J Mol Model (2005) 11: 457–467
DOI 10.1007/s00894-005-0237-x

ORIGINAL PAPER

**Victor E. Kuz'min · Anatoly G. Artemenko**
**Pavel G. Polischuk · Eugene N. Muratov**
**Alexander I. Hromov · Anatoly V. Liahovskiy**
**Sergey A. Andronati · Svetlana Yu. Makan**

# Hierarchic system of QSAR models (1D – 4D) on the base of simplex representation of molecular structure

Received: 15 April 2004 / Accepted: 8 December 2004 / Published online: 20 October 2005


**Abstract** In this work, a hierarchic system of QSAR models from 1D to 4D is considered on the basis of the simplex representation of molecular structure (SiRMS). The essence of this system is that the QSAR problem is solved sequentially in a series of the improved models of the description of molecular structure. Thus, at each subsequent stage of a hierarchic system, the QSAR problem is not solved ab ovo, but rather the information obtained from the previous step is used. Actually, we deal with a system of solutions defined more exactly. In the SiRMS approach, a molecule is represented as a system of different simplex descriptors (tetratomic fragments with fixed composition, structure, chirality and symmetry). The level of simplex-descriptor detail increases consecutively from 1D to 4D representations of molecular structure. It enables us to determine the fragments of structure that promote or interfere with the given biological activity easily. Molecular design of compounds with a given level of activity is possible on the basis of SiRMS. The efficiency of the method is demonstrated for the example of the analysis of substituted piperazines affinity for the 5-$HT_{1A}$ receptor.

**Keywords** 1D–4D QSAR · Simplex descriptors · Molecular design · Hierarchic system · 5-$HT_{1A}$ agonists

## Introduction

Presently, the technology of creation of new drugs, as a rule, includes the stage of QSAR (quantitative structure-activity relationship) researches. Nowadays there are many different QSAR methods [1–5]. They differ mainly by the principles and levels of representation and description of the molecular structure. The degree of adequacy of the molecular structure models varies from 1D to 4D representations.

The 1D models consider only the gross-formula of a molecule (for example, alanine—$C_3H_7NO_2$). Actually, such models reflect only the composition of the molecule. Obviously, it is impossible to solve adequately "structure-activity" tasks using such approaches. These models have an auxiliary role.

The 2D models contain information about the structure of the compound and are based on its structural formula [6].

Such models reflect only the topology of the molecule. These models are very popular. The capacity of such approaches is that the topology model of molecular structure in an implicit kind contains information about possible conformations of the compound.

The 3D QSAR models [1–4] give the full structural information, taking into account composition, topology and spatial shape of the molecule for one conformer only. These models are the most widespread. However, the choice of the analyzed conformer is mostly accidental.

The most adequate description of molecular structure will be given by 4D-QSAR models [5]. These models are similar to 3D models, but as compared to them, the structural information is considered for a set of conformers (conditionally, the fourth dimension), instead of one fixed conformation.

In this work, a hierarchic system of QSAR models from 1D to 4D has been considered on the basis of a simplex representation of molecular structure (SiRMS) [7–10]. This approach is based on QSAR problem solution via a sequence of the permanently improved molecular structure models. Thus, at each stage of the hierarchical system, the QSAR task is not solved ab ovo, but with the use of information received from a previous stage. In fact, it is

V. E. Kuz'min (✉) · A. G. Artemenko · P. G. Polischuk
E. N. Muratov · A. I. Hromov · A. V. Liahovskiy
S. A. Andronati · S. Y. Makan
A.V. Bogatsky Physico-Chemical Institute of the National Academy of Sciences, 7 Ukraine, Lustdorfskaya doroga 86, Odessa, 65080, Ukraine
E-mail: victor@farlep.net
Tel.: +38-04-82652012
Fax: +38-04-82225127

proposed to deal with a system of permanently improved solutions. This approach allows more effective interpretation of the QSAR models obtained because it reveals the molecular fragments/models for which the detailed development of structure is important.

The scheme of such a technology is shown in Fig. 1. Information from QSAR models of the lowest level is transferred (curve arrow) to models of a higher level after the corresponding statistical processing (block ≪Statistical models≫ in Fig. 1), during which the most significant structural parameters are chosen. It is necessary to point out that after 2D, the QSAR task is solved on the 4D level, because no *a priori* information about the "productive" conformation (the conformer which most effectively interacts for biological target) is available for 3D QSAR models. This information comes up only after the construction of 4D QSAR models and calculation of the activity of all conformers considered. Information about the "productive" (the most active) conformation is transferred to the 3D-QSAR level. At this stage, it is possible to construct the most adequate ≪structure-property≫ models. In all cases (1D–4D), the structural information is processed by the different statistical methods for obtaining the QSAR relations (block "Statistical models" in Fig. 1).

The principle feature of this strategy is that not only a *hierarchy of models* but also a *hierarchy of purposes* are taken into account (Fig. 1, unit - ≪Final Aims≫). Evidently, there cannot be only one model that will solve all the problems related to the influence of structure of the set of the studied molecules on the property examined. Hereby, for solving every concrete task, it is necessary to develop the set of different QSAR models, some of which are more suitable for the prognosis of the property studied, others for interpretation of the relations obtained, and yet others for molecular design. These models all together, in complex, work out the problem of creation of a new perspective of compounds and matters with the given set of properties. The important feature of such an approach is that the general results obtained by a few different independent models always are more relevant. Thus, we assume that the proposed strategy allows the solution of all problems relevant to virtual screening, modeling the functional

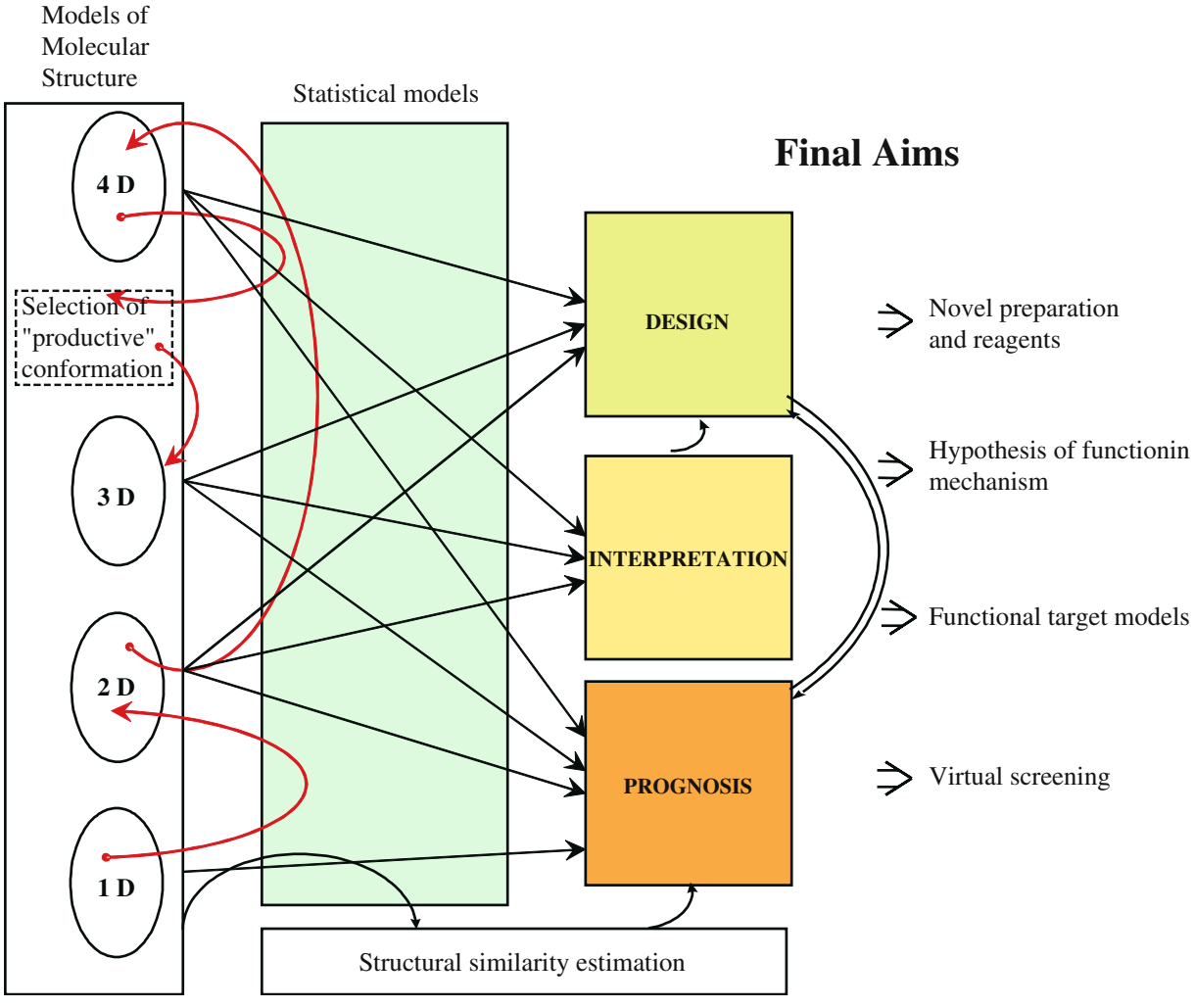



**Fig. 1** Hierarchical technology of solving QSAR and Drug-Design tasks

(biological) targets, advancing hypotheses about mechanisms of action, and finally, designing new compounds with a complex of useful properties.

This hierarchic strategy differs from the known hierarchic QSAR approach [11]. In this, the complication level of the structure representation rises among descriptors: topostructural, topochemical, geometrical, quantum-chemical, and physicochemical. In our approach, the system of descriptors is the same at all levels—these are simplexes (tetratomic fragments of molecular structure), and the hierarchic system consists of increasing the molecular structure representation adequacy from 1D to 4D models. Furthermore, taking into account the target hierarchy is also an important moment of our strategy (see Fig. 1).

## Materials and methods

### Hierarchic System on the Base of SiRMS

Any molecule can be represented as a system of different simplexes (tetratomic fragments of fixed composition, structure, chirality and symmetry) (Fig.2).

The total number ($N$) of all possible simplexes in an $n$-atomic structure is

$$N = \frac{n!}{(n-4)! \cdot 4!}$$

At the 1D level, a simplex is a combination of four atoms contained in the molecule (Fig. 3). The simplex descriptor (SD) at this level is the number of quadruples of atoms of definite composition. For compound ($A_aB_bC_cD_dE_eF_f...$), the value of SD ($A_iB_jC_lD_m$) is $K = f(i) \cdot f(j) \cdot f(l) \cdot f(m)$, where $f(i) = \frac{a!}{(a-i)! \cdot i!}$. The values of $f(j)$, $f(l)$, $f(m)$ have been calculated analogously. It is possible to define the number of smaller fragments (''pairs'', ''triples'') on the same scheme. In this case, some of parameters of $i$, $j$, $l$, $m$ are equated to zero.

At the 2D level, the connectivity of atoms in a simplex (11 base topological types, Table 1), atom type and bond nature (single, double, triple, aromatic) were considered. Atoms in a simplex can be differentiated on the basis of different characteristics, especially: (1) nuclear charge (nature of atom); (2) partial atom charge [12] (see Fig. 3) (3) lipophilicity; [13] (4) atomic refraction; (5) donor/acceptor of hydrogen in the potential H-bond; etc. For atom characteristics, which have real values (No. 2–4 in the list) at the preliminary stage, division of the value range into definite discrete groups is carried out. The number of groups (G) is a tuning parameter and can be varied (as a rule G = 3–7). For atom characteristic 5, the atoms have been divided into three groups: A (acceptor of hydrogen in H-bond), D (donor of hydrogen in H-bond), I (indifferent atom).

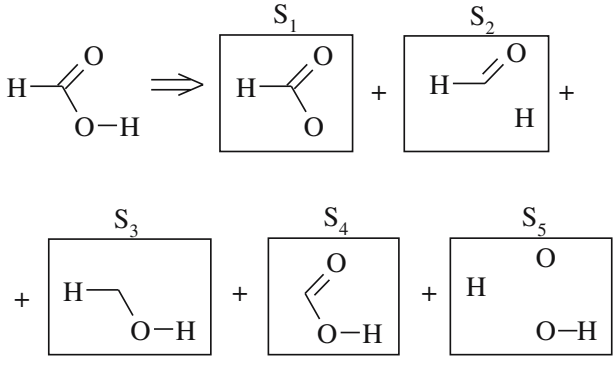



**Fig. 2** Scheme of molecule fragmentation to simplexes

The use of diverse variants of differentiation of simplex vertexes (atoms) represents the principle feature of our approach. We consider that specification of atoms by their nature (for example, C, N, O), as in many QSAR methods limits the possibilities of pharmacophore fragment selection. For example, if the –NH– group has been selected as the determining activity fragment (pharmacophore) and the ability to form H-bonds is the factor determining its activity, then we shall miss such donors of H-bonds as, for example, OH-groups, etc. The use of atoms differentiation by donor/acceptor of H-bond avoids the situation illustrated above. One can give analogous examples for other atom properties (lipophilicity, partial charge, refraction, etc).

Thus, the SD at 2D-level is the number of simplexes of fixed composition and topology. It is necessary to note that for a 2D-QSAR analysis, other structural parameters corresponding to molecular fragments of different size can be used together with simplex descriptors.

Actually, the 2D-descriptors generated by us are similar to the fragmentary parameters used in the HQSAR method (holographic QSAR). [2] The difference is that during the generation of descriptors we take into account both connected (Table 1, base types e, f, h, i, j, k) and unconnected fragments (Table 1, base types a, b, c, d, g), and also take into account not only the nature of atoms but also their different physical and chemical properties (charge, lipophilicity, etc.).

At the 3D-level, the stereochemistry of a molecule is taken into account. It is possible to differentiate all the simplexes as right (R), left (L), symmetrical achiral (S), or planar achiral (P) ones. The stereochemical configuration of simplexes is defined by modified Kahn-Ingold-Prelog rules [8]. The SD at this level is the number of simplexes of fixed composition, topology, chirality and symmetry (Fig. 3).

For 4D-QSAR models, each molecular structural parameter (MSP) is calculated by summing the products of the descriptor value for each conformer ($MSP_k$) and the probability of realization of the corresponding conformer ($P_k$).

$$MSP = \sum_{k=1}^{N} (MSP_k \cdot P_k),$$

where $N$ is the number of conformers considered and $MSP_k$ is the descriptor value for conformer $k$.

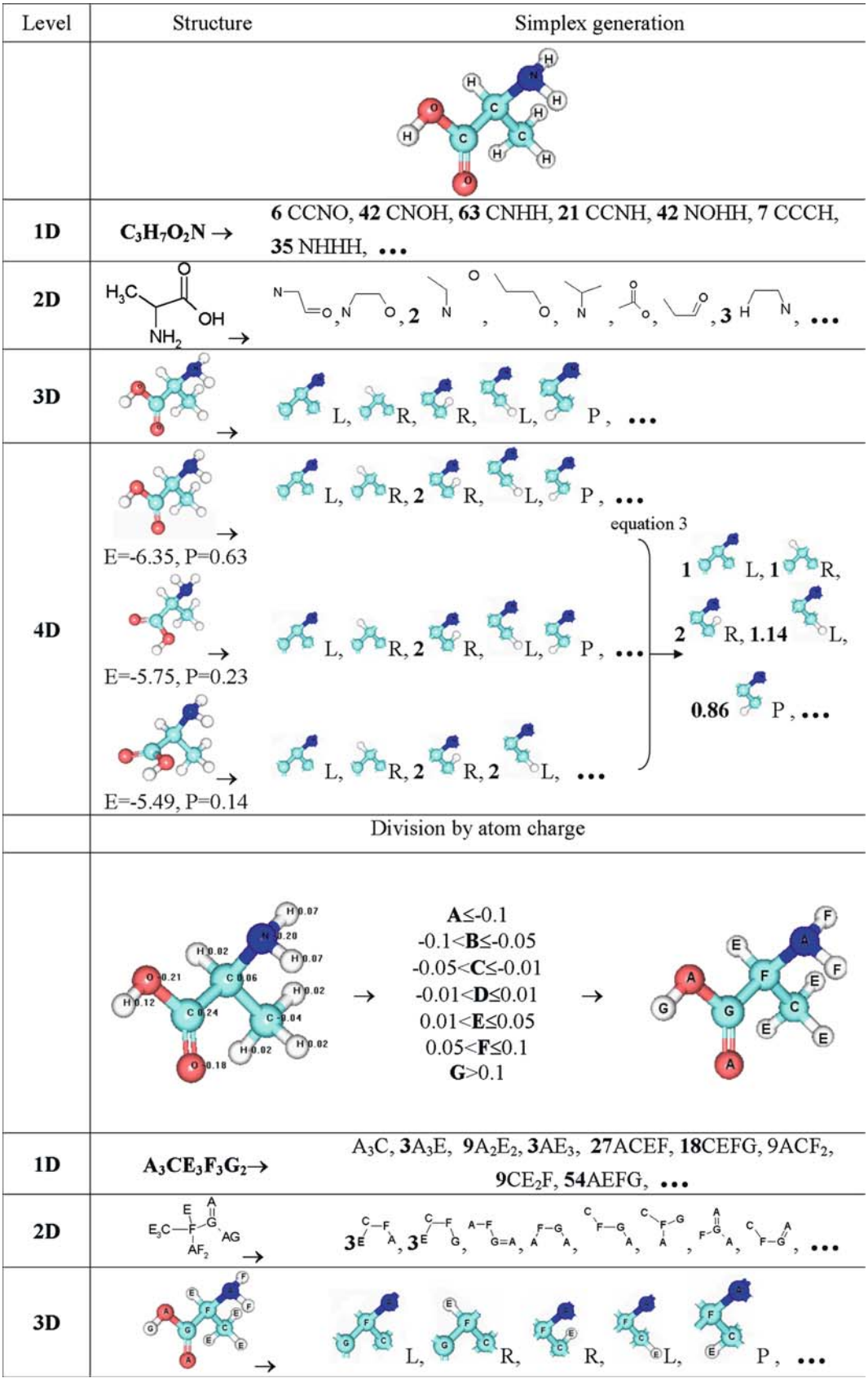
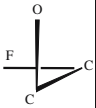



**Fig. 3** Examples of generation of simplex descriptors for alanine at 1D–4D levels

**Table 1** All possible topological types of molecular simplexes

| Base type | *a* | *b* | *c* | *d* | *e* | *f* | *g* | *h* | *i* | *j* | *k* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Simplex |  |  |  |  |  |  |  |  |  |  |  |
| Example |  |  |  |  |  |  |  |  | | | |

As is well known, [14] the probability of conformation $P_k$ is defined by its energy:

$$P_k=\left\{1+\sum_{i\neq k} EXP\left(\frac{-(E_i-E_k)}{RT}\right)\right\}^{-1},\sum_k P_k=1,$$

where $E_i$ and $E_k$ are the energies of conformations $i$ and $k$, respectively. The conformers are analyzed within an energy band of 5–7 kcal $mol^{-1}$. Thus, the molecular SD at the 4D-level takes the probability of realization of the 3D-level SD in the set of conformers into account. The subsequent examples (Fig. 3) demonstrate the representation of molecules as simplex sets at the different levels of structure detailing (1D–4D).

A large number of simplex descriptors has thus been generated in the method. The PLS-method proved efficient for working with a large number of variables [15]. It is well-known [16] that PLS-equation can be represented as

$$Y = b_0 + \sum_{i=1}^{N} b_i x_i,$$

where $Y$ is the appropriate activity, $b_i$ are PLS regression coefficients, $x_i$ is the $i$-th descriptor value (the number of simplexes of $i$-th type in the SiRMS) and $N$ is the total number of descriptors.

Using this equation it is not difficult to make a reverse analysis in the SiRMS approach. The contribution of each atom in the molecule can be defined as the sum of PLS regression coefficients ($b_i$) of all simplexes containing this atom divided by the number of atoms in the simplex.

Let the PLS regression coefficients for the simplexes $S_1$–$S_5$ (Fig. 2) accordingly be equal to $b_1$–$b_5$. Then it is easy to estimate the contribution of separate atoms to the activity studied. So, for example, the contributions (C) of carboxyl atoms will be equal:

$$C_{(o)} = \frac{b_1 + b_2 + b_4 + b_5}{4}; \quad C_{(C)} = \frac{b_1 + b_2 + b_3 + b_4}{4}$$

The results obtained can be represented on the molecule using color-coding according to the atoms' contribution. Realization of molecular design of compounds of a given activity level via the generation of the allowed combinations of simplexes that determine the property investigated, is possible on the base of SiRMS.

Thus, the proposed approach does not have a problem in optimal alignment of the set of considered molecules, which takes place in CoMFA analogues [1, 3, 4]. The SiRMS approach is similar to HQSAR [2] but has none of its restrictions (only topological representation of molecular structure) and deficiencies (ambiguity of descriptor formation when procedure of hashing of molecular holograms is realized). Besides, contrary to HQSAR, in SiRMS, different physical and chemical properties of atoms (charge, lipophilicity, etc.) can be taken into account.

## Results and discussion

The efficiency of the proposed method was demonstrated using as an example an affinity analysis of substituted piperazines for the serotonin 5-$HT_{1A}$ receptor. The values of $pK_i$ ($pK_i = -\lg(K_i)$, where $K_i$—inhibition constant) are shown in Table 2.

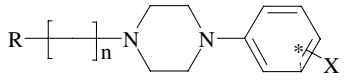
The multiple linear regression (MLR) [16] and partial least squares (PLS) [15] statistical methods have been used for QSAR analysis. Genetic algorithm [17] and trend-vector procedures [18] have been used for variable selection in PLS. The structures of the compounds for 3D-stages and 4D-stages were obtained using the MM + force field [19]. At the 4D-stage, for all molecules investigated the procedure of conformational search was applied, and the conformers within an energy band of 7 kcal $mol^{-1}$ above the optimal one were selected.

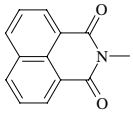
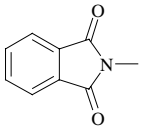
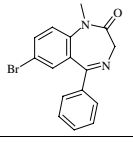
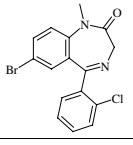
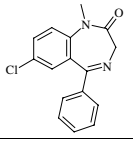
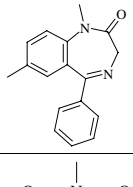
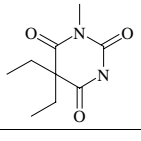
First, the 1D-QSAR task was solved. More than 400 SD were calculated. The result of the regression analysis is given by the equation:

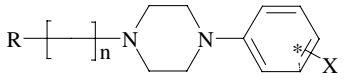
$$\begin{aligned} Pk_i &= +4.30 + 0.11n(NOOCl) + 0.22n(H) - 0.16n(C) \\ &\quad - 0.80n(NNNN), \\ R^2 &= 0.710, Q^2 = 0.664; SE = 0.67; F = 22.3, \end{aligned}$$

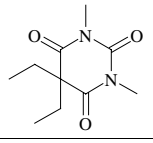
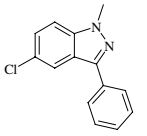
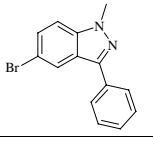
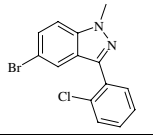
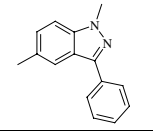
where $n(...)$ is the number of indicated combinations of atoms, $R$ the correlation coefficient, $Q$ the leave-one-out cross-validated $R$, SE the standard error of prediction, and $F$ the Fischer criterion. This model allows us to conclude that the simultaneous presence of the atoms N, O, O, Cl and a higher saturation of the compound enhance the affinity; the presence of a large number of carbon atoms

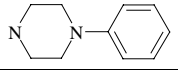
**Table 2** Experimental values of $pK_i$ for the compounds investigated

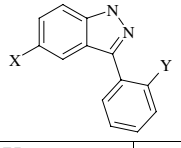


| Nr. | R | n | X | $pK_i$ |
|---|---|---|---|---|
| 1 |  | 4 | H | 8.54 |
| 2 | | 3 | o-$CH_3$ | 5.61 |
| 3 | | 3 | o-Cl | 7.14 |
| 4 | | 3 | p-$CH_3$ | 5.31 |
| 5 | | 3 | m-$CH_3$ | 5.33 |
| 6 | | 5 | o-$CH_3$ | 5.79 |
| 7 | | 5 | H | 7.36 |
| 8 | | 4 | o-Cl | 7.59 |
| 9 | | 3 | H | 7.27 |
| 10 | | 4 | p-$CH_3$ | 5.39 |
| 11 | | 4 | o-$CH_3$ | 5.89 |
| 12 | | 4 | m-Cl | 7.52 |
| 13 | | 4 | p-Cl | 7.41 |
| 14 | | 6 | H | 7.64 |
| 15 |  | 4 | H | 8.00 |
| 16 | | 4 | m-Cl | 6.74 |
| 17 | | 5 | m-$CH_3$ | 6.82 |
| 18 | | 5 | o-Cl | 8.17 |
| 19 | | 4 | o-Cl | 8.28 |
| 20 | | 5 | p-$CH_3$ | 6.79 |
| 21 | | 4 | p-$CH_3$ | 6.89 |
| 22 | | 5 | o-$CH_3$ | 7.88 |
| 23 | | 6 | o-$CH_3$ | 7.04 |
| 24 | | 6 | m-$CH_3$ | 6.61 |
| 25 | | 6 | p-$CH_3$ | 6.36 |
| 26 | | 6 | o-Cl | 7.85 |
| 27 | | 2 | o-$CH_3$ | 5.00 |
| 29 |  | 4 | H | 5.77 |
| 30 |  | 4 | H | 6.05 |
| 31 |  | 4 | H | 5.66 |
| 32 |  | 4 | H | 5.82 |
| 33 |  | 4 | H | 6.84 |



| Nr. | R | n | X | $pK_i$ |
|---|---|---|---|---|
| 34 |  | 4 | H | 7.02 |
| 35 |  | 4 | H | 5.55 |
| 36 |  | 4 | H | 5.26 |
| 37 |  | 4 | H | 5.21 |
| 38 |  | 4 | H | 6.35 |

| | | |
|---|---|---|
| 28 |  | 6.30 |

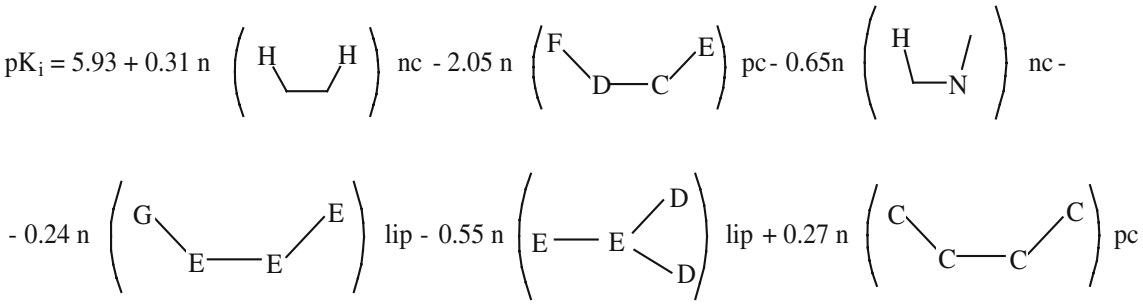


| Nr. | X | Y | $pK_i$ |
|---|---|---|---|
| 39 | $CH_3$ | H | 4.47 |
| 40 | Cl | H | 4.05 |
| 41 | Br | H | 3.91 |
| 42 | Br | Cl | 3.82 |

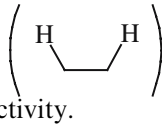
and four nitrogen atoms decrease the affinity. Obviously, this model has an auxiliary role only, and cannot be used for reliable prediction of activity of novel compounds.

Information about the constitution of compounds is taken into account in 2D-molecular models. At this stage the atoms in simplexes are differentiated on the basis of all the characteristics mentioned above (1–5). The total number of descriptors is more than 5 000. The MLR model obtained has quite good statistical characteristics:

$$R^2 = 0.910,\ Q^2 = 0.883,\ F = 56.7,\ SE = 0.39$$



where, nc is the differentiation of atoms by nature, pc by partial atomic charge, and lip by lipophilicity. The atoms are divided into seven groups corresponding to their partial charge: $\mathbf{A} < -0.1$; $-0.1 < \mathbf{B} < -0.05$; $-0.05 < \mathbf{C} < -0.01$; $-0.01 < \mathbf{D} < 0.01$; $0.01 < \mathbf{E} < 0.05$; $0.05 < \mathbf{F} < 0.1$; $\mathbf{G} > 0.1$. The atoms are also divided into seven groups corresponding to their lipophilicity: $\mathbf{A} < -1$; $-1 < \mathbf{B} < -0.5$; $0.5 < \mathbf{C} < -0.1$; $-0.1 < \mathbf{D} < 0.1$; $0.1 < \mathbf{E} < 0.5$; $0.5 < \mathbf{F} < 1$; $\mathbf{G} > 1$.

It is difficult to compare the equation obtained with the 1D-regression model, but in both cases increasing the degree of saturation (the  fragments in a 2D model) enhances the activity.

The QSAR task was also solved using the PLS-method because this method is more acceptable than MLR for a large number of structural parameters. A comparative analysis of the PLS models obtained is shown in Table 3. As expected, the quality of the models improves at the transition from 2D to 4D-levels. Taking 2D models as an example, one can see that different ways of differentiating atoms in simplexes lead to different results. The best models (6, 7) were obtained when all variants of the differentiation were used. It is obvious from Table 3 that the additional account of disconnected simplexes (Table 1, types b, c, d) improves the predicting ability ($Q^2$) of the QSAR models (see the models 6, 7 and 8, 9, respectively).

The most adequate model (Table 3, model 11) constructed used an analysis of the activity of different conformers (within the framework of a 4D-QSAR model) and selection of the most active ones. These "productive" conformers were used in the resulting 3D-QSAR model. As Fig. 4 indicates, "productive" conformers differ from the most favorable ones by shape and energy.

It is possible to trace the succession of the PLS models obtained from 2D to 4D from Table 4, which shows that the greater part of the simplexes selected (at the different levels of molecular structure modeling) have qualitatively similar influences on the activity explored. An example of the influence of different structural fragments on affinity is shown in Fig. 5. The resulting influence of all simplexes examined is reflected by the color-coding of the fragments of molecules both promoting (red) and interfering (green) affinity for the 5-$HT_{1A}$ receptor.

**Table 3** QSAR models obtained by the PLS-method

| | Structural descriptors | Atom differentiation type | Dimension | $R^2$ | $Q^2$ | Number of components |
|---|---|---|---|---|---|---|
| 1 | Simplexes[a] | Atom nature (1) | 2D | 0.925 | 0.758 | 9 |
| 2 | Simplexes[a] | Partial charge (2) | 2D | 0.966 | 0.766 | 9 |
| 3 | Simplexes[a] | Lipophilicity (3) | 2D | 0.851 | 0.750 | 4 |
| 4 | Simplexes[a] | Refraction (4) | 2D | 0.782 | 0.728 | 3 |
| 5 | Simplexes[a] | Donor/acceptor of H-bond (5) | 2D | 0.768 | 0.605 | 5 |
| 6 | Simplexes[b] | All mentioned (1–5) | 2D | 0.837 | 0.768 | 10 |
| 7 | Simplexes[a] | All mentioned (1–5) | 2D | 0.927 | 0.827 | 5 |
| 8 | Simplexes[b] | All mentioned (1–5) | 4D | 0.951 | 0.753 | 8 |
| 9 | Simplexes[a] | All mentioned (1–5) | 4D | 0.948 | 0.831 | 7 |
| 10 | Simplexes[b] | All mentioned (1–5) | 3D | 0.833 | 0.662 | 3 |
| 11 | Simplexes[a] | All mentioned (1–5) | 3D | 0.961 | 0.854 | 8 |
| 12 | Dragon | – | 2D | 0.845 | 0.729 | 4 |
| 13 | Dragon | – | 3D | 0.826 | 0.740 | 3 |

[a]Connected (e, f) and disconnected (b, c, d) simplexes
[b]Only connected (e, f) simplexes

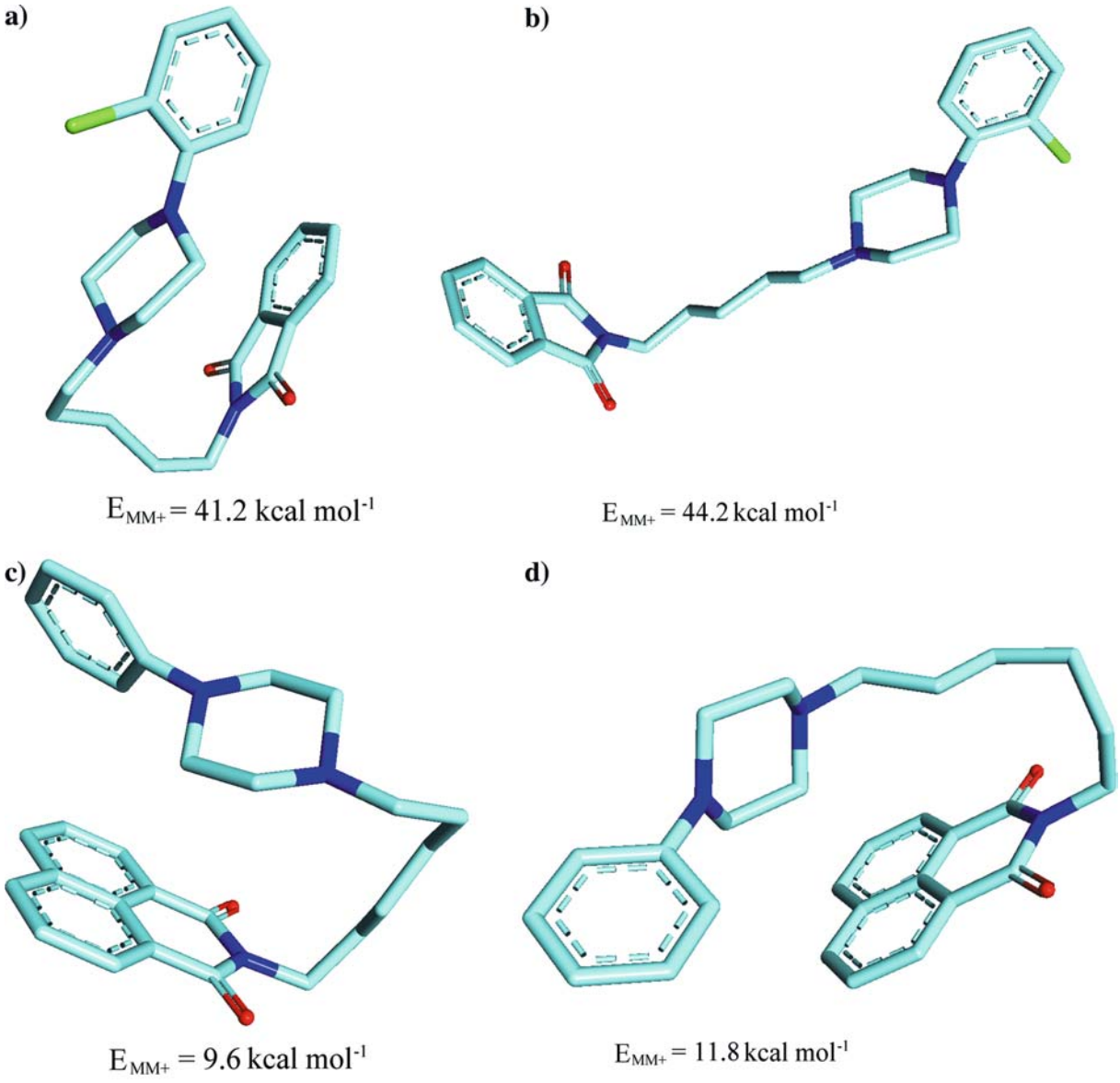


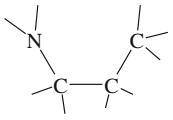
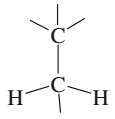
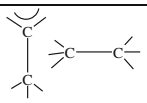
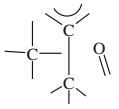
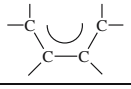
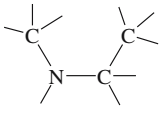
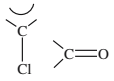
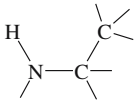
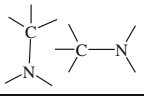

**Fig. 4** The most favorable (**a, c**) and "productive" (**b, d**) conformers of molecules 18 and 14

Within the framework of the SiRMS (like the CoMFA method [1]), it is possible to define the relative influence of the different physical and chemical factors on the character of the molecule's interaction with the biological target. For this purpose, it is necessary to sum and compare contributions of simplexes in a regression model separately for every differentiation group. Thus, the relative contribution of simplexes, where differentiation of vertexes corresponds to the partial charges on atoms, reflects the role of electrostatic factors; the relative contribution of simplexes, where atoms are differentiated by lipophilicity, reflects the role of hydrophobic factors, etc.

An analysis of the different QSAR models shows that, in all likelihood, electrostatic and hydrophobic interactions have roughly the same influence on the affinity of substituted piperazines for the 5-$HT_{1A}$ serotonin receptor (Fig. 6). The relative contribution of simplexes, where atoms are differentiated by types (nature), is also great (18–22%). This contribution in the implicit form reflects the influence of all physical and chemical characteristics on the interaction with the receptor. It is interesting to note that the role of H-bonds and dispersion forces (contribution of simplexes, where atoms are differentiated by refractions) into the ligand-receptor interaction is relatively small (3–9%).

On the basis of the most adequate models (Table 3, models No. 7, 9, 11), the molecular fragments with the maximal influence on affinity have been defined (Table 5). These results correspond roughly with the information of the 1D-QSAR model (see above). Thus, for example, fragments 1 and 3, which promote binding to the receptor, correspond to combinations of the NOOCl atoms,

**Table 4** Normalized influence of different simplexes on the affinity for the 5-$HT_{1A}$ receptor ("+","-" are positive and negative infuences on the affinity, respectively

| Simplex | atom differentiation type | 2D | 3D | 4D |
|---|---|---|---|---|
|  | lipophilicity | 0.15 | 0.12 | 0.15 |
|  | lipophilicity | 0.09 | 0.08 | 0.09 |
|  | nuclear charge | 0.10 | 0.08 | 0.07 |
|  | nuclear charge | -0.19 | -0.25 | -0.28 |
|  | partial charge | -0.14 | -0.20 | -0.23 |
|  | lipophilicity | -0.13 | -0.10 | -0.12 |
|  | refraction | 0.03 | | 0.02 |
|  | lipophilicity | -0.03 | | 0.01 |
|  | donor/acceptor of H-bond | -0.01 | 0.05 | |
|  | partial charge | 0.02 | 0.02 | |

and fragments 5 and 6, which along with the piperazine fragment prevent binding, correspond to the combinations of the NNNN atoms. Such correspondences of the different and independent QSAR models can demonstrate that the tendencies of the structure influence on affinity revealed are steady enough. Naturally, these tendencies can be modified after expansion of the set studied and increase of its structural diversity.

Molecular design of perspective molecules has been conducted on the basis of the information obtained (Table 6). Experimental research on some of them are planned in the future.

For the estimation of the efficiency of the proposed method, we tried to solve the task using other QSAR approaches. In particular, the solution of the QSAR task using Dragon [20] descriptors gives adequate models with worse statistical characteristics (Table 3, models 12,13). Moreover, the solution of the reverse task within the DRAGON method is difficult, because a lot of poorly interpreted parameters are involved in the model, for example: [21] different topological indices, 2D autocorrelations, RDF-descriptors, WHIM-descriptors, GETAWAY-descriptors, BCUT-descriptors, 3D MoRSE-descriptors, etc.

It should be noted that building adequate QSAR models using the Lattice Model approach [4] failed. Probably, this is related with the problem of molecular alignment of heterogeneous and conformationally flexible structures. Such a problem is also found in the popular 3D-QSAR approaches such as CoMFA, [1] and HASL [3]. Therefore, we propose that the above mentioned lattice methods are little-suited for this set.

In conclusion, it is important to note that the results obtained are not exhaustive, but represented for illus-

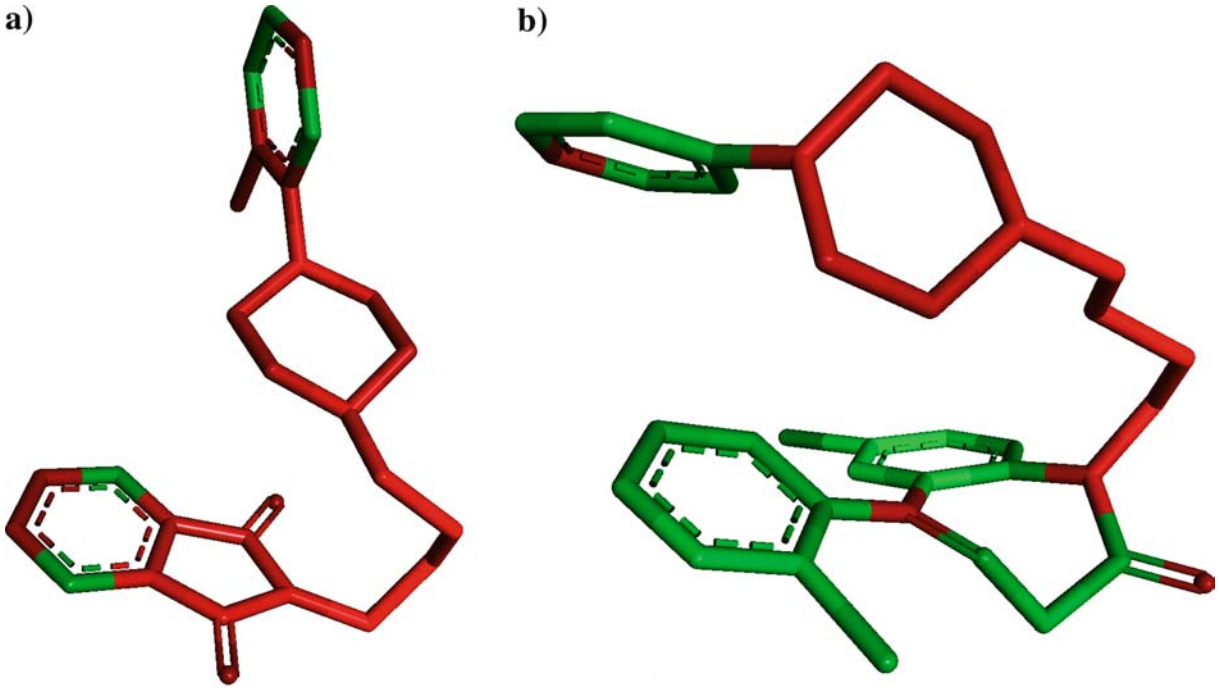



**Fig. 5** Color-coding of molecular fragments with standpoint of their influence on the activity for the compounds 19 (**a**) and 32 (**b**) (*red* fragment enhance the affinity, *green* —fragment decrease the affinity)

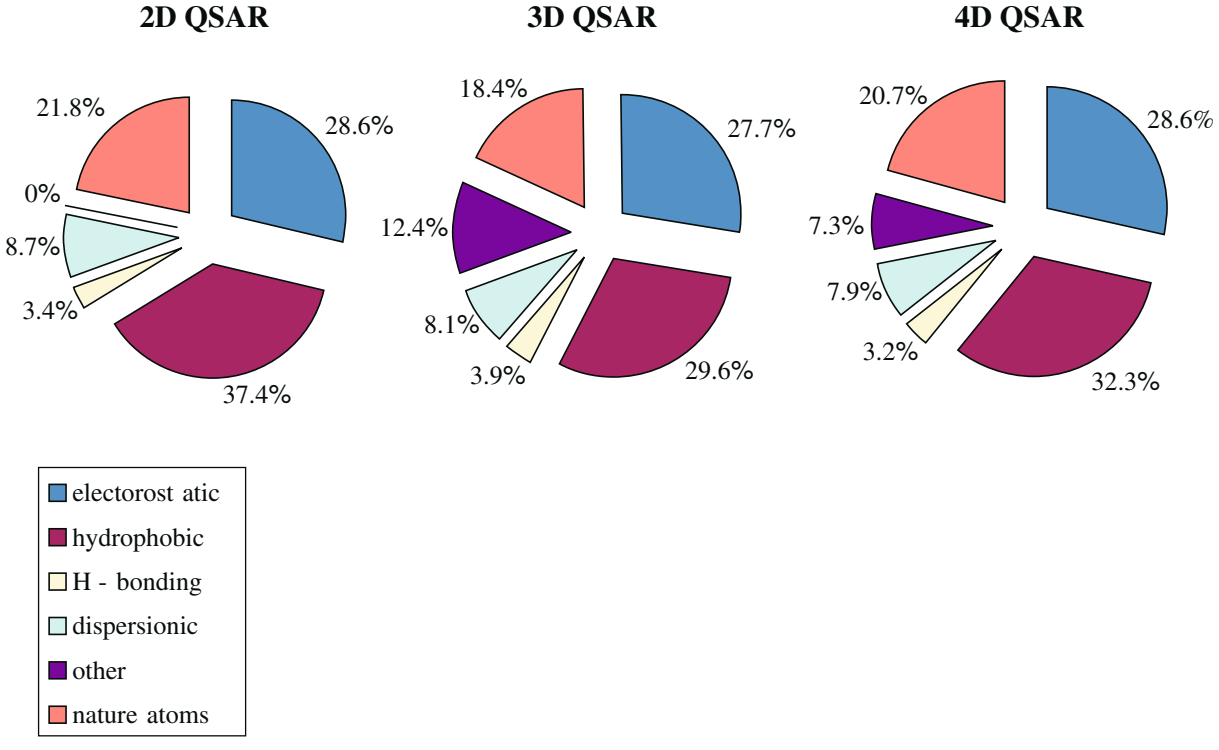



**Fig. 6** The relative influence of some physical and chemical factors into the peculiarity of ligand-receptor interaction estimated by the different QSAR models

tration of the possibilities and efficiency of the hierarchical QSAR strategy. Complete and detailed analysis of "structure-affinity to 5-$HT_{1A}$ receptor" relationship for substituted piperazines will be represented as a separate publication.

**Table 5** Examples of molecular fragments that promote and interfere with binding of 5-HT1A agonists to the serotonin receptor

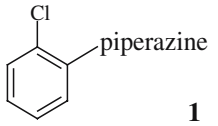
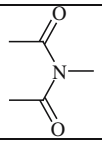
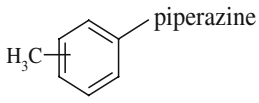
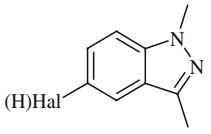
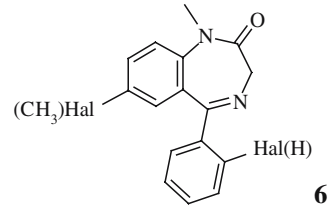
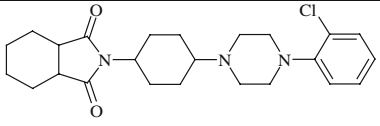
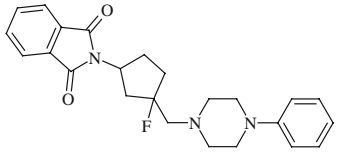
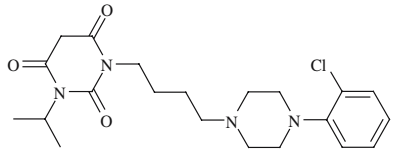
| **Promotes** | | |
|---|---|---|
|  **1** | $-(CH_2)_4-$, $-(CH_2)_5-$ **2** |  **3** |
| **Interfering** | | |
|  **4** |  **5** |  **6** |

**Table 6** Designed compounds and predicted value of their activities

| **Structure** | **Predicted $pK_i$** |
|---|---|
|  | 8.6 |
|  | 8.1 |
|  | 8.8 |

**Acknowledgments** This study was partially supported by INTAS (INTAS Grant 97-31528).

## References


1. Cramer RD, Patterson DE, Bunce JD (1988) J Am Chem Soc 110:5959–5967
2. Seel M, Turner DB, Wilett P (1999) QSAR 18:245–252
3. Doweyko AM (1988) J Math Chem 31:1396–1406
4. Kuz'min VE, Artemenko AG, Kovdienko NA, Tetko IV, Livingstone DJ (2000) J Mol Model 6:517–526
5. Vedani A, Dobler M (2000) Prog Drug Res 55:107–135
6. Suchachev DV, Pivina TS, Shliapochnikov VA, Petrov EA, Paliulin VA, Zefirov NS (1993) Dokl RAN (in Russian) 328:50–57
7. Kuz'min VE (1994) Zh Phys Khim (in Russian) 68:1037–1043
8. Kuz'min VE (1995) Zh Strucur Khim (in Russian) 36:871–878
9. Kuz'min VE, Chelombitko VA, Yudanova IV (1998) Zh Structur Khim (in Russian) 39:549–554
10. Kuz'min VE, Artemenko AG, Lozitsky VP, Muratov EN, Fedtchouk AS, Dyachenko NS, Nosach LN, Gridina TL, Shitikova LI, Mudrik LM, Chelombitko VA, Zheltvay AI, Vanden Eynde JJ (2002) Acta Biochim Pol 49:157–168
11. Basak SC, Gute BD, Grunwald GD (1997) J Chem Inf Comput Sci 37:651–655
12. Jolly WL, Perry WB (1973) J Am Chem Soc 95:5442–5450
13. Renxiao W, Ying Fu, Luhua L (1997) J Chem Inf Comput Sci 37:615–621
14. Burkert U, Allinger NI (1982) Molecular mechanics. ACS, Washington DC
15. Rännar S, Lindgren F, Geladi P, Wold S (1994) J Chemometrics 8:111–125
16. Ferster E, Renz B (1979) Methoden der Korrelations und Regressionanalyse. Verlag Die Wirtschaft, Berlin
17. Schwefel HP (1981) Numerical optimization of computer models. Chichester, Wiley
18. Carhart RE, Smith DH, Venkataraghavan R (1985) J Chem Inf Comput Sci 25:64–73
19. Hyperchem 7.5 software (2003) Hypercube Inc, 1115 NW 4th Street, Gainesville, FL 32601 USA. Trial version from http://www.hyper.com
20. Milano Chemometrics & QSAR Research Group (2001) DRAGON Web version from http://www.disat.unimib.it/chm
21. Todeschini R, Consonni V (2000) Handbook of molecular descriptors. In: Mannhold R, Kubinyi H, Timmerman H (eds) Methods and principles in medicinal chemistry, vol 11. Wiley-VCH