European Journal of Pharmaceutical Sciences 106 (2017) 94–101

Contents lists available at ScienceDirect

European Journal of Pharmaceutical Sciences

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

Development of a Sigma-2 Receptor affinity filter through a Monte Carlo based QSAR analysis

Antonio Rescifina[a], Giuseppe Floresta[a,b], Agostino Marrazzo[a], Carmela Parenti[a], Orazio Prezzavento[a], Giovanni Nastasi[c], Maria Dichiara[a], Emanuele Amata[a,*]

[a] Department of Drug Sciences, University of Catania, Viale A. Doria 6, 95125 Catania, Italy
[b] Department of Chemical Sciences, University of Catania, Viale A. Doria 6, 95125 Catania, Italy
[c] Department of Mathematics and Computer Sciences, University of Catania, Viale A. Doria 6, 95125 Catania, Italy

ARTICLE INFO

Keywords:
QSAR
Sigma receptor
Sigma-2 Receptor
Monte Carlo method
Repurposing
CORAL software
Virtual screening

ABSTRACT

For the first time in sigma-2 (σ_2) receptor field, a quantitative structure–activity relationship (QSAR) model has been built using pK_i values of the whole set of known selective σ_2 receptor ligands (548 compounds), taken from the Sigma-2 Receptor Selective Ligands Database (S2RSLDB) (http://www.researchdsf.unict.it/S2RSLDB/), through the Monte Carlo technique and employing the software CORAL. The model has been developed by using a large and structurally diverse set of compounds, allowing for a prediction of different populations of chemical compounds endpoint (σ_2 receptor pK_i). The statistical quality reached, suggested that model for pK_i determination is robust and possesses a satisfactory predictive potential. The statistical quality is high for both visible and invisible sets. The screening of the FDA approved drugs, external to our dataset, suggested that sixteen compounds might be repositioned as σ_2 receptor ligands (predicted p$K_i \geq 8$). A literature check showed that six of these compounds have already been tested for affinity at σ_2 receptor and, of these, two (Flunarizine and Terbinafine) have shown an experimental σ_2 receptor p$K_i > 7$. This suggests that this QSAR model may be used as focusing screening filter in order to prospectively find or repurpose new drugs with high affinity for the σ_2 receptor, and overall allowing for an enhanced hit rate respect to a random screening.

1. Introduction

Sigma (σ) receptors are recognized as a single receptor class implicated in a myriad of cellular functions, biological processes, and diseases. Two σ receptor subtypes are recognized and termed sigma-1 (σ_1) and sigma-2 (σ_2). Several findings indicate that the two receptor subtypes are distinguished by drug actions, pharmacological profiles, and molecular characteristics (Matsumoto et al., 2007; Quirion et al., 1992).

The σ_1 receptor has a MW of 25.3 kDa and was first cloned in 1996 from guinea pig liver (UniProtID Q60492, Gene names SIGMAR1, CHEMBL4153) and afterward from other sources (Hanner et al., 1996; Pan et al., 1998). The crystal structures of the human σ_1 receptor complexed with two ligands have recently been reported (PDB ID 5HK1 and 5HK2) (Schmidt et al., 2016). The σ_1 receptor agonists showed neuroprotective, anti-amnestic and antidepressant effects while σ_1 receptor antagonists are considered antiproliferative, antiangiogenic and to have modulatory effects on opioid analgesia (Chu and Ruoho, 2016; Marrazzo et al., 2011; Maurice, 2002; Mesangeau et al., 2011; Olivieri et al., 2016; Prezzavento et al., 2017).

The σ_2 receptor has been reported to have a MW between 18 and 21 kDa and, so far, it has not been cloned or crystallized. The σ_2 receptor ligands determine tumor cell death through apoptotic and non-apoptotic pathways, although their mechanisms of action have not been fully elucidated (Chu and Ruoho, 2016; Hellewell et al., 1994). The σ_2 receptor is a peculiar target overexpressed in several tumor cell lines and its ligands are actually under clinical evaluation as positron emission tomography radiotracer and indicated for the ligand-targeting therapy and as fluorescence imaging agents (ClinicalTrials.gov, 2009; ClinicalTrials.gov, 2014; ClinicalTrials.gov, 2016; Mach et al., 2013; Schinina et al., 2015; Srinivasarao et al., 2015). So far, few selective ligands have been found for the σ_2 receptor and in some cases, their finding occurred through an accidental discovery (Mach et al., 2013; Ronsisvalle et al., 2016).

In this scenario, and due to the lack of structural information about σ_2 receptor and its growing implication in cancer diagnosis and treatment, we recently proposed the Sigma-2 Receptor Selective Ligands Database (S2RSLDB, http://www.researchdsf.unict.it/

* Corresponding author.
E-mail address: eamata@unict.it (E. Amata).

http://dx.doi.org/10.1016/j.ejps.2017.05.061
Received 14 March 2017; Received in revised form 26 May 2017; Accepted 27 May 2017
Available online 29 May 2017

S2RSLDB/), a comprehensive manually curated, internet-accessible database of the sigma σ_2 receptor selective ligands. The database contains all the ligands that selectively bind the σ_2 receptor (i.e. $K_i\sigma_1/K_i\sigma_2 > 1$) (Nastasi et al., 2017). The aim of the present work is to build a quantitative structure – activity relationships (QSAR) analysis for binding affinity of the σ_2 receptor. QSAR models as well as other methods, are regression, classification or statistical methods used in the chemical and biological sciences helping in predict variables or in understanding patterns (Amata et al., 2016; Diaz et al., 2014; Romero-Parra et al., 2017; Toropova and Toropov, 2014; Veselinovic et al., 2013). The measure of affinity potential of different σ receptor ligands is their binding affinity. There have been a number of attempts to build up QSAR model for the determination of endpoint (binding affinity) for σ receptors. However, in previous attempts, QSAR models have been built using a limited number of compounds with similar chemical structure and thus useful for the determination of endpoints of a limited class of compounds (Abate et al., 2009; Laurini et al., 2010). In this work, we present a QSAR analysis of a whole set of σ_2 selective ligands over σ_1 receptor. The QSAR models for the determination of endpoints have been developed using the CORAL software which is a tool for creating models for an arbitrary endpoint using the Monte Carlo technique (Coral, 2016; Nesmerak et al., 2013; Toropov et al., 2011; Toropova et al., 2015a; Toropova et al., 2014).

2. Methods

2.1. Dataset generation and mining

We employed the whole set of ligands able to bind to the σ_2 receptor (650 compounds), retrieving only those compounds having a standard constant expressed as K_i and with selectivity over σ_1 receptor ($K_i\sigma_1/K_i\sigma_2$ ratio > 1). The simplified molecular input line entry system (SMILES) strings and the K_i values occurring for the CORAL input were directly retrieved from S2RSLDB (Nastasi et al., 2017). This set consisted of 554 compounds. From this set, 4 compounds were removed since they were quaternary ammonium salts. The dataset was thus reduced to 550 compounds. The vast majority of the experimental K_i values for these set of compounds (516 compounds) have been measured using the gold standard radioligand for determination of σ_2 receptor affinity [^3H]DTG in the presence of a selective σ_1 receptor ligand. While 15 compounds were evaluated using the σ_2 receptor selective radioligand [^3H]RHM-1, that is considered to give displacement constant in the same order of magnitude of [^3H]DTG (Xu et al., 2005). For additional 17 compounds, the original articles reported that for displacement assay [^3H]RHM-1 or [^3H]DTG were alternately used. Finally, for 2 compounds no information over the radioligand employed was reported or the radioligand used was not [^3H]RHM-1 or [^3H]DTG. For this reason, these two compounds were removed from the set. The final dataset was thus reduced to 548 compounds. The binding affinity data, expressed as K_i (in molar concentration), of the 548 σ_2 receptor selective ligands were converted into negative decimal logarithm pK_i (p$K_i = -\log K_i$). Collected pK_i values fall into a range from 5.10–11.21 for the σ_2 receptor. To achieve a paramount and consistent depiction of the system, the available data were three times, randomly split into four sets, training (≈ 38%), calibration (≈ 38%), test (≈ 12%) and validation (≈ 12%), and examined. The training set plays the role of builder of a model; the calibration set plays the role of preliminary critic of the model; the test is a visible estimator of the model, while the validation set is the invisible final estimator of the model.

2.2. Descriptors

To represent a molecular structure in a computer is the first step in developing a QSAR analysis. In this work, molecular structures have been represented with SMILES and molecular graph. Models produced with CORAL consist in a linear relationship between a predicted endpoint Y (pK_i) and a descriptor of correlation weights (DCW), namely, it has the following form:

$$Y = C_0 + C_1 \bullet DCW, \tag{1}$$

where C_0 and C_1 are the two regression coefficients evaluated by using the least squares method. The purpose is to build an optimal DCW model fitted on the dataset. CORAL software provides three kinds of DCW: graph-based, SMILES-based, and hybrid. Hybrid representation using SMILES together with the molecular graph may give better models with higher statistical quality respect those models with a unique representation of the molecular structure (Catelani et al., 2009; Toropova et al., 2013; Toropova et al., 2015b). In this work, we employed the hybrid DCW representation since single graph-based or SMILES-based DCW representation resulted in models with lower statistical quality (see Table S1 for the statistical characteristics and Table S2 for the regressions of QSAR models of pKi, for the σ_2 receptor). The hybrid optimal DCW used to build up models for the predicted Ki were calculated according to Eq. (1):

$$^{Hybrid}DCW = {}^{Graph}DCW + {}^{SMILES}DCW. \tag{2}$$

The molecular graphs are mathematical representation where each molecule is depicted by a number of nodes (atoms) that are encoded by simple vertices bearing certain properties and by bonds that are encoded by simple connections or edges between these points. We decided to use hydrogen-suppressed graph (HSG) also termed hydrogen-depleted chemical graphs where hydrogen atoms are represented as a property of non-hydrogen or heavy atoms. Indeed, for the enumeration of isomers, hydrogen-suppressed graph are good as hydrogen-filled graph but since hydrogen-suppressed graphs have a smaller number of vertices and a significantly simpler structure these have been in our case preferred (Ivanciuc, 2013). Moreover, in CORAL the analysis of cycles is available only for HSG (Coral, 2016).

In order to improve the variability of attributes, we have used several features calculated by CORAL within the graph-based descriptors. Thus, the graph-based optimal descriptors are calculated as reported below:

$$^{Graph}DCW = \sum_{i=0}^{2}\sum_{k} CW(^{i}EC_k) + \sum_{i=3}^{7}\sum_{k} CW(Ci_k) + \sum_{i=2}^{3}\sum_{k} CW(PTi_k) + \sum_{i=2}^{3}\sum_{k} CW(VSi_k) + \sum_{k} CW(NNC_k), \tag{3}$$

where A_k is a chemical element, $^{i}EC_k$, $i = 0, 1, 2$ is the hierarchy of the Morgan's extended connectivity, Ci_k, $i = 3, 4, \cdots, 7$ is the attribute related to the cycle containing a number of i atoms, PTi_k is the attribute taking into account paths of length i starting from the k-th vertex of the graph, VSi_k counts the valence shell of i-th range and NNC_k considers the nearest neighbors codes (Toropova et al., 2015b).

The SMILES-based optimal descriptors are calculated as the following:

$$^{SMILES}DW \sum_{k=1}^{N} CW(S_k) + \sum_{k=1}^{N-1} CW(SS_k) + \sum_{k=1}^{N-2} CW(SSS_k) + \sum_{k} CW(NOSP_k) + \sum_{k} CW(HALO_k) + \sum_{k} CW(BOND_k) + \sum_{k} CW(HARD_k), \tag{4}$$

where S_k is a symbol appearing into SMILES representation and N is their total number. Therefore SS_k and SSS_k are combinations of two and three symbols. $NOSP_k$, $HALO_k$, $BOND_k$ and $HARD_k$ are global SMILES attributes extracted from SMILES (each of them is codified in a certain number of attributes). Therefore, $NOSP_k$ indicates the relative presence of one or more of four chemical elements (nitrogen, oxygen, sulphur and phosphorus), $HALO_k$ considers the presence of fluorine, chlorine, bromine, and iodine, $BOND_k$ if there are or not one or more of three