PAPER

Check for updates

Machine learning *vs.* field 3D-QSAR models for serotonin 2A receptor psychoactive substances identification†

Giuseppe Floresta [ID] * and Vincenzo Abbate [ID] *

Serotonergic psychedelics, substances exerting their effects primarily through the serotonin 2A receptor (5HT2AR), continue to comprise a substantial portion of reported new psychoactive substances (NPS). In this paper five quantitative structure−activity relationship (QSAR) models for predicting the affinity of 5-HT2AR ligands have been developed. The resulting models, exploiting the accessibility of the QSAR equations, generate a useful tool for the investigation and identification of unclassified molecules. The models have been built using a set of 375 molecules using Forge software, and the quality was confirmed by statistical analysis, resulting in effective tools with respect to their predictive and descriptive capabilities. The best performing algorithm among the machine learning approaches and the classical field 3D-QSAR model were then combined to produce a consensus model and were exploited, together with a pharmacophorefilter, to explore the 5-HT2AR activity of 523 105 natural products, to classify a set of recently reported 5-HT2AR NPS and to design new potential active molecules. The findings of this study should facilitate the identification and classification of emerging 5-HT2AR ligands including NPS.

Introduction

The 5-HT2A receptor (5HT2AR) is a surface G protein-coupled receptor (GPCR) subtype of the 5-HT2 serotonin receptor family. The receptor was first discovered as a target of serotonergic psychedelic drugs such as LSD and psilocybin, and later it was proved to be a mediator of the action of many antipsychotic drugs.[1] Due to the wide expression of the receptor in the central nervous system (CNS) and other tissues, several physiological processes are mediated by the receptor: neuronal excitation, hallucinations, out-of-body experiences, and fear; its activation in the hypothalamus causes increases in several hormonal levels; activation of the receptor produces potent anti-inflammatory effects in several tissues including cardiovascular; it also has a role in memory and learning, in arthralgia, Alzheimer's disease, sleep paralysis, *etc.*[2–6]

Agonists and antagonists of this receptor are today used or being studied for several clinical application, *e.g.* methysergide (partial agonists) is used in treatment of migraine, AL-34662 (peripherally selective agonists) reduces the pressure inside the eyes without crossing the blood–brain barrier and producing hallucinogenic side effects;[7] atypical antipsychotic drugs such as risperidone, clozapine, quetiapine and asenapine

are relatively potent antagonists of 5-HT2A, and nelotanserin (inverse agonist) is studied for the treatment of insomnia.[8] 5HT2AR is also targeted for neuroimaging of patients with major depressive disorder using PET imaging.[9] However, due to the natural expression of this receptor in the CNS, the receptor is also targeted by new psychoactive substances (NPS) designed for recreational use, and every year hundreds of NPS are unearthed on the black market.

By early 2020, more than 950 NPS had been reported to the United Nations Office on Drugs and Crime (UNODC); in parallel, at the end of 2019 the European Monitoring Centre for Drugs and Drug Addiction (EMCDDA) was monitoring around 790 NPS, 53 of which were reported for the first time in Europe in 2019.[10,11] Over the time, the group of serotonergic psychedelics constituted a significant proportion of these recently reported NPS.[12] The term serotonergic psychedelics includes molecules that exert their pharmacological activity mainly by means of interacting with the serotonin 5-HT2A receptor (5-HT2AR).[13] The class of serotonergic psychedelics is formed by structurally diverse subclasses of compounds: the tryptamines (*e.g.* psilocybin), the ergolines (*e.g.* lysergic acid diethylamide (LSD)) and the phenylalkylamines (*e.g.* mescaline), Fig. 1.[14] Mystical experiences, empathic feelings, alterations in consciousness, sensory and somatic effects are the main effects looked among users of 5HT2AR NPS ligands. However, severe adverse effects, such as agitation, tachycardia, hyperthermia, rhabdomyolysis, hypertension and seizures, can also frequently occurr.[15,16] With the more recent group of NBOMes, an

Department of Analytical, Environmental and Forensic Sciences, King's College London, London, UK. E-mail: giuseppe.floresta@kcl.ac.uk; vincenzo.abbate@kcl.ac.uk
† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1ra01335a

emerging subclass of the phenylalkylamines psychedelics, cases requiring hospitalization, suicide attempts, deaths and mass poisoning (with *e.g.* BromoDragonFLY and 2C-E) have been reported.[17–19]

QSAR models are utilized to assist predicting or understanding molecular/drug design within the chemical and natural sciences.[20,21] Few attempts have been performed to build up QSAR models for the 5HT2AR, however all of them have been produced through a constrained number of compounds with comparable chemical structures and as a result, only the affinities of a restricted class of compounds could be determined.[22–24]

To facilitate the investigation of chemical datasets for 5HT2AR ligands capabilities and to potentially identify emerging or future NPS,[25,26] here we report the development of five quantitative structure–activity relationship (QSAR) models for predicting the affinity of 5-HT2AR ligands. Details of all the compounds having experimentally determined K_i values were retrieved from the literature and downloaded from CHEMBL. All models were built using a set of 375 5-HT2AR ligands. Four out of five such models were built by machine learning algorithms (*i.e.* support vector machine (SVM), *k*-nearest neighbors (*k*NN), random forest (RF) and relevance vector machine (RVM)) and one model was built with a field-based methodology.[27,28] Differently to previously published 5HT2AR QSAR models, the ones reported here include a wide range of chemically different (sub)classes of compounds. Moreover, the best performing tool from the generated machine learning (SVM) and the field based 3D-QSAR models were employed to rank a dataset of recently reported 5HT2AR NPS ligands and to screen a large dataset of natural products to identify potential active molecules against 5-HT2AR.

All the reported QSAR models were developed using the software Forge. Conversely to classical 3D-QSAR modelling, where molecular descriptors are calculated at the interception points of a 3D grid, which surrounds the entire space of the aligned molecules,[29–31] the modelling calculation in Forge is characterized by the use of probe positions that are defined directly from the field points of the aligned molecules in the training set, and only these positions are then used to describe the volume and the electrostatic potential of each molecule.[32–35]

Methods

Biological data & 3D structures generation

The chemical structures of the 375 5-HT2AR ligands were selected from the literature where the experimental K_i values are reported and retrieved from ChEMBL Database (https://www.ebi.ac.uk/chembl/g/). The biological data were all derived from similar and comparable cellular based experiments. Ki derived from displacement of [3H]-ketanserin, [125I]-DOI or [3H]INBMeO from human 5HT2AR expressed CHO or HEK293 cells were used. The binding affinity data of the selected dataset were converted into their negative decimal logarithm pK_i ($pK_i = -\log K_i$). Collected pK_i values fall into a range 5.19–10.40. The structures of the studied molecules were built using Marvin 17.21.0, ChemAxon (https://www.chemaxon.com).[36] The 2D structures were subjected to molecular mechanics energy minimization by Merck molecular force field (MMFF94) using the Marvin Sketch geometrical descriptors plugin.[37] The protonation states of the molecules were calculated assuming a pH = 7.0. The geometry of the obtained 3D structures was further optimized at semi-empirical level using the parameterized model number 3 (PM3) Hamiltonian as implemented in MOPAC package (vMOPAC2016).[38–40]

Compound alignment

All the 3D generated molecules, with their respective pK_i values, were imported into the computational chemistry software Forge (v10.4.2) for setting the machine learnings (SVM, *k*NN; RF and RVM) and the field-based 3D-QSAR model. Out of the 375 ligands for the 5HT2AR, we randomly selected (selection based on activity stratification) 300 molecules (80%) as a training set to build the models, while the remaining 75 compounds (20%) served as test set to evaluate the models.[41] For both training and test sets, the selected molecules covered a wide range of biological activities: from 10.40 to 5.19 pK_i for the training set and from 10.00 to 5.34 pK_i for the test set. All the molecules were aligned to the three co-crystallized ligands of 5HT2AR in their bioactive conformation inside the binding site retrieved from the protein data bank (https://www.rcsb.org/) PDB ID: 6a93, 6a94 and 6wgt.[42,43] 8NU (risperidone) was the ligand selected for 6a93, ZOT (zotepine) was the ligand selected for 6a94, 7LD ((8alpha)-N,N-diethyl-6-methyl-9,10-didehydroergoline-8-carboxamide) was the ligand selected for 6wgt.

The field points (used as a descriptor of negative and positive electrostatic, van der Waals shape, and hydrophobic areas)[28] of each molecule were generated using the extended electron distribution (XED) force field. The molecules in both training set and test set were aligned to the reference compounds by a most common substructure calculation (considering the calculated field points) employing a customized set-up.[44] The maximum number of conformations produced for each compound was set to 500. The root-mean-square deviation of nuclear positions (RMSD) cutoff for copy conformers was set to 0.5 Å. This parameter controls the likeness which two conformers are accepted indistinguishable. The gradient cutoff for conformer minimization was set to 0.1 kcal mol^{-1}. The energy window was set to 2.5 kcal mol^{-1}. Conformations that gave a minimized energy outside the energy window were discarded. All the alignments were manually checked to ensure the best possible alignment. All the field points of the molecules used for training the models were exploited to get a gauge
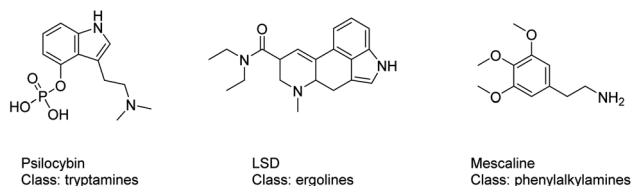

Fig. 1 Structures of 5HT2AR psychoactive substances: the tryptamines (*e.g.* psilocybin), the ergolines (*e.g.* lysergic acid diethylamide (LSD)) and the phenylalkylamines (*e.g.* mescaline).

Table 1 Models statistics

Model	r^2 training set	q^2 training set	r^2 test set	MSE^a training set	MSE^a test set	MAE^b training set	MAE^b test set	$MAPE^c$ training set	$MAPE^c$ test set
3D-field	0.88	0.75	0.73	0.13	0.29	0.31	0.45	4.21	6.24
SVM	0.99	0.74	0.73	0.01	0.30	0.06	0.44	0.84	6.21
kNN	0.66	0.66	0.49	0.38	0.56	0.48	0.56	6.75	7.70
RF	0.89	0.64	0.63	0.11	0.41	0.22	0.52	2.96	7.20
RVM	0.89	0.73	0.69	0.12	0.34	0.29	0.40	3.93	5.44

^{a} Mean squared forecast error. ^{b} Mean absolute forecast error. ^{c} Mean absolute percentage forecast error.
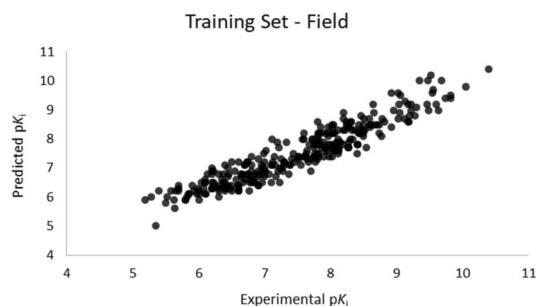
Fig. 2 Five component field based QSAR model experimental *vs.* predicted pK_i of the compounds in the training set.

invariant set of sampling points, which reduced the number of descriptors that must be considered. The sample values were calculated using a distance of 1 Å between the sample points, ensuring that all areas around the compounds that could contribute to the activity were effectively described. All the software's parameters used for the conformation hunt, alignment, and build model calculations are reported in the ESI.†

Results and discussion

Statistical analysis and results

For the calculation of the models, a partial least squares (PLS) regression method specifically employing the SIMPLS algorithm was used for the field model, whereas *k*-Nearest Neighbors, Support Vector Machine, Relevance Vector Machine and

Random Forest were used in the supervised machine learning models.[45,46] Detailed information for the assembly and the validation of all the models are reported in the ESI.† The 3D-QSAR models' statistics are reported in Table 1. The 5-component field based model shows both good predictive and descriptive capabilities, demonstrated by the good r^2 (0.88) and q^2 (0.75) values for the training and the cross-validated training set (Fig. 2).[47] The plots of experimental *vs.* predicted affinities for the compounds in the test set (Fig. 3) show accurate predictions with only a few outliers and an excellent cross-validated r^2 (0.73). Regarding the supervised learning models (SVM, *k*NN, RF and RVM), the r^2 of the models ranked between 0.99 and 0.66, the q^2 between 0.74 and 0.66 and the r^2 test between 0.73 and 0.79 (Table 1). According to these results, the best performing algorithm among the supervised machine learning models is the SVM. The model experimental *vs.* predicted pK_i of the compounds in the training set and in the test set for the SVM model are reported in Fig. 4 and 5.

The 3D visualizations of the field QSAR model is shown in Fig. 6, where the 3D-QSAR coefficients for the two models are superimposed to the most potent molecule in the training set and to risperidone, that was used in the alignment process (PDBID: 6a93). The 3D-QSAR model is portrayed by both steric and electrostatic impacts. The model outlines zones where the equation proposes that the nearby areas have a strong effect on ligand-receptor affinity (Fig. 6). The bigger the points (portrayed as octahedrons), the stronger is the relationship between the electrostatic and steric areas in that position. The higher affinity related to the electrostatic potential is delineated in red for the positive values and in blue for the negative ones. For the steric
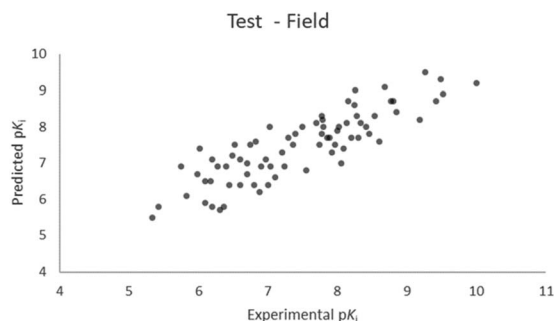

Fig. 3 Five component field based QSAR model experimental *vs.* predicted pK_i of the compounds in the test set.

Fig. 4 SVM QSAR model experimental *vs.* predicted pK_i of the compounds in the training set.
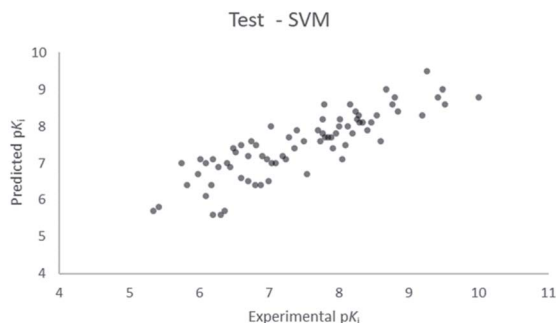
RSC Advances

Fig. 5 SVM QSAR model experimental *vs.* predicted pK_i of the compounds in the test set.

bulk, the green region leads to higher receptor affinity while the violet zone leads to lower affinity.

To uncover the key highlights of the examined set of compounds against the focused 5HT2AR, a structure–activity relationship (SAR) study was performed through activity-atlas (AA) visualization program. AA is a qualitative strategy valuable for summarizing structure–activity data into 3D maps, which helps within the design and optimization of new compounds. This strategy analyses the SAR of a set of aligned compounds as a function of their electrostatic, hydrophobic, and shape properties through a Bayesian approach to require a global see of the information in a qualitative way. Fig. 7 and 8 show the results of the AA calculations for the 5HT2AR. The model map is superimposed to risperidone in its bioactive conformation retrieved from the crystal structure (6a93). Electrostatic, hydrophobic, and shape features are highlighted by the different colors on the map. In the red area, a more positive electrostatic field increases the receptor-affinity, whereas in the blue region a more negative electrostatic field increases the

interaction energies. The violet and the green areas account for the steric and bulk/hydrophobic interactions. In the violet area, a steric/bulk interaction decreases the affinity, whereas in the green area a steric/bulk interaction improves the binding affinity.

Risperidone can be dissected in three different regions: (i) the fluorobenzisoxazol ring; (ii) the pyperidine nucleus; (iii) the tetrahydropyridopyrimidinone ring. Two main hydrophobic areas are described by the AA model for the fluorobenzisoxazol ring and the tetrahydropyridopyrimidinone ring, whereas the pyperidine nucleus is mainly located close to a red area, where a more positive electrostatic field increases the receptor-affinity. The occupancy of both hydrophobic area is fundamental for the affinity; indeed, ligands that do not bear any substituent in this area result in low affinity (*e.g.* (*S*)-1-(2,5-dimethoxyphenyl) propan-2-amine, pK_i = 5.28, training set; 5-hydroxy-8-(4-methyl-1,4-diazepan-1-yl)-2*H*-benzo[*b*][1,4]oxazin-3(4*H*)-one, pK_i
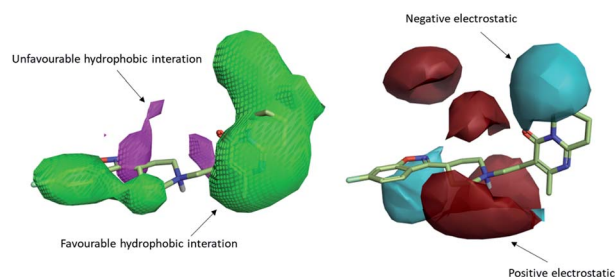

Fig. 7 The model map is superimposed to risperidone. Molecular insight of SAR mechanism models, revealing the different lead optimization sites of active compounds. Red color shows positive field region controlling the activity, and blue color the negative ones. Green color shows favorable shape/hydrophobic regions, and purple color the unfavorable ones.
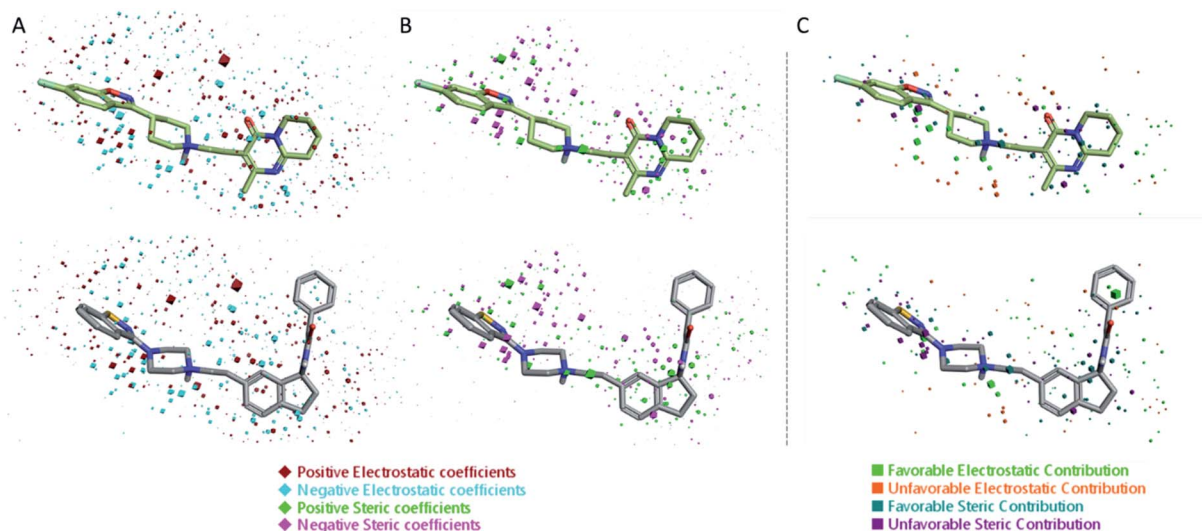

Fig. 6 Electrostatic and steric coefficients for the field model. (A) Electrostatic coefficients superimposed to risperidone (used for the alignment, PD BID: 6a93) (up) and the most potent compounds in the training set (down). (B) Steric coefficients superimposed to risperidone (up) and the most potent compounds in the training set (down). (C) Contributions to predicted affinity for risperidone (up) and the most potent compounds in the training set (down).

= 5.19, training set). Inside the binding pocket, risperidone adopts an extended conformation, as showed by the AA model. The basic nitrogen of the pyperidine nucleus forms a salt bridge with Asp155, this interaction being well highlighted in the AA model by the red area surrounding the nitrogen of the pyperidine. This salt bridge is strictly conserved in the structures of aminergic receptors and is thought to be stabilized by the conserved hydrogen bond between Asp155 and Tyr370. The fluorobenzisoxazol ring of risperidone is placed in the bottom hydrophobic cleft represented by a green, favourable hydrophobic interaction, by the AA model. This ring interacts with Ser159 by a CH–π interaction and with Ile163, Phe243 and Phe332 by hydrophobic interactions. As already stated, the occupancy of this area is fundamental for ligand affinity; indeed, mutation of these residues will result in a loss in affinity for 5HT2AR ligands.[42] The tetrahydropyridopyrimidinone ring is located in the other hydrophobic area of the AA model, interacting with Leu228 and Val366 of the protein.

NPS identification

Together with synthetic cannabinoids, cathinone derivatives, designer GABA-A/B receptor agonists and novel synthetic opioids, recently emerged and unregulated molecules with high 5HT2AR affinity can be considered NPS. These substances – once identified – are normally screened through cellular *in vitro* tests, that are expensive and time consuming. To exploit the

predictive capabilities of our models we selected 29 recently reported NPS (phenethylamines, phenylisopropylamines and *N*-benzylphenethylamines)[48] with measured EC50, by means of β-arrestin 2 recruitment in 5HT2AR and compared the *in vitro* results with the modelled results of the field based and best machine learning (SVM) based models. Despite the K_i measures the effective interaction of the molecule with the receptor whereas the EC50 measures the effective concentration at 50% of the total biological effect, the two K_i and EC50 are held together by the Cheng–Prusoff equation.[49] This means that for a giving set of agonist or antagonist agents for the 5HT2AR, the most potent agonist or antagonist will have the lowest K_i (highest pK_i). Of note, the models here reported are not able to distinguish between agonist or antagonist action but are able to rank sets of agonists or antagonists according to their receptor affinity.

The results of the experimental EC$_{50}$ *vs.* the calculated pK_i are reported in Fig. 9 and Table S3.† Interestingly, both the field based and the SVM models well performed in the ranking of the 29 compounds, with a slightly better performance for the machine learning model. Both models are able to rank the compounds with a good linearity between the calculated and the experimental values, with r^2 of 0.7105 and 0.7799 for the field and the SVM based models, respectively. More importantly, the models were able to distinctly classify potent compounds from not-potent ones *i.e.* the five most potent compounds (25C-NBOH, 25E-NBOMe, 25D-NBOMe, 25I-
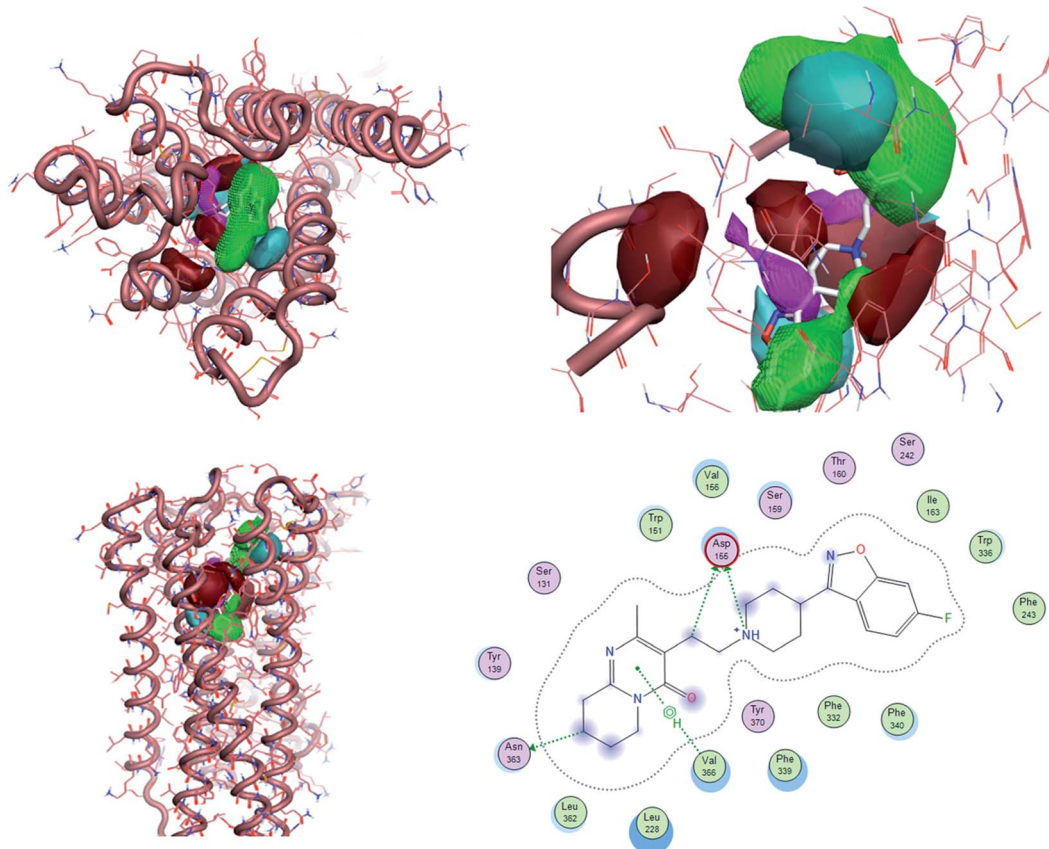

Fig. 8 The model AA map is superimposed to risperidone inside 5HT2AR (6a93). 2D interaction between risperidone and 5HT2AR.
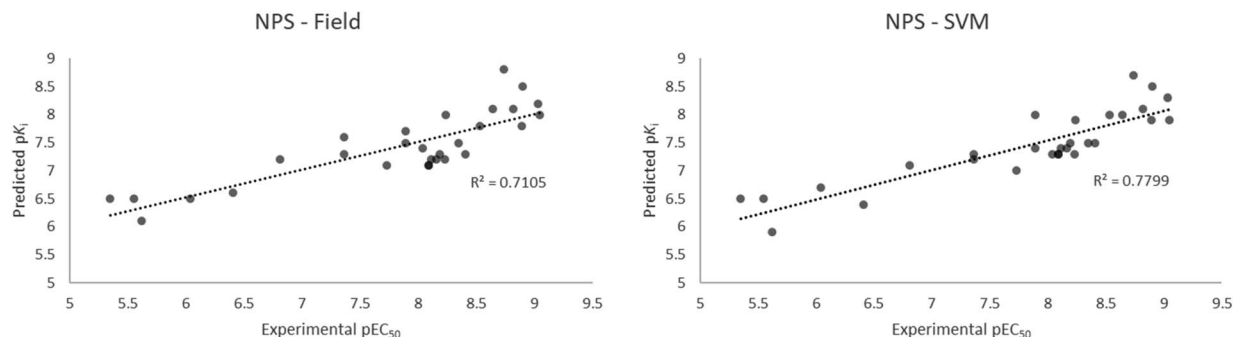
Fig. 9 Experimental EC_{50} vs. calculated pK_i of the selected 29 NPS with the field (left) and the SVM (right) models.

NBOMe, 25I-NBOH) in the set were identified among the most potent compounds for 5HT2AR activity by both models. Analogously, the five less potent molecules (25H-NBF, *b-k*-2C-B, 2C-H, DOH, N-Me-2C-H) were also identified as weak binders by both models.

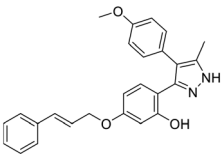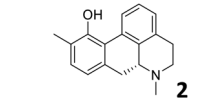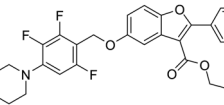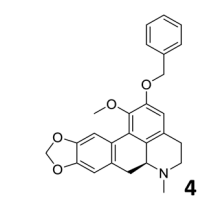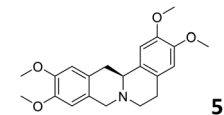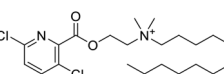### Finding NPS from natural databases and scaffold hopping analysis

In order to exploit the predicting capability of the models to retrieve novel 5HT2AR binders, the QSAR equations were used to screen for potential active molecules in three databases of natural compounds for a total of 523,105 molecules. Super Natural Product (SNP), ZINC and MolPort Natural (MPN) were selected as a database of natural compounds.[50] The 523 105 molecules were firstly filtered through a pharmacophore filter using pharmit (http://pharmit.csb.pitt.edu). The chemical database was screened against the pharmacophore models (generated from the three co-cristallyzed ligands of 5HT2AR in their bioactive conformation retrieved from the protein data bank PDB ID: 6a93, 6a94 and 6wgt) to find the best matches in terms of root mean square distance (RMSD) between pharmacophore query features and corresponding ligand points. 910 compounds with the lowest RMSD value (less than 0.75) were selected for further evaluation in our models. The field based and the SVM QSAR models were then used to score the filtered dataset of 910 compounds. With this objective in mind, all of the molecules were aligned to the models as described in the compound alignment paragraph. The six most potent molecules identified by means of the calculated pK_i from both models are reported in Table 2 and the whole set in Table S4.† Despite molecules **1**, **3** and **6** have not been reported as 5HT receptor binders (molecule **1** class was reported for the treatment of cancer[51] and molecule **6**[52] is derived from pyridyloxy carboxylic acid commonly used as herbicidal whereas no biological activity is reported for molecule **3**), molecule **2**, **4** and **5** have been tested for their activity against serotoninergic receptors showing the potentiality of the models to find new hits among library of compounds and would deserve further research investigation to better understand the potential 5HT2AR activity of high ranked compounds.[53,54]

All the molecules selected as most potent (Table 2) are able to interact with the 5HT2AR correctly inside its binding site. The

protonated nitrogen in molecules **2–6** is forming the fundamental salt-bridge with ASP155, moreover the aromatic part of the molecules are accommodated in the two hydrophobic pockets inside the receptor formed by SER131, TYR139, ASN363, LEU362, LEU228, VAL366 or PHE340, PHE243, TRP336, ILE163, SER242, THR160, SER159 gaining stability thought hydrophobic interactions.

The predictive capabilities of the model were then used for a scaffold-hopping study of the most potent compounds in the training set for the generation of a library of new compounds

Table 2 Top 6 ligands from the scoring of natural products

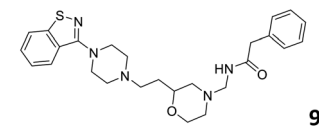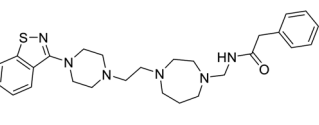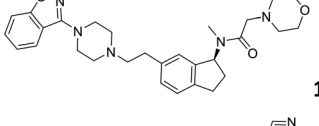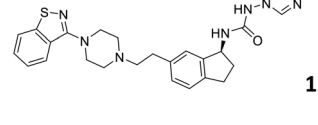| Molecule	Calculated pK_i
1	9.15
2	8.95
3	8.95
4	8.90
5	8.90
6	8.75

Paper
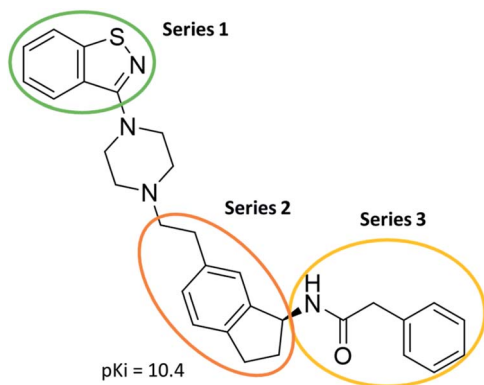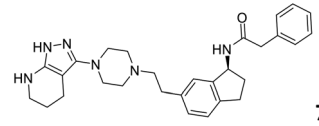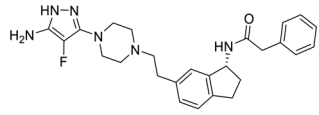
Fig. 10 Scaffold hopping series 1–3 derived from the most potent molecule in the training set.

with high affinity for the receptor. The molecule was divided in three parts and a scaffold hopping analysis was performed as already reported by us (ref. 21, 25 and 26) for each part (Fig. 10). In series 1 the benzo[*d*]isothiazole was substituted, in series 2 the ethyl-2,3-dihydro-1*H*-indene core was substituted and in series 3 the terminal phenylacetamide was substituted. The piperazine nucleus was not replaced due to the important

Table 3 Top 6 ligands from the scaffold replacement series 1–3. Molecules 7 and 8, series 1. Molecules 9 and 10, series 2. Molecules 11 and 12, series 3

Molecule	Calculated pK_i
7	11.2
8	11.1
9	11.3
10	11.3
11	11.1
12	11.0

interaction with ASP155. The resulting molecules were then evaluated by the superposition on the 3D-QSAR models. Overall, the results showed in Table 3 (best two molecules for each series) indicate that the scaffold replacement generated new structures with the appropriate chemical features for the binding to the 5HT2AR and the selected compounds resulted more potent than their precursor, showing again the potential of the models to identify new hits among library of compounds and would deserve further research investigation to better understand the potential 5HT2AR activity.

Conclusions

The present study deals with the development of different field based and machine learning based 3D-QSAR models for the prediction of the affinity for 5HT2AR ligands and their successive use as a tool for ranking 5HT2AR NPS, the screening of a large dataset of natural molecules and new designed compounds. Forge was used to build a statistically robust 3D-QSAR model using a set of 375 molecules for the 5HT2AR covering a whole range of different chemical classes of the ligands for 5HT2A protein. To the best of our knowledge, this is the first attempt of 3D-QSAR modelling for this subtype of serotoninergic receptor which includes such a wide number and range of molecular structures and accounts for the observed SAR. The 3D-QSAR models showed high statistical quality and robust predictive potential capability, particularly the field based and the SVM which were shown to possess the best statistical quality. Visualization of the model by means of AA allowed processing data in a 3D map format accounting for both steric and electrostatic effects and allow for a rationalization of both potency and selectivity. The field-based model and the SVM-based model were then used to rank a dataset of recently reported 5HT2AR NPS and screen a dataset of natural products for potential active molecules against 5HT2AR. Interestingly, the models were able to rank the experimental EC50 and the calculated K_i of the NPS in a linear manner and to identify alkaloids already tested for their activity against the 5HT2AR. The 3D-QSAR models here reported will guarantee, prospectively, fruitful applications to speed up the design of novel therapeutic molecules for 5HT2AR as well as the identification/classification of NPS acting on 5HT2AR.[25,26]

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 E. H. Cook Jr, K. E. Fletcher, M. Wainwright, N. Marks, S. Y. Yan and B. L. Leventhal, *J. Neurochem.*, 1994, **63**, 465–469.

2 B. Jalal, *Psychopharmacology*, 2018, **235**, 3083–3091.

3 A. Passier and E. van Puijenbroek, *Br. J. Clin. Pharmacol.*, 2005, **60**, 570–572.

4 J. A. Harvey, *Learn. Mem.*, 2003, **10**, 355–362.

5 F. Nau Jr, B. Yu, D. Martin and C. D. Nichols, *PLoS One*, 2013, **8**, e75426.

6 L. D. Van de Kar, A. Javed, Y. Zhang, F. Serres, D. K. Raap and T. S. Gray, *J. Neurosci.*, 2001, **21**, 3572–3579.

7 N. A. Sharif, M. A. McLaughlin and C. R. Kelly, *J. Ocul. Pharmacol. Ther.*, 2007, **23**, 1–13.

8 R. Rosenberg, D. J. Seiden, S. G. Hull, M. Erman, H. Schwartz, C. Anderson, W. Prosser, W. Shanahan, M. Sanchez, E. Chuang and T. Roth, *Sleep*, 2008, **31**, 1663–1671.

9 M. M. Herth and G. M. Knudsen, *J. Labelled Compd. Radiopharm.*, 2015, **58**, 265–273.

10 UNODC, *Current NPS Threats*, 2020.

11 EMCDDA, *European Drug Report 2020*, Trends and Developments, 2020.

12 F. Schifano, S. Chiappini, A. Miuli, J. M. Corkery, N. Scherbaum, F. Napoletano, D. Arillotta, C. Zangani, V. Catalani, A. Vento, M. Pettorruso, G. Martinotti, M. di Giannantonio and A. Guirguis, *Exp. Neurol.*, 2021, 113638, DOI: 10.1016/j.expneurol.2021.113638.

13 D. E. Nichols, *Pharmacol. Rev.*, 2016, **68**, 264–355.

14 D. E. Nichols, in *Behavioral Neurobiology of Psychedelic Drugs*, ed. A. L. Halberstadt, F. X. Vollenweider and D. E. Nichols, Springer Berlin Heidelberg, Berlin, Heidelberg, 2018, pp. 1–43, DOI: 10.1007/7854_2017_475.

15 B. V. Dean, S. J. Stellpflug, A. M. Burnett and K. M. Engebretsen, *J. Med. Toxicol.*, 2013, **9**, 172–178.

16 C. Kyriakou, E. Marinelli, P. Frati, A. Santurro, M. Afxentiou, S. Zaami and F. P. Busardo, *Eur. Rev. Med. Pharmacol. Sci.*, 2015, **19**, 3270–3281.

17 L. Hondebrink, A. Zwartsen and R. H. S. Westerink, *Pharmacol. Ther.*, 2018, **182**, 193–224.

18 D. Luethi and M. E. Liechti, *Arch. Toxicol.*, 2020, **94**, 1085–1133.

19 S. Iwersen-Bergmann, S. Lehmann, A. Heinemann, C. Schröder, A. Müller, H. Jungen, H. Andresen-Streichert, K. Pueschel, C. Vidal and K. Mercer-Chalmers-Bender, *Z. Rechtsmed.*, 2019, **133**, 123–129.

20 G. Floresta, E. Amata, M. Dichiara, A. Marrazzo, L. Salerno, G. Romeo, O. Prezzavento, V. Pittala and A. Rescifina, *ChemMedChem*, 2018, **13**, 1336–1342.

21 G. Floresta, V. Pittalà, V. Sorrenti, G. Romeo, L. Salerno and A. Rescifina, *Bioorg. Chem.*, 2018, **81**, 334–339.

22 A. Hamacher, M. Weigt, M. Wiese, B. Hoefgen, J. Lehmann and M. U. Kassack, *BMC Pharmacol.*, 2006, **6**, 11.

23 J. Brea, J. Rodrigo, A. Carrieri, F. Sanz, M. I. Cadavid, M. J. Enguix, M. Villazón, G. Mengod, Y. Caro, C. F. Masaguer, E. Raviña, N. B. Centeno, A. Carotti and M. I. Loza, *J. Med. Chem.*, 2002, **45**, 54–71.

24 M. Radan, D. Ruzic, M. Antonijevic, T. Djikic and K. Nikolic, *J. Biomol. Struct. Dyn.*, 2020, 1–19, DOI: 10.1080/07391102.2020.1738961.

25 G. Floresta, O. Apirakkan, A. Rescifina and V. Abbate, *Molecules*, 2018, **23**, 2183.

26 G. Floresta, A. Rescifina and V. Abbate, *Int. J. Mol. Sci.*, 2019, **20**, 2311.

27 *Intelligent and Cloud Computing*, ed. D. Mishra, R. Buyya, P. Mohapatra and S. Patnaik, Springer, Singapore, pp. 495-507, DOI: 10.1007/978-981-15-5971-6.

28 T. Cheeseright, M. Mackey, S. Rose and A. Vinter, *J. Chem. Inf. Model.*, 2006, **46**, 665–676.

29 X. Zhao, M. Chen, B. Huang, H. Ji and M. Yuan, *Int. J. Mol. Sci.*, 2011, **12**, 7022–7037.

30 B. Q. Cai, H. X. Jin, X. J. Yan, P. Zhu and G. X. Hu, *Acta Pharmacol. Sin.*, 2014, **35**, 151–160.

31 G. Floresta, E. Amata, C. Barbaraci, D. Gentile, R. Turnaturi, A. Marrazzo and A. Rescifina, *Mar. Drugs*, 2018, **16**, 384.

32 S. Alam and F. Khan, *Sci. Rep.*, 2017, **7**, 6019.

33 S. Fallarini, A. Massarotti, A. Gesu, S. Giovarruscio, G. C. Zabetta, R. Bergo, B. Giannelli, A. Brunco, G. Lombardi, G. Sorba and T. Pirali, *MedChemComm*, 2016, **7**, 409–419.

34 J. W. Lee, T. Hirota, A. Kumar, N. J. Kim, S. Irle and S. A. Kay, *ChemMedChem*, 2015, **10**, 1489–1497.

35 G. Floresta, A. Cilibrizzi, V. Abbate, A. Spampinato, C. Zagni and A. Rescifina, *Bioorg. Chem.*, 2019, **84**, 276–284.

36 T. Barf, F. Lehmann, K. Hammer, S. Haile, E. Axen, C. Medina, J. Uppenberg, S. Svensson, L. Rondahl and T. Lundback, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 1745–1748.

37 R. G. Pertwee, *Pharmacol. Ther.*, 1997, **74**, 129–180.

38 J. J. Stewart, *J. Mol. Model.*, 2004, **10**, 155–164.

39 C. Alemán, F. J. Luque and M. Orozco, *J. Comput. Chem.*, 1993, **14**, 799–808.

40 F. Qiao, L. Luo, H. Peng, S. Luo, W. Huang, J. Cui, X. Li, L. Kong, D. Jiang, D. J. Chitwood and D. Peng, *PLoS One*, 2016, **11**, e0160003.

41 P. P. Roy, J. T. Leonard and K. Roy, *Chemom. Intell. Lab. Syst.*, 2008, **90**, 31–42.

42 K. T. Kimura, H. Asada, A. Inoue, F. M. N. Kadji, D. Im, C. Mori, T. Arakawa, K. Hirata, Y. Nomura, N. Nomura, J. Aoki, S. Iwata and T. Shimamura, *Nat. Struct. Mol. Biol.*, 2019, **26**, 121–128.

43 K. Kim, T. Che, O. Panova, J. F. DiBerto, J. Lyu, B. E. Krumm, D. Wacker, M. J. Robertson, A. B. Seven, D. E. Nichols, B. K. Shoichet, G. Skiniotis and B. L. Roth, *Cell*, 2020, **182**, 1574–1588.

44 S. S. Chaudhaery, K. K. Roy and A. K. Saxena, *J. Chem. Inf. Model.*, 2009, **49**, 1590–1601.

45 S. de Jong, *Chemom. Intell. Lab. Syst.*, 1993, **18**, 251–263.

46 S. Wold, M. Sjöström and L. Eriksson, *Chemom. Intell. Lab. Syst.*, 2001, **58**, 109–130.

47 A. Golbraikh and A. Tropsha, *J. Mol. Graphics Modell.*, 2002, **20**, 269–276.

48 E. Pottie, A. Cannaert and C. P. Stove, *Arch. Toxicol.*, 2020, **94**, 3449–3460.

49 E. Maréchal, in *Chemogenomics and Chemical Genetics: A User's Introduction for Biologists, Chemists and Informaticians*, ed. E. Marechal, S. Roy and L. Lafanechère, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 55–65, DOI: 10.1007/978-3-642-19615-7_5.

50 D. Gentile, G. Floresta, V. Patamia, R. Chiaramonte, G. L. Mauro, A. Rescifina and M. Vecchio, *Int. J. Mol. Sci.*, 2020, **21**, 9470.

51 G. E. Schiltz, A. Jain, H. Han and S. A. Abdulkadir, *US Pat.*, US20200392116, U.S. Pat. Appl. Publ., 2020.

52 P. Schmitzer, J. Epp, R. Gast, W. Lo and J. Nelson, *Bioactive Carboxylic Compound Classes: Pharmaceuticals and Agrochemicals CHAPTER 20 Herbicidal Carboxylic Acids as Synthetic Auxins*, 2016, pp. 281-292, DOI: 10.1002/9783527693931.

53 N. R. Kapadia and W. Harding, *Med. Chem.*, 2016, **6**, 241–249.

54 J. S. Marchant, W. W. Harding and J. D. Chan, *Int. J. Parasitol.*, 2018, **8**, 550–558.