted pK_i values and residual values for the training set using Model 2

Training set	Experimental pK_i	Predicted pK_i	Residual
1	9.699	9.653	−0.046
2	9.301	9.278	−0.023
3	9.301	9.235	−0.066
5	9.097	9.113	0.016
6	9.301	9.213	−0.088
7	9.398	9.583	0.185
8	10.000	9.835	−0.165
9	7.886	7.938	0.052
12	9.155	9.385	0.230
13	8.658	8.628	−0.030
14	8.585	8.618	0.033
15	8.921	8.812	−0.109
17	8.398	8.531	0.133
18	9.155	9.098	−0.057
19	8.301	8.510	0.209
20	8.456	8.569	0.113
21	8.310	8.545	0.235
22	8.921	8.865	−0.056
24	8.824	8.846	0.022
25	8.959	9.069	0.110
26	8.921	8.710	−0.211
27	8.796	8.730	−0.066
28	9.000	8.896	−0.104
29	8.721	8.961	0.239
30	8.921	8.822	−0.099
32	8.319	8.543	0.214
33	8.721	8.451	−0.270
34	8.398	8.531	0.133
35	8.328	8.336	0.008
36	8.824	8.876	0.052
37	9.097	9.097	−0.000
39	8.699	8.690	−0.009
40	8.347	8.618	0.271
42	8.456	8.638	0.182
43	8.482	8.545	0.064
45	9.699	9.504	−0.195
46	8.638	8.693	0.055
47	8.553	8.503	−0.050
48	8.585	8.598	0.013
49	8.569	8.338	−0.231
50	8.638	8.822	0.184
52	9.097	8.827	−0.270

External validation

Therefore, calibration parameters were used to predict the bioactivities of ten test set compounds (**4**, **10**, **11**, **16**, **23**, **31**, **38**, **41**, **44**, and **51**), since they were not included in the development of the 4DQSAR models. The observed and predicted pK_i values and the residues of the fit for the test set are shown in Table 3. A high correlation (r^2_{pred} = 0.87)

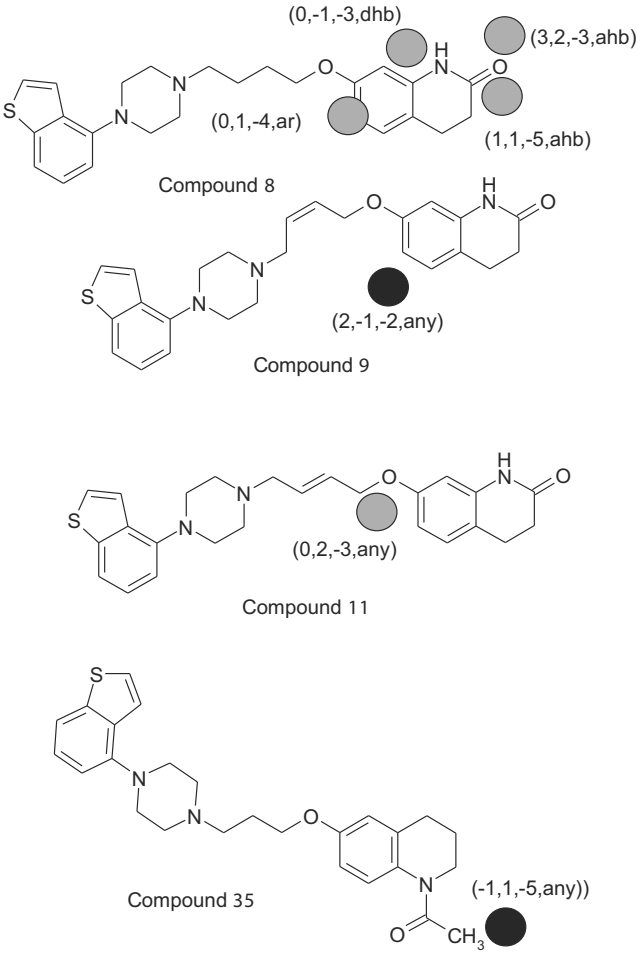


Fig. 2 Graphical representation of descriptors obtained of model 2. *Light spheres* indicate activity-enhancing pharmacophore sites and *dark spheres* indicate activity-decreasing pharmacophore sites

Table 3 Experimental and predicted pK_i values and residual values for the test set using model 2

Test set	Experimental pK_i	Predicted pK_i	Residual
4	9.222	8.845	−0.377
10	8.495	8.461	−0.034
11	9.700	9.719	0.019
16	8.824	8.744	−0.080
23	9.155	8.864	−0.291
31	8.921	8.879	−0.042
38	8.620	8.510	−0.110
41	8.328	8.336	0.008
44	8.886	8.958	0.072
51	8.796	8.813	0.018

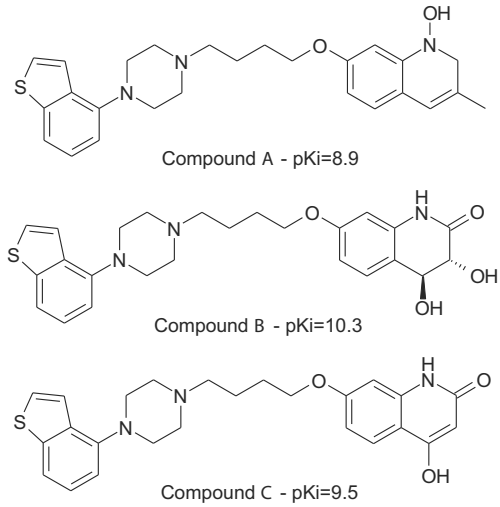


Fig. 3 Structures of the proposed compounds and calculated potencies based on the 4D-QSAR model 2

between experimental and predicted values was found, showing the satisfactory predictive potency of the 4D-QSAR model.

It is interesting to see how well the experimentally observed activity agrees with the predicted one. Considering only the test set compounds, the r2 and standard deviation (SD) of the residuals computed by Eq. 2 were equal to 0.87 and 0.11, respectively.

Compounds based on model 2 analysis

In medicinal chemistry, the optimization of lead compounds proceeds along two main methods: based on chemical modifications of the molecular structure or the application of conformational constraints that change the molecular flexibility. Based on the results obtained in this study, modifications to the structure of compound **8** (quinolinone ring) are suggested. A search in the PubChem Database [24, 25] was carried out and three compounds were selected. The idea is to increase the frequencies of occupation of descriptors with positive regression coefficients. The bioactivity of the proposed compounds was predicted using the model 2 equation. The structure of the three compounds and their predicted pK_i values are shown in Fig. 3. These results were comparable to the best experimental candidate, compound **8**. Compound **B** has shown the highest calculated pK_i, corroborating the idea that donors/acceptors hydrogen bond substituents in the ring increase the potency of D2R inhibitors.

Conclusions

In the present study, 4D-QSAR models for D2R inhibitors were built and evaluated, based on a series of benzothiophene analogs. The best models were obtained from alignments 1–4, using a grid cell size of 1.0 Å, from a training set of 42 compounds. The best model showed r^2, q^2, R_m^2, and R_p^2 values equal to 0.84, 0.76, 0.72, and 0.66, respectively, confirming the acceptability of a predictive QSAR model; in other words, model 2 can be considered robust. In addition, a test set of ten compounds was used in the external validation and a r^2_{pred} value equal to 0.87 was found. Based on these results, the activity of three compounds that were not included in the development of the 4DQSAR models was predicted, based on the model 2 equation. Compound **B** showed greater inhibitory potency than compound **8**, the most active of the series of benzothiophene inhibitors studied.

Acknowledgments The authors are thankful to Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the financial support.

References

1. Silva DR, Ramalho TC, da Cunha EFF (2014) 4D-QSAR model for compounds with binding affinity towards dopamine D2 receptors. Lett Drug Des Discov 11(5):1–16
2. Froimowitz M, Bellott EM (1995) Structural factors that distinguish dopamine D1 and D2 agonists. J Mol Model 1:36–45
3. Estevinho MF, Fortunato JS (2003) Dopamine e Receptores. Rev Port Psicossomática 5:21–31
4. Zhang A, Neumeyer JL, Baldessarini RJ (2007) Recent progress in development of dopamine receptor subtype-selective agents: potential therapeutics for neurological and psychiatric disorders. Chem Rev 107:274–302
5. Seeman P, Schwarz J, Chen J, Szechtman H, Perreault M, McKnight GS, Roder JC, Quirion R, Boksa P, Srivastava LK, Yanai K, Weinshenker D, Sumiyoshi T (2006) Psychosis pathways converge via D2High dopamine receptors. Synapse 60(4):319–346
6. Yamashita H, Ito N, Miyamura S, Oshima K, Matsubara J, Kuroda H, Takahashi H, Shimizu S, Tanaka T (2011) Piperazine-substituted benzothiophenes for treatment of mental disorders. United Sates Patent n° US8,349,840
7. Damale MG, Harke SN, Khan FAK (2014) Recent advances in multidimensional QSAR (4D-6D): a critical review. Mini-Rev Med Chem 14(1):35–55
8. Freitas MP, Ramalho TC (2013) Employing conformational analysis in the molecular modeling of agrochemicals: insights on QSAR parameters of 2,4-D. Cienc e Agrotecnologia 37(6):485–494
9. Gangwal RP, Bhadauriya A, Damre M (2013) 38 mitogen-activated protein kinase inhibitors: a review on pharmacophore mapping and QSAR studies. Curr Top Med Chem 13(9):1015–1035
10. The Chem21 Group Inc. (1997) 4D-QSAR user's manual. The Chem21 Group Inc., Lake forest, IL
11. Rogers D, Hopfinger AJ (1994) Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. J Chem Inf Comput Sci 34:854–866
12. Romeiro NC, Albuquerque MG, Alencastro RB, Hopfinger AJ (2004) Construction of 4D-QSAR models for use in the design of novel p38-MAPK inhibitors. J Comput Aided Mol Des 19(6):385–400

13. da Cunha EFF, Albuquerque MG, Antunes OAC, Alencastro RB (2005) 4D-QSAR models of HOE/BAY-793 analogues as HIV-1 protease inhibitors. QSAR Comb Sci 24:240–153
14. Sodero ACR, Romeiro NC, da Cunha EFF, Magalhães UO, Alencastro RB, Rodrigues CR, Cabral LM, Castro HC, Albuquerque MG (2012) Application of 4D-QSAR studies to a series of raloxifene analogs and design of potential selective estrogen receptor modulators. Molecules 17:7415–7439
15. Hypercube, Inc. (2007) *HyperChem® 7.0* software. Hypercube, Inc., Gainesville, FL
16. Tokarski JS, Hopfinger AJ (1997) Prediction of ligand-receptor binding thermodynamics by free energy force field (FEFF) 3D-QSAR analysis: application to a set of peptidometic renin inhibitors. J Chem Inf Comput Sci 37:792–811
17. Doherty DC & The Chem21Group Inc. (1997) MOLSIM Users Guide v.3.0. Doherty DC & The Chem21Group Inc., Lake Forest, IL
18. Weiner SJ, Kollman PA, Nguyen DT (1986) An all atom force field for simulations of proteins and nucleic acids. J Comput Chem 7:230–252
19. Kiralj R, Ferreira MMC (2009) Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. J Braz Chem Soc 20(4):770–787
20. Kar S, Roy K (2011) Development and validation of a robust QSAR model for prediction of carcinogenicity of drugs. Indian J Biochem Biophys 48:111–122
21. Silva DG, Freitas MP, da Cunha EFF, Ramalho TC, Nunes CA (2012) Rational design of small modified peptides as ACE inhibitors. Med Chem Commun 3:1290–1293
22. Roy PP, Paul S, Mitra I, Roy K (2009) On two novel parameters for validation of predictive QSAR models. Molecules 14: 1660–1701
23. Hopfinger AJ, Wang S, Tokarski JS, Jin B, Albuquerque M, Madhav PJ, Duraiswami CJ (1997) Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. J Am Chem Soc 119: 10509–10524
24. Bolton E, Wang Y, Thiessen PA, Bryant SH (2008) PubChem: Integrated platform of small molecules and biological activities Chap. 12 In: Annual Reports in Computational Chemistry, Vol. 4. American Chemical Society, Washington, DC
25. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Zhou Z, Han L, Karapetyan K, Dracheva S, Shoemaker BA, Bolton E, Gindulyte A, Bryant SH (2011) PubChem's BioAssay Database. Nucleic Acids Res 40(1):400–412