European Journal of Medicinal Chemistry 44 (2009) 5045–5054

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Application of desirability-based multi(bi)-objective optimization in the design of selective arylpiperazine derivates for the 5-HT_{1A} serotonin receptor

A. Machado [a,*], E. Tejera [a], M. Cruz-Monteagudo [b,c], I. Rebelo [a]

[a] IBMC, Department of Biochemistry, Faculty of Pharmacy, University of Porto, 4150-047 Porto, Portugal
[b] Applied Chemistry Research Center, Faculty of Chemistry and Pharmacy, Central University of "Las Villas", Santa Clara 54830, Cuba
[c] Department of Organic Chemistry, Faculty of Pharmacy, University of Porto, 4150-047 Porto, Portugal

ARTICLE INFO

Article history:
Received 29 January 2009
Received in revised form
6 July 2009
Accepted 6 September 2009
Available online 18 September 2009

Keywords:
Multiobjective optimization
Desirability
Inverse QSAR
Arylpiperazine
Drug design
Serotonin receptor

ABSTRACT

The multiobjective optimization technique based on the desirability estimation of several interrelated responses (MOOP-DESIRE) has been recently applied to quantitative structure–activity relationship (QSAR) studies. However, the advantage of applying this new methodology to the study of selectivity and affinity to competitive targets has been little explored. We used the MOOP-DESIRE methodology and a variation of this, to study the arylpiperazine derivates that could interact with 5-HT_{1A} and 5-HT_{2A}, serotonin receptor subtypes with the objective of designing more selective molecules for the 5-HT_{1A} receptor. We did show that the model results are in agreement with the available pharmacophore descriptions, guaranteeing an appropriate structural correlation and proving the methodology, as a useful tool for the important problem of selective drug design.

1. Introduction

The potential multireceptor activity of arylpiperazine derivates has become a source of new agents with different therapeutic properties. This wide spectrum of possible interactions is associated with a variety of pharmacological and toxicological effects [1–3]. One important branch is related to the capacity of these molecules to interact with serotonin receptors (5-hydroxytryptamine, 5-HT) 1A (5-HT_{1A}) and 2A (5-HT_{2A}), being the principal subtypes [3–5].

Serotonin 5-HT_{1A} subtype is frequently a therapeutic target for depression and anxiety [6–9] while 5-HT_{2A} subtype causes various adverse effects, resulting from several psychoactive and hallucinogenic drugs of abuse, such as ecstasy and LSD, by excessive stimulus of serotonin secretion and inhibition of its enzymatic degradation in the reward system [2,5,10–13].

The independent construction of models to increase or reduce the selectivity to one or other of the targets could be mutually affected due to the receptor and pharmacophore similarity. For this reason a simultaneous approach could be more appropriate for selectivity improvement using the multiobjectives optimization (MOOP) methods as a way to design more selective molecules.

Multiobjective optimization method (MOOP) allows the identification of the best solution by optimizing several dependent properties simultaneously. The advantage of MOOP is creating a local optimal solution through a group of different objectives. So, this method allows the use of different goals in the same study. Each objective may contribute in a positive or a negative way to the final solution [14–18]. This methodology has been applied in several cases: for example, the focus on molecular alignment and pharmacophore identification, through the genetic algorithm method for multiple molecular alignment [19] or the genetic algorithm with linear application for the hypermolecular alignment of datasets [20].

MOOP techniques have only recently been applied to the building of QSAR models, in particular, for the interpretation of physicochemical properties responsible for selectivity [14,21,22]. A novel multiobjective optimization technique based on the desirability estimation of several interrelated responses (MOOP-DESIRE) was recently proposed [23]. This methodology simplified multiple objectives to a single one using independent desirability criteria

* Corresponding author.
E-mail addresses: machadobq@gmail.com (A. Machado), edutp00@yahoo.com (E. Tejera), gmailkelcm@yahoo.es (M. Cruz-Monteagudo), irebelo@ff.up.pt (I. Rebelo).

0223-5234/$ – see front matter © 2009 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2009.09.008

and was used for the optimization of direct pharmaceutical profiles through multiple QSAR model combinations.

In the present study we applied, for the first time, the multi(bi)-objective desirability function of the MOOP-DESIRE method to a selectivity problem and simultaneously we present an alternative improvement of the preceding methodology [23]. Our aims are the design of new arylpiperazine candidates with high selectivity for the 5-HT_{1A} receptor and simultaneously low selectivity for the 5-HT_{2A} receptor, as well as the comparison of two methodologies presented here and their correspondence with the available pharmacophore models.

2. Materials and methods

2.1. Data set

Prediction models were developed from a group of forty-three arylpiperazine derivates characterized by Grundt et al. [24], Leopoldo et al. [22] and Bojarski et al. [25], which were used in previous QSAR studies (Tables 1–3).

2.2. Computational methods

Structures of all arylpiperazine derivates were drawn in the ChemDraw [26] software package, pre-optimized by molecular mechanics using the MM2 force field. The final structure was obtained by posterior optimization with AM1 semi empirical Hamiltonian, implemented in the MOPAC 6.0 program [27].

Molecular descriptors ($n = 1666$) were calculated for each molecular structure using the eDragon software [28–31] while the appropriate descriptor selection was made by means of the genetic algorithm (GA) program designed in Matlab v7.0 for this purpose. The GA procedure was composed of three principal blocks [32,33]: initial population generation, crossover and mutation procedures (see Fig. 1). In this methodology two parameters needed to be considered: the crossover and mutation probabilities. In our work the initial population was randomly generated and composed of 600 chromosomes (representing 600 initial not optimized models) fixing the maximal number of variables at five. However, we had already explored the number of variables from four until eight, with the objective to analyse two important aspects: the predictability and interpretability of the model. The problem of the selected number of variables will be discussed later (see Results and discussion).

The crossover procedure is an iterative process where any two chromosomes (parents) are combined to produce new child chromosomes conserving the common genes (variables) present in the parents. In our application the crossover was performed uniformly and with a probability of 50%, which means that any two parents are selected and the children information is composed of 50% of the information of each parent. The children obtained were evaluated with the objective function of leave-one-out internal cross-validation (Q^2_{LOO}). If the (Q^2_{LOO}) obtained was higher that any chromosome of the initial population, then the chromosomes with less (Q^2_{LOO}) were replaced by the generated children. This procedure was fixed at 1000 effective steps (Fig. 1).

The mutation procedure was carried out after the crossover and basically consisted of changing the genes (variables) of the chromosomes with a fixed probability. In our case the probability was 50%, therefore the mutation was almost random. The goal of this procedure was to expand the search capability in the variable space. The entire population was mutated and the (Q^2_{LOO}) function was evaluated for each mutated chromosome. Similar to the crossover procedure, the chromosome with less (Q^2_{LOO}) value in the population was replaced by the accepted mutated chromosome (which (Q^2_{LOO}) was greater that any other in the population). So, the initial population was updated after the crossover and mutation procedures with the most promising chromosomes. The crossover/mutation steps (GA generations) were repeated several times until they had reached a fixed stop criteria. In our model, the stop criteria was fixed at 100 iterations (Fig. 1); however, after approximately 10–15 GA generation (which represent 10,000–15,000 effective crossovers) a convergence was generally reached.

It is known that the (Q^2_{LOO}) criterion on its own tends to overestimate the predictability of the model. Several approaches were available [34,35] for model validation and, so, we calculated (beside the Q^2_{LOO}) the determination coefficient (R^2 and $Q^2_{-adjusted}$), the leave-multiple-out internal cross-validation (Q^2_{LMO}) and the bootstrap internal cross-validation (Q^2_{BOO}). The Q^2_{LMO} was calculated considering only a group of molecules (around 50%) for the model construction and predicted the remaining ones. On the other hand, the Q^2_{BOO} is a more accepted internal cross-validation [36] and is calculated taking samples randomly and repeatedly, of size N (where N is the molecules number) for the construction of the model and predicting the remaining molecules. This procedure was repeated several times (we used 5000 repetitions) and the Q^2_{BOO} represented the mean predictability coefficient. For the GA calculation and model validation we used the Matlab v7.0 and the multiple linear regression module of the Statistica Package [37,38].

A further important point that we had to take into consideration was the applicability domain (AD; the space of molecules that could affect the stability of the model). The most common way to do this is using the leverage [39] of each compound as well as its respective standardized residues of the cross-validation procedure. The leverage values were calculated as:

$$H = X(X'X)^{-1}X' \tag{1}$$

where X is the model matrix and H is the leverage matrix. The diagonal elements h_{ii} of H correspond to the interest parameter that must be less than a critical leverage (h_c) defined as:

$$h_c = 3\frac{p}{n} \tag{2}$$

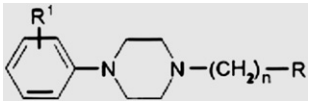
where p and n are the number of model coefficients and the molecules number, respectively. All the molecules i with $H_{ii} > h_c$, represent an important influence which points to the regression model and should be analyzed carefully as well as those with standardized residues greater than ± 2.

2.3. MOOP based on the desirability estimation

From the previously constructed model, for each of the properties (pK_{1A} and pK_{2A}) and according to the previous methodology [23], each predicted property response ($\hat{P}_i$), corresponding in our case to $\hat{p}K_{1A}$ and $\hat{p}K_{2A}$, was translated to a desirability value, d_i, using the desirability functions proposed by Derringer and Suich [18,23]. Therefore, pK_i values related to 5HT_{2A} and 5HT_{1A} receptors were minimized and maximized using the following eqs. (3) and (4), respectively, defined as:

$$d_i = \begin{cases} 1 \text{ if } \hat{P}_i \le T_i = L \\ \left[\dfrac{\hat{P}_i - U}{T_i - U}\right]^s \text{ if } U < \hat{P}_i \le T_i \\ 0 \text{ if } \hat{P}_i \ge U \end{cases} \tag{3}$$

Table 1
5-HT_{1A} and 5-HT_{2A} receptor binding affinities.



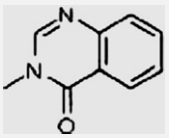
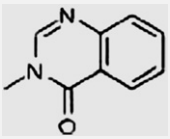
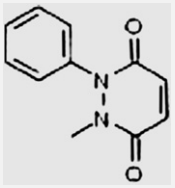
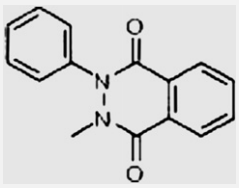
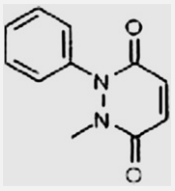
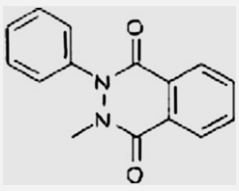
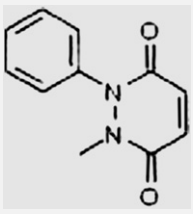
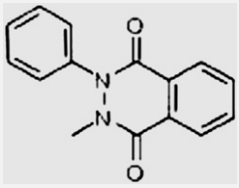
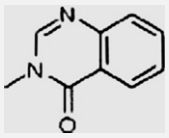
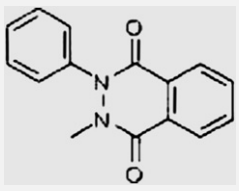
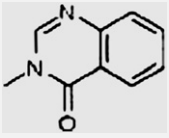
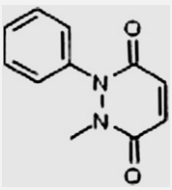
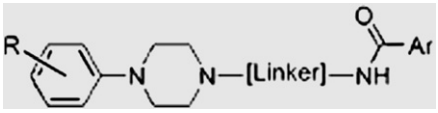
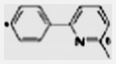
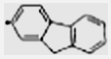
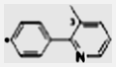
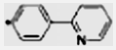
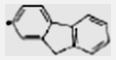
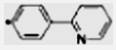
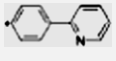
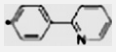
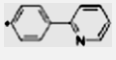
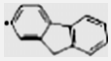
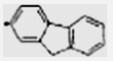
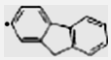
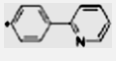
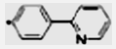
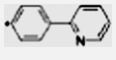
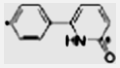
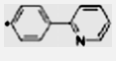
Mol.	N	$-R^1$	$-R$	pK_{1A}	pK_{2A}	Mol.	n	$-R1$	$-R$	pK_{1A}	pK_{2A}
S1	3	*m*-Cl		6.63	7.80	S7	4	*o*-OCH_3		7.44	6.25
S2	4	*m*-Cl		7.30	5.74	S8	3	*m*-Cl		6.38	6.18
S3	3	*o*-OCH_3		7.27	5.67	S9	4	*m*-Cl		6.40	6.24
S4	4	*o*-OCH_3		7.96	5.84	S10	3	*o*-OCH_3		7.52	6.52
S5	4	*m*-Cl		7.30	7.17	S11	4	*o*-OCH_3		7.37	6.43
S6	3	*o*-OCH_3		7.00	6.34	S12	3	*m*-Cl		6.69	6.57

Table 2
5-HT_{1A} and 5-HT_{2A} receptor binding affinities.



Mol.	Ar	[Linker]	–R	pK_{1A}	pK_{2A}	Mol.	Ar	[Linker]	–R	pK_{1A}	pK_{2A}
S13			2,3-diCl	6.51	7.03	S22			2,3-diCl	5.69	6.17
S14			2,3-diCl	7.04	7.33	S23			2,3-diCl	7.23	7.33
S15			2-OCH_3	7.66	7.12	S24			2,3-diCl	6.01	6.05
S16			2-OCH_3	7.14	7.18	S25			2,3-diCl	7.63	7.64
S17			2,3-diCl	7.46	7.37	S26			2,3-diCl	6.03	6.48
S18			2,3-diCl	5.74	6.26	S27			2-OCH_3	7.47	6.81
S19			2-OCH_3	7.44	6.16	S28			2-OCH_3	8.17	6.64
S20			2,3-diCl	6.93	7.05	S29			2,3-diCl	7.53	7.81
S21			2,3-diCl	7.82	7.81	–	–	–	–	–	–

$$d_i = \begin{cases} 0 \text{ if } \hat{P}_i \geq L \\ \left[\dfrac{\hat{P}_i - L}{T_i - L}\right]^s \text{ if } L < \hat{P}_i \leq T_i \\ 1 \text{ if } \hat{P}_i \geq T_i = U \end{cases} \quad (4)$$

where $\hat{P}_i$ is the predicted response ($\hat{p}K_{1A}$ and $\hat{p}K_{2A}$), s corresponds to an exponent that determine how important is to hit the target value T_i and U, L and T are the upper, lower and target values that must be chosen for each response, respectively. For $s = 1$, the desirability function increases linearly towards T_i. So, it was easy to note that eqs. (3) and (4) performed a linear minimization and maximization of the responses respectively and $0 \leq d_i \leq 1$. In our case, [L; U] values were [5; 8.5] for the 5-HT_{1A} receptor and [4.5; 7.5] for the 5-HT_{2A} receptor. The idea was previously maximized the 5-HT_{1A} receptor interaction instead of 5-HT_{2A} receptor, so, we consider that a molecule with a pK_{2A} value of 4.5 as tolerant and pK_{1A} value of 8.5 as a good selection. On the other hand, fixing a 0 or 1 desirability value outside of the minimum or maximum value of the set would affect the linearity of the transformation and, therefore, the predictability of the models.

Finally, all individual desirability was combined into a single total desirability D according to eq.(5) and then was maximized as a function of the descriptor's space (X_1,X_2,…,X_n) of each independent QSAR model.

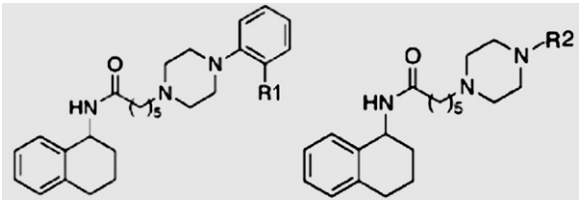
$$D = (d_1 \times d_2 \times \ldots \times d_k)^{1/k} \quad (5)$$

where k are the number of objectives, in our case two objectives more exactly a maximize pK_{1A} value and a minimize pK_{2A} value for each molecule.

An important topic in this methodology was the descriptors' space (X_1,X_2,…,X_n) used to build QSAR equations. It is evident that if the equations share the same descriptors the final interpretation is easy; however, if all the models did not have the same descriptors, even more if the number of properties is high, the final interpretation is more difficult.

In our particular case, the molecular aspects (and molecular descriptors) that tend to increase the selectivity for the 5-HT_{1A} receptor ($X_1^1, X_2^1, \ldots, X_n^1$) could be different from the 5-HT_{2A} receptor

Table 3
5-HT_{1A} and 5-HT_{2A} receptor binding affinities.



Mol.	–R1	–R2	pK_{1A}	pK_{2A}	Mol.	–R1	–R2	pK_{1A}	pK_{2A}
S30	–S–CH_3	–	7.28	6.49	S40	–S(CH_3)O_2	–	5.51	5.40
S31	–O–CH_3	–	8.07	5.90	S41	–CH_2–CH_3	–	7.10	5.47
S32	–CH_3	–	6.55	6.58	S42	–(CH_2)$_2$–CH_3	–	6.78	5.50
S33	–C(CH_3)O	–	8.42	4.91	S43	–CH(CH_3)$_2$	–	6.78	5.32
S34	–OH	–	7.62	5.47	–	–	–	–	–
S35	–H	–	6.89	7.11	–	–	–	–	–
S36	–NO_2	–	6.74	6.55	–	–	–	–	–
S37	–Cl	–	7.02	6.52	–	–	–	–	–
S38	–CN	–	7.78	6.15	–	–	–	–	–
S39	–C(CH_2)O	–	6.31	5.40	–	–	–	–	–

$(X_1^2, X_2^2, \ldots, X_m^2)$ and if we considered the simultaneous case of maximal 5-HT_{1A} and minimal 5-HT_{2A} interactions, even other molecular aspects could be pointed out (different descriptors and/or different QSAR equations) instead of the lineal combination $X_1^1, X_2^1, \ldots, X_n^1$ and $X_1^2, X_2^2, \ldots, X_m^2$. The essential advantage of this methodology was transforming the inconvenience of multiple variable optimizations to a single one, by applying a desirability criterion; as we can see, in the described methodology was necessary as QSAR equations as properties involved, because the eqs. (3) and (4) were obtained using the predicted response. However, we will use and compare an alternative approach using the same equations (eqs. (3) and (4)) with the experimental activities, therefore, the overall desirability resulted from experimental values using eq. (5) and it was considered as the target response (D_{exp}). Using this approach the models were reduced to a single QSAR equation independent of the number of initials properties; consequently, the errors involved in the complete methodology were minimized and the descriptor space to future interpretation was reduced. The two different approaches of methodologies are illustrated in Fig. 2. As we can see, in the alternative methodology (see Fig. 2; Methodology II), the GA procedure was used to select a set of variables that should contain the information relative to an increase and reduce affinity to 5-HT_{1A} and 5-HT_{2A} receptors, respectively.

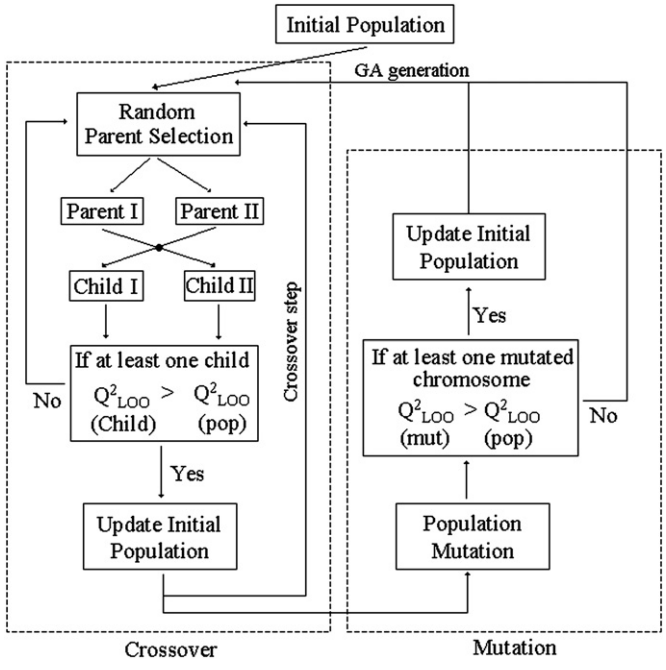


Fig. 1. Diagram of the implemented GA procedure. In the present procedure, a number of 600 chromosomes were generated for the initial population. Then, the crossover step was fixed to 1000 while the GA generation was fixed to 100 iterations.

3. Results and discussion

According to the previously described methodology, the selected models for 5-HT_{1A}, 5-HT_{2A} and D_{exp} are presented in Table 4. As previously referred, we explored four until eight variables in the models constructions. However, the results in the Table 4 show only five variables. In the construction of the Model III with four variables, the value of Q^2_{LOO} was lower than 0.80 and a maximal Q^2_{LOO} value of 0.89 was noted when we fixed the number of variables to seven. In the construction of Model I and II considering four variables, the Q^2_{LOO} values were 0.73 and 0.65, respectively, while maximal Q^2_{LOO} values of 0.84 and 0.82 were reached with 8 variables, respectively. We considered that even when the predictability with 8 variables appears to be higher that 5 variables consideration, these did not justify the use of 8 variables due to the difficulty increment degree in the models' interpretation and the future molecular design. Nevertheless, the desirabilities of the designed molecules were recalculated one more time with eight variables in all the models and the selected molecules with higher desirability still remained as the same promising ones saw in the five variables' models.

The leverage variation of the training data (Fig. 3) is always less than the warning leverage and the cross-validation standardized residual values are within the ±2 range indicating acceptable models, but to obtain that kind of results we had to exclude the following molecules: S5; S17; S21; S26; S29; S34 and S36.

The desirability predicted values (cross-validation) calculated from $\hat{p}K_{1A}$ and $\hat{p}K_{2A}$ (using eq. (3)–(5)) as well as through Model III (Table 4) have the same Q^2 values (Fig. 3) that confirm the possibility of the use of Model III for desirability prediction, simplifying the methodology and minimizing the source of errors by the reduction of the number of models. However, analysing the partial models (Models I and II in Table 4), we could note that the model obtained using the D_{exp} values demonstrate superior predictability, even when using a bootstrap procedure. The possible implication of these differences will be discussed later.

As already said, the desirability was maximized directly over independent variables presented in Table 4 (supplemented material). Afterwards, predicted values for each receptor were used to fit a model containing all the independent variables from which a desirability profile was built (Fig. 4). Alternatively, we actually performed an optimization of D_{exp} (Fig. 5).

The most important contribution to overall desirability comes from descriptors MATS2v, SHP2, nBnz, X3A and AMW and, to a lesser degree E1s, nBM, RDF085v and EEig06d descriptors. The marked values in red (Figs. 4 and 5) represent the optimal values to reach a maximal overall desirability. The appropriate transformation of these descriptor values to the chemical structure (inverse QSAR) is a very difficult task (if possible, at all) as well as its correct interpretation. However, the possibility to obtain the optimal descriptors values is one of the main contributions of the

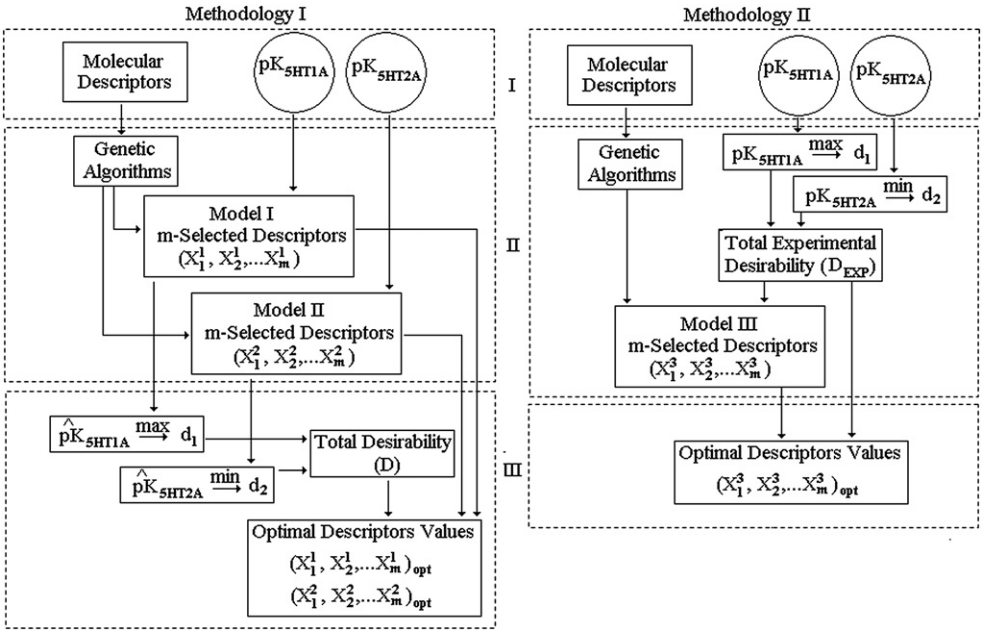


Fig. 2. Diagram of the methodologies used in the present work. The blocks I, II and III represent: the initial data, the model construction and the optimization procedures, respectively. In the first methodology a QSAR equation was necessary for each receptor and the desirabilities (see Table 4; 5-HT_{1A} Mod. I and Mod. II) were evaluated in the predictive activities. Instead, in the second methodology, the desirability was calculated using the experimental activities and before the GA procedure. For those reasons only one QSAR equation (see Table 4; D_{exp} Mod. III) is needed in the second methodology.

Table 4
Predictive models and statistical parameters obtained by GA-MLR.

	R	R^2	F	P	Q^2_{LOO}	Q^2_{LMO}	Q^2_{BOO}
5-HT_{1A} Mod. I	$pK_{1A} = 21.3232(1.49) - 11.187(1.23)$ MATS2v $- 52.301(5.05)$ SHP2 $- 0.080(0.02)$ RDF045m $- 0.802$ (0.15) Mor15m $+ 4.491(0.86)$ E1s						
	0.926	0.857	36.063	<0.01	0.807	0.866	0.795
5-HT_{2A}	$pK_{2A} = -26.163(4.14) + 0.060(0.03)$ nBM $+ 1.042(0.30)$ nBnz $+ 157.156(20.15)$ X3A $+ 0.116(0.02)$ RDF080m $- 0.103(0.04)$ RDF085v						
Mod. II	0.898	0.806	24.921	<0.01	0.721	0.652	0.710
D_{exp}	$D_{exp} = 0.681(0.23) - 0.326(0.02)$ AMW $- 2.625(0.24)$ MATS2v $+ 0.453(0.09)$ EEig02d $+ 0.298(0.04)$ EEig06d $+ 0.123(0.04)$ Mor18m						
Mod. III	0.948	0.899	53.592	<0.01	0.852	0.866	0.842

The number in (…) is the standard error of the coefficient; Q^2_{LMO} is the mean value from a group of 15 molecules leave-multiple-out; finally, Q^2_{BOO} is the median value of 5000 run of boostrap validation.

desirability proposed methodology, especially in the drug design research by using screening or combinatorial libraries [40].

The easiest interpretable descriptors family includes the constitutional descriptors, functional groups, atom-centred fragments and a few other descriptors. However, all of them are often weakly correlated even though several of them are needed to build a desirable model. Therefore, we explored the correlation matrix between model descriptors and "interpretable ones" (data not shown), as a way to understand our models of future design.

All the descriptors used are strongly related to flexibility (number of rotatable bonds), molecular weight, aromatic substructure and atom type involved in the aromatic substitution;

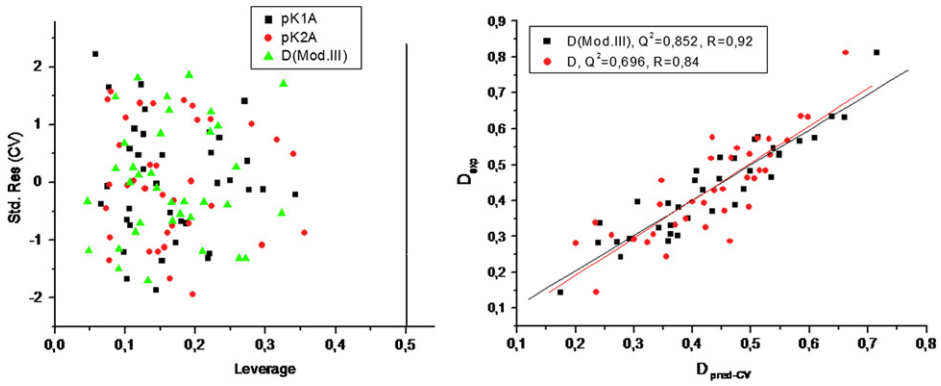


Fig. 3. Leverage plot of all the models (Table 4). The critical leverage of all the models (5-HT_{1A}, 5-HT_{2A} and D_{exp}) were 0.50 (h_c).

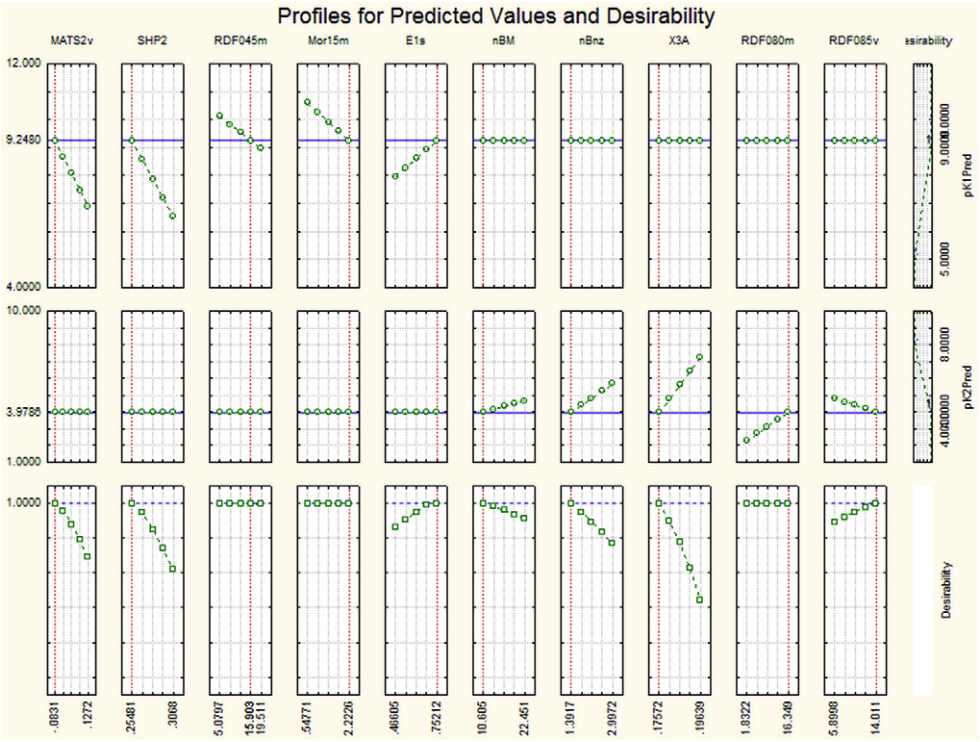


Fig. 4. Multiple response desirability function due to maximization of 5-HT_{1A} and minimization of 5-HT_{2A} affinities for the data set molecules (S1-43), along with the individual desirability functions coming from the indicated descriptors.

and all they represent the principal factors involved in 5-HT receptor selectivity and affinity. Correlation data, together with the result of the desirability optimization, revealed that the number of benzene substitutions, the absence of halogen elements (such as fluorine and chlorine) and the increment of hydrogen attached to alpha carbons and others unsaturated structures (C–C(═X)–; R-C≡X; X═C═X; X: O, N), allow a referent affinity to 5-HT_{1A} receptor. The presence of halogen elements and saturated structures, such as the number of benzene rings, permit the promotion of affinity to 5-HT_{2A} receptors. However, the increment in the polar

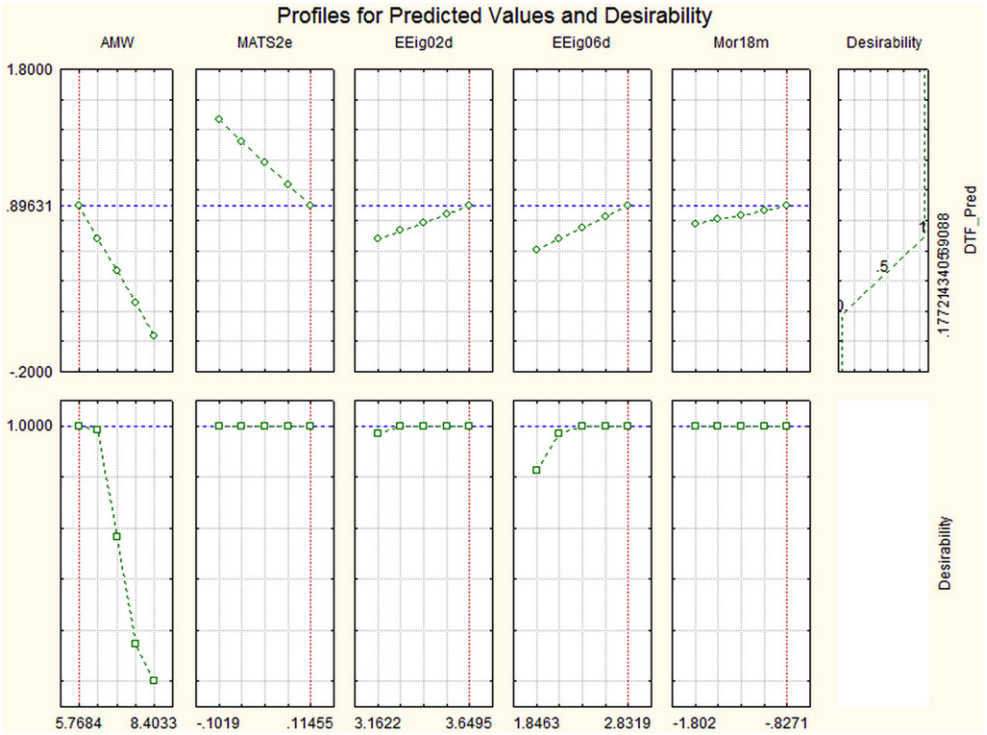
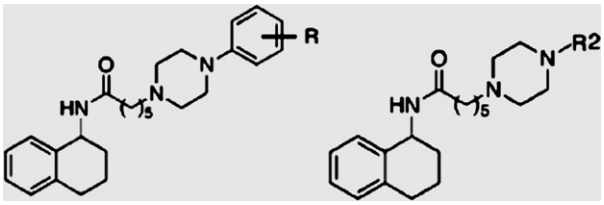


Fig. 5. Optimization results of desirability (D_{exp}) obtained by experimental p*Ki*s presented in Table 2.

Table 5
5-HT_{1A} and 5-HT_{2A} receptor binding affinities predicted for designed compounds.



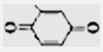
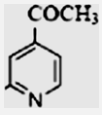
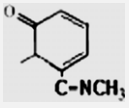
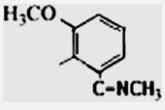
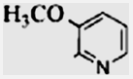
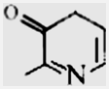
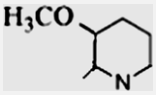
Comp.	–R	–R2	$\hat{p}K_{1A}$	$\hat{p}K_{2A}$
M1	–COOH	–	7.10	6.49
M2	–C(CH_2CH_3)=O	–	8.40	6.99
M3	–Cl; 5-Cl	–	7.40	5.72
M4	–F; 5-F	–	5.84	5.54
M5	–		7.58	5.53
M6	–OH; 5-OH	–	6.88	5.74
M7	–		8.46	5.90
M8	–		7.59	6.18
M9	–OCH_3; 4-OCH_3	–	8.09	5.65
M10	–OCH_3; 5-OCH_3	–	6.84	4.52
M11	–C(CH_3)=S	–	7.34	5.82
M12	–		6.85	5.29
M13	–		6.50	5.31
M14	–C=N–CH_3	–	7.45	5.46
M15	–		8.21	6.01
M16	–		6.21	5.35
M17	–		8.10	5.45
M18	–OCH_3; 6-OCH_3	–	7.74	6.50
M19	–OH; 6-OH	–	8.27	5.75
M20	–		8.58	5.06
M21	–		8.78	4.96

Table 5 (*continued*)

Comp.	–R	–R2	$\hat{p}K_{1A}$	$\hat{p}K_{2A}$
M22	–		7.08	4.86
M23	–		8.01	5.40
M24	–		7.18	5.49
M25	–		7.83	5.87

area associated with N and the increment of rotatable bonds in the benzene substitute are important to both receptors with a slight preference for the 5-HT_{2A} receptor. This is one of the reasons why even when the RDF085v descriptor is present in Model II (Table 4) with a negative coefficient, in the overall desirability shows a positive increment but in narrow range (Fig. 4).

Our previous interpretations are in agreement with the well-known Mellin's two-point pharmacophore models [41] and three-point Chilmonczyk's model [42]. Both of them consist of an aromatic ring and basic nitrogen (of the piperazine ring), but the three-point model also includes an oxygen atom forming a triangular configuration. Therefore, the basic nitrogen is surrounded by: an aromatic or hydrophobic group on one side and polar (or H-bond) group in the other. On the other hand, even when the basic nitrogen presence is questionable [43], the decrease of 5-HT_{2A} receptor affinity to the detriment of 5-HT_{1A} affinity requires the minimization of structure donor hydrogen (in accordance with Chidester et al. [44]), aromatic unsaturated structures and the absence of halogen elements. The negative effect of halogen elements is probably related to the spatial arrangement of the group in the aromatic ring and its influence in the final arylpiperazine conformation, as observed by López-Rodríguez et al. [9] and Gaillard et al. [7].

3.1. Molecular design

On the basis of the previous discussion, we designed a group of arylpiperazine derivates by substitution of the $R1$ or R^2 position, exploring several aspects of the molecular space (Table 5). The predictability of the QSAR models obtained was always restricted to the structural spaces contained in the initial molecular family used for the models' constructions. In the present work, the structural variation was almost limited to modifications on the aromatic groups, steric and H-bond properties. In addition, the model predictability would be low if the arylpiperazines derivates have a poor similarity in respect to the initial compounds family. So, all these aspects were taken in consideration for the molecular design of new arylpiperazines.

The source of leverage increment in the desirability calculated with $\hat{p}K_{1A}$ and $\hat{p}K_{2A}$ (Table 5 and shown in Fig. 6) is associated with the 5-HT_{2A} receptor model by the minimum induced from the selected descriptors. However, it can be noted that in the case of M4 molecule, the prediction is closer to both models, while in the M3

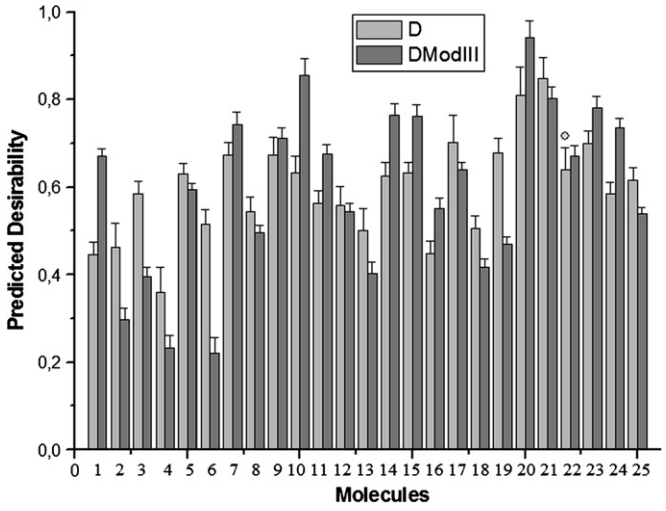


Fig. 6. Predicted desirability for the designed molecules (Table 3) using both methodologies. Columns marked represent molecules with leverage higher than critical values.

there could have been a lower estimation of pK_{2A} values increasing, incorrectly, the desirability value with respect to the other model.

The results indicate that M22 is outside of the AD and the predicted values may not be reliable. The effects of aromatic substitutions in the *ortho* position by an oxygen with a double bond in the plane of the aromatic ring or other bond with a methyl group as sterically favourable to 5-HT$_{1A}$ affinity is observed in molecules M20 and M23 or M21 and M22, respectively, in accordance with the experimental results of Gaillard and co-workers [7]. Another validation is noted comparing M4, S22, S23, S24 and S25 molecules with respect to the desirability reduction by the halogen (e.g., chlorine) presence and saturated structure fragment. In general even when some discordant desirability values are present between the two models, the general pattern is conserved and the results show that, at least in the AD study, as well as in the target molecule selections, could be more useful the combination of both methods.

In general, molecules with an electrostatic or H-bond interaction (substituent) near the *ortho* position and a little flexibility, without affecting the rotation of benzene, demonstrate high affinity and selectivity at 5-HT$_{1A}$ receptor with increased desirability in both models. In these groups the new molecules M7, M9, M10, M14, M15, M17, M19, M20 M21 and M23 are the promising ones. In the case of M20 and M21 the affinity for 5-HT$_{1A}$ is approximately 3300 fold higher and 6600 fold (Table 5), respectively.

The limitations of the models presented are conditioned to the quality of fit and predictive capabilities of the initial equations. The methodology followed to obtain the individual QSAR equations for each receptor or any other properties of interest is the decisive factor in the desirability estimation of a higher confidence. Although, only a subgroup of arylpiperazines was explored, it is possible to use the same methodology in a more general chemical compounds group to expand the structural diversity on molecular design and to understand the 5-HT$_{1A}$ and 5-HT$_{2A}$ receptors' interaction.

4. Conclusion

In this work, we applied the MOOP-DESIRE method as well as an alternative variation of the same one, with the purpose of facilitating the design of new compounds more selective for the 5-HT$_{1A}$ receptor with respect to 5-HT$_{2A}$ subtype. The validation of these methodologies is capable of reflecting the pharmacophore aspects of both receptors.

Higher values of desirability are related with molecules without H-donor groups or halogen atoms and with electronic density close to the benzene *ortho* substitution position in agreement with the principal pharmacophore aspects associated with 5-HT$_{1A}$ receptor.

Modifying the previous methodology using only one equation to predict the desirability simplifies the overall procedures conserving the same chemical meaning. However the combination of models can be helpful in the final design process. We propose two molecules with potentially selective capabilities: M20 and M21 with 80–94% desirability values and highest differences between pK_{1A} and pK_{2A} values. The designed molecules are not available in the PubChem molecular database and therefore experimental work is pursued to validate these results.

Appendix A. Supplemental material

Supplementary information for this manuscript can be downloaded at doi:10.1016/j.ejmech.2009.09.008.

References

[1] W. Kuipers, I. Van Wijngaarden, C.G. Kruse, M. Ter Horst-Van Amstel, M. Th. M. Tulp, A.P. IJzerman, J. Med. Chem. 38 (1995) 1942–1954.
[2] J.P. Capela, E. Fernandes, F. Remião, M.L. Bastos, A. Meisel, F. Carvalho, NeuroToxicology (2007).
[3] N.M. Barnes, T. Sharp, Neuropharmacology 38 (1999) 1083–1152.
[4] D. Hoyer, J.P. Hannon, G.R. Martin, Pharmacol. Biochem. Behav. 71 (2002) 533–554.
[5] M.D. Tricklebank, Trends Pharmacol. Sci. 6 (1985) 403–407.
[6] S. Jurczyk, M. Kołaczkowski, E. Maryniak, P. Zajdel, M. Pawłowski, E. Tatarczynska, A. Kłodzinska, E. Chojnacka-Wójcik, A. Bojarski, S. Charakchieva-Minol, B. Duszynska, G. Nowak, D. Maciaü, J. Med. Chem. 47 (2004) 2659–2666.
[7] P. Gaillard, P.-A. Carrupt, B. Testa, P. Schambel, J. Med. Chem. 39 (1996) 126–134.
[8] L. Orús, S. Pérez-Silanes, A. Oficialdegui, J. Martínez-Esparza, J. Castillo, M. Mourelle, T. Langer, S. Guccione, G. Donzella, E. Krovat, K. Poptodorov, B. Lasheras, S. Ballaz, I. Hervías, R. Tordera, J. Río, A. Monge, J. Med. Chem. 45 (2002) 4128–4139.
[9] M.L. López-Rodríguez, D. Ayala, B. Benhamú, M.J. Morcillo, A. Viso, Curr. Med. Chem. 9 (2002) 443–469.
[10] D. Centurión, M. Ortiz, P. Saxena, C. Villalón, Br. J. Pharmacol. 135 (2002) 1531–1539.
[11] J. Brea, J. Rodrigo, A. Carrieri, F. Sanz, M. Cadavid, M. Enguix, M. Villazón, G. Mengod, Y. Caro, C. Masaguer, E. Ravina, N. Centeno, A. Carotti, M. Loza, J. Med. Chem. 45 (2002) 54–71.
[12] J. Chambers, D. Nichols, J. Comput. Aided Mol. Des. 16 (2002) 511–520.
[13] A. Hamacher, M. Weigt, M. Wiese1, B. Hoefgen, J. Lehmann, M. Kassack, BMC. Pharmacol. 6 (2006) 11.
[14] J. Handl, D. Kell, J. Knowles, IEEE/ACM Trans. Comput. Biol. Bioinform. 4 (2007) 279–292.
[15] O. Nicolotti, V. Gillet, P. Fleming, D. Green, J. Med. Chem. 45 (2002) 5069–5080.
[16] N. Brown, B. McKay, J. Gasteiger, J. Comput. Aided Mol. Des. 20 (2006) 333–341.
[17] S. Janson, D. Merkle, M. Middendorf, Appl. Soft Comput. 8 (2008) 666–675.
[18] G. Derringer, R. Suich, J. Qual. Technol. 12 (1980) 214–219.
[19] J. Singh, M. Ator, E. Jaeger, M. Allen, D. Whipple, J. Soloweij, S. Chowdhary, A. Treasurywala, J. Am. Chem. Soc. 118 (1996) 1669–1676.
[20] R. Sheridan, S. Kearsley, J. Chem. Inf. Comput. Sci. 35 (1995) 310–320.
[21] S. Rao, T. Stockfisch, J. Chem. Inf. Comput. Sci. 43 (2003) 1614–1622.
[22] M. Leopoldo, E. Lacivita, M. Contino, N.A. Colabufo, F. Berardi, R. Perrone, J. Med. Chem. 50 (2007) 4214–4221.
[23] M. Cruz–MonteAgudo, F. Borges, M. Cordeiro, J. Comput. Chem. 29 (2008) 2445–2459.
[24] P. Grundt, K.M. Prevatt, J. Cao, M. Taylor, C.Z. Floresca, J. Choi, B.G. Jenkins, R.R. Luedtke, A.H. Newman, J. Med. Chem. 50 (2007) 4135–4146.
[25] A.J. Bojarski, P. Kowalski, T. Kowalska, B. Duszynska, S. Charakchieva-Minol, E. Tatarczynska, A. Klodzinska, E. Chojnacka-Wójcik, Bioorg. Med. Chem. 10 (2002) 3817–3827.
[26] Chem Draw Ultra 9.0. Cambridge Soft. 2004.
[27] J. Frank, MOPAC 2.0. Seiler Research Laboratory, US Air Force Academy, Colorado Springs, 1993, version 2.0.
[28] I.V. Tetko, Drug. Discov. Today. 10 (2005) 1497–1500.

[29] I.V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V.A. Palyulin, E.V. Radchenko, N.S. Zefirov, A.S. Makarenko, V.Y. Tanchuk, V.V.J. Prokopenko, Comput. Aid. Mol. Des 19 (2005) 453–463.
[30] I.V. Tetko, V.V. Kovalishyn, D.J. Livingstone, J. Med. Chem. 44 (2001) 2411–2420.
[31] I.V. Tetko, A.I. Luik, G.I. Poda, J. Med. Chem. 36 (1993) 811–814.
[32] R. Cavill, H.C. Keun, E. Holmes, J.C. Lindon, J.K. Nicholson, T.M.D. Ebbels, Bioinformatics 25 (2009) 112–118.
[33] A. Yasri, D. Hartsough, J. Chem, Inf. Comput. Sci. 41 (2001) 1218–1227.
[34] R. Todeschini, V. Consonni, A. Mauri, M. Pavan, Anal. Chim. Acta 515 (2004) 199–208.
[35] R. Todeschini, V. Consonni, M. Pavan, Chemom. Intell. Lab. Syst. 70 (2004) 55–61.
[36] R. Wehrens, H. Putter, L.M.C. Buydens, Chemom. Intell. Lab. Syst. 54 (2000) 35–52.
[37] STATISTICA 6.0 Statsoft_Inc.; 2001.
[38] D.C. Young, Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems. John Wiley & Sons, Inc., 2001.
[39] P. Gramatica, QSAR Comb. Sci. 5 (2007) 694–701.
[40] I. Tudor, Opera, Chemoinformatics in Drugs Discovery, vol. 13, Wiley, Weinheim, 2005.
[41] C. Mellin, J. Vallgárda, D.L. Nelson, L. Bjork, H. Yu, N.E. Andén, I. Csoregh, L.E. Arvidsson, U. Hacksell, J. Med. Chem. 34 (1991) 497–510.
[42] Z. Chilmonczyk, A. Szelejewska-Wozniakowska, J. Cybulski, M. Cybulski, A. Kozio, M. Gdaniec, Arch. Pham. Pham. Med. Chem. 330 (1997) 146–160.
[43] T. Ladduwahetty, A.L. Boase, A. Mitchinson, C. Quin, S. Patel, K. Chapman, A.M. MacLeod, Bioorg Med Chem Lett 16 (2006) 3201–3204.
[44] C.G. Chidester, C. Lin, R.A. Lahti, S.R. Haadsma-Svensson, M.W. Smith, J. Med. Chem. 36 (1993) 10.