J Comput Aided Mol Des (2008) 22:489–505
DOI 10.1007/s10822-008-9203-x

A new paradigm for pattern recognition of drugs

Vladimir A. Potemkin · Maria A. Grishina

Received: 6 September 2007 / Accepted: 20 February 2008 / Published online: 21 March 2008

Abstract A new paradigm is suggested for pattern recognition of drugs. The approach is based on the combined application of the 4D/3D quantitative structure–activity relationship (QSAR) algorithms BiS and ConGO. The first algorithm, BiS/MC (multiconformational), is used for the search for the conformers interacting with a receptor. The second algorithm, ConGO, has been suggested for the detailed study of the selected conformers' electron density and for the search for the electron structure fragments that determine the pharmacophore and antipharmacophore parts of the compounds. In this work we suggest using a new AlteQ method for the evaluation of the molecular electron density. AlteQ describes the experimental electron density (determined by low-temperature highly accurate X-ray analysis) much better than a number of quantum approaches. Herein this is shown using a comparison of the computed electron density with the results of highly accurate X-ray analysis. In the present study the desirability function is used for the first time for the analysis of the effects of the electron structure in the process of pattern recognition of active and inactive compounds. The suggested method for pattern recognition has been used for the investigation of various sets of compounds such as DNA-antimetabolites, fXa inhibitors, 5-HT_{1A}, and α_1-AR receptors inhibitors. The pharmacophore and antipharmacophore fragments have been found in the electron structures of the compounds. It has been shown that the pattern recognition cross-validation quality for the datasets is unity.

V. A. Potemkin · M. A. Grishina (✉)
Chemical Department, Chelyabinsk State University,
Br. Kashirianich, 129, Chelyabinsk 454021,
Russian Federation
e-mail: maria_grishina@modelchem.ru

Keywords Antipharmacophore · Desirability function · Electron density · Enantioconformers · Multiconformational analysis · Pattern recognition · Pharmacophore · Superimposition · X-ray · 3D and 4D QSAR

Introduction

Pattern recognition methods are used for analysis and prediction of biological activity of drugs [e.g., 1–3]. There are many alternative variations of these methods, such as neural network, discriminant analysis, cluster analysis, etc. [1, 3–6]. Most of the research results obtained by the specific methods describing various kinds of biological activity allow one to identify the molecular fragments most often found in drugs. Energy, geometrical, orbital, topological, and other characteristics have been determined in a number of studies. They permit compounds to be classified as active or inactive. It is possible to make a supposition about the binding mechanism of the described agents [e.g., 2, 4, 6]. However, the given approaches are usually used for the consideration of sets with a unique scaffold [e.g., 5, 8–10]. At the same time there is a problem of quantum correlation or quantum conformity for the various classes of compounds acting by similar mechanisms. Besides, most of the existing approaches to pattern recognition of drugs fail to account for the conformational state of the drugs, which determines the shape, size, and energy characteristics of a molecule and hence its potential ability to bind with a receptor [5, 8–10]. Therefore, the present study offers a new paradigm for pattern recognition of drugs taking into consideration these problems.

Description of the paradigm suggested for pattern recognition of drugs

The suggested approach is based on the combination of 4D/3D QSAR algorithms—BiS/MC (multiconformational)[1] [11–14] and ConGO (www.modelchem.ru) [14, 15]. The first algorithm, BiS/MC, permits multiconformational study of biologically active compounds and selection of conformers responsible for the binding of a molecule with the receptor [e.g., 16–18] (Figs. 1 and 2). This consideration is necessary even in the cases of molecules which appear to have minimal conformational flexibility. For example, for the molecule **1bl6**, which has just four rotatable bonds, 12 conformations have been found (six pairs of mirror antipodes—enantioconformers, e.g., Fig. 2) within a 3 kcal/mol energy range of the lowest energy conformer. It often happens that only one conformer (the most active one) can interact with the target. The BiS/MC algorithm determines this conformer.

To date many sets of substances possessing anti-tumor, anti-inflammatory, and other kinds of activity have been studied using the mentioned algorithm [11–13, 16–24]. It has been demonstrated that the conformers selected by this algorithm and determined by X-ray analysis [e.g., 25–27] in receptor-ligand complexes are in good agreement [12, 20, 21, 24] (Figs. 1 and 2).

The BiS/MC algorithm, as well as the results of its use, has been described in a number of published papers and is also briefly described in the next section. The second algorithm, ConGO, has been suggested for the detailed study of the electron structure of the selected conformers, and for the search for the electron structure fragments that determine patterns of the most active compounds. ConGO algorithm is described in the section entitled "ConGO algorithm".

BiS/multiconformational (BiS/MC) algorithm

In the BiS algorithm it is supposed that the drug orientation in a receptor depends on the whole spectrum of van der Waals, Coulomb, and specific interactions with the receptor [28]. All these interactions determine a molecular field which must provide the maximal complementarity of the biologically active compounds to the receptors. Various molecules may bind to different parts of the active site of the receptor. Therefore, the universe of active molecules permits the reconstruction of the receptor field complementary to the field of the superposed active molecules. Hence, to solve the problem of the molecular orientation within a receptor cavity, it is necessary to determine a generalized field for the set of molecules.

[1] www.modelchem.ru

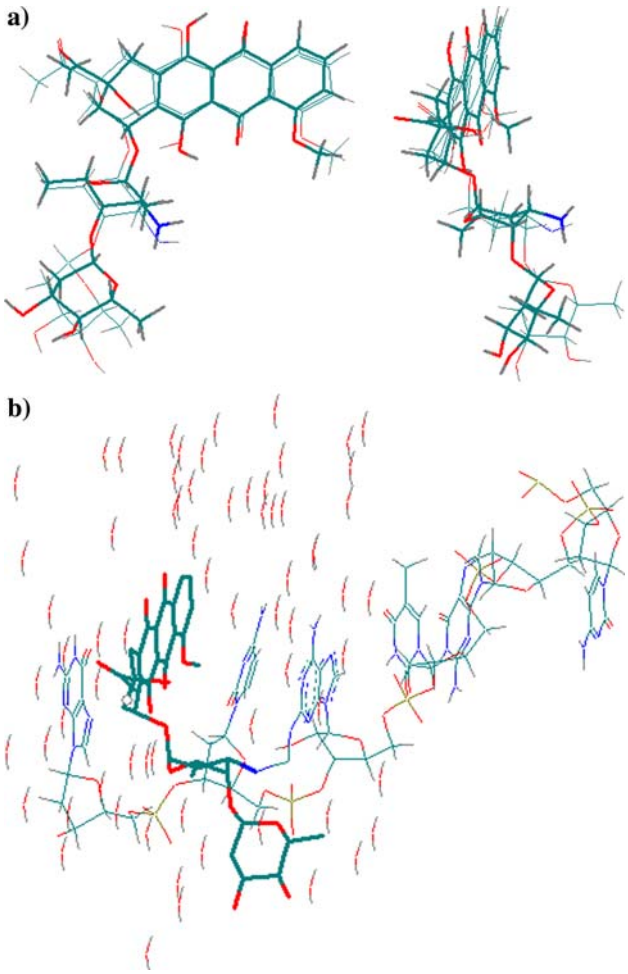


Fig. 1 Anti-tumor DNA-antimetabolite **1d35**: (**a**) the orthogonal views of the actual [25, 27] and the predicted conformers overlaid by least-squares optimization; (**b**) **1d35** complexed with DNA, obtained by X-ray [25, 26]. (**1d35**: the names of molecules are taken from the Protein Data Bank [25])

Within the BiS algorithm, the molecular field is determined, taking into account the Coulomb potential and the potential of van der Waals interactions affecting a given point m of the molecular surface. The Coulomb potential is calculated from the canonical equation 1 [28]:

$$\varphi_m^q = \sum_{i=1}^{N} \frac{q_i}{R_{im}} k \qquad (1)$$

where N is the number of atoms in the structure; q_i is the charge of the i atom; R_{im} is the distance from a point m to the atom i; k is a scaling coefficient.

To calculate the van der Waals potential at the point m, similarly to Coulomb interactions, the following equation was suggested:

$$\varphi_m^{\mathrm{VDW}} = -2 \sum_{i=1}^{N} V_{im} \frac{2^3 r_i^3}{R_{im}^6} \qquad (2)$$

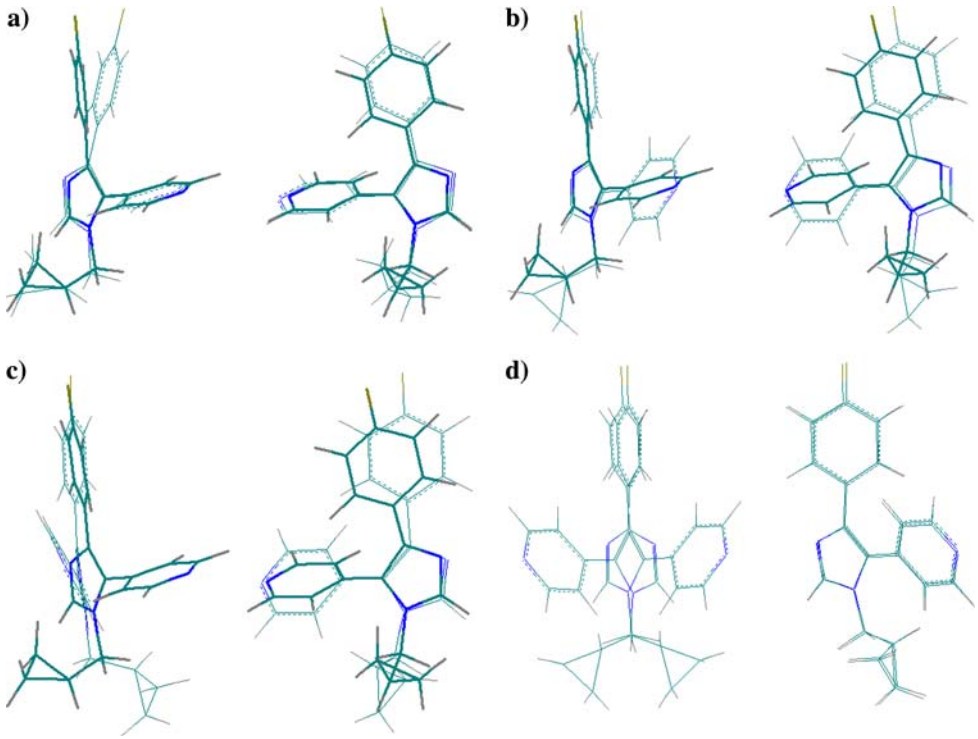


Fig. 2 P38 MAP-kinase inhibitor **1bl6** (the names of molecules are taken from the Protein Data Bank [25]). Orthogonal views of the pairs of conformers overlaid by least-squares optimization: (**a**) the actual [25, 27] and the predicted most active conformers (RMS = 0.30 Å); (**b**) the predicted most active conformer and the first lowest energy enantioconformer (RMS = 0.70 Å); (**c**) the predicted most active conformer and the second lowest energy enantioconformer (RMS = 1.35 Å); (**d**) the lowest energy enantioconformers (mirror antipodes). The predicted conformer is marked by a bold line

where V_{im} is the potential energy minimum of the Lennard-Jones equation; r_i is the Van der Waals radius of the i atom. V_{im}, r_i, and q_i values can be calculated with the model of effective radii of atoms (MERA) model [29]. To simplify the calculations, equation (2) includes only the attractive term of the Van der Waals interactions.

The complementary receptor field can be represented as a set of pseudo-atoms (probe spheres with some charge and radius) placed on the circumscribed ellipsoid of the considered molecule. To determine the characteristics of the complementary field, the Coulomb and Van der Waals potentials are calculated for the first molecule of the set. The derived potentials permit the characteristics of the pseudo-atom located at the point m (charge q_m and radius r_m) to be found, yielding the maximum complementarity to the molecule in the given point of the field. The set of pseudo-atoms represents the model of the receptor.

The characteristics of each probe sphere m can be calculated as follows:

$$q_m = -\frac{\varphi_m^q}{\sum_{i=1}^{N} \frac{k}{R_{im}}}; \quad r_m = \sqrt[3]{\frac{\varphi_m^{\mathrm{VDW}}}{-2^3 \sum_{i=1}^{N} \frac{2V_{im}}{R_{im}^6}}}$$

Then the orientation of the second molecule is optimized in the obtained field by the combined simplex and quasi-Newton methods to reach the minimal overall probability (P) of the contact of its atoms with all the pseudo-atoms:

$P = 1 - \prod_{m=1}^{M} (1 - p_m)$, where $p_m = \exp\left(-\frac{E_m}{RT}\right)$ and M is the number of probe spheres;

$$E_m = \sum_{i=1}^{N} \left(\frac{kq_i q_m}{R_{im}} - 2V_{im} \frac{(r_m + r_i)^6}{R_{im}^6} + V_{im} \frac{(r_m + r_i)^{12}}{R_{im}^{12}} \right)$$

Then the complementary receptor field can be improved by the addition of the field potentials of the second molecule to the previously obtained potentials:

$$\varphi_m^q = \varphi_m^q + \varphi_m^{q\prime} \text{ and } \varphi_m^{\mathrm{VDW}} = \varphi_m^{\mathrm{VDW}} + \varphi_m^{\mathrm{VDW}\prime}$$

The field potentials $\varphi_m^{q\prime}$ and $\varphi_m^{\mathrm{VDW}\prime}$ of the second molecule are calculated similarly to the first molecule in Eqs. 1, 2. The same procedures are carried out for the third, fourth, and all the following molecules of the set. When the orientation of all the molecules is completed, the positions of the first, second, and the following molecules are reanalyzed in the improved complementary field.

The iterations of the orientation are terminated when the difference in the atomic coordinates at the current and previous steps is below a predetermined threshold.

After the optimization of the orientation is completed, a linear relationship between the biological activity (BA) and the parameters of the interactions between the pseudo-atom ad the molecule is derived using backward and forward stepwise procedures. The parameters can include interaction energies calculated according to Eq. (3) and forces F_m

$$F_m = \sum_{i=1}^{N} \frac{dE_m}{dR_{im}}$$

$$BA = a_1 + a_2 \sum_{l=1}^{L} E_l + a_3 \sum_{o=1}^{O} F_o,$$

where BA is the biological activity,

E_l is the interaction energy of pseudo-atom l with a drug,
L is the total number of the pseudo-atoms whose interaction energy with the drugs defines the biological activity,
F_k is the interaction force of pseudo-atom o with a molecule.
O is the number of the pseudo-atoms whose interaction force with the drugs defines the biological activity.
a_1, a_2, and a_3 are the parameters of the expression

The computation of forces permits the dynamics of interactions in the receptor-ligand system to be taken into account. The direction of the force vectors determines the direction of a molecule's movement in the process of interaction with the receptor.

Thus, the BiS approach, considering the molecular field, analyses the exterior of the molecules, taking into account most of the factors determining the biological activity: most possible kinds of interactions (van der Waals, Coulomb, and specific interactions), flexibility of the receptor, and dynamics of the interactions in the receptor-ligand system. However, molecules can very often exist in many conformations, only one of which can define the interaction of the compound with the receptor. In order to take into account the conformational flexibility of molecules, the possible conformers whose probability of existence is more than 0.0001 can be found by the MultiGen algorithm [28] using the MM3 force field.

In the multiconformational version of BiS algorithm (BiS/MC) all of the described procedures (molecular field computation, orientation of structures in the complementary field, the computation of interaction forces and energies with pseudo-atomic receptor) are carried out for each of the possible conformers of a molecule.

The biological activity of a drug is represented as a superposition of partial activities of the conformers, taking into account the probabilities of their existence (p_c) [13, 28]:

$$BA = a_1 + a_2 \sum_{c=1}^{C} p_c \sum_{l=1}^{L} E_{cl} + a_3 \sum_{c=1}^{C} p_c \sum_{o=1}^{O} F_{co}$$

$$p_c = \frac{\exp(-E_c/kT)}{\sum_{d=1}^{D} \exp(-E_d/kT)}$$

where E_{cl} is the interaction energy of pseudo-atom l with conformer c, F_{co} is the interaction force of pseudo-atom o with conformer c, and E_c is the total energy of conformer c of a molecule represented by C conformers.

The best relationship describing the biological activity value with maximal leave-one-out cross-validation quality is found. Then the biological activity of conformer c of a molecule can be computed using the following formula:

$$BA_c = a_1 + a_2 p_c \sum_{l=1}^{L} E_{cl} + a_3 p_{kc} \sum_{o=1}^{O} F_{co},$$

which allows the determination of the most active conformer responsible for the interaction of the compound with the receptor [13].

ConGO algorithm

The ConGO algorithm enables a detailed analysis of the electron structures of the most active conformers selected by the BiS/MC algorithm. In order to describe the electron structure of molecules, the ConGO algorithm deals with the values of the electron density function at lattice points, separated by a distance of d Å (Fig. 3).

Then, the algorithm compares the electron density of the molecules of the set at the lattice points, superimposing each molecule of the set in turn until the greatest similarity of electron density values at the lattice points is achieved, according to the following algorithm.

First, ranking of the structures is carried out according to decreasing activity. Then, the second molecule of the set is superimposed on the first using rotation and translation operations. The minimal value of the following expression is chosen as the criterion of maximal similarity of the electron structures:

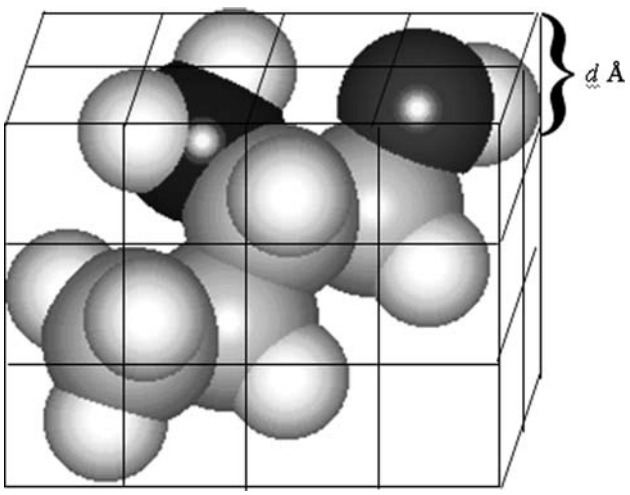


Fig. 3 The lattice spanning the molecular structure with resolution d Å

$$\Delta\Phi_i = \frac{1}{M-1}\sum_{m=1}^{M}\frac{(\bar{\rho}_{mi}-\rho_{mi})^2}{(\bar{\rho}_{mi}+\rho_{mi})^2} \quad (3)$$

where i is the number of the considered molecule (here the number of the molecule $i = 2$), ρ_{mi} is the electron density of molecule i at point m of the lattice; $\bar{\rho}_{mi} = \frac{\sum_{j=1}^{i}\rho_{mj}}{i}$ is the average value of the electron density determined for the considered molecules at lattice point m, and M is the total number of lattice points.

The function (3) is the optimizable function used for the alignment optimization using the combination of simplex and quasi-Newton methods.

the molecular electron structures. Then, on the basis of the experimental assignment of molecules as active (assigned a probability of activity of 1) or inactive (assigned a probability of activity of 0), it is possible to analyze the electron structure fragments which are characteristic features of the drugs. To do this, a quantitative relationship expressed as a desirability function [30, 31] as shown in formula (4) between the probability of activity (P_k) and the electron density values at the lattice points is carried out. The relationship is obtained using forward stepwise and backward stepwise procedures. The quality of the predicted P_k value is estimated using the leave-one-out cross-validation technique.

$$P_k = \exp\left[-\exp\left\{0.834 - 3.08 \times \left(c_1 + c_2\sum_{t=1}^{T}\rho_{kt} - c_3\sum_{u=1}^{T}\rho_{ku}\right)\right\}\right] \quad (4)$$

The same superimposing procedure is carried out for the third, the fourth, etc. molecule of the set. Then, the whole superimposing cycle is repeated until the initial electron density at the grid points is unchanged. The iterative superimposing procedure is completed when all the $\overline{\rho}_{mi}$ become constant with an accepted accuracy, i.e., the difference between the total electron density values at all the lattice points $\left(\overline{P} = \sum_{i=1}^{N}\sum_{m=1}^{M}\overline{\rho}_{mi}\right)$ at the current and preceding steps becomes less than 0.1. Then, a more accurate orientation is fulfilled using ten times greater precise resolution (using a lattice with $\frac{d}{10}$ Å threshold).

Formula (3) for the evaluation of the structure superposition quality has been chosen in order to decrease the influence of the electron density at the lattice points located close to the atomic nuclei on the $\Delta\Phi_i$ value. It is evident that these zones determine the possibility of intermolecular interactions, e.g., between a molecule and a receptor, to a lesser degree, although the electron density is much greater than the electron density at distances comparable to the Van der Waals radii. For example, for the elements of the second period the electron density is about 10^2–10^3 e/Å^3 at the atomic centers, while for distances comparable to Van der Waals radii it is about $10^{-2} \div 10^{-3}$. Therefore the choice of the absolute difference used as the criterion $\left(\frac{1}{M-1}\sum_{m=1}^{M}(\bar{\rho}_{mi}-\rho_{mi})^2\right)$ would have led to inappropriately large addition of the electron density values of the zones close to the nuclei to the absolute difference.

The determined superimposition corresponds to the maximal coincidence of the most resembling fragments of

- where the index k is the number of the examined molecule in the dataset,
- ρ_{kt} is the lattice point whose electron density increases the P_k-value, t is its number (T is the total number of points whose electron density increases the P_k value),
- ρ_{ku} is the lattice point whose electron density decreases P_k-value, u is its number (U is the total number of points whose electron density decreases the P_k value),
- c_1, c_2, and c_3 are parameters.

The analysis of the determined quantitative model can help us find pharmacophore and antipharmacophore parts: the t points should be regarded as a pharmacophore part of the structure, while the u lattice junctions should be called an antipharmacophore part.

The advantages of the desirability function (Eq. 4) are the following:

- a small change of the argument leads to a large change in the function value, which permits the precise distinction of active and inactive molecules.
- the value area of the function is [0;1], therefore the function does not predict a quantitative value for bioactivity but gives a probability of activity (P_k). If $P_k > 0.5$ a molecule is considered to be active (class = 1), in other cases a molecule is considered inactive (class = 0). Thus, the method performs a categorical classification into active and inactive compounds. This is a radically new approach for solving the problem of pattern recognition of drugs.

The consideration of the interior of a molecule according to the characteristics calculated at the lattice point is

used in many well-known methods, e.g., in hypotetical active site lattice (HASL) and comparative molecular field analysis (CoMFA) [32–35]. However, in the ConGO algorithm the structures are considered on the basis of the calculated electron density agreeing closely with the results of highly accurate X-ray analysis, as shown in the next section. Therefore, the results of the superposition of the structures obtained with the ConGO algorithm are more easily interpretable compared to the results obtained by means of many other widely used methods. The algorithm permits the selection of the pharmacophore and antipharmacophore fragments of the molecular electron structure. Besides, this is the first time that the desirability function has been used to assign molecules to the active or inactive groups, which makes the ConGO algorithm absolutely different from current methods for pattern recognition of drugs.

A new approach to the evaluation of the molecular electron density distribution in the ConGO algorithm

It is well known that the study of the molecular electron structure requires the evaluation of the electron density distribution function in molecules. One of the variants for the solution of this problem is the use of quantum electron density functions. However, the electron density observed in low-temperature highly accurate X-ray analysis is in a poor agreement with the density attained in quantum approaches, as has been shown in a number of molecules. Thus, Fig. 4 gives the relation of the electron density logarithm, determined by low-temperature highly accurate X-ray analysis [36, 37], with the quantum electron density logarithm at the same spatial points, using the organic (ethoxy-(2-mercapto-4,6-dimethylhexahydropyrimidine-5-yl)-methanol—**I-1**) and the inorganic ($ZrCl_2$—**I-2**) compounds as examples. The calculations were carried out at quite a high level of theory: for the molecule **I-1** - DFT B3LYP/6-311G(d,p), for the molecule **I-2** Hartree-Fock computation (6-311G(d,p) for the chlorine atoms and 3-21G for the zirconium atoms).

When experimental and calculated values of the electron density are equal, the coefficients in the equation:

$$\lg\ \rho = b + k \cdot \lg\ \rho_{\text{quant}}$$

should be $b = 0$ and $k = 1$.

However, in Fig. 4 it can easily be seen that the coefficients b and k differ considerably from these values, indicating an error of computation of the electron density in close proximity to the atom nuclei (the calculated values are low). The spatial zones with high relative error in the electron density calculation $\left(\frac{|\rho_m - \rho_{m_{\text{quant}}}|}{\rho_m} \geq 10\right)$ are shown in Fig. 5. These zones exceed 20% of the molecular space. Figure 5 illustrates that the maximal error of the electron density calculation typical occurs for distances comparable

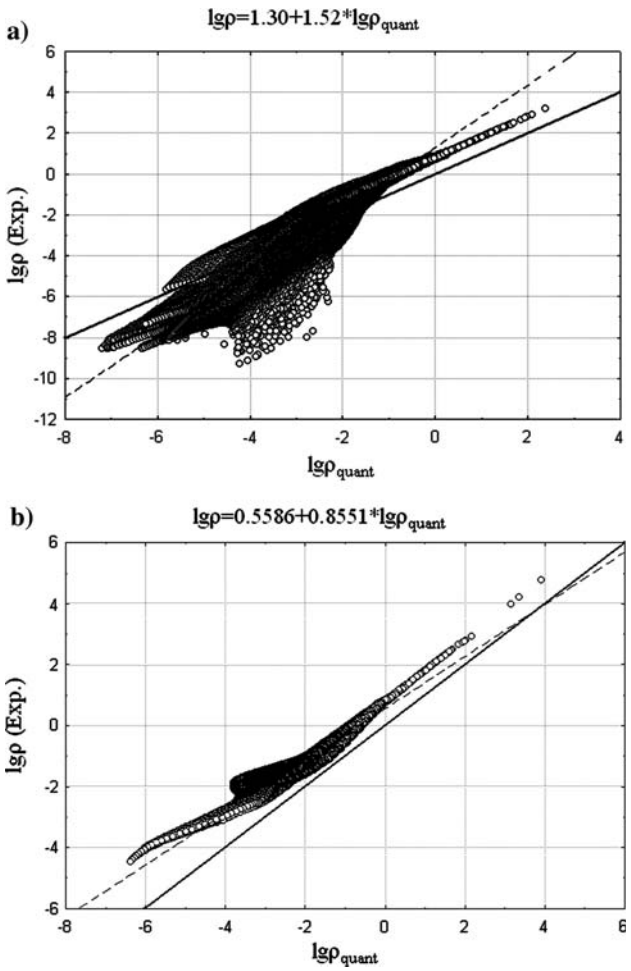


Fig. 4 Correlations of the logarithms of the electron density determined by low-temperature highly accurate X-ray analysis (lg ρ (exp.)) and computed by quantum methods (lg ρ_{quant}): (**a**) for the molecule **I-1;** (**b**) for the molecule **I-2**

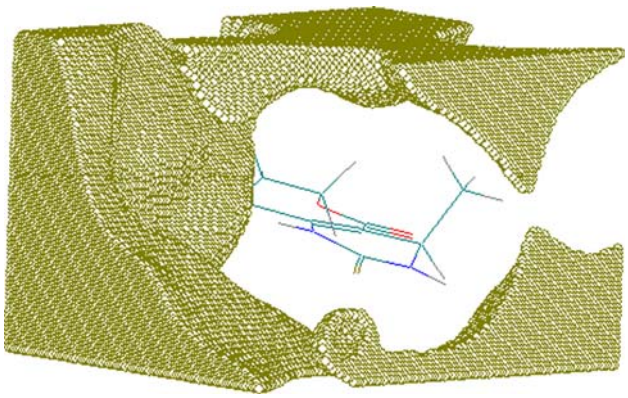


Fig. 5 The spatial zones of the high relative error of the electron density calculation within the quantum approach for the example of the **I-1** molecule

to or greater than the Van der Waals radii. It is quite clear that the electron density values in these zones determine the possibility of intermolecular interactions. Therefore, incorrectly calculated values of the electron density may result in incorrect prediction of the molecular association centers, reaction sites, etc.

For these reasons as well as to save the time it is suggested to represent the electron density function of an atom in a simpler form:

$$\rho_A = \sum_{i=1}^{n} a_{Ai} \cdot 10^{b_{Ai} \cdot R_A},$$

where a_{Ai} and b_{Ai} are parameters for each element of the periodic system, n is the period number of atom A, and R_A is the distance from the nucleus of the atom A.

This approach to the electron density evaluation is called AlteQ.

This function (Eq. 5) has been suggested on the basis of the relationship analysis between the experimental electron density logarithm (determined by highly accurate X-ray analysis) and the distance to the nuclei of the atoms. Examples of such relationships for H, O, and S are illustrated in Fig. 6. It was assumed that the lower part of the diagrams represents the logarithm of the atom electron density without taking into account their overlaps with the neighbor atoms. The higher electron density values correspond to overlapping of the electron density of a given atom with the electron density of neighboring atoms. It is possible to mark the areas on the diagrams with the number of the period number of the element in the periodic table. Therefore, the number of functions in Eq. 5 is chosen as equal to the period of the element in the periodic table.

The spatial integration of the electron density function should give the whole number of electrons in atom A (N_A), equal to the atomic number in the periodic table.

$$\int_0^{\pi} \sin\theta \cdot \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\varphi \int_0^{\infty} \rho_A(R) R^2 \mathrm{d}R = 4\pi \int_0^{\infty} \left(\sum_{i=1}^{n} a_{Ai} \cdot 10^{b_{Ai} R_A} \right) R^2 \mathrm{d}R = \sum_{i=1}^{n} \frac{8\pi \cdot a_{Ai}}{b_{Ai}^3 (\ln 10)^3} = N_A,$$

where a_{Ai} and b_{Ai}, the parameters of atom A, were estimated based on the fact that every exponential function i in Eq. 5 should describe the contribution of m_{Ai} electrons of the corresponding quantum level i. Therefore, the space integral of every exponential function expressed in spatial spherical coordinates (θ,φ,R) should be equal to the number of electrons m_{Ai} constituting the corresponding quantum level i.

$$4\pi \int_0^{\infty} a_{Ai} \cdot 10^{b_{Ai} R_A} R_A^2 \mathrm{d}R_A = m_{Ai}$$

Therefore, $\sum_{i=1}^{n} m_{Ai} = N_A$.

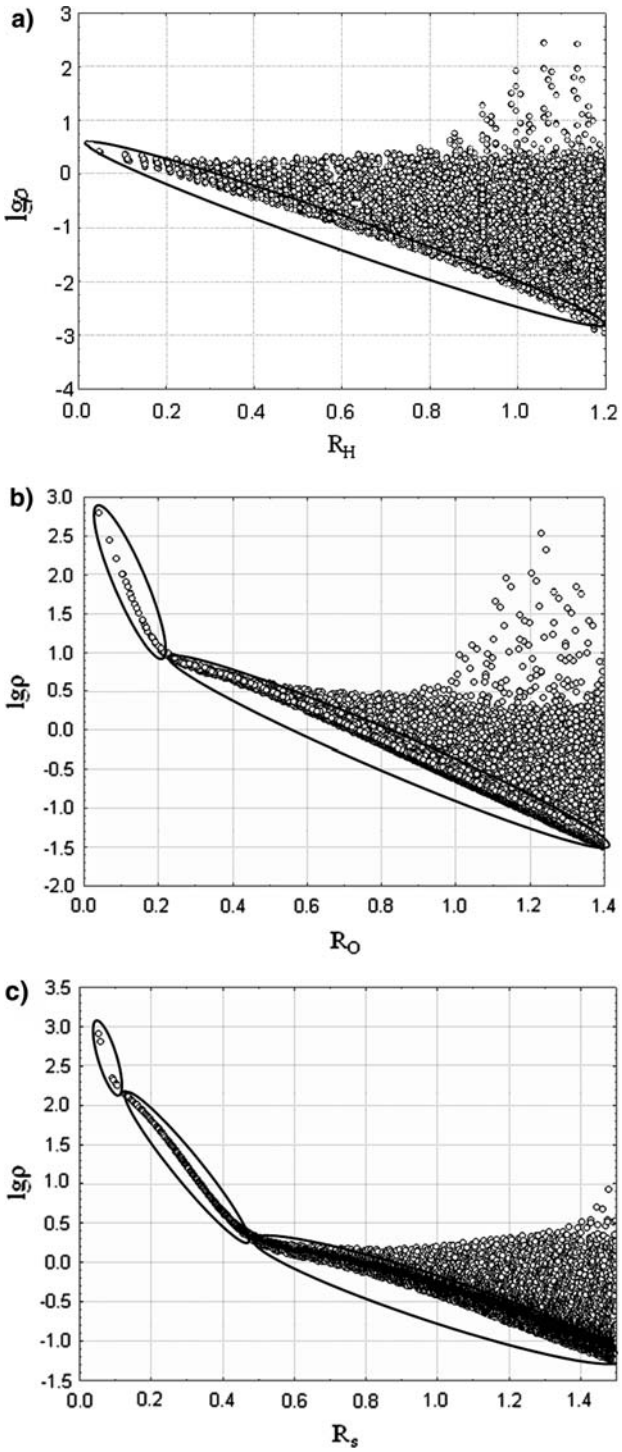


Fig. 6 The relationship between experimental electron density logarithm (determined by highly accurate X-ray analysis) and the distance to nuclei for: (**a**) hydrogen (R_H); (**b**) oxygen (R_O); and (**c**) sulfur (R_S). The regions of different shells are marked by ovals

For example, for sulfur the total number of functions in Eq. 5 is three ($n = 3$): the first exponential function ($i = 1$) describes the contribution of two electrons ($m_{(S)1} = 2$), the second exponential function ($i = 2$) describes the

contribution of eight electrons ($m_{(S)2} = 8$), and the third exponential function ($i = 3$) describes the contribution of six electrons ($m_{(S)3} = 6$).

$$N_S = m_{(S)1} + m_{(S)2} + m_{(S)3} = 16$$

$$4\pi \int_0^\infty a_{(S)1} \cdot 10^{b_{(S)1} R_{(S)}} R_{(S)}^2 \mathrm{d}R_{(S)} = \frac{8\pi a_{(S)1}}{b_{(S)1}^3 (\ln 10)^3} = 2$$

$$4\pi \int_0^\infty a_{(S)2} \cdot 10^{b_{(S)2} R_{(S)}} R_{(S)}^2 \mathrm{d}R_{(S)} = \frac{8\pi a_{(S)2}}{b_{(S)2}^3 (\ln 10)^3} = 8$$

$$4\pi \int_0^\infty a_{(S)3} \cdot 10^{b_{(S)3} R_{(S)}} R_{(S)}^2 \mathrm{d}R_{(S)} = \frac{8\pi a_{(S)3}}{b_{(S)3}^3 (\ln 10)^3} = 6$$

The parameters a_{Ai} and b_{Ai} for some elements are presented in Table 1. The parameters for H, C, N, O, S, and Cl were computed using highly accurate X-ray analysis of the compounds **I-1** and **I-2**. The parameters for He, Li, Be, B, F, Ne, and Br presented in Table 1 are supposed and require experimental accurate determination. Internal exponential functions for these atoms cannot be adequately predicted now, therefore, they are not presented in Table 1.

It was found that the atomic parameter of the highest quantum level ($i = n$) b_{An} divided by the radius of maximal electron density ($R_{\max}$) correlated with the ionization potential of atom A (I_A) (Table 2) with a correlation coefficient of $K = 0.93$ (Fisher's coefficient $F = 70$, standard error $S = 2.0$); see Fig. 7. This allows one to suppose that b_{An} is the effective nucleus charge with an opposite sign.

Therefore, the supposed expression (5) can be used for each atom of a molecule. Then, the total electron density of a molecule in the point m with coordinates x_m, y_m, z_m is the sum of the electron density functions determined by each atom. So, the electron density in the point m can be calculated using the following equation:

Table 2 Number of atom A in the periodic table (N_A), the ionization potential of the atom (I_A), the radius of maximal electron density (R_A), and the parameter b_{An} of the AlteQ external exponential function

N_A	I_A, eV	$R_{\max}$, Å	$\frac{b_{An}}{R_{\max}}$
1	13.599	0.5262	−3.5485
2	24.588	0.30953	−7.0278
3	5.392	1.55911	−0.98852
4	9.323	1.07938	−1.3730
5	8.298	0.80954	−1.9376
6	11.26	0.64763	−2.6659
7	14.534	0.53969	−3.5980
8	13.618	0.46259	−4.7857
9	17.423	0.40477	−6.2944
10	21.565	0.35979	−8.2056
16	10.36	0.86895	−1.7357
17	12.968	0.77636	−2.0973
35	11.84	1.10779	−2.5480

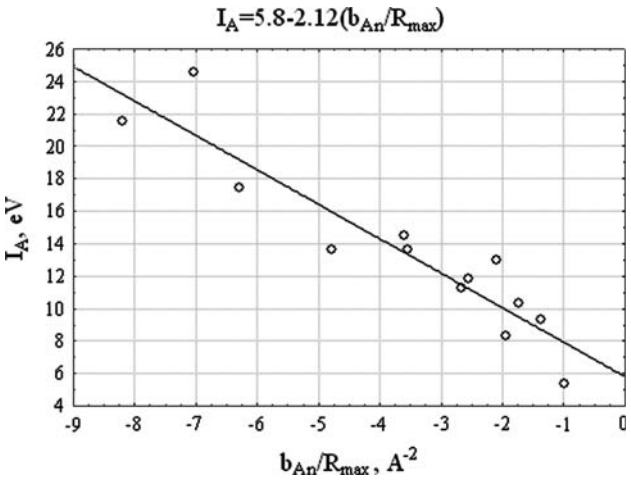


Fig. 7 The correlation between the ionization potential of atom A (I_A) and the AlteQ parameter b_{An} divided into the radius of the maximal electron density ($R_{\max}$)

Table 1 The parameters of the exponential functions used in AlteQ

N_A	A	a_{A1}, e/Å^3	b_{A1}, Å	a_{A2}, e/Å^3	b_{A2}, Å	a_{A3}, e/Å^3	b_{A3}, Å	a_{A4}, e/Å^3	b_{A4}, Å
1	H	3.1623	−1.8672						
2	He	10	−2.1753						
3	Li			1.7783	−1.5412				
4	Be			3.1623	−1.4820				
5	B			5.6234	−1.5685				
6	C	848.42	−9.5585	10.000	−1.7265				
7	N	1362.7	−11.194	17.783	−1.9418				
8	O	2083.8	−12.897	31.623	−2.2138				
9	F	2938.0	−14.461	56.234	−2.5478				
10	Ne			100.00	−2.9523				
16	S	16706	−25.812	1000.0	−6.3606	10.000	−1.5083		
17	Cl	18926	−26.908	1584.9	−7.4160	14.678	−1.6283		
35	Br	102530	−47.258	48338	−23.170	1000.0	−6.3606	133.35	−2.8226

$$\rho(x_m, y_m, z_m) = \sum_{A=1}^{N} \left(\sum_{i=1}^{n_A} a_{Ai} \cdot 10^{b_{Ai} \cdot R_{Am}} \right) \quad (6)$$

where $R_{Am} = \sqrt{(x_A - x_m)^2 + (y_A - y_m)^2 + (z_A - z_m)^2}$ is the distance from the center of the nucleus of an atom A to the point m, N is the total number of atoms in a molecule.

Figure 8 illustrates the comparison of the experimental electron density logarithm with the computed density logarithm (lg ρ_{comp}), evaluated by expression (6) with the compounds **I-1** and **I-2** as examples. It was found that lg ρ_{comp} agrees well with the experimental value (lg ρ). The areas that are characterized by the greatest discrepancy between calculated and experimental values (the relative error of electron density calculation, $\frac{|\rho_m - \rho_{m_{quant}}|}{\rho_m} \geq 10$) occupy approximately 0.4% of molecular space (Fig. 9), which significantly exceeds the quality of the quantum computation results (cf. Figs. 5 and 9). The figures clearly

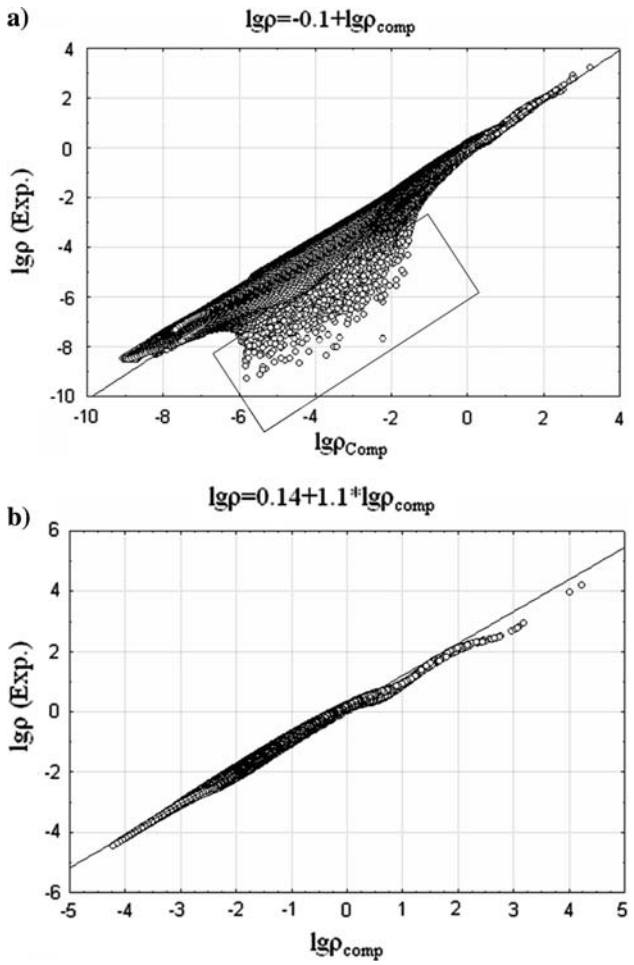


Fig. 8 The correlations of logarithms of electron density determined by low-temperature highly accurate X-ray analysis (lg ρ (exp.)) and computed within AlteQ (lg ρ_{comp}): (**a**) for molecule **I-1**; (**b**) for molecule **I-2**

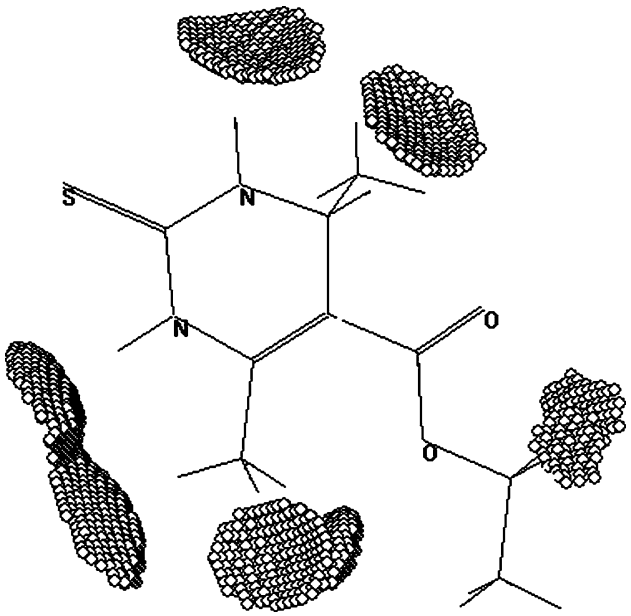


Fig. 9 The regions that are characterized by the greatest discrepancy between calculated (AlteQ) and experimental values of the electron density for **I-1** (the relative error of the electron density calculation, $\frac{|\rho_m - \rho_{m_{quant}}|}{\rho_m} \geq 10$)

illustrate that the greatest difference between the calculated and the experimental value of the electron density is observed for the points which are near hydrogen atoms at distances exceeding the Van der Waals radius. Moreover, when this method is used, the coefficients b and k in the equation describing the relationship between the experimental and the computed electron density logarithm are equal to -0.1 and 1, respectively:

$$\lg \rho = b + k \cdot \lg \rho_{comp},$$

which proves a good correlation of absolute densities.

The points corresponding to the relative difference of the logarithms of the calculated and experimental electron densities $\frac{|\rho_m - \rho_{m_{comp}}|}{\rho_m} \geq 10$ are marked for the molecule **I-1** by a rectangle in Fig. 8a.

The reason for the considerable deviation is that the description of electron density with the use of the expression (6) does not take into account atomic vibrations occurring even at the temperature of highly accurate X-ray experiments (-70°C). In the case of hydrogens with low electron density, the vibrations of the atoms influence the ρ value greatly. Another reason for the deviation of the calculated value from the experimental one might be the error which occurs in the subtraction of the contribution of neighboring molecules in the crystal from the total electron density value. Besides, one should take into account that in the framework of the multipole model formalism the interpretation of reflexes on an experimental X-ray diffraction pattern is based on Slater-type quantum functions.

Thus, it has been shown that AlteQ provides a satisfactory description of the molecular electron structure. It uses the exponential functions that were suggested many years ago, e.g., in the methods of quantum chemistry. It is well known that $\phi = A \cdot 10^{-BR}$ is the Slater-type wavefunction of the atoms of the first period [38]. Moreover, exponential functions have been used for description of the electron or matter densities in some other works [e.g., 39, 40]. However, none of the existing approaches has compared the computational results of the density with the experimental high-accuracy X-ray day. Therefore, their computational results are poorly interpretable. In this work we are the first to suggest an approach which adequately evaluates the real electron density. Therefore, the AlteQ computations are interpretable very well: the method gives the real electron density. Moreover, the parameters of the AlteQ functions are not dependent on the atom neighborhood (for example, the electron density function for carbon is the same for the carbon atoms of CO, COOH, CH_3, CN, etc. groups). That makes the method absolutely different from the classical quantum-chemical concept and from the other methods for electron density evaluation [38–40].

The results of the pattern recognition of drug electron structures using the suggested paradigm

Consequently, analysis of anti-tumor DNA-antimetabolites, fXa inhibitors, inhibitors of 5-HT_{1A}, and α_1-AR receptors was carried out using the suggested paradigm, which includes:

(1) the multiconformational representation of the molecules, the search for the most active conformers using 3D/4D QSAR algorithm BiS
(2) the analysis of AlteQ electron density within 3D QSAR algorithm—ConGO, including:
   - superimposing of the conformers
   - pattern recognition of drug electron structures
   - search for pharmacophore and antipharmacophore fragments

The study of the drugs was done using a lattice with 2 Å resolution, subsequently improved to 0.2 Å. It has been shown that for all of the considered sets the obtained cross-validation quality Q^2 of the models describing the probability of activity P(Exp.) is extremely close to 1.0 (Q^2 obtained for DNA-antimetabolites is 0.996, for fXa inhibitors −0.992, for 5-HT_{1A} inhibitors −0.999, and for α_1-AR inhibitors −0.998). Moreover, in all the cases P(Exp.) concurs with the predicted class of compounds (Class). Therefore, the suggested paradigm provides the pattern recognition of drugs with a quality level of 1.00.

Table 3 The considered DNA-antimetabolites, the experimental probability of the activity P(Exp.) [P(Exp.) = 1 for the compounds with $pGI_{50\%} > 5.0$ and P(Exp.) = 0 for the compounds with $pGI_{50\%} \leq 5.0$], the predicted probability of the activity, P(Pred.), and the predicted class of the compounds

Drug	P(Exp.)	P(Pred.)	Class
1d35[a]	1	1.00	1
1d37[a]	1	1.00	1
1ims[a]	1	1.00	1
NSC 752[b]	1	0.96	1
NSC 755	1	1.00	1
NSC 1895	0	0.00	0
NSC 2764	1	1.00	1
NSC 32065	0	0.00	0
NSC 51143	0	0.01	0
NSC 663878	0	0.00	0
NSC 71261	1	0.95	1
NSC 71851	0	0.02	0
NSC 95678	1	0.97	1
NSC 107392	0	0.00	0
NSC 118994	0	0.02	0
NSC 127716	0	0.00	0
NSC 145668	0	0.10	0
NSC 303812	0	0.00	0
NSC 330500	1	1.00	1

[a] The names of molecules are taken from the Protein Data Bank [27]

[b] Hereafter in the Table the numbers of the molecules correspond to the numbers of the National Cancer Institute http://dtp.nci.nih.gov/docs/3d_database/Structural_information/structural_data.html

The study of DNA-antimetabolites

The DNA-antimetabolites [41, 42] presented in Table 3 were considered by using the described approach. The investigation of the molecules within BiS/MC included not only the multiconformational search but also multitautomeric analysis because many compounds of the set can exist in different tautomeric forms. The conformational 'n' tautomeric forms interacting with the receptor were found within BiS/MC approach. Then, the found structures were studied with the ConGO algorithm.

The superimposed structures as well as the lattice junctions are given in Fig. 10. The analysis of the drugs shows that, in the case of the most active molecules, the antipharmacophore fragments of the electron structure are, as a rule, located at distances exceeding the Van der Waals radii of atoms. The electron density is extremely low at such distances. Therefore, its negative influence on the activity level of high active compounds is reduced to a minimum. The points determining the antipharmacophore effect can be mostly attributed to methylene fragments of the structures. Probably these fragments of the molecules

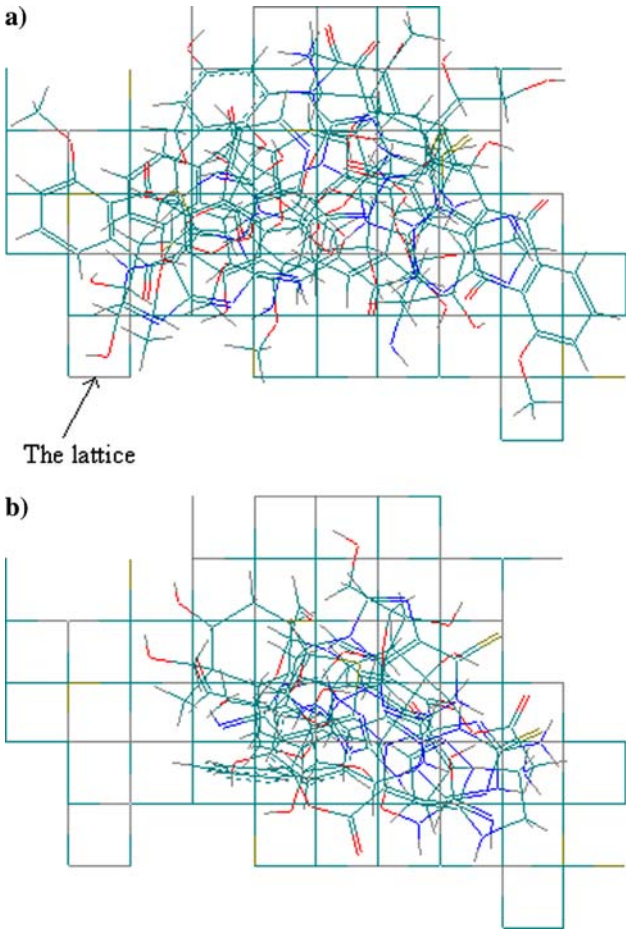


Fig. 10 The superimposed set of DNA-antimetabolites (the superimposition corresponds to the best coincidence of the electron density values in the lattice point shown)

increase the structure volume, interfering with their intercalation into the receptor (DNA). Moreover, DNA is noticeably polar, so its interaction with lipophilic methylene fragments of DNA-antimetabolites can be ineffective. The points that determine the pharmacophore fragments are located mostly in the areas of the aromatic rings, as well as oxygen atoms of oxymethyl, hydroxyl, carbonyl groups (Fig. 11a). The reason for this is that the aromatic system of the DNA-antimetabolite molecules allows them to fit between two aromatic systems of DNA base pairs, causing effective stacking interactions with them [26] (Fig. 11b). Oxymethyl, hydroxyl, and carbonyl groups produce a net of hydrogen bonds and electrostatic interactions with the polar fragments of the DNA structure. Inactive compounds are characterized by a smaller volume compared to the active structures. Therefore, most of the points which electron densities provide the activity of the compounds are located out of the Van der Waals spheres of the atoms (Fig. 11b). In the case of inactive molecules the small volume may prevent the effective interactions with the DNA.

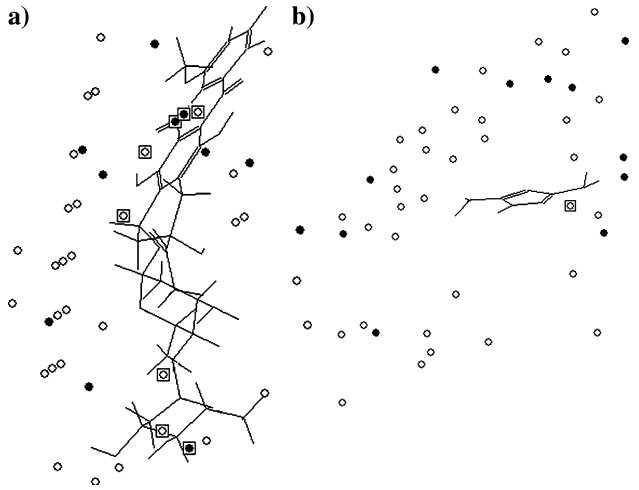


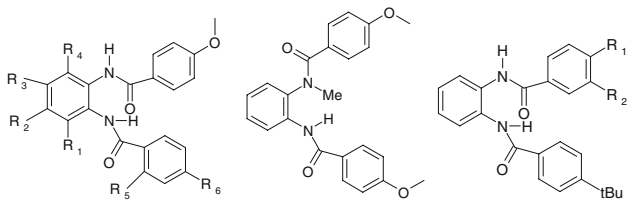
Fig. 11 The junctions attributed to the pharmacophore (marked by black circles ●) and antipharmacophore (marked by white circles ○) parts of the DNA-antimetabolites: (**a**) for active compound (**1d35**[2]); (**b**) for inactive compound (**NSC 1895**; the number of the molecule correspond to the number of the National Cancer Institute http://dtp.nci.nih.gov/docs/3d_database/Structural_information/structural_data.html). The points located within the Van der Waals spheres of the nearest atoms are marked by squares

It has been proved that the cross-validation Q^2 of the model obtained for DNA-antimetabolites using the suggested approach is equal to 0.996. The active compounds of the set have been recognized with quality equal to 1.0. The predicted values of the probability of the activity, P(Pred.), and the found class are presented in Table 3.

The study of fXa inhibitors

The analysis of fXa inhibitors (Table 4) [43] with the ConGO algorithm shows significant superimposition of the most active molecules due to closely similar electron structures (Fig. 12a). Less active molecules are characterized by considerable differences in the electron structures, preventing good structure superimposition (Fig. 12b). Therefore, the reasons of the deactivation of the molecules are determined by various factors. One of them can be molecular geometry that is inconvenient for adsorption, causing anticoagulative activity reduction and therefore decreasing factor Xa inhibition [43, 44]. Indeed, for a number of the less active molecules in comparison to the active ones mostly orthogonal orientation of aromatic systems is evident (cf. Fig. 13a, b). The described phenomenon leads to a decrease in adsorption ability. Another factor possibly causing the deactivation of the substances is inadequate polarity of a molecule, which results in lower solubility of the compound and ability to form stable hydrogen-bonded complexes during absorption on receptors, etc. The absence of polar groups is typical for many inactive compounds.

Table 4 fXa inhibitors [43], experimental probability of the activity P(Exp.) [P(Exp.) has been accepted as equal to 1 for compounds with $pIC_{50\%} > 6.8$ and equal to 0 for compounds with $pGI_{50\%} \leq 6.8$], the predicted probability of the activity, P(Pred.), and the predicted class of the compounds



Com.	R_1	R_2	R_3	R_4	R_5	R_6	P(Exp.)	P(Pred.)	Class
II-1	H	OMe	H	H	H	OMe	0	0.00	0
II-2	H	H	H	H	H	OMe	0	0.01	0
II-3	OMe	H	H	OMe	H	OMe	0	0.00	0
II-4	H	OH	H	H	H	OMe	0	0.00	0
II-5	H	Me	H	H	H	OMe	0	0.00	0
II-7	OH	H	H	H	H	OMe	0	0.00	0
II-8	H	Cl	H	H	H	OMe	0	0.02	0
II-9	H	CO_2H	H	H	H	OMe	0	0.00	0
II-10	–	–	–	–	–	–	0	0.01	0
II-12	H	OH	H	H	H	NMe_2	1	0.99	1
II-13	H	H	$MeSO_2NH$	H	H	t-Bu	1	1.00	1
II-16	H	CO_2H	H	H	H	t-Bu	1	0.82	1
II-21	H	$MeSO_2NH$	H	H	H	t-Bu	1	1.00	1
II-22	H	H	H	H	H	t-Bu	1	0.98	1
II-23	H	OH	H	H	H	t-Bu	1	1.00	1
II-24	H	H	H	H	H	i-Pr	0	0.00	0
II-27	H	H	CO_2H	H	H	t-Bu	0	0.05	0
II-30	H	H	OH	H	H	t-Bu	0	0.20	0
II-32	H	OH	H	H	H	Et	0	0.03	0
II-33	H	H	H	H	NMe_2	H	1	0.96	1
II-36	H	H	H	H	CMe_2OMe	H	0	0.05	0
II-37	H	H	H	H	SMe	H	0	0.00	0
II-38	H	H	H	H	O-i-Pr	H	0	0.02	0
II-39	H	H	H	H	Ph	H	0	0.01	0
II-40	H	H	H	H	O-t-Bu	H	0	0.00	0
II-41	H	H	H	H	OMe	Cl	0	0.08	0
II-42	H	H	H	H	SO_2NMe_2	H	0	0.04	0
II-43	H	H	H	H	OEt	H	0	0.02	0
II-44	H	H	H	H	SO_2Me	H	0	0.02	0
II-45	H	H	H	H	OMe	OMe	0	0.01	0
II-46	H	H	H	H	OPr	H	0	0.01	0
II-47	H	H	H	H	SOMe	H	0	0.03	0
II-48	H	H	H	H	OPh	H	0	0.00	0
II-50	H	H	H	H	OBu	H	0	0.01	0
II-51	H	H	H	H	$COCF_3$	H	0	0.00	0
II-53	$CHCH_2$	H	–	–	–	–	0	0.03	0
II-54	Cl	H	–	–	–	–	0	0.00	0
II-55	H	$CHCH_2$	–	–	–	–	0	0.00	0
II-56	F	H	–	–	–	–	0	0.03	0

Table 4 continued

Com.	R_1	R_2	R_3	R_4	R_5	R_6	P(Exp.)	P(Pred.)	Class
II-57	H	F	–	–	–	–	0	0.00	0
II-60	H	H	–	–	–	–	0	0.00	0
II-61	CH_3	H	–	–	–	–	0	0.00	0
II-62	H	OCH_3	–	–	–	–	0	0.00	0
II-63	Et	H	–	–	–	–	0	0.04	0

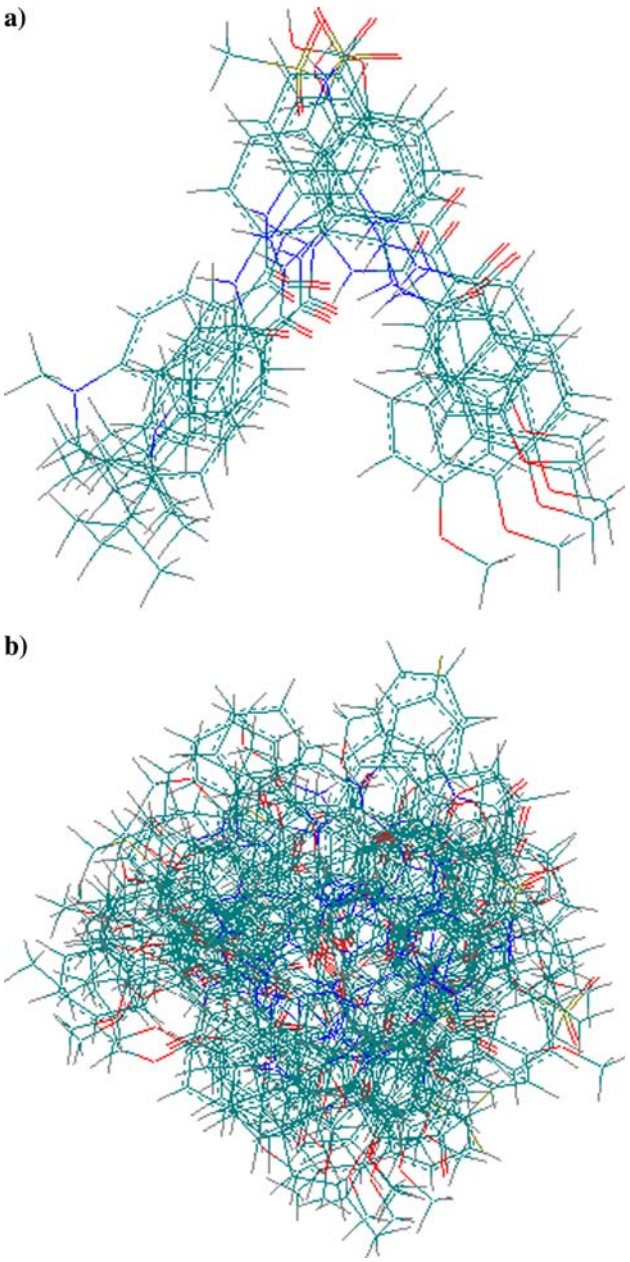


Fig. 12 The superimposed fXa inhibitors: (**a**) active molecules; (**b**) inactive molecules

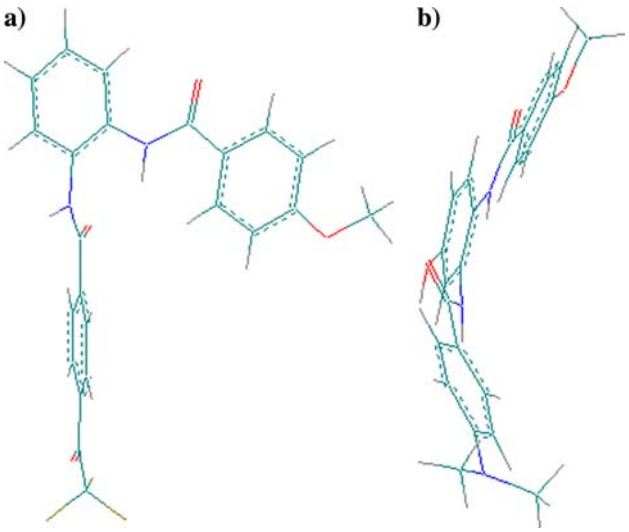


Fig. 13 fXa inhibitors: geometry of an inactive compound (**II-51**) (**a**) and an active one (**II-12**) (**b**)

These assumptions are proved by the pharmacophore and antipharmacophore fragments in the electron structures of the fXa inhibitors. It was found that the most active molecules are characterized by an insignificant antipharmacophore part located at the methyl of the OMe and NMe_2 groups. For the less active structures the antipharmacophore fragments are present in aromatic rings, oxymethyl and NH fragments, etc. and occupy a significant part of the electron structure.

It was found that the cross-validation Q^2 value of the obtained quantitative model, which was carried out for the fXa inhibitors using the suggested approach, was 0.992. The recognition quality was 1.0. The predicted probability of activity, P(Pred.), and the obtained class are presented in Table 4.

The study of inhibitors of 5-HT_{1A} and α_1-AR receptors

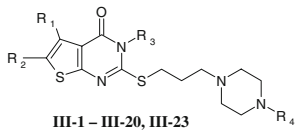
One of the most important problems in modern research is the selectivity of a drug's action with respect to the desired target. Thus, experimental data from the paper [45] are

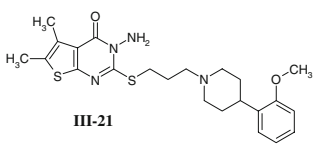
presented for the activity of a number of substances with respect to 5-HT_{1A} and α_1-AR receptors (Table 5). The goal of investigations such as this is to find structural features that may determine the selectivity of molecules upon 5-HT_{1A} in the presence of the α_1-AR receptor. In the paper [45] it was shown that the differences in $pIC_{50\%}$ values in regard to 5-HT_{1A} and α_1-AR receptors are insignificant as a rule. Therefore, in the present study we made a detailed analysis of the electron structure for the described compounds and a search for significant differences in the compound structures that determine their preferential binding with the 5-HT_{1A} receptor compared to the α_1-AR receptor. It was found that the cross-validation quality of the model for the 5-HT_{1A} inhibitors was 0.999 and for α_1-AR inhibitors it was 0.998. The recognition quality of the active and inactive compounds was 1.0 for both receptors (Table 6).

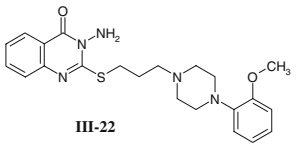
It was shown that the binding with 5-HT_{1A} and α_1-AR receptors is determined by different conformers of compounds in approximately 50% of cases, e.g., in the case of **III-03** the mirror-image conformers (enantioconformers) are the most active in relation to 5-HT_{1A} and α_1-AR receptors (Fig. 14). As for the other 50%, the same conformers are active in the both receptors.

At the same time preferential action of the highly active molecules was revealed for both receptors, concerning the conformers characterized by the compact geometry (Fig. 15a). This allows their fitting into a receptor cavity without any problem. The binding with these receptors for inactive compounds is determined by conformers with the

Table 5 Inhibitors of 5-HT_{1A} and α_1-AR receptors [45]







No.	R_1	R_2	R_3	R_4
III-1	Me	Me	H	2-ClPhenyl
III-2	Me	Me	H	3-ClPhenyl
III-3	Me	Me	H	2-OMePhenyl
III-4	Me	Me	H	1-Naphthyl
III-5	Me	Me	H	2-Pyrimidyl
III-6	$-(CH_2)_4-$		H	2-ClPhenyl
III-7	$-(CH_2)_4-$		H	2-OMePhenyl
III-8	$-(CH_2)_4-$		H	1-Naphthyl
III-9	$-(CH_2)_4-$		H	2-Pyrimidyl
III-10	H	Ph	H	2-OMePhenyl
III-11	H	Ph	H	1-Naphthyl
III-12	–(CH=CH)–		H	2-OMePhenyl
III-13	H	H	NH_2	2-OMePhenyl
III-14	$-(CH_2)_4-$		Me	2-OMePhenyl
III-15	$-(CH_2)_4-$		NH_2	2-OMePhenyl
III-16	Me	Me	NH_2	Ph
III-17	Me	Me	Me	2-OMePhenyl
III-18	Me	Me	NHPhenyl	2-OMePhenyl
III-19	Me	Me	Me	2-Pyrimidyl
III-20	Me	Me	NH_2	2-Pyrimidyl
III-21	–	–	–	–
III-22	–	–	–	–
III-23	Me	Me	NH_2	2-OMePhenyl

Table 6 Experimental probabilities of the activities P(Exp.) [for 5-HT_{1A} inhibitors P(Exp.) = 1 if $pIC_{50\%} > 7.62$ and P(Exp.) = 0 if $pIC_{50\%} \leq 7.62$; for inhibitors of α_1-AR receptor P(Exp.) = 1 if $pIC_{50\%} \geq 7.57$ and P(Exp.) = 0 if $pGI_{50\%} < 7.57$], the predicted probabilities of the activities, P(Pred.), and the predicted classes of the compounds

Drug	$P(Exp.)_{5\text{-}HT1A}$	$P(Pred.)_{5\text{-}HT1A}$	$Class_{5\text{-}HT1A}$	$P(Exp.)_{\alpha1\text{-}AR}$	$P(Pred.)_{\alpha1\text{-}AR}$	$Class_{\alpha1\text{-}AR}$
III-1	0	0.03	0	0	0.00	0
III-2	0	0.00	0	0	0.05	0
III-3	0	0.00	0	0	0.00	0
III-4	0	0.00	0	0	0.00	0
III-5	0	0.01	0	0	0.00	0
III-6	0	0.03	0	0	0.00	0
III-7	0	0.01	0	0	0.00	0
III-8	0	0.00	0	0	0.01	0
III-9	0	0.00	0	0	0.00	0
III-10	0	0.01	0	0	0.00	0
III-11	0	0.00	0	0	0.01	0
III-12	0	0.02	0	0	0.00	0
III-13	1	1.00	1	1	1.00	1
III-14	1	1.00	1	0	0.00	0
III-15	1	1.00	1	0	0.02	0
III-16	1	1.00	1	0	0.00	0
III-17	1	1.00	1	1	1.00	1
III-18	0	0.00	0	0	0.00	0
III-19	0	0.03	0	0	0.01	0
III-20	1	0.95	1	0	0.00	0
III-21	1	1.00	1	0	0.00	0
III-22	1	1.00	1	1	0.98	1
III-23	1	1.00	1	1	1.00	1

expanded structure (Fig. 15b). This can prevent their fitting into the 5-HT_{1A} and α_1-AR receptor cavities.

The investigation of differences in the inhibition process for the 5-HT_{1A} and α_1-AR receptors showed that there is a significant difference in the antipharmacophore parts of the molecules with respect to these receptors. Thus, for the example of **III-11** compound, a dotted line marks the antipharmacophore fragments with respect to the 5-HT_{1A} receptor and an undotted line marks the same for the α_1-AR receptor (Fig. 15b).

The greatest difference in $pIC_{50\%}$ values during inhibition of 5-HT_{1A} and α_1-AR receptors is observed for the **III-20** compound, which is why this compound was examined as the reference substance for recommendations of the selective action increase for the 5-HT_{1A} receptor. It was found that the whole structure of the **III-20** compound has a positive influence upon the action for the given receptor (i.e., absence of antipharmacophore fragments), while the interaction with the α_1-AR receptor results in antipharmacophore fragments (Fig. 16). Therefore, in order to decrease the activity of the given compound concerning the α_1-AR receptor, it is necessary to introduce heavier atoms into the described areas, e.g., to change the hydrogen atoms marked with crosses into halogen atoms. This change will strengthen the electron density in the antipharmacophore part of the compound concerning the α_1-AR receptor, at the same time decreasing the probability of activity to the given receptor. Besides, similar changes must lead to the inhibition of the 5-HT_{1A} receptor in a positive way, since they will reinforce the pharmacophore part of a molecule for the 5-HT_{1A} receptor.

Conclusion

In the present study a new paradigm based upon the combined application of the 4D/3D QSAR algorithms, BiS and ConGO, has been suggested for pattern recognition of drugs. The first algorithm, BiS/MC (multiconformational), is used for the search for conformers binding with a receptor, while the second algorithm, ConGO, is applied for the study of the electron structure of the conformers, as well as for defining the pharmacophore and antipharmacophore parts of a molecule for pattern recognition of biologically active compounds. A new AlteQ method has been suggested for the evaluation of the molecular electron

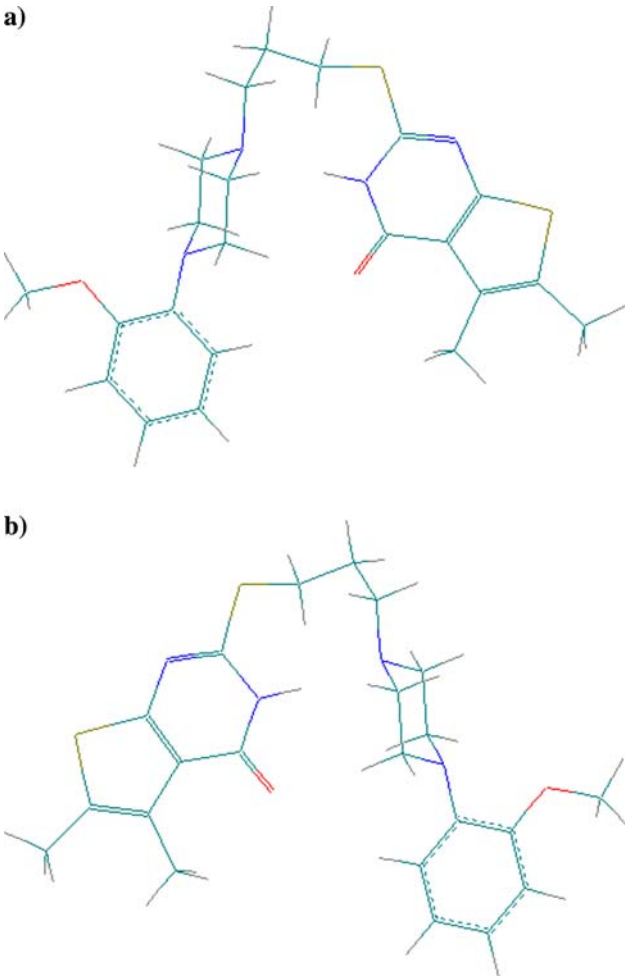


Fig. 14 Enantioconformers of **III-03** (the conformers of **III-03** are mirror antipodes characterizing by the same total energy): one of them is selected by BiS/MC as the most active to the α_1-AR receptor (**a**), another is selected by BiS/MC as the most active to the 5-HT_{1A} receptor (**b**)

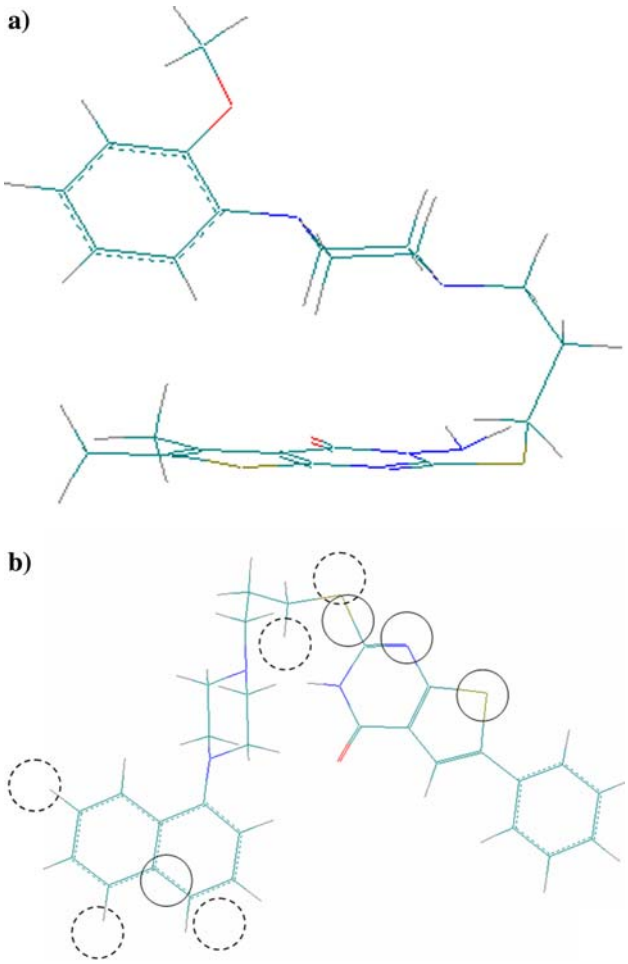


Fig. 15 The conformers that determine binding with the 5-HT_{1A} and α_1-AR receptors: (**a**) for the active compound (**III-23**); (**b**) for the inactive compound (**III-11**) (antipharmacophore parts of **III-11** with respect to 5-HT_{1A} are marked by dotted circles; antipharmacophore parts with respect to α_1-AR are marked by undotted circles)

density. It has been shown that the described method noticeably surpasses a number of widely used quantum methods in terms of the quality of the electron density evaluation. In the present study this was demonstrated by means of a comparison of the computed electron density with the results of high-accuracy X-ray analysis. Additionally, the desirability function was used for the first time for the analysis of the effects of the electron structure for pattern recognition of active and inactive compounds. The quality of the suggested approach has been demonstrated in datasets of DNA-antimetabolites, fXa inhibitors, 5-HT_{1A}, and α_1-AR receptors inhibitors. The conformers binding with the receptors have been found in all the examples. Pharmacophore and antipharmacophore fragments have been defined in the electron structure. It has been shown that the pattern recognition cross-validation quality for all of the datasets is equal to 1.0.

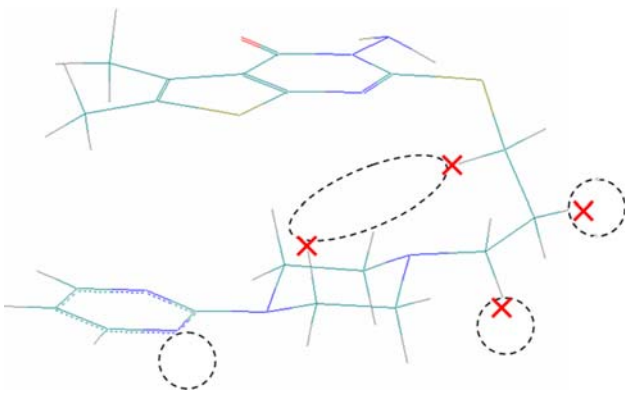


Fig. 16 Antipharmacophore fragments of **III-20** for the α_1-AR receptor. The hydrogen atoms, which can be changed by heavier atoms (for example, halogen) in order to strengthen the antipharmacophore part, are marked by crosses

Acknowledgements The work is supported by RFBR (grants 07-03-96041, 07-04-96053) and Human Capital Foundation. Besides, we are thankful to Tsirelson V.G. and Stash A.I. for the low-temperature highly accurate X-ray results for a number of molecules.

References

1. Khalil M, Weaver DF (1990) J Pharm Pharmacol 42:349
2. Weaver DF, Knight JL (1998) Seizure 7:347
3. Khlebnikov AI (1997) Pharm Chem J 31:147
4. Kasabov N (2007) Pattern Recognit Lett 28:673
5. Bartzatt R (2005) Eur J Pharm Biopharm 59:63
6. Keith Davies E, Glick M, Harrison KN, Graham Richards W (2002) Inc J Comput Chem 23:1544
7. Kowalski BR, Bender CF (1975) Naturwissenschaften 62:10
8. Bartzatt R, Donigan L (2006) AAPS PharmSciTech 7:E35
9. Tetko IV, Aksenova TI, Patiokha AA, Villa AEP, Welsh WJ, Zielinski WL, Livingstone DJ (1999) Anal Chem 71:2431
10. Potashnikov PF, Fetisov VI, Ezhov VV, Dan'shin BI, Sokol'skii GA (1980) Pharm Chem J 14:61
11. Potemkin VA, Bartashevich EV, Grishina MA, Guccione S (2001) In: Holtje H-D, Sippl W (eds) Rational approaches to drug design. Proceedings of the 13th European symposium on quantitative structure-activity relationships, QSAR 2000. Prous Science Publishers, Barcelona, Spain, pp 349–353
12. Potemkin VA, Grishina MA, Bartashevich EV (2007) J Struct Chem 48:155
13. Potemkin VA, Arslambekov RM, Bartashevich EV, Grishina MA, Belik AV, Perspicace S, Guccione S (2002) J Struct Chem 43:1045
14. http://www.modelchem.ru
15. Grishina MA, Potemkin VA (2007) Drugs Future 32(Suppl A): 112
16. Pereyaslavskaya ES, Potemkin VA, Grishina MA, Bartashevich EV (2007) Drugs Future 32(A):87
17. Potemkin VA, Grishina MA, Fedorova OV, Rusinov GL, Ovchinnikova IG, Ishmetova RI (2003) Pharm Chem J 37:468
18. Grishina MA, Pogrebnoi AA, Potemkin VA, Zrakova TYu (2005) Pharm Chem J 39:509
19. Potemkin VA, Grishina MA, Belik AV, Chupakhin ON (2002) Pharm Chem J 36:22
20. Grishina MA, Potemkin VA, Rusinov GL, Bartashevich EV, Guccione S, Perspicace S, Chupakhin ON (2002) In: From genes to drugs via crystallography, 33rd crystallographic course at the E. Majorana Centre, Erice, Italy, 23 May–2 June
21. Mikuchina K, Potemkin V, Grishina M, Laufer S (2002) Arch Pharm Pharm Med Chem 335:C74
22. Grishina MA, Potemkin VA (2003) In: IXth International seminar on inclusion compounds (ISIC-9), Novosibirsk, Russia, 23–27 June
23. Grishina MA, Pogrebnoi AA, Potemkin VA, Zrakova TYu (2005) Pharm Chem J 39:509
24. Grishina MA, Mikushina KM, Potemkin VA (2006) In: The 16th European symposium on quantitative structure-activity relationships and molecular modelling, Mediterranean Sea, Italy, 10–17 September
25. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235
26. Gao Y-G, Wang AH-J (1991) Anticancer Drug Des 6:137
27. Wang Z, Canagarajah BJ, Boehm JC, Kassisi S, Corb MH, Young PR, Abdel- Meguid S, Adams JL, Goldsmith EJ (1998) Structure 6:1117
28. Bartashevich EV, Potemkin VA, Grishina MA, Belik AV (2002) J Struct Chem 43:1033
29. Potemkin VA, Bartashevich EV, Belik AV (1996) Russ J Phys Chem A 70:411
30. Harrington EC (1965) Ind Qual Control 21:494
31. Hsieh K-L, Tong L-I, Chiu H-P, Yeh H-Y (2005) Comput Ind Eng 49:556
32. Doweyko AM (1988) J Med Chem 31:1396
33. Doweyko AM (1991) J Math Chem 7:273
34. Doweyko AM (1994) J Med Chem 37:1769
35. Kaminski JJ, Doweyko AM (1997) J Med Chem 40:427
36. Tsirelson VG, Stash AI, Potemkin VA, Rykounov AA, Shutalev AD, Zhurova EA, Zhurov VV, Pinkerton AA, Gurskaya GV, Zavodnik VE (2006) Acta Crystallogr B 62:676
37. Tsirelson V, Stash A, Zhurova E, Zhurov V, Pinkerton AA, Zavodnik V, Shutalev A, Gurskaya G, Rykounov A, Potemkin V (2005) Acta Crystallogr A 61:C427
38. Clark T (1985) Handbook of computational chemistry; a practical guide to chemical structure and energy calculations. Wiley-Interscience, 352 pp
39. Grant JA, Pickup BT (1995) J Phys Chem 99:3503
40. Vaz RJ (1997) Quant Struct Act Relat 16:303
41. Weinstein et al (1992) Science 258:447
42. van Osdol et al (1994) J Natl Cancer Inst 86:1853
43. Herron D, Goodson Th, Wiley MR, Weir LC, Kyle JA, Yee YK, Tebbe AL, Tinsley JM, Mendel D, Masters JJ, Franciskovich JB, Sawyer JS, Beight DW, Ratz AM, Milot G, Hall SE, Klimkowski VJ, Wikel JH, Eastwood BJ, Towner RD, Gifford-Moore DS, Craft TJ, Smith JF (2000) J Med Chem 43:859
44. Yee YK, Tebbe AL, Linebarger JH, Beight DW, Craft TJ, Gifford-Moore DS, Goodson Th, Herron D, Klimkowski VJ, Kyle JA, Sawyer JS, Smith JF, Tinsley JM, Towner RD, Weir LC, Wiley MR (2000) J Med Chem 43:873
45. Guccione S, Doweyko AM, Chen H, Barretta GU, Balzano F (2000) J Comput Aided Mol Des 14:647