European Journal of Pharmaceutical Sciences 48 (2013) 532–541

Contents lists available at SciVerse ScienceDirect

European Journal of Pharmaceutical Sciences

journal homepage: www.elsevier.com/locate/ejps

SMILES-based QSAR model for arylpiperazines as high-affinity 5-HT_{1A} receptor ligands using CORAL

Aleksandar M. Veselinović [a,*], Jovana B. Milosavljević [a], Andrey A. Toropov [b], Goran M. Nikolić [a]

[a] University of Niš, Faculty of Medicine, Department of Chemistry, Niš, Serbia
[b] Istituto di Ricerche Farmacologiche Mario Negri, Milano, Italy

ARTICLE INFO

Article history:
Received 13 October 2012
Received in revised form 6 December 2012
Accepted 22 December 2012
Available online 31 December 2012

Keywords:
5-HT_{1A} receptor
QSAR
Arylpiperazines
SMILES
CORAL software

ABSTRACT

A predictive quantitative structure – activity relationships model of arylpiperazines as high-affinity 5-HT_{1A} receptor ligands was developed using CORAL software (http://www.insilico.eu/CORAL). Simplified molecular input-line entry system (SMILES) was used as representation of the molecular structure of the arylpiperazines. The balance of correlations was used in the Monte Carlo optimization aimed to build up optimal descriptors for one-variable models. The robustness of this model has been tested in four random splits into the sub-training, calibration, and test set. The obtained results reveal good predictive potential of the applied approach: correlation coefficients (r^2) for the test sets of the four random splits are 0.9459, 0.9249, 0.9473 and 0.9362.

1. Introduction

The most effective class of antidepressants in current clinical use are selective serotonin reuptake inhibitors (SSRIs). However, they present the serious drawback of a delay of two to six weeks in the onset of therapeutic effect, what can be attributed to the need of SSRI to overcome the inhibitory influence of 5-HT_{1A} receptor, which reduces neuronal firing rate and neurotransmitter release (Celada et al., 2004). 5-HT_{1A} receptor antagonists have been found to accelerate the onset of therapeutic effect of SSRIs by blocking the inhibition of neuronal firing rate mediated by 5-HT_{1A} somatodendritic autoreceptors (Albert and Lemonde, 2004). Arylpiperazine derivatives represent one of the most important chemical classes of 5-HT_{1A} receptor ligands (Lopez-Rodriguez et al., 2002). Arylpiperazine compounds with a dual mode of action, including serotonin reuptake inhibition and 5-HT_{1A} receptor affinity, have also been investigated as new antidepressants (Martinez-Esparza et al., 2001a,b; Orus et al., 2002).

The importance of quantitative structure–activity relationship (QSAR) methods in modern drug design is well established since they can make the early prediction of activity-related characteristics of drug candidates and eliminating molecules with undesired properties (Hansch et al., 1996). Arylpiperazine derivatives as 5-HT_{1A} receptor antagonists have been the aim of many QSAR studies (Dessalew, 2008; Habibi-Yangjeh, 2009; Weber and da Silva, 2008; Weber et al., 2008, 2010).

The main hypothesis involved in any QSAR is that the variation of the behavior of chemical compounds, as expressed by any experimentally measured biological or physicochemical property, can be correlated with numerical entities related to some aspect of the chemical structure termed molecular descriptors. In QSAR studies, thousands of molecular descriptors have been defined to encode chemical and structural features of molecules (Karelson, 2000; Todeschini and Consonni, 2000).

QSAR analysis widely uses descriptors calculated with molecular graphs (Duchowicz et al., 2008; Katritzky et al., 2001). The simplified molecular input line entry system (SMILES) is an alternative to molecular graph and it can be used for elucidation of molecular structures (Daylight Chemical Information Systems, Inc., 2008). Recent researches have demonstrated the applicability of SMILES in QSAR analysis (Benfenati et al., 2011; de Melo, 2012; Martinez et al., 2011; Toropov and Benfenati, 2008; Toropov et al., 2008, 2011, 2012; Toropov and Nesmerak, 2012; Toropova et al., 2011, 2012).

CORAL software is designed to construct one-variable QSPR/QSAR models built up by the Monte Carlo method (Benfenati et al., 2011; Toropov et al., 2011, 2012; Toropov and Nesmerak, 2012; Toropova et al., 2011, 2012). This is combined with the representation of the molecular structure by SMILES (Daylight Chemical Information Systems, Inc., 2008). Recently, the predictability of the SMILES-based models made with CORAL software has been reported (Toropov et al., 2011, 2012; Toropov and Nesmerak, 2012; Toropova et al., 2012).

* Corresponding author. Address: Faculty of Medicine, Department of Chemistry, Bulevar Dr. Zorana Djindjića 81, Niš, Serbia. Tel.: +381 18 45 70 029; fax: +381 18 42 38 770.
E-mail address: aveselinovic@medfak.ni.ac.rs (A.M. Veselinović).

0928-0987/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejps.2012.12.021

The aim of this study is to evaluate the ability of SMILES-based optimal descriptors with using of CORAL software in QSAR modeling of the arylpiperazine derivatives as 5-HT_{1A} receptor antagonists.

2. Method

2.1. Data

A dataset of 88 arylpiperazines as high affinity 5-HT_{1A} receptor ligands, to which the in vitro affinity values (as measured by inhibition constants, Ki) were collected from the literature (Martinez-Esparza et al., 2001a,b). These structures were used to generate canonical SMILES. There are a number of software systems that can generate canonical SMILES; however, different software packages generate different canonical SMILES (ACD/ChemSketch v. 11.0; MDL QSAR v. 2.2). For that reason only SMILES which are generated by a selected software package (not a mixture of SMILES generated by several software systems) should be used for QSAR analysis. Accordingly, the SMILES used in the present study were generated with the ACD/ChemSketch program (ACD/ChemSketch v. 11.0). For the QSAR analyses, the K_i values were expressed in negative logarithmic units, pK_i ($-\log K_i$). General chemical structures of used arylpiperazines are represented at Fig. 1. The chemical structures represented with SMILES notation, the activity (pK_i) experimental pK_i^{exp} data, calculated pK_i^{cal} with CORAL (split 3) and difference between pK_i^{exp} and pK_i^{cal} are listed in Table 1.

2.2. Optimal SMILES-based descriptors

SMILES is a representation of the molecular structure by sequence of symbols. Some symbols represent molecular fragments, such as atoms or bonds (e.g. 'C', 'N', '=', '#', etc.). Some of these fragments are represented by two symbols (e.g. 'Br', 'Cl', '@@', etc.) which cannot be separated. Optimal SMILES-based descriptors were calculated with CORAL software as the following (Toropova et al., 2012):

$$\begin{aligned} DCW(Treshold,N_{epoch}) = {} & CW(ATOMPAIR) + CW(BOND) \\ & + CW(NOSP) + CW(HALO) \\ & + \alpha \sum CW(S_k) + \beta \sum CW(SS_k) \\ & + \gamma \sum CW(SSS_k) \end{aligned} \quad (1)$$

where ATOMPAIR is defined in the following manner. Nine SMILES elements are considered: F, Cl, Br, N, O, S, P, double bond, triple bond. Coral software checks for the simultaneous presence of two of these SMILES elements. Similarly, the software searches for the occurrence of these bonds in the BOND index: double, triple, or stereochemical bonds, and if they are present at the same time in the molecule. The NOSP index looks specifically for the occurrence of these atoms: N, O, P, S, and if they are present together or not. Finally, the HALO index searches in the molecule the occurrence of halogens: F, Cl, Br, and if they are present simultaneously in the molecule. Table 2 represents the calculation of BOND, NOSP and HALO indices (Toropova et al., 2012). $CW(S_k)$, $CW(SS_k)$ and $CW(SSS_k)$ are correlation weights of the SMILES fragments where S_k, SS_k, and SSS_k are one-, two-, and three-components SMILES attributes, respectively; the component of SMILES is one symbol (e.g., C, c, N, n, =, #, etc.) or two symbols which cannot be separated (e.g., Cl, Br, @@, etc.). α, β, And γ are coefficients which can be 1 or 0, one can select model which is based on the attributes of one-element ($\alpha = 1$, $\beta = 0$, $\gamma = 0$), or model based on the S_k and SS_k ($\alpha = 1$, $\beta = 1$, $\gamma = 0$), etc. We build model where we take into consideration all SMILES indices and fragments (ATOMPAIR, BOND, NOSP, HALO, S_k, SS_k and SSS_k).

Two parameters should be defined for the Monte Carlo optimization: Threshold (T) and the number of epochs (N_{epoch}) (in Eq. (1)). The criterion for the classification of components of the representation of the molecular structure into two classes: rare and active is defined with the T. The correlation weight of a rare component is fixed as zero, because this component brings noise to the model, hence rare component is not involved in the building up of the model. The N_{epoch} is the number of epochs of the Monte Carlo optimization (one epoch is the cycle of modifications of all correlation weights involved in the model). The predictive potentials of the SMILES-based model are a mathematical functions the T and N_{epoch} of the Monte Carlo optimization. The search for most predictive combination of T and N_{epoch} for all splits were concluded from values 0–7 for T and 0–30 for N_{epoch} for all splits, according to previously published methodology (Benfenati et al., 2011; Toropov et al., 2011, 2012; Toropov and Nesmerak, 2012; Toropova et al., 2011, 2012). Having numerical data on these CW(ATOMPAIR), CW(BOND), CW(NOSP), CW(HALO), $CW(S_k)$, $CW(SS_k)$ and $CW(SSS_k)$ one can calculate DCW(Threshold, N_{epoch}) for compounds of sub-training, calibration, and test set. These data can used for calculation by Least Squares method model of view:

$$pK_i = C_0 + C_1 \times DCW(Threshold,N_{epoch}) \quad (2)$$

The predictability of the model calculated with Eq. (2) should be checked with the external test set. It is to be noted that statistical quality of the model for test set is a mathematical function of the Threshold and the number of epochs of the Monte Carlo optimization.

2.3. Validation of QSAR model

The main purpose of QSPR modeling is developing a robust model capable of predicting the property of new molecules in objective, reliable and precise manner (Roy, 2007). Roy et al.,

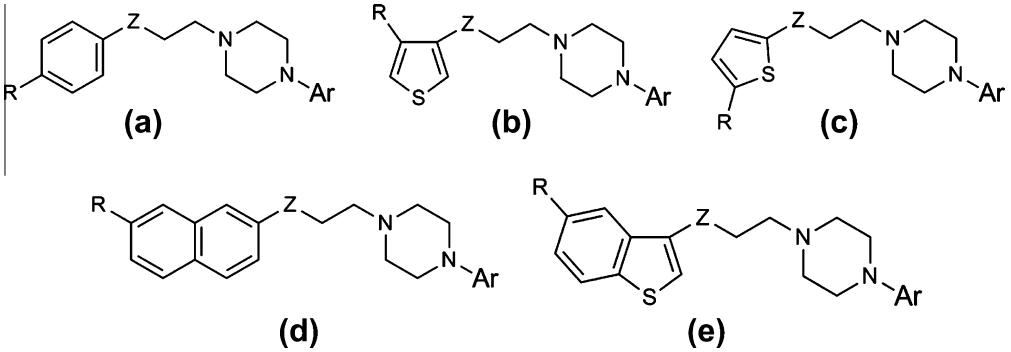

Fig. 1. General chemical structures of used molecules.