



Kent Academic Repository

da Silva, Aldineia P., Chiari, Laise P. A., R. Guimaraes, Amanda, Honorio, Kathia M. and da Silva, Albérico B. F. (2021) *New D2R partial agonist candidates: an in silico approach from statistical models, molecular docking, and ADME/Tox properties*. *Structural Chemistry*, 32 (5). pp. 2019-2033. ISSN 1040-0400.

Downloaded from

<https://kar.kent.ac.uk/115836/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1007/s11224-021-01742-w>

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>).



New D2R partial agonist candidates: an in silico approach from statistical models, molecular docking, and ADME/Tox properties

Aldineia P. da Silva¹ · Laise P. A. Chiari¹ · Amanda R. Guimarães¹ · Kathia M. Honório² · Albérico B. F. da Silva¹ 

Received: 22 December 2020 / Accepted: 25 January 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Schizophrenia is a psychiatric disorder that affects 20 million people worldwide and the mortality rate is two or three times higher than the average overall population mainly due to high frequency of suicide and other associated comorbidities. The therapeutic approaches used for its treatment, besides do not combat this illness, trigger several side effects, being imperative the development of new medical countermeasures. In the search for new agents, we constructed a statistical model (model **A**) with a series of aripiprazole-derived using the partial least squares technique, and posteriorly we applied several validation tests to guarantee consistency of model **A** and its forecasting ability. In addition, molecular docking simulations were employed to extract important information on structural elements involved in the molecular recognition process from receptor-ligand complex. We designed new compounds whose biological activity values were predicted by our model. Among new designed molecules, we highlight compounds **2**, **5**, **9**, and **11** with predicted pK_i values over **8.5**. We underline that the presence of ester moieties attached to the aromatic ring from region **A** is well tolerated in this position and contributes to increase the pK_i values; otherwise it is essential that there are no bulky groups in region **C**. Finally, ADME/Tox properties evaluated via in silico approach for the proposed compounds shed light on their drug-likeness characteristics, indicating that they could be a reliable starting point as potent candidates to further experimental exploration (chemical synthesis, in vitro and in vivo analyses).

Keywords Schizophrenia · Molecular docking · Drug design · ADME/Tox

Introduction

According to World Health Organization, schizophrenia is a severe mental disorder that affects approximately 20 million people worldwide [1]. People suffering severe mental illnesses such as schizophrenia have mortality rate of two to three times higher than the general population [2, 3].

Schizophrenia—word that means splitted or dissociated mind [4]—is a mental disorder characterized by a neurochemical imbalance in which the affected individuals experience a hallucinatory reality [5].

It is extremely difficult to understand the pathophysiological mechanism of schizophrenia as its cause remains totally unknown, even so researches have shown that schizophrenia symptoms may be related with alteration in dopamine levels in the brain [6–17].

Basically, three types of symptoms are associated with schizophrenia: positive symptoms attached to delusions, hallucinations and disorganized speech; negative symptoms related to apathy, anhedonia, avolition, and affective blunting; and finally, cognitive deficits that include concentration problems and memory changes [18–23].

Within the hypotheses formulated to explain schizophrenia, the dopaminergic has been considered the most plausible one as it is related to the excess of dopaminergic activity in the mesolimbic system that is accountable for positive symptoms, and in the mesocortical pathway which is responsible for the negative symptoms [24].

Regarding drugs used in the treatment of schizophrenia, atypical antipsychotics have been shown the best results, which are characterized by a significant reduction of extrapyramidal side effects (EPS) while acting in both positive and

✉ Albérico B. F. da Silva
alberico@iqsc.usp.br

¹ Departamento de Química e Física Molecular, Instituto de Química de São Carlos, Universidade de São Paulo, C.P 780, São Carlos, SP 13560-970, Brazil

² Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo 03828-000, Brazil

negative symptoms. Among the antipsychotic drugs with such feature, aripiprazole appears as the first D2 dopamine receptor partial agonist used in mental health disorders [22, 23, 25].

As partial agonist, aripiprazole binds with high-affinity to the dopaminergic D2 [26, 27] and D3 [28] receptors and serotonergic 5-HT_{1A} receptor [29] and, as antagonist it acts at 5-HT_{2A} [28] receptor. Besides, aripiprazole has been reported as an efficient drug against positive, negative, and cognitive symptoms of schizophrenia with reduced adverse effect including metabolic and endocrine alterations [30]. The functional profiles as partial agonist and antagonist added to all the advantageous in terms of side effects turns aripiprazole as a good starting point to design of new antipsychotics [31].

In this work, quantitative structure-activity relationship (QSAR) [32] techniques are applied to a series of potential partial D2 receptor agonists aripiprazole-derived [33] with the aim to correlate chemical features (from their molecular structure) to biological activity by applying partial least square (PLS). Next, we perform docking molecular simulation to obtain information about the main interaction types from the D2 receptor active pocket. Finally, from statistical models as well as molecular docking simulations analyses, we designed new compounds (whose biological activity values were predicted by our model), and afterward their drug-likeness properties were assessed via *in silico* approach.

Materials and methods

Data set

In this work, a set of molecules were selected to construct the QSAR models. These compounds were synthesized from exploration of four regions of aripiprazole [33] that exhibit structural diversity and biological activity values represented by their inhibitory constants (pK_i) as illustrated in Chart 1.

Calculation and selection of descriptors

The electronic descriptors of the compound set were obtained through the optimization and the harmonic vibrational frequency calculations of the compounds (Chart 1) run in Gaussian 09 [34] with B3LYP [35, 36] /6-31G++(d,p) [37–39].

The molecular descriptors were generated using the optimized structure of all compounds in the platform e-dragon [40]. The Pearson's correlation coefficients (*r*) between each descriptor and the pK_i values were calculated in order to exclude the descriptors with low correlation with the biological activity.

A plethora of descriptors was excluded (those with coefficient values <0.4 in modulus) [41]. Next, genetic algorithm (GA) method as implemented in BuildQSAR [42] was applied as a second screening to get the descriptors that mostly contributed to the biological activity.

The PLS technique was employed to generate the multivariate model using Pirouette 3.10 [43]. Thereafter the molecular docking technique was applied to analyze the main interactions between D2 receptor and the selected ligands using Glide package [44]. Finally, from both analyses, the statistical model and the docking results, new compounds were designed whose biological activity values were predicted by our model.

Model construction and validation

Initially, the molecular properties of all compounds (Chart 1) were autoscaled [41] in order to construct the PLS models. Next, the most consistent and accurate model was chosen based on the optimal number of latent variables.

The PLS model was internally validated via *leave-one-out* cross-validation procedure, in which one sample is omitted during the construction of the model, and subsequently its biological activity value is predicted by PLS model. This procedure is repeated up to that all samples have been excluded at least once.

Afterwards, statistical parameters such as SEC (standard error of calibration), PRESS (*prediction error sum of squares*), and SEV (standard error of validation) were calculated and from them information on the model's quality were obtained and analyzed along with the values of *q*² (correlation coefficient of cross validation) and *r*² (coefficient of multiple determination) [45].

In addition, the *y*-randomization tests [45] were applied to assure that there is no chance correlation between dependent (pK_i) and independent (descriptors) variables by scrambling the biological activity values from 12 runs using the QSAR modeling software [46]. Lastly, an external compound set [47] (Chart 2) was used as ultimate proof to certify the predictability of our PLS model.

Molecular docking simulations

The 3D coordinates of the D2 receptor (DRD2) were obtained from the RCSB Protein Data Bank (PDB ID 6CM4). To fix some issues inherent to the PDB structure (such as missing side-chain, loop and hydrogens), the D2 receptor was prepared using the Protein Preparation Wizard (PPW) module using the OPLS-3 force field which are included in Maestro 17.1 suite [44]. PPW was set up to exclude the crystallographic water molecules away from the active site, to add hydrogens, to fill the missing side-chain, to maintain the pH at 7, and to assign the charges as well as protonation.

Next, the refined structure was submitted to process of energy minimization at MacroModel module [44] to attenuate the steric clashes among the amino acid residue, and posteriorly the D2 receptor active site was defined using the Receptor Grid Generation panel [44] from the co-crystallized ligand coordinates, for which a grid box of 10 × 10 × 10 Å³ was constructed.

Region A			
Compound	R	pK _i	
1	2,3-(Cl) ₂	8.41	
2	2-OCH ₃	9.52	
3	3-OEt	8.60	
4	2-OiPr	8.51	
5	2-H	8.40	
6	2-F	8.30	
7	2-Cl	8.43	
8	2-CN	8.54	
9	2-CH ₃	8.23	
10	2-CF ₃	8.40	
11	3-OEt	7.70	
12	4-OEt	7.30	
13	2,3-(CH ₃) ₂	8.10	
Region B			
Compound	B	pK _i	
14		8.00	
15		8.44	
Region D			
Compound	D	pK _i	
22		7.1	
23		8.5	
24		7.6	
25		7.2	
26		7.8	
27		7.8	
Region C			
Compound	C	pK _i	
16	n = 2	7.7	
17		6.8	
18		9.0	
19		6.6	
20		7.0	
21		7.0	

Chart 1 Structure and biological activity of the compounds used as training set, synthesized from the exploration of four regions of aripiprazole

The validation protocol was applied to assess the performance of the selected scoring function by their ability to reproduce the co-crystallized ligand (risperidone, PDB code 8NU) conformation using the same setup aforementioned at Glide module [44]. In this protocol, we chose the XP extra precision as a scoring function to accomplish the molecular docking simulations with compounds C2 and C17 (Chart 1). Afterwards, some parameters were evaluated, for example, ligand orientation, interaction among the ligands with the amino acid residues within the active site, the scores associated with each binding pose, as well as the differences in the binding mode for the compounds under analysis.

In silico ADME/Tox analyses

It is essential to evaluate the ADME/Tox [48] (absorption, distribution, metabolism and elimination/ toxicity) profile of new drug candidate molecules. Nowadays, the big pharmaceutical companies use computer-based methods as it is less expensive and to reduce the amount of consuming-time to select lead molecules to advance experimental evaluation

[49]. In this paper, we used the QikProp from the Schrodinger suite [44] to assess the physicochemical and pharmacokinetic properties of new compounds here designed.

Results and discussion

One thousand two hundred seventeen descriptors were calculated and posteriorly selected from the Pearson's correlation coefficient (*r*) between the X (descriptor) and Y (biological activity) variables. In this correlation matrix, the variables in the interval -0.4 to 0.4 were promptly excluded, but an expressive number is still remaining (248 descriptors). Therefore, other screening was performed using the genetic algorithm method to select the most relevant descriptors. For this purpose, 100 models were constructed, whose descriptors selection was based on the R and q^2 values (Table 1 presents the combination of descriptors for which the best *r* and q^2 values were obtained).

The generation equals to **4000** refers to the number of times that the genetic algorithm searched for the best combination of descriptors (fittest individual). The

N°	Molecular Structure	pK _i	N°	Molecular Structure	pK _i
1		8.79	7		6.54
2		7.80	8		8.03
3		7.38	9		7.30
4		7.31	10		7.06
5		7.57	11		8.07
6		6.63			

Chart 2 Molecular structure and biological activity values of the compounds used as test set

criterion **0.8** indicates that the descriptors more correlated with each other were cut off.

Table 2 exhibits the class of each descriptor and its definitions (the values of the descriptors and pK_i are displayed in [Supplementary Material](#)).

Model validation

Our statistical model was built with all compounds (Chart 1), wherein the data set (X and Y vectors) were

autoscaled and internally validated by the leave-one-out tests. Table 3 shows the statistical parameters arising from the PLS technique, i.e., latent variable, SEP_{val}, SEP_{cal}, q² and r². The best model (**model A**) was obtained with five latent variables, which described **88.80%** of the original data information and presented the lowest values of SEP_{validation} (**0.31**), PRESS_{validation} (**3.82**), PRESS_{calibration} (**2.58**), and the highest values of q² (**0.77**) and r² (**0.84**). These optimum values are the initial proof of the model's statistical significance since,

Table 1 Variables Selected by the genetic algorithm technique

Generation	Criterion	Descriptors	R	q ²
4000	0.8	BEHe3, RDF145u, E1e, HATS7v, R2v+, E _{LUMO+4} (eV)	0.94	0.83

Table 2 Class and description of the selected descriptors

Descriptor	Class	Definition
BEHe3	BCUT (Burden- CAS-University of Texas Eigenvalues)	Highest eigenvalue n. 3 of Burden matrix/weighted by atomic Sanderson electronegativities
RDF145u	RDF (radial distribution function)	Radial distribution function—14.5/unweighted
E1e	WHIM (weighted holistic invariant molecular)	First component accessibility directional WHIM index/weighted by atomic Sanderson electronegativities
HATS7v	GETAWAY (geometric topology and atom weights assembly)	Leverage-weighted autocorrelation of lag 7/weighted by atomic van der Waals volumes
R2v+		R maximal autocorrelation of lag 2/weighted by atomic van der Waals volumes
E_{LUMO + 4}	Quantum	Energy of the LUMO ₊₄ orbital.

Table 3 Statistical results used to select the best PLS model

Latent variable	Variance	SEP _{validation}	PRESS _{validation}	SEP _{calibration}	q ²	PRESS _{calibration}	r ²
1	41.89	0.49	9.22	0.47	0.45	7.85	0.53
2	55.43	0.45	7.74	0.38	0.55	4.96	0.70
3	67.80	0.37	5.15	0.30	0.70	3.16	0.81
4	79.12	0.33	4.09	0.29	0.76	2.72	0.84
5 (model A)	88.80	0.31	3.82	0.28	0.77	2.58	0.84
6	100.00	0.32	3.83	0.29	0.77	2.58	0.85

and for a model to be considered trustworthy, the values of q^2 and r^2 must be greater than 0.5 and 0.6, respectively [50–52].

To ensure the model robustness, it is expected that the difference between the values of r^2 and q^2 be less than 0.30 [52, 53]. As presented in Table 3 (data in bold), the variance between the values of these parameters is undermost (0.07), singled out that the model has not been overfitted and so it can be considered robust.

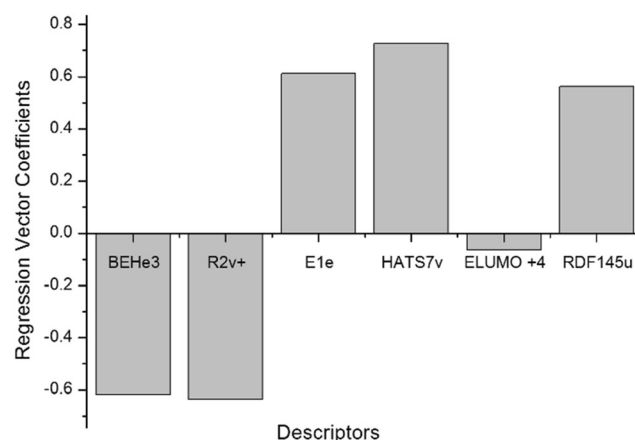
Figure 1 shows the contribution of each descriptor for the model A (BEHe3, R2v+, E1e, HATS7v, E_{LUMO + 4}, RDF145u). The descriptors E1e, HATS7v, and RDF145u contribution positively to the model, i.e., the greater values of these ones are attributed to compounds with the greater biological activity values, whereas the descriptors BEHe3m R2V+ and E_{LUMO + 4} are inversely correlated with the pK_i values. In the light of these results, we can identify the chemical features of the compounds (Charts 1 and 2) that most contribute to increase their partial agonist activity in the presence of D2R, as discussed in the following section (Descriptors analysis).

Detection of outlier objects was also performed to ensure the quality of the model A. It can be seen from Fig. 2 that the compound C2 has studentized residual value away from the expected limit [49]. However, this deviation is insignificant, and thereupon C2 cannot be considered as an atypical object.

Moreover, the leverage value for this compound is within the range of acceptable values [49].

From Fig. 3 (experimental versus calculated pK_i), it is apparent that there is a good correlation, evidencing that our model has significative power prediction.

The model robustness was tested via cross-validation procedure, whereupon 1–12 compounds were excluded with 50 repetitions for each value of N left out. The results (Fig. 4) showed q^2 remained with its excellent values, presenting

**Fig. 1** Regression vector coefficients for each descriptor used in the model construction

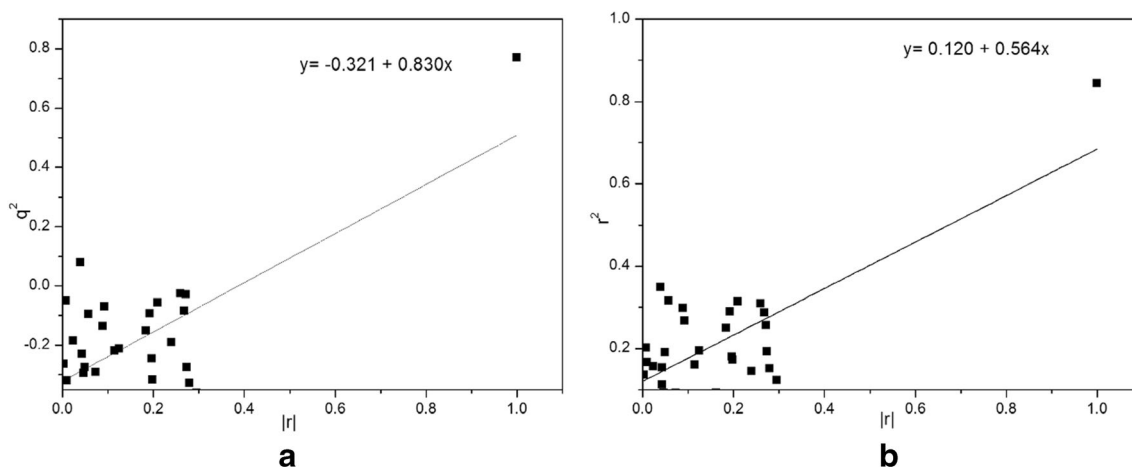


Fig. 5 Y-scrambling test applied to model A

identified: all compounds with higher pK_i values have substituents in the position 1 of the aromatic ring (region A—Chart 1); among these substituents, the ester moieties are more tolerable in this position. As we will see later, chemical substituents of this nature are involved in hydrophobic interactions with amino acid residues that are responsible for the D2 receptor inactivation, showing that the volume factor is very important for increasing the affinity of the compounds.

E1e: the highest values of this descriptor (corresponding to the range of 0.6 to 0.62) are assigned to compounds C2, C7, C10, C12, C16, and C18 (Fig. 7). For compounds C2, C7, C10, and C12, structural differences occur in region A, whereas for the compounds C16 and C18, the structural variation is observed in region C, and besides they are the only ones with changes in this region that do not have bulky groups as substituents. Compound C12 has substituents at position 4 of the ring (region A, Chart 1), reason why the biological activity decrease. This indicates that the groups in this position do not

contribute to increase the pK_i values because it prevents that the aromatic ring interacts with **Trp100**, which is essential for the D2R inactivation (as we will see from the molecular docking analyses).

HATS7v: analyses from this descriptor indicated that the concomitant presence of chlorine and sulfur atoms in the molecular structure of the compounds studied did not result in a significantly increase of the biological affinity (Fig. 8), showing that bulky substituents can influence negatively the pK_i values.

R2v+: higher values of this descriptor are attributed to the compounds with the lower values of pK_i . It is also noted that compounds with **R2v+** value equals to 0.12 (compounds C25, C27, C28, C29, C30, and C32) exhibited intermediate biological activity, and all of them have sulfur atoms in region D, showing once again that the presence of bulky moieties in this region is not suitable.

E_{LUMO+4}: this descriptor refers to the LUMO (Lowest Unoccupied Molecular Orbital) energy of a compound and shows that the lower the LUMO energy, the greater the ability to accept electron [53, 54]. The analyses of this descriptor reveal the highest electronic density regions where the interaction among the compounds and the D2R receptor (regions A and D—Fig. 9) are likely to occur. Subsequent analyses will show (*docking analyses*) that these regions are essential to participate in hydrophobic interactions with the D2R receptor resulting in its inactivation.

Docking analyses

Once the results from the redocking (see [Supplementary Material](#)) using the co-crystallized ligand validated the employed protocol (section Material and Methods), we chose the most (C2) and the least (C17) active compounds to proceed to molecular docking simulations, allowing to evaluate

Table 4 Predicted biological activity by model A for the test set compounds

Compound	Experimental pK_i	Predicted pK_i	Absolute residual
1	8.79	7.06	1.73
2	7.80	7.01	0.79
3	7.38	6.74	0.64
4	7.31	6.61	0.70
5	7.57	7.76	-0.19
6	6.63	7.65	-1.02
7	6.54	7.08	-0.54
8	8.03	7.26	0.77
9	7.30	6.04	1.26
10	7.06	7.50	-0.44
11	8.07	8.21	-0.14

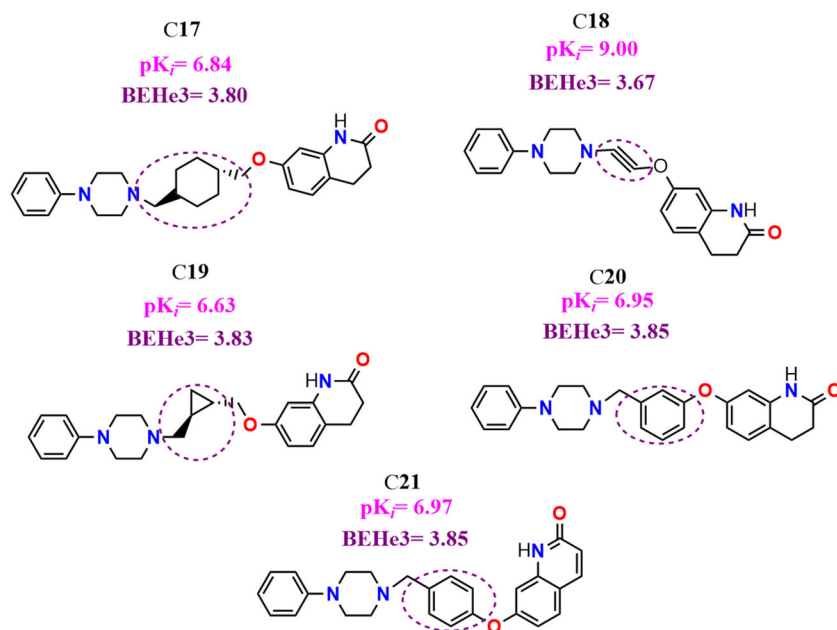


Fig. 6 Chemical structure of compounds that had modifications in region C

the binding mode and the types of ligand-receptor interactions (Fig. 10).

From Fig. 10, we can observe that several hydrophobic contacts stabilize the binding of C2 within the binding pocket in D2 (Fig. 10b - amino acids residues in green), such as **Phe198**, **Ala122**, **Trp386**, **Phe389**, **Phe390**, **Phe189**, **Phe110**, **Ile184**, **Tyr416**, **Trp100**, **Val91**, **Trp413**, **Leu94**, **Leu41**, **Val115**, **Cys118**.

The results from molecular docking simulation indicate the 3,4-dihydroquinolin-2(1H)-one moiety at the compound **C2** interacts both by hydrogen bonding (dotted line in magenta) with **Trp413** and pi-pi stacking interactions (dotted line in blue). In the other extreme, the methoxybenzene group also interacts by pi-pi stacking contacts with **Trp386** (double interactions) and **Phe390**; besides it participates of the aromatic H-bond

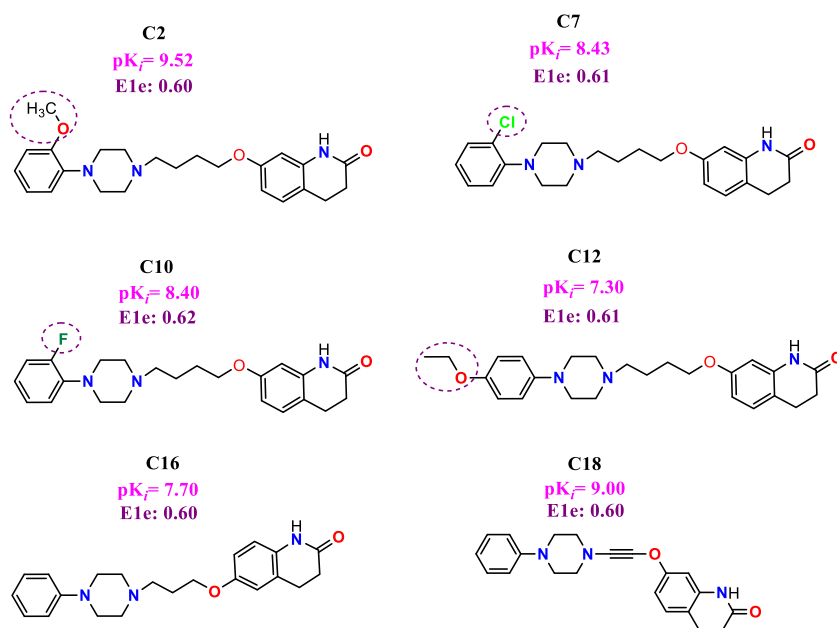


Fig. 7 Compounds that presented the highest descriptor E1e values

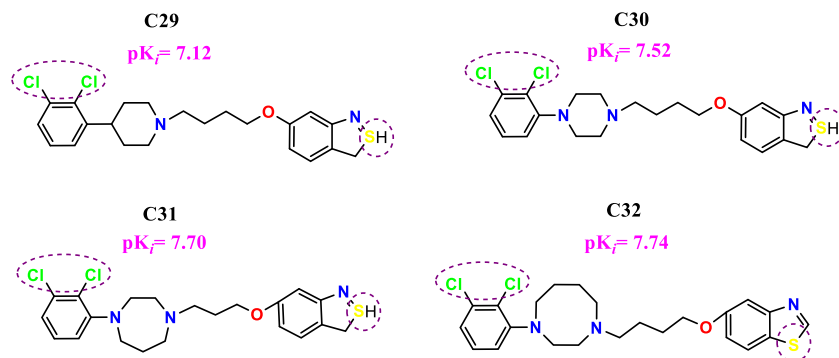


Fig. 8 Compounds with HATS7v values equal to 0.1

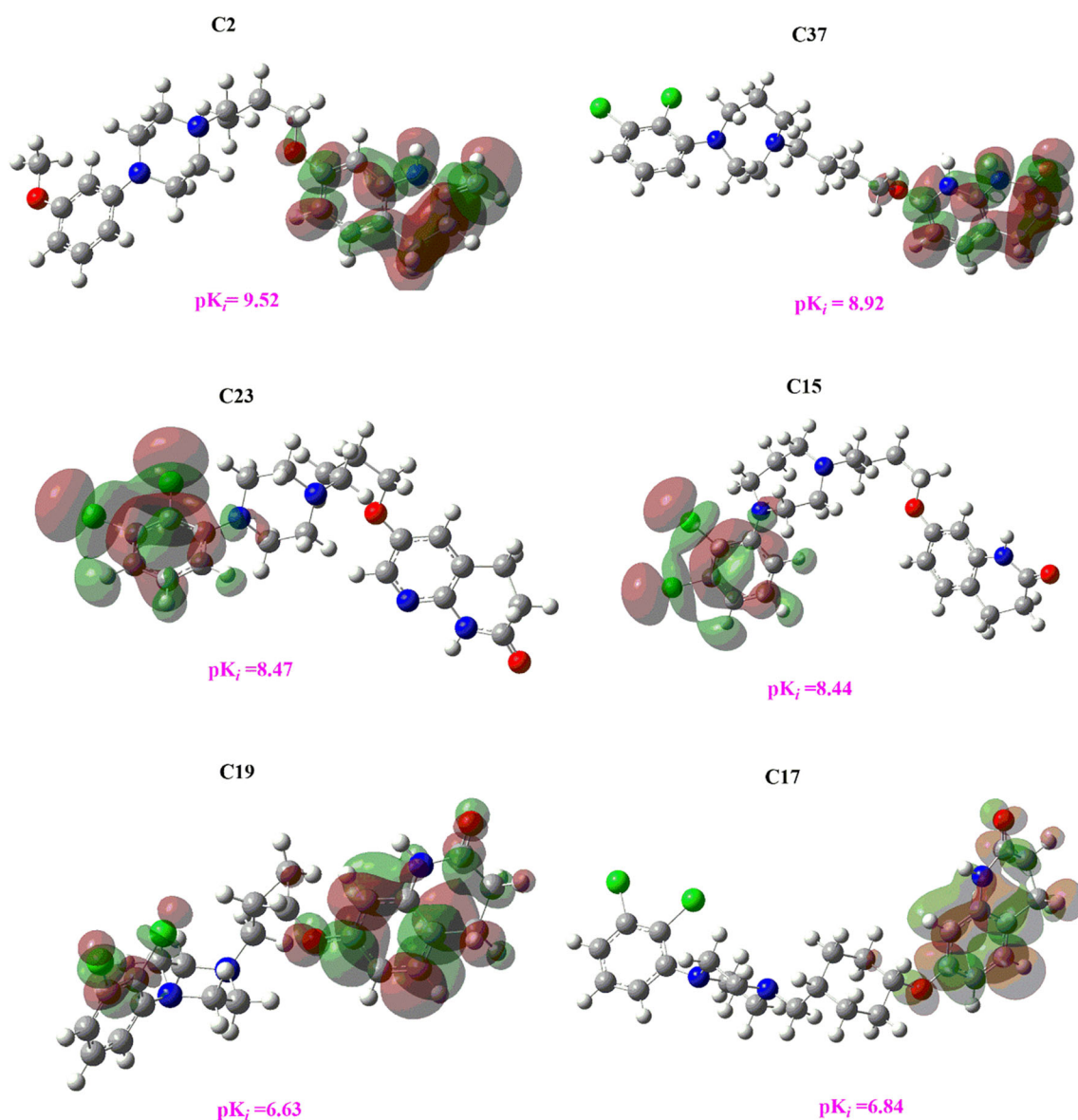


Fig. 9 Regions of the molecules that will participate in hydrophobic interactions with the D2R receptor

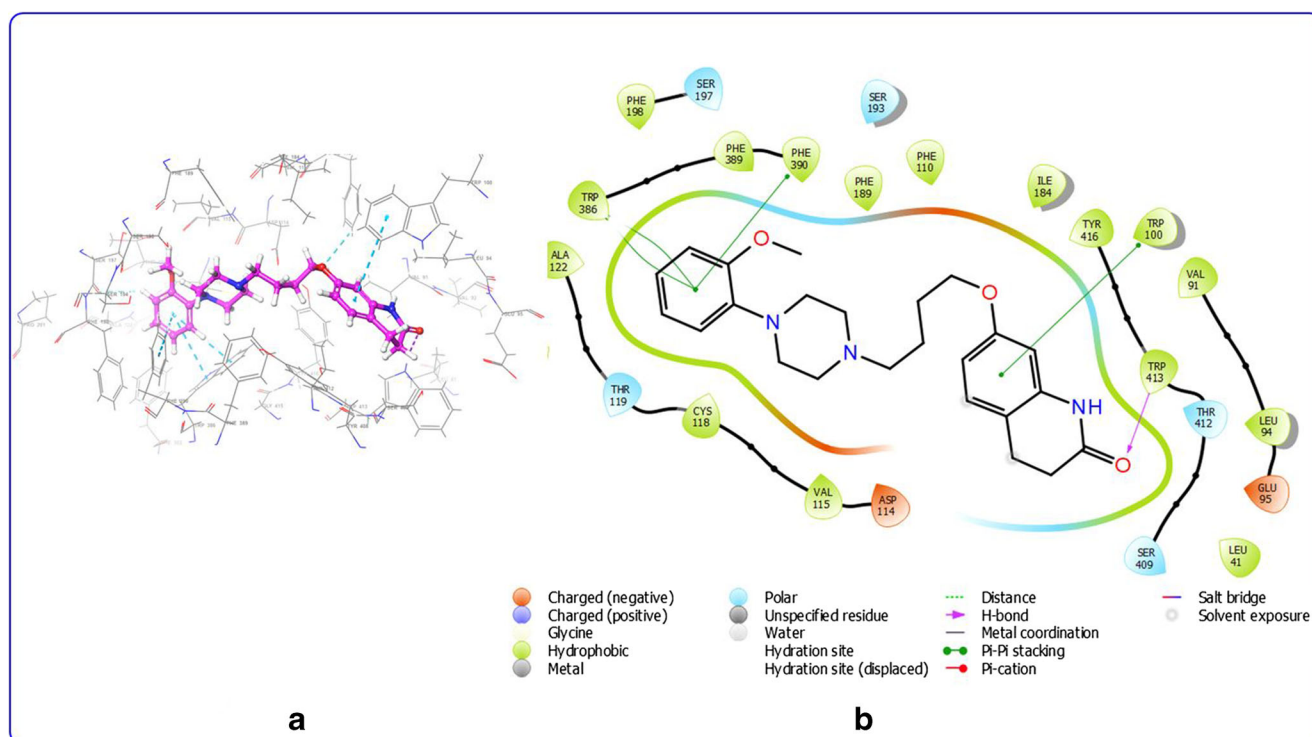


Fig. 10 Binding pose of compound C2 within D2R active pocket. **a)** 3D representation and **b)** bidimensional diagram

interactions (dotted line in green) with **Ser197**, as we can see in Fig. 11.

It is known that the residues **Leu94**, **Trp100**, e **Ile194** form a hydrophobic pocket that contributes to the ligand to remain longer within the D2R active site [55]. These residues interact strongly with compound

C2, and consequently a significant score value (-9.91) was obtained, which explains the great biological affinity ($pK_i = 9.52$).

The molecular docking results indicate that the least active compound (**C17**) also participates in hydrophobic interactions (residues in green cycles, Fig. 12), but the

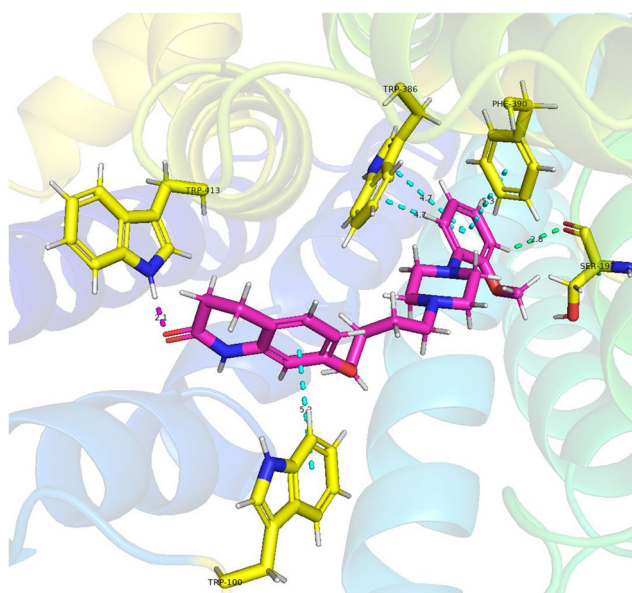


Fig. 11 Main binding interactions between the compound C2 and the D2 receptor

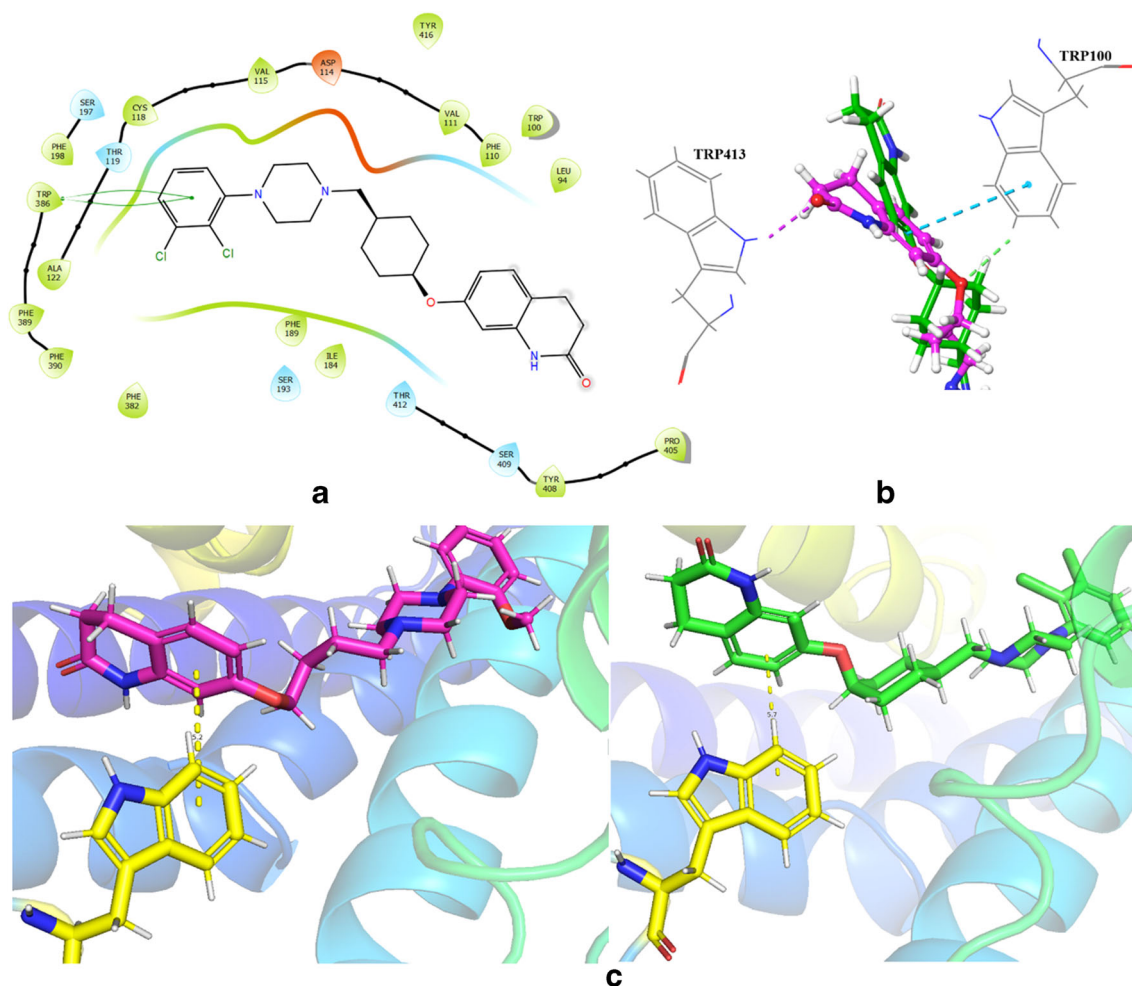


Fig. 12 **a** Bidimensional representations of amino acid residues involved in the interaction with compound C17; **b** binding pose comparison between compounds C2 (in magenta) and C17 (in green) - the rigidity of the methoxycyclohexane moiety in C17 prevents the 3,4-dihydroquinolin-2(1H)-one moiety from the region **D** orienting itself in the proper position to interact by H-bonding with Trp413 and pi-pi

stacking interactions with Trp100. Even so, the residue Trp100 interacts by hydrophobic contacts with C17; **c** as the ring centroids distance (Trp100 and C17) is greater than 5.5 Å (5.7 Å, edge-to-face), there are no Pi-Pi stacking interaction, while for C2 the distance between the ring centroids is 5.2 Å and thereupon there is the occurrence of this type of interaction

bulky group at region **C** (methoxycyclohexane) prevents 3,4-dihydroquinolin-2(1H)-one moiety (region **D**) to participate in hydrophobic interactions with essential amino acid residues (Fig. 12). These residues do not allow that the ligand to leave the active site, resulting in a strong receptor-ligand interaction. The absence of this type of hydrophobic contacts may be the reason why C17 has a lower score value (−8.20) when compared with the C2 ones and, consequently, the lowest pK_i value (6.84).

Structural modification and design of new compounds

Based upon the information extracted from the selected descriptors and the molecular docking simulation

analyses, we made structural modifications in the compounds, in particular at regions **A** and **C** (Chart 1), in order to propose new molecules.

The new molecules were optimized following the same protocol previously described (*methodology* section) and, subsequently, the descriptors were obtained. Lastly, we subjected the new compounds to our PLS model to predict their pK_i values.

From Chart 3, we can see that the new proposed compounds presented relevant biological activity values, even larger than most of the compounds used to construct the model.

The new compounds obtained (Chart 3) indicate that hydrocarbons, ester moieties, and halogen atoms are well tolerated when attached to the benzene ring (region

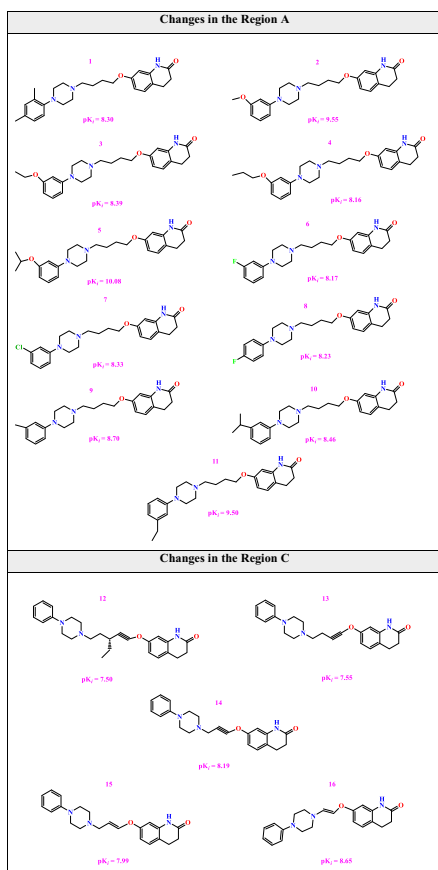


Chart 3 Design of new compounds from the structural modifications in regions A and C

A) since the biological activity values predicted by our model are quite expressive. In particular, among the compounds with ester moiety in their structures, it is worthy to mention that compounds **2** and **5** presented the highest pK_i values (Chart 3). Indeed, analysis from the descriptor **RDF145u** had already indicated that ester moieties attached to the benzene ring (region A) are well tolerated in this position. In addition, molecular docking simulations singled out that chemical entities with these features contribute to increase the hydrophobicity of the molecules. As aforementioned, this type of interaction is essential so that the compounds stay longer within the D2 receptor active site, which may result in a slow dissociation of the compounds from the D2 receptor.

Lastly, the analyses from the descriptors **BEHe3** and **E1e** indicated that bulky alkenyl and alkynyl groups are profitable in position C as long as they do not increase considerably the distance between B and D regions (Chart 1). From this observation, we conducted structural modifications at region C resulting in compounds with expressive biological affinity, highlighting the compounds **14**, **15**, and **16**.

Evaluating of the ADME/Tox profile

The new proposed compounds (Chart 3) were submitted to drug-likeness analyses (physicochemical and pharmacokinetic properties) via QikProp module from the Schrödinger package [44]. The results indicated that these molecules exhibit ADME/Tox profile within the desired range (Table 5), and hence the values of these parameters shed light on their drug-likeness characteristics, which could configure an indispensable starting point to further experimental procedures.

Conclusions

Computational tools applied to drug design have been widely employed by both pharmaceutical company and universities because they reduce time and costs regarded with the development of one marketable product. In this work, we constructed a statistical model and performed molecular docking simulations to extract information on structural elements involved in the receptor-ligand complex interaction. From this information, new D2R partial agonist candidates were designed.

We applied several validation procedures to assure the model A quality, such as detection of outlier objects, cross-validation test and y-scrambling. Lastly, as an ultimate proof, we used an external set to ensure the forecasting ability of our model and certify that it can be employed to predict the biological activity of new molecules reliably. Additionally, molecular docking simulations played an important role in the design of new compounds, whose pK_i values were predicted by our model.

Among the proposed molecules, we highlight compounds **2**, **5**, **9**, and **11** that presented expressive biological affinity values, overcoming even the molecules used to construct the model A. Furthermore, we underline that the presence of ester moieties attached to the aromatic ring at region A is well tolerated in this position and contributes to increase the biological activity values. Likewise, it is essential to emphasize that there are no bulky groups at region C; otherwise interactions with important amino acid residues (e.g. Trp100) will not be conceived. As a consequence, there would be a faster dissociation of the compounds from the D2 receptor active site, and, therefore, the non-occurrence of the expected partial agonist effects.

The significant biological activity values obtained from the new proposed compounds, as well as their optimum physicochemical and pharmacokinetic properties values achieved via in silico ADME/Tox prediction,

Table 5 ADME/Tox properties calculated for the new proposed compounds

Compound	MW	HB _{don}	HB _{accp}	QPlogP _{o/w}	QPlogS	QPPCaco	QPlogBB	PHOA	R05	stars
1	407.56	1.00	6.25	4.69	−6.292	432.11	−0.32	100.00	0	0
2	409.53	1.00	7.00	4.22	−5.974	270.17	−0.66	95.19	0	1
3	423.55	1.00	7.00	4.62	−6.064	419.94	−0.47	100.00	0	1
4	437.58	1.00	7.00	5.00	−6.468	420.18	−0.55	90.23	1	1
5	437.58	1.00	7.00	4.96	−6.606	404.95	−0.51	100.00	0	2
6	397.49	1.00	6.25	4.32	−5.581	420.63	−0.18	100.00	0	0
7	413.95	1.00	6.25	4.60	−6.097	402.26	−0.17	100.00	0	0
8	413.95	1.00	6.25	4.60	−5.981	419.60	−0.13	100.00	0	0
9	393.53	1.00	6.25	4.41	−5.936	402.22	−0.35	100.00	0	0
10	421.58	1.00	6.25	5.08	−6.736	404.69	−0.44	90.40	1	1
11	407.56	1.00	6.25	4.78	−6.333	403.17	−0.43	100.00	0	0
12	417.55	1.00	6.00	5.05	−6.688	306.90	−0.59	88.05	1	1
13	375.47	1.00	6.00	3.77	−5.282	270.58	−0.46	92.54	0	0
14	361.44	1.00	6.00	3.60	−5.190	263.50	−0.42	91.36	0	0
15	363.46	1.00	6.00	3.74	−5.100	377.97	−0.21	94.96	0	0
16	349.43	1.00	6.00	3.96	−5.970	163.79	−0.496	100.00	0	0

MW– Molecular weight (130.0 to 725.0 Da);

HB_{don} – Recommended number of hydrogen bond donor group (0.0–6.0)

HB_{accpt} – Recommended number of hydrogen bond accept group (2.0–20.0)

QPlogP_{o/w} – Predicted octanol/ water partition coefficient (−2.0 to 6.5)

QPlogS – Predicted aqueous solubility, log S. (−6.5 to 0.5)

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

QPlogBB – Predicted brain/blood partition coefficient. (−3.0 to 1.2)

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

R05 – Rule of five (Mol_MW < 500, QPlogP_{o/w} < 5, donor HB ≤ 5, and accptHB ≤ 10);

Stars – Number of property or descriptor values that are incongruous when compared with analogous values (95% interval) for known drugs (0–5)

singled out that these new molecules have drug-likeness characteristics and could be a reliable starting point to further with experimental studies (chemical synthesis, in vitro and in vivo analyses) to certify their drugability parameters.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11224-021-01742-w>.

Authors' contributions Aldineia P da Silva and Laise P. A. Chiari performed the analytic calculation and discussion, Aldineia P da Silva and Laise P. A. Chiari, Amanda R. Guimarães, Kathia M. Honorio and Albérico B. F. da Silva contributed to the final version of the manuscript. Alberico B. F. da Silva supervised the project.

Funding 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).

Availability of data and material Data included in Supplementary material.

Code availability No applicable

Declaration

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

References

- James SL et al (2018) Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. *Lancet* 392: 1789–1858. [https://doi.org/10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7)

2. Guloksuz S, van Os J (2018) The slow death of the concept of schizophrenia and the painful birth of the psychosis spectrum. *Psychol Med* 48:229–244. <https://doi.org/10.1017/S0033291717001775>
3. Pardiñas AF, Holmans P, Pocklington AJ et al (2018) Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* 50:381–389. <https://doi.org/10.1038/s41588-018-0059-2>
4. van Os J, Kenis G, Rutten BPF (2010) The environment and schizophrenia. *Nature* 468:203–212. <https://doi.org/10.1038/nature09563>
5. Johnson DS, Choi C, Fay LK et al (2011) Discovery of PF-00217830: aryl piperazine naphthyridinones as D2 partial agonists for schizophrenia and bipolar disorder. *Bioorg Med Chem Lett* 21:2621–2625. <https://doi.org/10.1016/j.bmcl.2011.01.059>
6. von Hausswolff-Juhlin Y, Bjartveit M, Lindström E, Jones P (2009) Schizophrenia and physical health problems. *Acta Psychiatr Scand* 119:15–21. <https://doi.org/10.1111/j.1600-0447.2008.01309.x>
7. Tran E, Rouillon F, Loze JY et al (2009) Cancer mortality in patients with schizophrenia: an 11-year prospective cohort study. *Cancer* 115. <https://doi.org/10.1002/cncr.24383>
8. Townsend LA, Malla AK, Norman RM (2001) Cognitive functioning in stabilized first-episode psychosis patients. *Psychiatry Res* 104:119–131. [https://doi.org/10.1016/S0165-1781\(01\)00302-X](https://doi.org/10.1016/S0165-1781(01)00302-X)
9. Rössler W, Joachim Salize H, van Os J, Riecher-Rössler A (2005) Size of burden of schizophrenia and psychotic disorders. *Eur Neuropsychopharmacol* 15:399–409. <https://doi.org/10.1016/j.euroneuro.2005.04.009>
10. Roshanaei-Moghaddam B, Katon W (2009) Premature mortality from general medical illnesses among persons with bipolar disorder: a review. *Psychiatr Serv* 60:147–156. <https://doi.org/10.1176/ps.2009.60.2.147>
11. Marder SR, Essock SM, Miller AL et al (2004) Physical health monitoring of patients with schizophrenia. *Am J Psychiatry* 161:1334–1349. <https://doi.org/10.1176/appi.ajp.161.8.1334>
12. Leucht S, Burkard T, Henderson J et al (2007) Physical illness and schizophrenia: a review of the literature. *Acta Psychiatr Scand* 116:317–333. <https://doi.org/10.1111/j.1600-0447.2007.01095.x>
13. Laursen TM, Munk-Olsen T, Nordentoft M, Mortensen PB (2007) Increased mortality among patients admitted with major psychiatric disorders: a register-based study comparing mortality in unipolar depressive disorder, bipolar affective disorder, schizoaffective disorder, and schizophrenia. *J Clin Psychiatry* 68. <https://doi.org/10.4088/JCP.v68n0612>
14. De Hert M, Dekker JM, Wood D et al (2009) Cardiovascular disease and diabetes in people with severe mental illness position statement from the European psychiatric association (EPA), supported by the European Association for the Study of diabetes (EASD) and the European Society of Cardiology (ESC). *Eur Psychiatry* 24:412–424. <https://doi.org/10.1016/j.eurpsy.2009.01.005>
15. Brown S, Kim M, Mitchell C, Inskip H (2010) Twenty-five year mortality of a community cohort with schizophrenia. *Br J Psychiatry* 196:116–121. <https://doi.org/10.1192/bjp.bp.109.067512>
16. Brown S, Inskip H, Barraclough B (2000) Causes of the excess mortality of schizophrenia. *Br J Psychiatry* 177. <https://doi.org/10.1192/bjp.177.3.212>
17. Beyer J, Kuchibhatla M, Gersing K, Krishnan KRR (2005) Medical comorbidity in a bipolar outpatient clinical population. *Neuropsychopharmacology* 30:401–404. <https://doi.org/10.1038/sj.npp.1300608>
18. Suzuki K, Kimura H (2018) TAK-063, a novel PDE10A inhibitor with balanced activation of direct and indirect pathways, provides a unique opportunity for the treatment of schizophrenia. *CNS Neurosci Ther* 24:604–614. <https://doi.org/10.1111/cns.12798>
19. Sahakyan L, Kwapil TR (2018) Episodic memory retrieval is impaired in negative schizotypy under fast response deadline. *Schizophr Res* 201:167–171. <https://doi.org/10.1016/j.schres.2018.05.039>
20. Kurtz MM (2006) Symptoms versus neurocognitive skills as correlates of everyday functioning in severe mental illness. *Expert Rev Neurother* 6:47–56. <https://doi.org/10.1586/14737175.6.1.47>
21. Dienel SJ, Lewis DA (2019) Alterations in cortical interneurons and cognitive function in schizophrenia. *Neurobiol Dis* 131:104208. <https://doi.org/10.1016/j.nbd.2018.06.020>
22. Meltzer HY, Huang M (2008) In vivo actions of atypical antipsychotic drug on serotonergic and dopaminergic systems. *Prog Brain Res* 172
23. Roth BL, Sheffler DJ, Kroeze WK (2004) Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat Rev Drug Discov* 3:353–359. <https://doi.org/10.1038/nrd1346>
24. Kondej M, Stepnicki P, Kaczor AA (2018) Multi-target approach for drug discovery against schizophrenia. *Int J Mol Sci* 19:3105. <https://doi.org/10.3390/ijms19103105>
25. Shapiro DA, Renock S, Arrington E et al (2003) Aripiprazole, a novel atypical antipsychotic drug with a unique and robust pharmacology. *Neuropsychopharmacology* 28:1400–1411. <https://doi.org/10.1038/sj.npp.1300203>
26. Kikuchi T, Tottori K, Uwahodo Y et al (1995) 7-(4-[(2,3-Dichlorophenyl)-1-piperazinyl]butyloxy)-3,4-dihydro-2(1H)-quinolinone (OPC-14597), a new putative antipsychotic drug with both presynaptic dopamine autoreceptor agonistic activity and post-synaptic D2 receptor antagonistic activity. *J Pharmacol Exp Ther* 274:329
27. Burris KD, Molski TF, Xu C et al (2002) Aripiprazole, a novel antipsychotic, is a high-affinity partial agonist at human dopamine D2 receptors. *J Pharmacol Exp Ther* 302:381–389. <https://doi.org/10.1124/jpet.102.033175>
28. Lawler CP, Prioleau C, Lewis MM et al (1999) Interactions of the novel antipsychotic aripiprazole (OPC-14597) with dopamine and serotonin receptor subtypes. *Neuropsychopharmacology* 20. [https://doi.org/10.1016/S0893-133X\(98\)00099-2](https://doi.org/10.1016/S0893-133X(98)00099-2)
29. Jordan S, Koprivica V, Chen R et al (2002) The antipsychotic aripiprazole is a potent, partial agonist at the human 5-HT1A receptor. *Eur J Pharmacol* 441:137–140. [https://doi.org/10.1016/S0014-2999\(02\)01532-7](https://doi.org/10.1016/S0014-2999(02)01532-7)
30. Stip E, Tourjman V (2010) Aripiprazole in schizophrenia and schizoaffective disorder: a review. *Clin Ther* 32:S3–S20. <https://doi.org/10.1016/j.clinthera.2010.01.021>
31. Zajdel P, Marciniec K, Maślankiewicz A et al (2013) Antidepressant and antipsychotic activity of new quinoline- and isoquinoline-sulfonamide analogs of aripiprazole targeting serotonin 5-HT1A/5-HT2A/5-HT7 and dopamine D2/D3 receptors. *Eur J Med Chem* 60:42–50. <https://doi.org/10.1016/j.ejmech.2012.11.042>
32. Shahlaei M (2013) Descriptor selection methods in quantitative structure–activity relationship studies: a review study. *Chem Rev* 113:8093–8103. <https://doi.org/10.1021/cr3004339>
33. Chen X, Sassano MF, Zheng L et al (2012) Structure-functional selectivity relationship studies of β -arrestin-biased dopamine D2 receptor agonists. *J Med Chem* 55:7141–7153. <https://doi.org/10.1021/jm300603y>
34. Frisch MJ, Trucks GW, Schlegel HB et al (2009). Gaussian:09
35. Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ (1994) Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J Phys Chem* 98:11623–11627. <https://doi.org/10.1021/j100096a001>
36. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98:5648. <https://doi.org/10.1063/1.464913>

37. Hehre WJ, Ditchfield R, Pople JA (1972) Self-consistent molecular orbital methods. XII Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J Chem Phys* 56:2257–2261. <https://doi.org/10.1063/1.1677527>
38. Hariharan PC, Pople JA (1973) The influence of polarization functions on molecular orbital hydrogenation energies. *Theor Chim Acta* 28:213–222. <https://doi.org/10.1007/BF00533485>
39. Ditchfield R, Hehre WJ, Pople JA (1971) Self-consistent molecular-orbital methods IX An extended gaussian-type basis for molecular-orbital studies of organic molecules. *J Chem Phys*:54. <https://doi.org/10.1063/1.1674902>
40. Tetko IV, Gasteiger J, Todeschini R et al (2005) Virtual computational chemistry laboratory – design and description. *J Comput Aided Mol Des* 19:453–463. <https://doi.org/10.1007/s10822-005-8694-y>
41. Kiralj R, Ferreira MMC (2009) Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. *J Braz Chem Soc* 20:770–787
42. de Oliveira DB, Gaudio AC (2000) BuildQSAR: a new computer program for QSAR analysis. *Quant Struct Relationships* 19:599–601. [https://doi.org/10.1002/1521-3838\(200012\)19:6<599::AID-QSAR599>3.0.CO;2-B](https://doi.org/10.1002/1521-3838(200012)19:6<599::AID-QSAR599>3.0.CO;2-B)
43. Infometrix Inc (2002) Pirouette 3.10
44. Schrödinger (2018) Maestro- Schrödinger Release 2018–1
45. Ferreira MMC (2002) Multivariate QSAR. *J Braz Chem Soc* 13: 742–753
46. Martins JPA, Ferreira MMC (2013) QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR. *Quim Nova* 36:554–560. <https://doi.org/10.1590/S0100-40422013000400013>
47. Bonifazi A, Yano H, Guerrero AM et al (2019) Novel and potent dopamine D₂ receptor go-protein biased agonists. *ACS Pharmacol Transl Sci* 2:52–65. <https://doi.org/10.1021/acspsci.8b00060>
48. Dagan-Wiener A, Nissim I, Ben Abu N et al (2017) Bitter or not? BitterPredict, a tool for predicting taste from chemical structure. *Sci Rep* 7:12074. <https://doi.org/10.1038/s41598-017-12359-7>
49. Bocci G, Carosati E, Vayer P et al (2017) ADME-space: a new tool for medicinal chemists to explore ADME properties. *Sci Rep* 7: 6359. <https://doi.org/10.1038/s41598-017-06692-0>
50. Tropsha A, Gramatica P, Gombar V (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb Sci* 22:69–77. <https://doi.org/10.1002/qsar.200390007>
51. Golbraikh A, Tropsha A (2002) Beware of q²! *J Mol Graph Model* 20:269–276. [https://doi.org/10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1)
52. Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. *Mol Inform* 29:476–488. <https://doi.org/10.1002/minf.201000061>
53. Arroio A, Honório KM, da Silva ABF (2010) Propriedades químico-quânticas empregadas em estudos das relações estrutural-atividade. *Quim Nova* 33:694–699. <https://doi.org/10.1590/S0100-40422010000300037>
54. Honório KM, Da Silva ABF (2003) An AM1 study on the electron-donating and electron-accepting character of biomolecules. *Int J Quantum Chem* 95:126–132. <https://doi.org/10.1002/qua.10661>
55. Wang S, Che T, Levit A et al (2018) Structure of the D2 dopamine receptor bound to the atypical antipsychotic drug risperidone. *Nature* 555:269–273. <https://doi.org/10.1038/nature25758>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.