

# Technique for Energy Decomposition in the Study of “Receptor-Ligand” Complexes

Vladimir A. Potemkin, Alexander A. Pogrebnoy, and Maria A. Grishina\*

Chelyabinsk State Medical Academy, Pharmaceutical Chemistry, Vorovskiy St., 64, Chelyabinsk,  
Russian Federation 454048

Received November 6, 2008

A new methodology to describe the interactions in “receptor-ligand” complexes is presented. The methodology is based on a combination of the 3D/4D QSAR BiS/MC and CoCon algorithms. The first algorithm performs the restricted docking of compounds to receptor pockets. The second determines the relationships between the bioactivity and the parameters of interactions in the “receptor-ligand” complexes, including a new formalism for estimating hydrogen bond energies.

## 1. INTRODUCTION

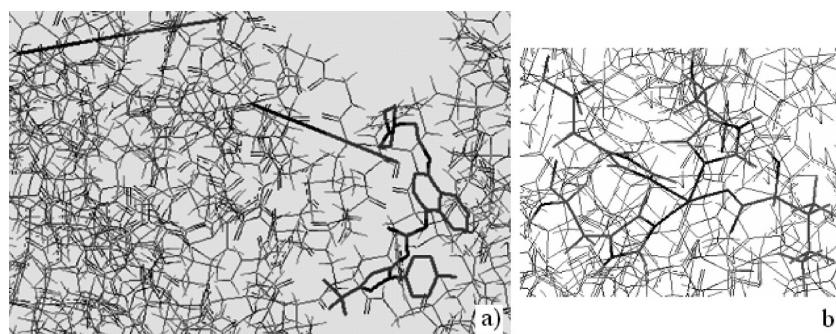
The obligatory stage of the biological action of compounds is their interaction with the target. The highest level of investigation of the biologically active compounds, determination of the mechanism of their action, is the consideration of “biological receptor-ligand” complexes. There are different experimental methods (such as NMR, X-ray crystallography, the quantitative analysis of biological activity (pIC, pGI, pMIC, etc.)) used to describe the “receptor-ligand” complexes. Structural investigations of the “receptor-ligand” complexes (e.g., ref 1) are of utmost importance because they allow for the determination of the 3D structure of a target and its binding site and help to identify the parameters of the “receptor-ligand” complexes that should be used for studying the biological activity (e.g., refs 2 and 3). Obviously, such investigations have practical applications since the obtained results can be used for the development of the new perspective drugs. Therefore, docking methods modeling “receptor-ligand” complexes using the receptor structures determined by X-ray crystallography or NMR spectroscopy are widely used for the investigation, development, and characterization of the biologically active compounds, e.g., refs 2, 4, and 5. However, the researchers providing the docking models must account for weaknesses of the experimental data, which can significantly lower the quality of the molecular docking. First of all, it must not be ruled out that the structure of a protein in an aqueous environment in a living organism can differ from its structure in a monocrystal, as analyzed by X-ray crystallography. Moreover, it is important to keep in mind that X-ray crystallography and NMR spectroscopy are semicomputational methods; both of them compare the computed diffraction pattern or chemical shifts of a simulated structure with the experimentally generated data. If the characteristics of the simulated structure are in close agreement with the experimental data, the structure is considered to be true. It is obvious that the determined 3D structure is dependent on the computational approaches used for the decoding of the experimental X-ray crystallography or NMR data.

NMR shifts are often not completely sensitive to the conformational state of the molecules.<sup>6</sup> Therefore, the 3D structures determined by NMR spectroscopy are less reliable than the 3D structures determined with X-ray crystallography. Nevertheless, X-ray crystallography analysis sometimes does not permit the structural determination of some fragments because of their partial disorder in crystals. Therefore, the data taken from the Protein Data Bank often contain undetermined fragments of structures<sup>7–12</sup> (Figure 1).<sup>13</sup> Moreover, in order to reduce the time, the resolution of the X-ray crystallography experiment is very often chosen to be lower than it is necessary for the reliable localization of the hydrogen atoms. Thus, the obtained data do not usually determine which tautomeric forms of the protein and ligand are interacting with each other, e.g., refs 14 and 15.

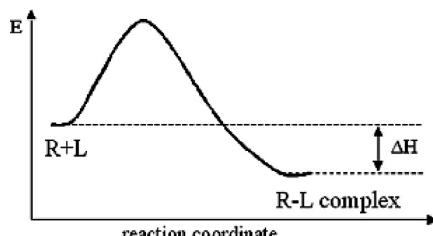
Nevertheless, the experimental data on biological receptors and their complexes with ligands are very important and useful for determining their mechanisms of the biological action. The reliable docking methods must take into account the weaknesses of the experimental data, and their influence on the predictivity of the obtained models must be at least minimized in the analysis and prognosis of biological active compounds. Besides, the computational methods are complementary to the experimental ones because they give us the additional information and help us answer the fundamental questions: what the reason of formation of such a complex is and which intermolecular interactions in the complex are the most important for biological activity.

The existing docking methods usually use scoring functions that are represented as the criteria for searching for the optimal “receptor-ligand” complexes. However, the scoring functions are often not examined for macroscopic physicochemical properties, and sometimes they have no clear physicochemical meaning, e.g., the functions given in the literature.<sup>16</sup> They consist of discrete values dependent on the types of interactions in the “receptor-ligand” systems (hydrophobic... hydrophobic, H-donor... H-acceptor, etc.) The scoring functions of some other docking methods characterize the energy of the “receptor-ligand” interactions, e.g., Gibbs free energy ( $\Delta G$ ) or the potential energy ( $\Delta E$ ).<sup>16–18</sup> All of the mentioned parameters characterize the thermodynamics of the interaction of a ligand with the receptor, but are these

\* Corresponding author phone +79028636961; fax: +7(351)2340236;  
e-mail: maria\_grishina@modelchem.ru.



**Figure 1.** The weaknesses of X-ray data of the “receptor-ligand” complexes: a) undetermined fragments in **1kv2** and b) strange structure of the heme in **1n8q**.



**Figure 2.** The diagram shows the energy depending on the reaction coordinate.

criteria correct? According to physical chemistry theories, the free energy surface can be represented by a diagram as demonstrated in Figure 2.

Do the receptor and ligand know what the result of their interaction will be; is it a stable (with low  $\Delta G$ ) or not stable (with higher  $\Delta G$ ) complex? They just interact. Therefore, the kinetic factors can play a more important role in the interaction. Moreover, the binding of a ligand with a receptor and the structure of the “receptor-ligand” complex depend on the interaction of some active centers of the molecules and do not depend on the total interaction energy. This is the reason why the interaction energies, scoring functions, Gibbs free energy, etc. of the “receptor-ligand” complexes are often poorly related to the biological activities of the compounds, e.g., refs 19 and 20. Another problem of the current widely used docking methods is insufficient consideration of the receptor and ligand flexibilities.<sup>21</sup> For example, in the FlexX<sup>21,22</sup> method, a ligand is decomposed into small rigid fragments, which are reassembled to a low-energy conformation fitting the binding site. At the same time, the ligand decomposition may have subjective errors. Besides, there are many variants of adjustment of the fragments to each other. That is why this problem may not have an unambiguous solution. In accordance with this, it is necessary to estimate the conformational flexibility of the whole system.

All of the above-mentioned aspects - the disadvantages of the experimental data and the weaknesses of the docking methods - are the reasons for the low predictivity of the docking (e.g., refs 23–27): the rms deviation of the coordinates of the atoms of the computed and experimentally determined ligand structures in complexes is high and often significantly exceeds 2 Å, e.g., refs 17, 19, 20, and 28. Is it possible to make a reliable prognosis considering such modeled complexes?

To achieve good reproducibility of the computational results, we suggest a new approach for docking, which takes into account most of the above-mentioned problems. The approach could be named restricted docking because it is

based on the orientation of molecules in the receptor pocket determined by X-ray crystallography of the “receptor-ligand” complex. This complex is used in the approach as a template.

## 2. A DOCKING APPROACH IN THE FRAMEWORK OF BiS/MC (MULTICONFORMATIONAL) ALGORITHM

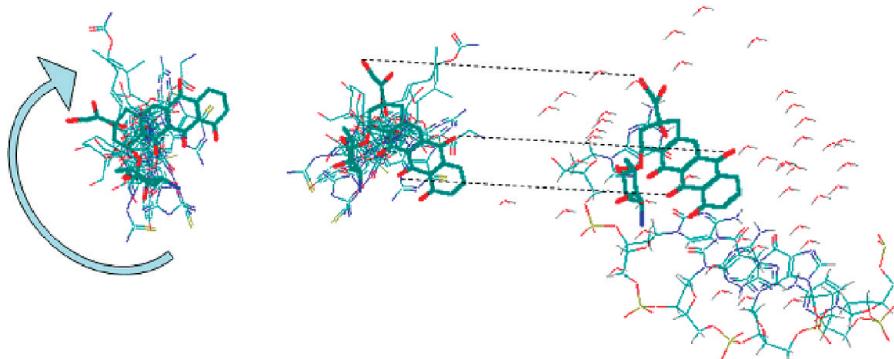
In order to solve the problem of the imperfections of the experimental data, the following procedures were suggested in this approach. First of all, 3D structures of the “receptor-ligand” complexes are used but not receptors without a ligand. In our opinion, these experimental data are more suitable for the restricted docking procedure of the other compounds because the receptor geometry in the complexes is prepared for the interaction with a ligand. Even from one ligand to another, the receptor geometry changes to a lesser degree than from the geometry of a free receptor (receptor not complexed with a drug). This approach uses PDB data that have no serious errors in the areas of the active receptor sites (such as undetermined sequences, anomalously long bonds, etc.). In other cases, the errors are corrected before docking.

The influence of the imperfections of the experimental investigations (such as nonlocalized hydrogens, changes in receptor geometry in water as compared to the crystal, etc.) on the quality of the docking can be lowered by using some additional information about the interaction. In our approach, we used the information on the structures of all ligands that can interact with the receptor (although there can be no X-ray crystallography data of the “receptor-ligand” complex for the docked compound) and the activities characterizing the effectiveness of the interaction. This can be done with the combined use of 3D QSAR methods and X-ray crystallography data for docking.

We suggest the 3D/4D QSAR algorithm BiS/Multiconformational<sup>1</sup> (BiS/MC) for the restricted docking.<sup>29–32</sup> It aligns molecules onto each other by considering their fields as being represented by Coulomb ( $\varphi_m^q$ ) and van der Waals ( $\varphi_m^{VDW}$ ) potentials at points on the molecular surface

$$\varphi_m^q = \sum_{i=1}^N \frac{q_i}{R_{im}} k \quad \varphi_m^{VDW} = -2 \sum_{i=1}^N V_{im} \frac{2^3 r_i^3}{R_{im}^6}$$

where  $m$  is the number of the point on the molecular surface,  $N$  is the total number of atoms in the considered molecule,  $q_i$  is the charge of the  $i$ th atom of the molecule,  $R_{im}$  is the distance from the  $m$ th point to the  $i$ th atom,  $k$  is a scaling coefficient,  $V_{im}$  is the potential energy minimum of the



**Figure 3.** The docking of a molecular set superimposed using BiS/MC: procedures of rotation and translation of the set for superposition of the molecule whose complex with the biological receptor is determined by X-ray crystallography.

Lennard-Jones equation, and  $r_i$  is the van der Waals radius of the  $i$ th atom.  $V_{im}$ ,  $r_i$ , and  $q_i$  values can be calculated with the nonparametric MERA model<sup>33</sup> (see item 3). To simplify the calculations, the equation for van der Waals potential includes only the attraction term of the van der Waals interactions.

The algorithm reconstructs the model receptor as being complementary to the field of the generalized molecular set. The receptor is represented as a set of pseudoatoms whose parameters can be calculated from the complementarity formalism

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

Here,  $q_m$  and  $r_m$  are the charge and radius of the pseudoatom located in the  $m$ th point.

The molecules are oriented in the complementary receptor model and the maximal total probability ( $P$ ) (not energy!) of the interaction of a molecule with the model receptor is optimized

$$P = 1 - \prod_{m=1}^M (1 - p_m)$$

where  $p_m = \exp(-E_m)/(RT)$  is the probability of the interaction of the considered molecule with the  $m$ th pseudoatom,  $M$  is the total number of the pseudoatoms, and  $E_m$  is the interaction energy of the molecule with the  $m$ th pseudoatom

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

The use of probability greatly increases the influence of the most important "receptor atom-ligand atom" interactions (the interaction of the active centers) on the molecular orientation. Moreover, the model receptor is flexible in the BiS/MC algorithm, and every ligand is represented as a set of conformers found with the MultiGen algorithm interfaced with the MM3 force field. This permits taking into account the flexibilities of both the receptor and ligands.

The reconstruction of the complementary model receptor is performed by means of the BiS/MC algorithm using characteristics computed with the MERA force field.<sup>33</sup>

which is described in the next item. All types of characteristics computed with MERA have been examined on different macroscopic physicochemical properties. This is demonstrated in the item devoted to MERA and shows that the characteristics computed with MERA are quite reliable.

Thus, the BiS/MC algorithm gets additional information on the molecular orientation in the receptor cavity. Using the results of the BiS/MC alignment for the set of the molecules and X-ray data of a "receptor-ligand" complex for one molecule of the set as a template, it is possible to predict the orientation of the other molecules in the real receptor pocket as shown in Figure 3. Next, the docking of the structures was refined by optimization with the MM3 force field.

The approach demonstrates good predictivity of the restricted docking results. Approximately 70 "receptor-ligand" complexes (thermolysine, elastase, CDK2, p38 MAP, DHFR inhibitors, inhibitors of rhinovirus HRV14, DNA antimatabolites with the receptors) were modeled, and the rms was usually less than 1 Å (Tables 1 and 2). The docking quality exceeds the results presented in previous work.<sup>28</sup> Moreover, the conformers found by BiS/MC as the most active ones are in a good agreement with the conformers found in "receptor-ligand" complexes studied by X-ray crystallography and NMR.<sup>30,32,34,35</sup>

### 3. MERA FORCE FIELD

MERA (Model of Effective Radii of Atoms) is a nonparametric model for simulating the effective atomic radii, geometry, and energy molecular characteristics. In this model each atom fills the space allowed by the other atoms, so the atom is consequently represented as an expanding balloon. The expansion of the atomic sphere is limited by other expanding spheres. The expansion of the balloon ceases when its internal pressure is equal to the external pressure provided by the other atoms, while also taking into account the rigidity of the balloons. The internal density of the atom  $i$  is proportional to its inverse volume. The external density equals the sum of the inverse volumes of spheres with radii equal to the distances between atoms  $i$  and  $j$

$$\frac{1}{V_i^{(0)}} \cong \sum_{\substack{j=1 \\ j \neq i}}^N \frac{1}{4 \pi R_{ij}^3} \quad \text{or} \quad \frac{1}{V_i^{(0)}} \cong \sum_{\substack{j=1 \\ j \neq i}}^N \frac{3}{4 \pi R_{ij}^3} \quad (1)$$

**Table 1.** RMS Deviations and Maximal Distance between the Atoms of the Docked Compounds and the Structures Determined by X-ray in the “Receptor-Ligand” Complexes<sup>1a</sup>

	molecule	rms, Å	R <sub>MAX</sub> , Å	rms, Å <sup>28</sup>	R <sub>MAX</sub> , Å <sup>28</sup>
Rhinovirus HRV14 inhibitors (the template is <b>1ruc</b> )	<b>1hri</b>	0.14	0.15	0.38	0.73
	<b>1hrv</b>	0.29	0.41	0.93	1.17
	<b>1r09</b>	0.29	0.48	4.71	4.77
	<b>1ruc</b>	0.00	0.00	0.00	0.00
	<b>1rue</b>	0.02	0.04	0.37	0.57
	<b>1rug</b>	0.00	0.00	0.02	0.03
	<b>1vrh</b>	0.35	0.64	0.74	1.50
	<b>2r04</b>	0.18	0.22	0.04	0.06
	<b>2r06</b>	0.00	0.00	0.49	0.68
	<b>2r07</b>	0.08	0.15	1.46	2.30
	<b>2rs5</b>	0.14	0.15	0.65	0.75
	<b>1r08</b>	0.31	0.42	0.97	1.48
	<b>2hwb</b>	0.18	0.26	0.80	1.65
	<b>2hwc</b>	0.14	0.24	0.62	1.04
	<b>1rud</b>	0.29	0.35	0.79	1.47
	<b>1ruh</b>	0.30	0.38	1.04	1.88
	<b>1rui</b>	0.29	0.35	0.79	1.47
	<b>2rm2</b>	0.32	0.42	1.01	1.45
	<b>2rr1</b>	0.28	0.32	0.80	1.41
	<b>2rs1</b>	0.29	0.36	0.80	1.46
	<b>2rs3</b>	0.36	0.40	0.86	1.34
	<b>1thl</b>	0.87	1.93	0.83	1.43
thermolysin inhibitors (the template is <b>5tmm</b> )	<b>1tmn</b>	0.45	0.92	0.77	1.50
	<b>1tlp</b>	0.80	1.71	0.54	0.75
	<b>5tln</b>	1.26	2.25	7.33	11.14
	<b>4tmm</b>	0.18	0.31	1.85	3.65
	<b>3tmm</b>	0.94	1.66	0.79	1.19
	<b>5tmm</b>	0.00	0.00	0.00	0.00
	<b>6tmm</b>	0.01	0.02	0.12	0.16
	<b>1agw</b>	0.80	1.24	3.34	4.74
	<b>1aq1</b>	0.66	1.23	1.13	1.96
	<b>1atp</b>	0.70	1.36	9.25	16.51
CDK2 inhibitors (the template is <b>1bkx</b> )	<b>1bkx</b>	0.00	0.00	4.85	8.67
	<b>1ckp</b>	0.26	0.56	5.81	9.93
	<b>2csn</b>	0.74	1.09	0.00	0.00
	<b>1fgi</b>	0.81	1.64	3.68	6.20
	<b>2heck</b>	0.81	1.63	4.93	8.60
	<b>1ian</b>	0.24	0.43	9.72	18.70
	<b>1stc</b>	0.72	1.46	0.00	0.00
	<b>1ydr</b>	0.97	1.81	0.43	0.72
	<b>1yds</b>	0.68	1.11	0.00	0.00
	<b>1ydt</b>	1.14	2.17	1.04	2.18
	<b>1b0e</b>	0.82	1.30	5.14	9.04
	<b>1bma</b>	0.86	1.86	0.31	0.41
	<b>1ela</b>	0.44	0.87	0.72	1.32
	<b>1elb</b>	0.61	0.97	4.66	7.81
	<b>1elc</b>	1.10	1.92	5.51	9.41
the elastase inhibitors (the template is <b>1eau</b> )	<b>1eld</b>	0.45	0.86	0.52	0.97
	<b>1ele</b>	0.00	0.00	0.00	0.00
	<b>1inc</b>	0.77	1.36	6.75	10.01
	<b>7est</b>	0.50	0.90	0.28	0.47
	<b>1eas</b>	0.41	0.86	0.33	0.43
	<b>1eat</b>	0.43	0.96	0.00	0.00
	<b>1eau</b>	0.80	1.30	0.81	1.64
	<b>4est</b>	0.58	0.88	0.00	0.00

<sup>a</sup> The results of this paper in comparison with a previous work.<sup>28</sup>

where  $V_i^{(0)}$  is the volume of the  $i$ th atom (zero-order approximation of MERA model),  $R_{ij}$  is the distance between atoms  $i$  and  $j$ , and  $N$  is the total number of atoms in the molecule or the molecular system.

**Table 2.** RMS Deviations and Maximal Distance between the Atoms of the Docked Compounds and the Structures Determined by X-ray in the “Receptor-Ligand” Complexes

	molecule	rms, Å	max. dist, Å
DNA-antimetabolites (the template is <b>1lims</b> )	<b>1d35</b>	0.235	0.353
	<b>1d37</b>	0.613	1.017
	<b>1lims</b>	0.000	0.000
DHFR inhibitors (the template is <b>1rg7</b> )	<b>1rg7</b>	0.0000	0.0000
	<b>1hfp</b>	1.0519	2.0160
p38 MAP-kinase inhibitors (the template is <b>1bl7</b> )	<b>1a9u</b>	0.77	1.37
	<b>1bl6</b>	0.51	0.92
	<b>1bl7</b>	0.00	0.00
	<b>1bmk</b>	0.50	1.00
	<b>1d19</b>	0.80	1.28
	<b>1ian</b>	1.36	2.74
	<b>1kv1</b>	0.00	0.00
	<b>1kv2</b>	1.07	2.03

The coefficient of proportionality was found empirically to be equal to  $(4/9)\pi\sqrt{n_i}$  (where  $n_i$  is a period number of the atom  $i$ ). In this case, formula (1) appears as

$$\frac{1}{V_i^{(0)}} = \sum_{j=1}^N \frac{\sqrt{n_i}}{3R_{ij}^3} \quad j \neq i$$

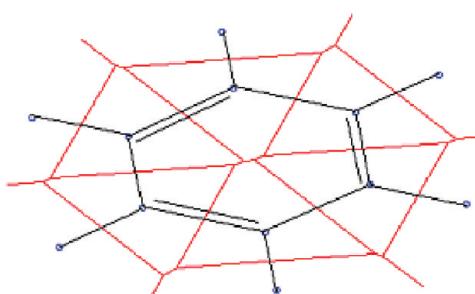
Now it is possible to calculate the zero-order approximation volume of each atom

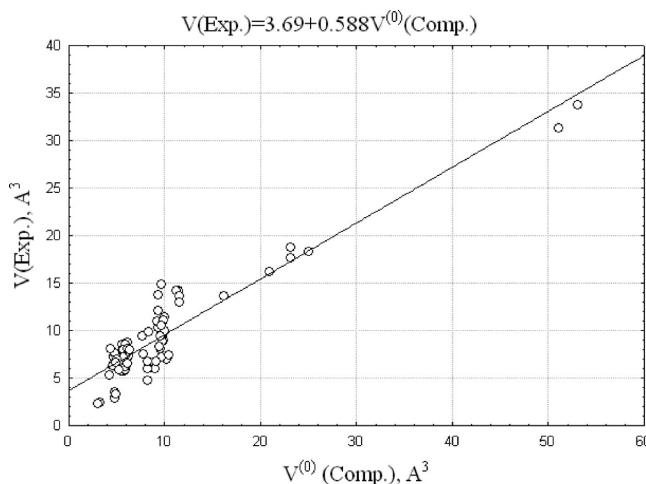
$$V_i^{(0)} = \left( \sum_{j=1}^N \frac{\sqrt{n_i}}{3R_{ij}^3} \right)^{-1}$$

This is the volume of nonspherical atoms, which is analogous to the presentation of the Voronoi polyhedrons as shown in Figure 4. The sum of the atomic volumes must be equal to the molecular volume ( $V_M^{(0)} = \sum_{i=1}^N V_i^{(0)}$ ), and this permits the calculation of the density of the compound

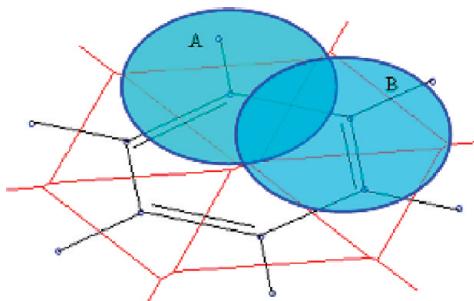
$$\rho^{(0)} = K_p \frac{M}{N_A V_M^{(0)}}$$

This approach was used for the computation of the densities of 137 organic liquids of various classes, such as alkanes, alkenes, aromatics, alcohols, carboxylic acids, amines, etc., and including C, H, N, O, F, Cl, Br, and I.<sup>33</sup> The computed density of the compounds correlated with the

**Figure 4.** Nonspherical atoms analogous to the Voronoi polyhedrons representation.



**Figure 5.** The relationship between the atomic volumes calculated with the zero-order MERA approximation and the atomic volumes determined by low temperature, highly accurate X-ray structures<sup>36</sup> determined by low temperature, highly accurate X-ray structures<sup>36</sup> for organic compounds containing H, C, N, O, F, and S.



**Figure 6.** The atomic spheres limited with the chord forming Voronoi atomic polyhedron.

experimental density ( $d_4^{20}$ ) with a correlation coefficient of  $R = 0.9893$  (standard error  $S = 0.072$ ; Fisher F-criterion  $F = 6200$ ).

The zero-order MERA atomic volumes were compared to the atomic volumes determined by low temperature, highly accurate X-ray crystallography<sup>36</sup> and were found to correlate well with each other (Figure 5). However, the molecular model obtained with the zero-order MERA approximation (Figure 4) does not permit the computation of the van der Waals radii of atoms, which are very important for the estimation of intermolecular distances, energies of intermolecular interactions, and steric effects.

The most common presentation of a molecule is the space-filling model with spherical atoms. Thus, it is necessary to expand a nonspherical atom to a spherical form. Figure 6 shows that the border of the Voronoi polyhedron is limited with the chord of the proposed atomic spheres. At the same time, a part of sphere A is located in the area of atom B as well as a part of sphere B being located in the area of atom A. Taking into account that the "atomic density" is proportional to the Gaussian function (analogously to the Grant and Pickup approach<sup>37,38</sup> and the quantum theory), we can suppose that the volume of atom B located in the area of atom A can be presented as  $V_B^{(0)} \exp(-R_{AB}^2/2a_0^2\pi)$  (where  $a_0$  is the Bohr radius), and the volume of atom A located in the area of atom B can be presented as  $V_A^{(0)} \exp(-R_{AB}^2/2a_0^2\pi)$ . Since each atom  $i$  of the molecule can be intersected with every atom of the molecule, its volume  $V_i^{(1)}$  can be represented as

$$V_i^{(1)} = V_i^{(0)} + \sum_{\substack{j=1 \\ j \neq i}}^N V_j^{(0)} \exp(-R_{ij}^2/2\pi a_0^2)$$

where  $V_i^{(1)}$  is the volume of the  $i$ th atom (the first-order approximation of the MERA model).

Now it is possible to calculate the radii of the spherical atoms with the first-order MERA approximation

$$r_i^{(1)} = \left( \frac{3V_i^{(1)}}{4\pi} \right)^{1/3}$$

It is obvious that the molecular volume is not equal to the sum of the atomic volumes in such representation. The molecular volume of the first-order approximation ( $V_M^{(1)}$ ) can be computed by accounting for the intersection volumes either with numerical integration or the Connolly method or the Grant-Pickup method<sup>37–40</sup>

$$V_M^{(1)} = \sum_i V_i^{(1)} - \sum_{i < j} V_{ij} + \sum_{i < j < k} V_{ijk} - \dots$$

where  $V_{ij}, V_{ijk}, \dots$  are the second-, third-, etc. order intersection volumes.

Then it is possible to compute densities with the first-order approximation

$$\rho^{(1)} = K_p \frac{M}{N_A V_M^{(1)}}$$

The value for  $\rho^{(1)}$  correlates with the experimental data for the same set of the compounds with a correlation coefficient equal to  $R = 0.9953$  ( $S = 0.047$ ;  $F = 14300$ ).<sup>33</sup>

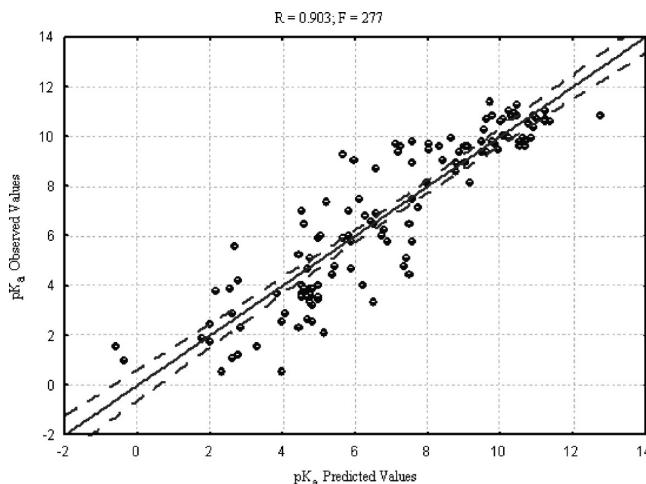
The atomic radii computed with MERA can be used for the estimation of the energy characteristics. The interaction energy is presented as the sum of the atom–atomic potentials, including van der Waals and Coulomb interactions

$$E_{ij} = -2U_{ij} \left( \frac{R_{ij}^e}{R_{ij}} \right)^6 + U_{ij} \left( \frac{R_{ij}^e}{R_{ij}} \right)^{12} + \frac{q_i q_j}{4\pi \epsilon_0 R_{ij}} \quad (2)$$

where  $U_{ij}$  is the depth of the potential;  $R_{ij}^e$  is the equilibrium distance, which can be considered as  $R_{ij}^e = r_i + r_j$ ;  $q_i$  and  $q_j$  are the charges of the atoms  $i$  and  $j$ ; and  $\epsilon_0$  is the vacuum permittivity.

The MERA atomic charges are computed with the formalism of the Full Equalization of Orbital Electronegativities<sup>41</sup> using some modifications. A set of equations for the calculation of the electronegativities is presented in the MERA model as follows

$$\begin{cases} \eta_1 = \eta_1^0 + \frac{q_1}{4\pi \epsilon_0 r_1} + \sum_{i \neq 1} \frac{q_i}{4\pi \epsilon_0 R_{1i}} \\ \eta_2 = \eta_2^0 + \frac{q_2}{4\pi \epsilon_0 r_2} + \sum_{i \neq 2} \frac{q_i}{4\pi \epsilon_0 R_{2i}} \\ \dots \\ \eta_n = \eta_n^0 + \frac{q_n}{4\pi \epsilon_0 r_n} + \sum_{i \neq n} \frac{q_i}{4\pi \epsilon_0 R_{ni}} \\ \sum q_i = Q \\ \eta_1 = \eta_2 = \dots = \eta_n \end{cases}$$



**Figure 7.** The calculated pK<sub>A</sub> using ProK and experimental ones found in water solutions (ionic force equals 0).

where  $\eta_i^0$  is the initial electronegativity of atom  $i$ ,  $\eta$  is a molecular characteristic derived from the full equalization of the electronegativities, and  $Q$  is the charge of the considered structure (e.g.,  $Q = 0$  for molecules,  $Q > 0$  for cations, and  $Q < 0$  for anions).

The MERA presentation does not include empirical orbital electronegativities ( $\eta_i^0$ ), which are calculated as follows

$$\eta_m^0 = I_H \left( \frac{a_o}{r_m} \right)^2$$

where  $a_o$  is the Bohr radius, and  $I_H$  is the ionization potential of hydrogen.

This approach calculates atomic charges of carbons that correlate with NMR C<sup>13</sup> chemical shifts with a correlation coefficient of  $R = 0.977$ . Moreover, the molecular values of  $\eta$  are related to the pK<sub>A</sub> values of compounds measured in water solutions with the ionic force equal to 0 and the temperature equal to 298 K by

$$pK_A = a + b \cdot \eta$$

where  $a$  and  $b$  are the parameters of the regression.

The correlation coefficient of the linear regression for the set of 150 compounds (including amines, heterocycles, etc.) is equal to  $R = 0.903$  ( $F = 277$ ) (Figure 7). The algorithm for the prognosis of pK<sub>A</sub> values was named ProK.

Equation 2 includes  $U_{ij}$  values, and MERA uses a nonparametric approach for the estimation of these values. They are calculated with the equation for the thermal coefficient of expansion. From classical physics, it is well-known that the coefficient can be calculated as follows

$$\alpha = \frac{\chi_{ij}k}{2\kappa_{ij}^2} \quad (3)$$

where  $\kappa_{ij}$  is a force constant for the interaction of atoms  $i$  and  $j$ ;  $\kappa_{ij} = ((\partial^2 E_{ij})/(\partial R_{ij}^2))$ ;  $\chi_{ij}$  is the anharmonicity of the interaction of the atoms  $i$  and  $j$ ,  $\chi_{ij} = -((\partial^3 E_{ij})/(\partial R_{ij}^3))$ ; and  $k$  is the Boltzmann constant.

By considering just the van der Waals interactions, the interaction energy is

$$E_{ij}^V = -2U_{ij} \left( \frac{R_{ij}^e}{R_{ij}} \right)^6 + U_{ij} \left( \frac{R_{ij}^e}{R_{ij}} \right)^{12}$$

The force constant at the point where  $R_{ij} = R_{ij}^e$  is equal to

$$\kappa_{ij} = \frac{\partial^2 E_{ij}^V}{\partial R_{ij}^2} = \frac{72U_{ij}}{(R_{ij}^e)^2} \quad (4)$$

The anharmonicity is equal to

$$\chi_{ij} = -\frac{\partial^3 E_{ij}^V}{\partial R_{ij}^3} = \frac{1512U_{ij}}{(R_{ij}^e)^3} \quad (5)$$

Therefore,  $\alpha$  can be found from the expressions (3), (4), and (5)

$$\alpha = \frac{7kR_{ij}^e}{48U_{ij}}$$

and

$$U_{ij} = \frac{7kR_{ij}^e}{48\alpha}$$

Since  $R_{ij}^e = r_i + r_j$ , then

$$U_{ij} = \frac{7k(r_i + r_j)}{48\alpha}$$

Thus, the van der Waals interaction energy is equal to

$$E_{ij}^V = -\frac{7k(r_i + r_j)}{24\alpha} \left( \frac{r_i + r_j}{R_{ij}} \right)^6 + \frac{7k(r_i + r_j)}{48\alpha} \left( \frac{r_i + r_j}{R_{ij}} \right)^{12}$$

and the total interaction energy  $E_{ij}$ , including Coulomb interactions, is

$$E_{ij} = E_{ij}^V + E_{ij}^q = -\frac{7k(r_i + r_j)}{24\alpha} \left( \frac{r_i + r_j}{R_{ij}} \right)^6 + \frac{7k(r_i + r_j)}{48\alpha} \left( \frac{r_i + r_j}{R_{ij}} \right)^{12} + \frac{q_i q_j}{4\pi \epsilon_0 R_{ij}}$$

This formula does not take into account the probabilities of the intermolecular atom–atomic interaction which are presented in MERA as follows

$$p_{ij} = \frac{n_{ij} S_i \exp(-E_{ij}/kT)}{1 + \sum_{l=1}^N \exp(-E_{il}/kT)}$$

where  $n_{ij}$  is a coordination number of  $i$ th atom with respect to the  $j$ th atom, and  $S_i$  is the solvent-accessible area of the  $i$ th atom.

Thus, the real interaction energy of the  $i$ th atom of the molecule with the  $j$ th atom of a neighboring molecule in the substance is the sum of energies  $E_{ij}$  multiplied by  $p_{ij}$

$$E_{ij} = \sum_{j=1}^N p_{ij} \left[ -\frac{7k(r_i + r_j)}{24\alpha} \left( \frac{r_i + r_j}{R_{ij}} \right)^6 + \frac{7k(r_i + r_j)}{48\alpha} \left( \frac{r_i + r_j}{R_{ij}} \right)^{12} + \frac{q_i q_j}{4\pi\epsilon_0 R_{ij}} \right] \quad (6)$$

The equilibrium van der Waals distance ( $r_i + r_j$ ) in the substance is reduced by the influence of the electrostatic interactions. Therefore, eq 6 was suggested for the estimation of their influence on the van der Waals atomic radii. It is evident that the exact equilibrium distance can be calculated as follows

$$\frac{\partial E_{ij}}{\partial R_{ij}} = 0 \quad \text{i.e.} \quad \sum_{j=1}^N p_{ij} \left[ \frac{7k(r_i + r_j)(r_i + r_j)^6}{4\alpha} - \frac{7k(r_i + r_j)(r_i + r_j)^{12}}{R_{ij}^{13}} - \frac{q_i q_j}{4\pi\epsilon_0 R_{ij}^2} \right] = 0$$

Thus, the correction of the van der Waals atomic radii for the electrostatic interaction  $\Delta r_i^{(A)}$  is equal to half of the mean deviation of the true distance  $R_{ij}$  from the equilibrium van der Waals distance

$$\Delta r_i^{(A)} = \frac{1}{2N} \sum_{j=1}^N (R_{ij} - r_i - r_j)$$

Therefore, it is possible to use the temperature correction, which can be calculated as follows

$$\Delta r_i^{(T)} = \gamma + \alpha T$$

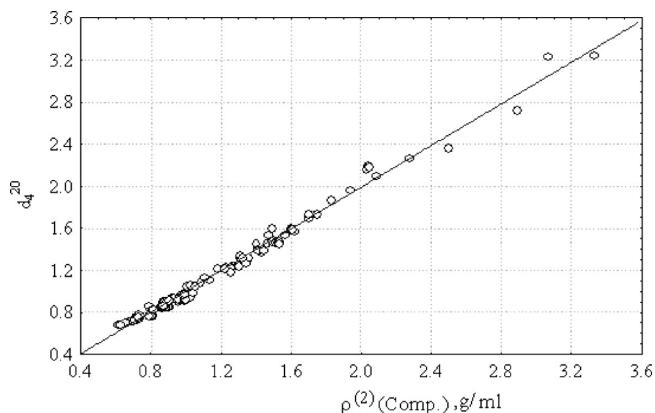
where  $T$  is the absolute temperature;  $\alpha = 6.662 \cdot 10^{-4} \text{ \AA/K}$ ; and  $\gamma = -0.1935 \text{ \AA}$ .

All the procedures permit the calculation of the second-order MERA with the approximated van der Waals radii ( $r_i^{(2)}$ ) of atoms presented as

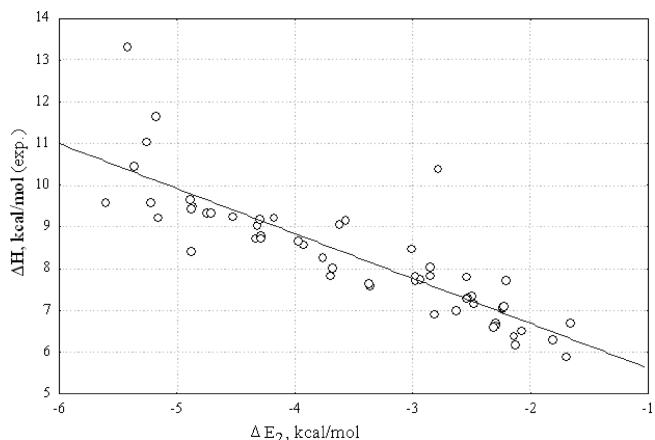
$$r_i^{(2)} = r_i^{(1)} + \Delta r_i^{(A)} + \Delta r_i^{(T)}$$

Using these radii, it is possible to calculate the molecular volume of the second-order approximation ( $V_M^{(2)}$ ) while taking into account the intersection volumes computed either with numerical integration or the Connolly method or the Grant-Pickup method. It was found that the density computed with MERA is in good agreement with the experimental results. The correlation coefficient of this relationship is a little bit higher than was previously found using the first-order approximation (see above). For the second-order approximation, the correlation coefficient was  $R = 0.9963$  ( $S = 0.043$ ;  $F = 17600$ ) for the same set of the compounds (Figure 8).

The MERA model was used for the estimation of the enthalpy of evaporation of 80 organic compounds of various classes, such as hydrocarbons (nonaromatic, aromatic, paraffin), alcohols, carbonyl compounds, carboxylic acids, amines, ethers, halogensubstituted hydrocarbons, etc., including C, H, N, O, F, Cl, Br, and I.<sup>42</sup> The MERA interaction energy calculated with eq 6 for the dimers and trimers of the compounds modeled with the algorithm Mech<sup>42</sup> correlates with the experimental enthalpy of the evaporation with a correlation coefficient of 0.890 ( $S = 0.67 \text{ kcal/mol}$ ;  $F = 190$ ) (Figure 9).



**Figure 8.** The densities of liquids calculated with the second-order MERA approximation and the experimental densities.



**Figure 9.** The relationship between the energy of dimer formation computed with MERA and the enthalpy of evaporation determined experimentally.

Thus, the MERA force field determines different physicochemical characteristics of compounds with high quality.<sup>33,42,43</sup> This shows good predictive capabilities of the MERA model and the reliability of the calculated descriptors. Therefore, the MERA model can be used for the correct estimation of the parameters of interactions in the "receptor-ligand" complexes.

#### 4. THE COCON ALGORITHM FOR THE STUDY OF THE "RECEPTOR-LIGAND" COMPLEXES

The modeled "receptor-ligand" complexes whose parameters are computed with MERA can be studied with a new CoCon approach for the determination of the mechanisms of the biological action of compounds and for the search of active centers of receptors and ligands.

The CoCon approach is based on the idea that biological activity of a compound mainly depends on the interaction ability of the active centers of the ligand and the receptor site, whereas the other atoms of the receptor site play less important roles due to the conformational flexibility and tautomerism at these locations. The proposed method creates relationships between the biological activity and the parameters of interactions in the "receptor-ligand" complexes.

The possibility of the interaction depends on the interaction energy, which can be represented as atom–atomic potentials of van der Waals and Coulomb interactions calculated with the MERA model. Such representation permits the decom-

position of the atom–atomic interaction energies in search for the best linear relationship between the biological activity and the parameters of the interaction. The receptor and ligand atoms, whose parameters determine biological activity, are supposed to be active centers in CoCon. Other atoms play less important roles in the interaction or belong to the changeable fragments of the molecules.

The decomposition is realized in the CoCon algorithm by using forward and backward stepwise procedures of the regression analysis. The quality of the relationships is estimated with the leave-one-out cross validation technique.

This approach could be realized without any problem if the number of atoms in the considered ligands were the same, but ligands have different numbers of atoms. Therefore, the number of factors used for the determination of the relationship is different.

To solve this problem, the CoCon algorithm uses two principles.

**4.1. The First CoCon Principle To Solve the Problem of Different Numbers of Atoms in Ligands.** The first CoCon principle is based on the supposition that the biological activity of a compound depends on its general ability to interact with the active receptor centers. Therefore, the interaction energies of all ligand atoms with the active centers of the receptor are considered.

**The Idea Is Realized in CoCon in Three Ways.** The first way is based on the calculations of the parameters of an interaction of a ligand with each atom of the receptor site, including water. The relationship is presented as a multifactor model whose factors are interaction energies and forces, i.e.

$$BA = a + \sum_i b_i E_i + \sum_j c_j F_j \quad (7)$$

where  $a$ ,  $b_i$ , and  $c_j$  are the parameters of a linear regression (the parameters are determined using forward and backward stepwise procedures of the regression analysis);  $E_i = \sum_{n=1}^N E_{in}$  is the interaction energy of the  $i$ th receptor active center with all the atoms of the ligand under consideration;  $F_j = \sum_{n=1}^N F_{jn}$  is the interaction force of the  $j$ th receptor active center with all atoms of the ligand under consideration; and  $N$  is the number of atoms in the ligand.

The consideration of the forces permits taking into account the dynamics of the interaction because the force is a vector quantity. The vector shows the direction of the molecular movement in the process of binding with the receptor active centers. Because the forces are the first derivatives of the energy function, the force of the interaction of the ligand atom  $i$  and the receptor atom  $m$  is calculated with CoCon as follows

$$F_{im} = \frac{dE_{im}}{dR_{im}} = 12U_{im} \left( \frac{(r_m + r_i)^6}{R_{im}^7} - \frac{(r_m + r_i)^{12}}{R_{im}^{13}} \right) - \frac{q_i q_m}{4\pi\epsilon_0 R_{im}^2}$$

As mentioned above, the use of backward and forward stepwise procedures of the regression analysis allows for a linear relationship between biological activity and the parameters to be obtained. The receptor atoms, whose parameters influence the biological activity, are the receptor active centers.

Interaction energies and forces of the molecules with  $i$ -th receptor atom are the factors for the regression analysis. The procedure of the decomposition starts from the search of the best relationship between biological activity and interaction energy/force of a molecule with  $i$ -th atom of the receptor. As soon as it is found, the next interaction energy/force of a molecule with  $j$ -th receptor atom is added to the relationship in order to improve the quality. This forward stepwise procedure is repeated iteratively while the cross-validation quality is increasing. After that the backward stepwise procedure is used. It excludes the factors from the obtained relationship while the cross-validation quality is increasing.

The second way is to present the relationship as a single-factor model. The factor is the sum of the interaction energies of the ligand with the  $i$ th active center of the receptor site including water molecules

$$BA = a + b \sum_i E_i \quad (8)$$

where  $E_i = \sum_{n=1}^N E_{in}$  is the interaction energy of the  $i$ th receptor active center with all atoms of the ligand under consideration.

In this case, receptor active centers can be determined by their contribution in the total interaction energy using backward and forward stepwise procedures as described above.

3) In the two previously considered ways, the receptor site includes water molecules. Since the interactions with the receptor and water can play different roles in the biological activity, the contributions of the “receptor-ligand” interactions and “water-ligand” interactions can be considered separately. Therefore, the relationship can be represented as a two-factor model, where the first factor depends on the interactions with receptor atoms, and the second depends on the interaction with water atoms included in the receptor cavity

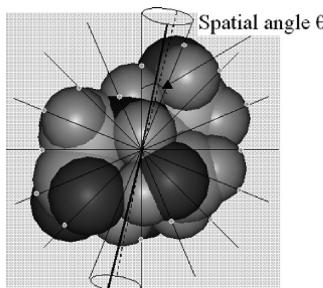
$$BA = a + b \sum_i E_{i(R)} + c \sum_j E_{j(W)} \quad (9)$$

where  $E_{i(R)} = \sum_{n=1}^N E_{in(R)}$  is the interaction energy of the  $i$ th receptor active center with all atoms of the ligand under consideration, and  $E_{j(W)} = \sum_{n=1}^N E_{in(W)}$  is the interaction energy of the  $j$ th water active center with all atoms of the ligand under consideration.

Similar to the two previous ways, the active centers can be determined by their contribution to the sum of interaction energies using backward and forward stepwise procedures.

The above-mentioned CoCon idea permits the determination of the active centers of the receptor site and surrounding water. However, in the “receptor-ligand” interactions, just some of the ligand atoms, named active centers (or pharmacophore atoms), play the most important role. Therefore, the second principle has been suggested in order to solve the problem of the different number of analyzed factors for various ligands and to search for ligand active centers.

**4.2. The Second CoCon Principle To Solve the Problem of Different Numbers of Atoms in Ligands.** The principle is based on the consideration of interaction parameters with points on the ligand surface, and the number of the points has to be constant for all ligands. The points



**Figure 10.** The principle of choice of the points location on the molecular surface for the determination of the parameters of the "receptor-ligand" interactions.

can be chosen by rays drawn from the mechanical center of a ligand with a constant spatial angle value  $\theta$  between them and intersecting the molecular surface as shown in Figure 10.

The number of points is constant for all the molecules of the set. The parameters of the interaction of each ligand atom, containing the point with the receptor site, were calculated.

The second principle is realized in CoCon in two ways.

1) The relationship is presented as a two-factor model. One of the factors is the sum of interaction energies of all the atoms of the receptor site (without water) with the active centers of the ligand. Another factor is the sum of interaction energies of all the water molecules included in the site with the active centers of the ligand

$$BA = a + b \sum_l E_{l(R)} + c \sum_k E_{k(W)} \quad (10)$$

where  $E_{l(R)} = \sum_{i=1}^{N_{(R)}} E_{li(R)}$  is the interaction energy of all the receptor atoms (the total number of the receptor site atoms is  $N_{(R)}$ ) with the  $l$ th ligand active center, and  $E_{k(W)} = \sum_{j=1}^{N_{(W)}} E_{kj(W)}$  is the interaction energy of all the water molecules of the receptor cavity (the total number of water molecules is  $N_{(W)}$ ) with the  $k$ th ligand active center.

Active centers can be found by energy decomposition using backward and forward stepwise procedures.

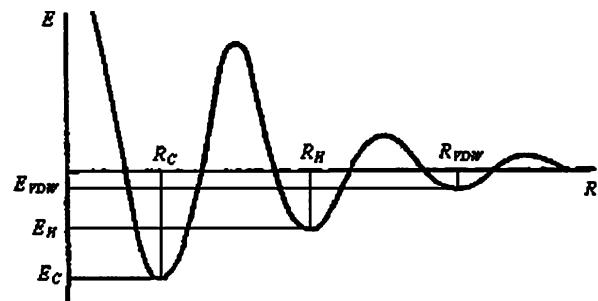
2) The second way is also realized by presenting the relationship as a two-factor model. The factors are the sums of interaction energies of active centers of the receptor site with the active centers of the ligand. The interaction energy with the receptor is the first factor, and water is the second factor

$$BA = a + b \sum_{i=1}^{N_{(R)}} \sum_{j=1}^{N_{(ray)}} E_{ij(R)} + c \sum_{k=1}^{N_{(W)}} \sum_{l=1}^{N_{(ray)}} E_{kl(W)} \quad (11)$$

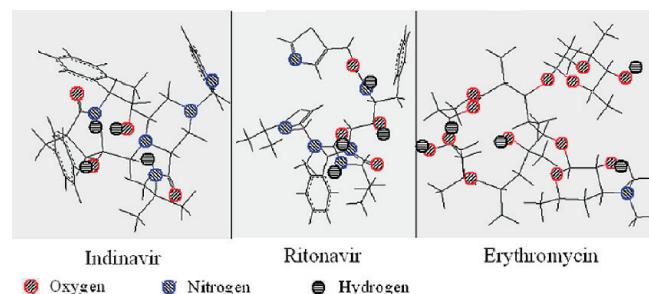
where  $E_{ij(R)}$  is the energy of interaction of the  $i$ th receptor active center with the  $j$ th ligand active center;  $E_{kl(W)}$  is the energy of interaction of the water active center  $k$  with the ligand active center  $l$ ;  $N_{(ray)}$  is the number of rays; and  $a$ ,  $b$ , and  $c$  are the parameters of the linear regression.

The active centers are determined using backward and forward stepwise procedures of the regression analysis as in the approaches listed above. The atoms can be considered as ligand and receptor active centers if their interaction energies are included in the sum.

**4.3. CoCon Energies of the Hydrogen Bonds.** Hydrogens in organic compounds mostly determine the molecular surface and consequently intermolecular interactions.



**Figure 11.** The distribution function.<sup>44</sup>



**Figure 12.** The potential centers for the hydrogen binding of the substrates of 3A4 isoform.

One of the most important types of intermolecular interactions is a hydrogen bond. Therefore, the problem of the correct consideration of the hydrogen bonds must be solved for the proper study of intermolecular interactions, e.g., in the "receptor-ligand" complexes. In this paper, new approaches for the estimation of the hydrogen bonds have been suggested. The approaches consider not only classical hydrogen bonds (between a typical H-acceptor and a hydrogen atom of a typical H-donor) but also other A···H intermolecular contacts (such as H···H, C···H, S···H, etc.) which often occur in "receptor-ligand" complexes (e.g., see item 5).

In most cases a hydrogen bond is characterized by the distances equal to the intermediate value between covalent A-H and van der Waals A···H contact. Moreover, it has energies less than covalent bond but greater than van der Waals noncovalent contact.

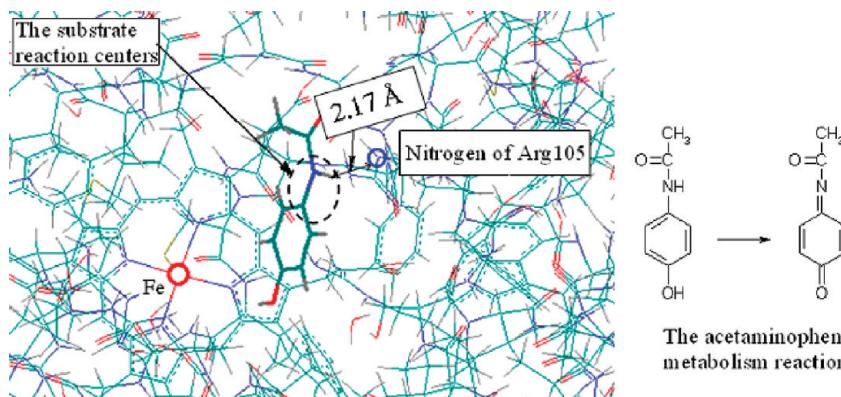
The first variant of the function of the hydrogen bond energy ( $E_H$ ) is presented in this approach as the function similar to the distribution function (Figure 11)<sup>44</sup>

$$E_H = \frac{a \cdot \exp(-b \cdot R) \cdot \cos(c \cdot R + \varphi)}{R} \quad (12)$$

where  $a$ ,  $b$ ,  $c$ , and  $\varphi$  are the parameters of the estimation. They can be calculated with the following equations set

$$\begin{cases} cR_C + \varphi = \pi \\ cR_{VDW} + \varphi = 5\pi \\ a \cdot \exp(-bR_C) = E_C \cdot R_C \\ a \cdot \exp(-bR_V) = E_{VDW} \cdot R_{VDW} \end{cases}$$

where  $R_C$  is an equilibrium covalent distance,  $E_C$  is a covalent bond energy,  $R_{VDW}$  is a van der Waals equilibrium distance, and  $E_{VDW}$  is the energy of van der Waals interactions.



**Figure 13.** Acetoaminophen (bold line) docked to the 3A4 isoform; the metabolized reaction centers are covered with a substitute from the Fe atom of the heme; the acetaminophen hydrogen forms a hydrogen bond with the nitrogen of Arg105 (the distance equals 2.17 Å).

Another variant in calculating the energy of hydrogen bonds is to present the interaction energy as follows

$$E_H = -a \cdot \frac{R_H^{10}}{R^{10}} + b \cdot \frac{R_H^{12}}{R^{12}} \quad (13)$$

where  $R_H$  is the equilibrium distance of the hydrogen bond.

The parameters  $a$  and  $b$  are determined in CoCon based on the hypothesis that  $E_H$  corresponds to the minima of the interaction energy. Therefore, the parameters are calculated with the following set of equations

$$\begin{cases} \left( \frac{dE}{dR} \right)_{R=R_H} = 10a \cdot \frac{R_H^{10}}{R^{11}} - 12b \cdot \frac{R_H^{12}}{R^{13}} = 0 \\ E_H = E(R = R_H) \\ E_H = \sqrt{E_C \cdot E_{VDW}} \end{cases} \quad \text{or}$$

$$\begin{cases} a = 1.2b \\ E_H = b - a \\ E_H = \sqrt{E_C \cdot E_{VDW}} \end{cases}$$

The third variant for estimating the hydrogen bonds interactions is based on the presentation of the interaction energy in the “receptor-ligand” complexes as

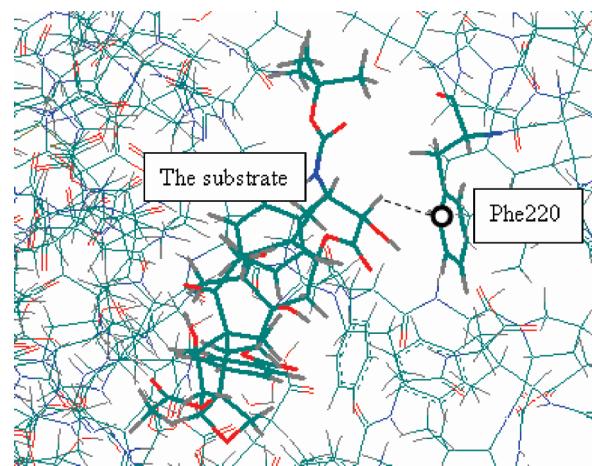
$$E_H = -a \cdot \frac{R_H^6}{R^6} + b \cdot \frac{R_H^{10}}{R^{10}} \quad (14)$$

The parameters of this function (14) are similarly calculated with the following set of equations

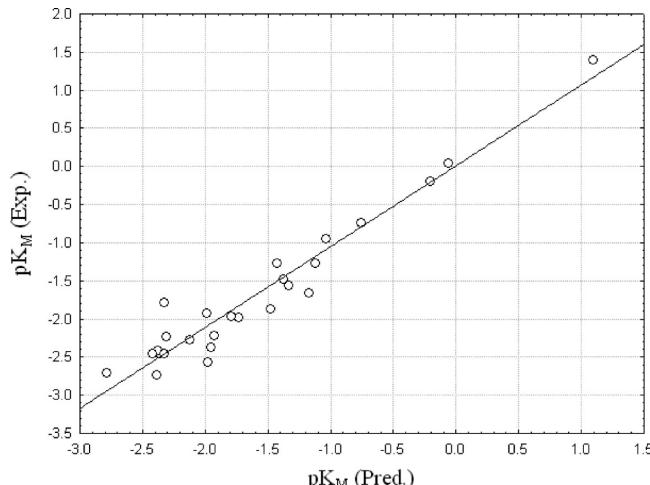
$$\begin{cases} \left( \frac{dE}{dR} \right)_{R=R_H} = 6a \cdot \frac{R_H^6}{R^7} - 10b \cdot \frac{R_H^{10}}{R^{11}} = 0 \\ E_H = E(R = R_H) \\ E_H = \sqrt{E_C \cdot E_{VDW}} \end{cases} \quad \text{or}$$

$$\begin{cases} b = 0.6a \\ E_H = b - a \\ E_H = \sqrt{E_C \cdot E_{VDW}} \end{cases}$$

Therefore, CoCon creates five variants of biological activity relationships (eqs 7–11) with the parameters of interactions in “receptor-ligand” complexes. Their analysis permits the finding of the active centers of the receptor site and the ligands. CoCon determines all types of intermolecular



**Figure 14.** The lipophilic contact of the Phe220 carbon with the hydrogen of the Taxotere docked to the 3A4 isoform.



**Figure 15.** Observed and computed (with CoCon)  $pK_M$ .

interactions: Coulomb, van der Waals, and hydrogen bonds. The hydrogen bonds energies in the “receptor-ligand” complexes are calculated with three equations (eqs 12–14). The CoCon software was used for the study of different sets of molecules: substrates of the 3A4 isoform of cytochrome P450, neuraminidase, DHFR, and p38 MAP kinase inhibitors. The sets included different types of molecules: one of the sets included the compounds with high conformational flexibility (DHFR inhibitors), the second- with low conformational flexibility (p38 MAP kinase inhibitors), the third set included both high and low flexible compounds (sub-

**Table 3.** CYP3A4 Substrates, Their  $pK_M$  Values, the Number of Lipophilic Contacts and Hydrogen Bonds in Modeled "Receptor-Ligand" Complexes, Molecular Volumes Computed with the Second-Order MERA Approximation (See Item 3)

compound	number of lipophilic contacts	number of hydrogen bonds	$V_M$ , Å <sup>3</sup>	$pK_M$ (obs)
Acetaminophen	7	2	196.50	-2.45 <sup>47</sup>
Amitriptyline	11	0	445.02	-1.92 <sup>53</sup>
Buprenorphine	7	1	690.19	-1.56 <sup>69</sup>
Celecoxib	7	3	427.34	-1.26 <sup>63</sup>
(S)-Citalopram	6	0	476.87	-2.23 <sup>51</sup>
Delavirdine	18	4	640.21	-0.73 <sup>45</sup>
Haloperidol	3	0	494.80	-1.26 <sup>68</sup>
Indinavir	5	1	909.33	1.40 <sup>55</sup>
Estrone	6	0	388.20	-1.26 <sup>54</sup>
Midazolam	6	0	372.56	-0.19 <sup>49</sup>
Pranidipine	15	2	637.95	-0.94 <sup>46</sup>
(R)-Propafenone	5	0	547.46	-1.86 <sup>66</sup>
9-cis-Retinal	5	1	502.21	-2.56 <sup>60</sup>
trans-Retinal	11	0	488.95	-2.41 <sup>60</sup>
Ritonavir	10	0	1080.99	1.22 <sup>55</sup>
Salicylate	5	2	159.45	-2.71 <sup>58</sup>
Taxotere	20	1	1151.03	0.04 <sup>56</sup>
Terfenadine	8	0	767.29	-1.48 <sup>70</sup>
Tirilazad	15	2	1030.19	-0.32 <sup>48</sup>
Verapamil	15	1	727.67	-1.97 <sup>59</sup>
Ziprasidone	14	1	528.59	-2.37 <sup>62</sup>
CCl <sub>2</sub> BrH	3	0	106.08	-1.78 <sup>67</sup>
Desacetyl-diltiazem	15	0	533.38	-2.73 <sup>65</sup>
(R)-Citalopram	9	0	476.87	-2.21 <sup>51</sup>
2-Cl-3-Py-3-yl-5,6,7,8-tetra-H-indolizine-1-carboxamide	6	2	320.53	-2.45 <sup>64</sup>
Ambroxol	2	1	359.19	-2.46 <sup>52</sup>
Erythromycin	10	1	1157.21	-1.66 <sup>71</sup>
(S)-Propafenone	8	0	537.97	-2.27 <sup>66</sup>

strates of 3A4 isoform), and the forth set included organic salts (neuraminidase inhibitors) the consideration of which included the modeling of cation–anion complexes. This permits the showing that the approach can be used for various classes of organic compounds including neutral molecules and salts, with different flexibility and association capacity, for the sets with homologous and nonhomologous compounds. The applications of the method for the investigation of the biological activity of the compounds and their corresponding mechanisms are demonstrated in the next item. The leave-one-out plots are presented in Figures 15, 19, 21, and 22.

## 5. COCON USED FOR THE ANALYSIS OF BIOACTIVITIES OF COMPOUNDS DOCKED WITH BIS/MC

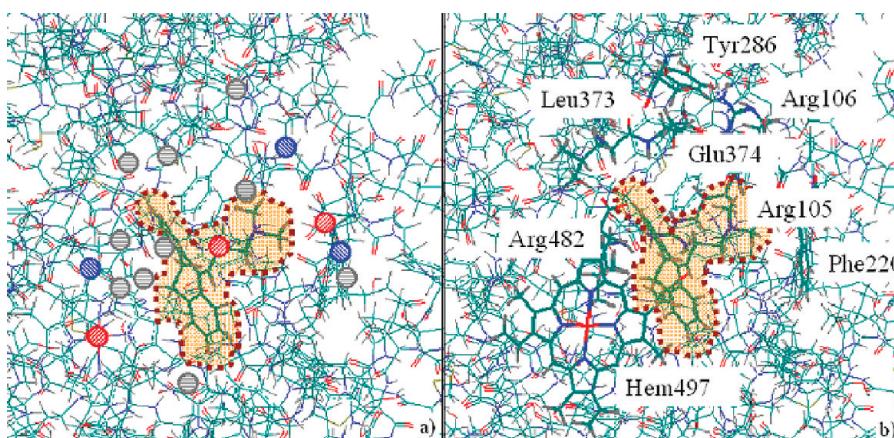
**5.1. The Study of the Substrates of 3A4 Isoform of the Cytochrome P450.** A prognosis of new prospective drugs must be accompanied by the prognosis of their metabolism for partial solution of the ADMET problem. It is well-known that the isoforms of cytochrome P450 play key roles in the metabolism of compounds. Therefore, we performed a detailed theoretical analysis of the substrates of the cytochrome P450 3A4 (CYP3A4) with known  $pK_M$  values<sup>45–71</sup> ( $K_M$  - Michaelis–Menten constant) with the CoCon algorithm. The substrates are presented in Table 3.

The "CYP3A4-erythromycin" (2j0d<sup>72</sup>) complex taken from the Protein Data Bank<sup>73</sup> was used as a template for modeling of the complexes with other substrates. A theoretical study of the complexes showed that there are many short contacts (the distance between contacting atoms is less than the sum of van der Waals radii) of the substrate atoms with

the atoms of the receptor site and water, and most are lipophilic H...H contacts (Table 3). It is interesting that the lipophilic contacts are formed even though these molecules contain a considerable number of atoms capable of forming hydrogen bonds, such as indinavir, ritonavir, and erythromycin (Figure 12).

In most cases, the metabolized substrate reaction centers are directed to the iron heme atom, but the distances between them are rather large, usually more than 8 Å. It is generally agreed that the metabolism process of cytochrome occurs due to the coordination of a substrate with an oxygen bonded with the heme iron atom. However, X-ray crystallography data show that there is not any oxygen molecule in the receptor site.<sup>72,74</sup> Moreover, the docking demonstrated that the distances between the substrate reaction centers and the heme iron atom are too large for the interaction. In this study, we account only for water molecules found by X-ray crystallography in the template ("CYP3A4-erythromycin" complex<sup>72</sup>). The space between a substrate and the heme can be filled with water molecules (e.g., in the cases of a substrate smaller than erythromycin) in a living organism. The influence of the heme iron atom can be transmitted, such as through the hydrogen-bond net of water molecules according to the relay-race mechanism, to the substrate.<sup>75</sup>

Sometimes the reaction center is covered with the substituents, which prevent its interaction with the iron atom. Figure 13 demonstrates an example of an acetaminophen complex. In such cases, the metabolic process can start from the hydrogen bond formation of the substrate with CYP3A4. In the case of acetaminophen, the hydrogen bond can be between a substrate hydrogen and the nitrogen of Arg105.



**Figure 16.** Active centers of the 3A4 isoform found by CoCon.

The interaction of the substrates with a carbon atom on the aromatic ring of Phe220 is very important for the metabolic processes of the compounds in the isoform 3A4. It was demonstrated that the  $pK_M$  value depends on the energy of the interaction  $E_{C(Phe220)}$  of the molecules with the active center of the receptor. The following expression was obtained using eq 7. It determines  $pK_M$  value of the compounds with a correlation coefficient of  $R = 0.774$  ( $S = 0.15$ ).

$$pK_M = -2.09 - 1.80 \cdot E_{C(Phe220)}$$

The amino acid residue is located at the inlet of the 3A4 isoform cavity. This explains why the most effective interactions with the atom will take place only with large-sized molecules (Indinavir, Taxotere) that fill the isoform cavity and have the strongest interactions with the atoms of the inlet part. It is true that the bigger molecules, Indinavir, Taxotere, Ritonavir, and Tirilazad (see the molecular volumes computed with the second-order MERA approximation (Table 3)), fill the cavity and form contacts with the atoms of the inlet part to interact with the carbon atom of Phe220. Lipophilic intermolecular C...H contacts (Figure 14, carbon atom is circled) are usually formed as a result of such interactions. For the complexes with the large molecules, the C...H distance is equal to 2.3–2.6 Å.

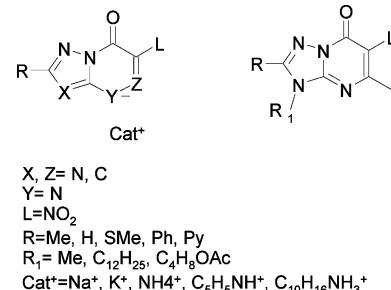
The analysis with CoCon showed that the  $pK_M$  value of the compounds can be calculated with eq 9 with a cross-validation quality of  $Q = 0.92$  ( $R = 0.97$ ,  $S = 0.05$ )

$$pK_M = -3.18 - 0.83 \sum_i E_{i(R)} + 0.53 \sum_j E_{j(W)}$$

The relationship is presented in Figure 15. The first regressor shows that the more effective interactions of the substrates with the active centers of the protein promote the metabolic process, while the second regressor shows that the molecules must be hydrophobic enough to metabolize effectively. The atoms, whose interaction energies are given in the eq 9, are demonstrated in Figure 16a, and the critical amino acid residues are shown in Figure 16b.

**5.2. The Study of Neuraminidase Inhibitors.** Viral neuraminidase plays an important role during the first stages of contamination. Because this protein functions in the distribution of gripe virions to cells, it is a prospective target for prophylaxis and therapy of viral infections.

**Scheme 1.** Neuraminidase Inhibitors

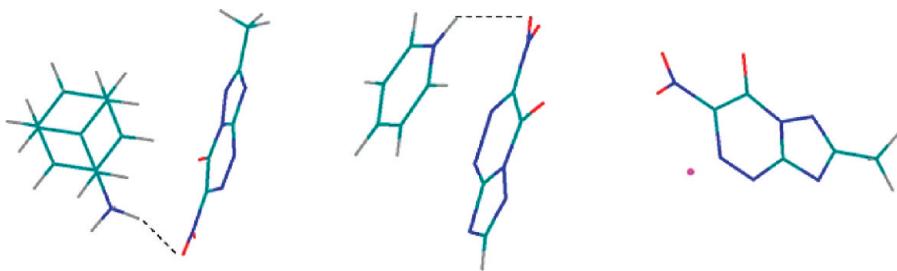


The set of the investigated inhibitors includes azoloanulated triazinone and pyrimidinone derivatives (Scheme 1). Most of them are organic salts (with cations such as sodium, potassium, ammonium, and pyridinium).

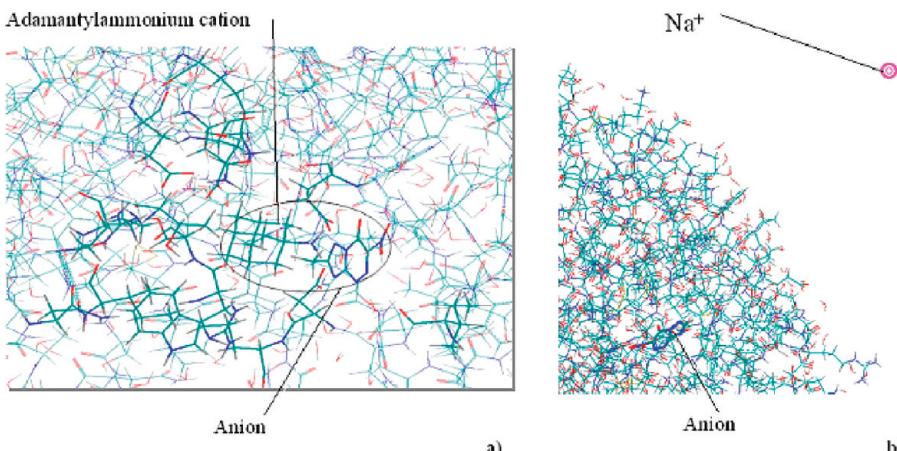
The first stage of the study of the neuraminidase inhibitors included the modeling of the “cation-anion” complexes with the MOPS algorithm,<sup>76</sup> with examples presented in Figure 17. The obtained complexes have been superimposed within the BiS/MC algorithm and then docked to the neuraminidase receptor pocket using **2qwh**<sup>77</sup> as a template.

The investigation of the “neuraminidase-inhibitor” complexes revealed the following:

There are a lot of amino, carboxy, hydroxy, and ketonic groups lining the site. The high hydrophilicity provides a coordination of a great number of water molecules that are of great importance in the “enzyme-inhibitor” complex stabilization. More effective interactions contribute to stronger inhibition. The negatively charged atoms/fragments of the inhibitors are located closely to the positively charged hydrogens of neuraminidase, whereas the positively charged parts are situated at the negatively charged nitrogen and oxygen atoms of neuraminidase carboxy, hydroxy and amino groups. Furthermore, it was found out that the cation plays an important role in the bioactivity of the inhibitors. In most cases, organic cations are inserted in the receptor cavity where they interact with Asp151 (Figure 18a). Their low solvation ability in comparison to inorganic cations may contribute to their association with the neuraminidase site. The methylene fragments of the cations form effective hydrophobic interactions with the hydrophobic groups of Ile222, Arg224, and Ala246. The most effective interactions are formed in the case of the adamantylammonium cation due to its suitable sizes and lipophilic surface. As a rule, sodium and potassium salts have low activity because their



**Figure 17.** The complexes of the cation and anion of organic salts (neuraminidase inhibitors) modeled using MOPS.<sup>76</sup>



**Figure 18.** The complexes of the neuraminidase inhibitors with the enzyme; the organic cation is inserted in the receptor site and the inorganic cation is out of the pocket.

cations associate with water molecules, consequently preventing their insertion into and effective filling of the receptor site (Figure 18b).

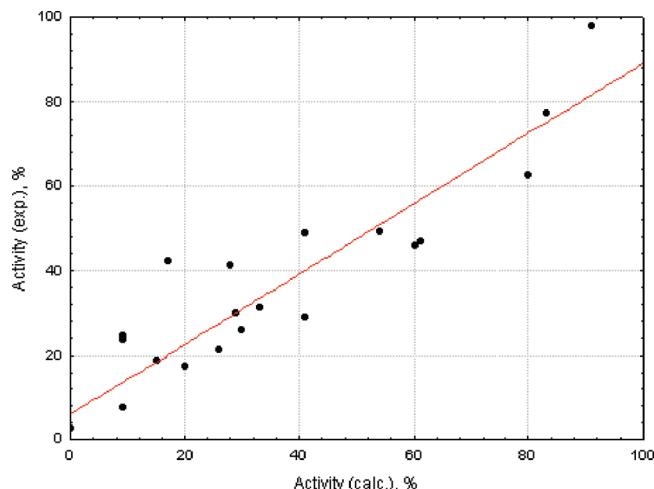
A characteristic feature of the most active compounds is the interaction of nitro groups with Pro431 and with the hydroxyl group of Tyr406. Water molecules located closely to the nitrogen and oxygen atoms of the inhibitors play important roles in the "receptor-ligand" complex formation.

The study of the complexes of the inhibitors with the neuraminidase showed that the active center of the most active compounds is the substituent L (in most cases  $L=\text{NO}_2$ ), which is located close to Glu119, Ile149, and Asp151. The less active compounds are also located near Glu119 and Asp151, but their binding centers are carbonyl oxygen or the heterocyclic nitrogen or hydrogen of the triazole fragment. Such an orientation prevents the effective interactions with Ile149. Moreover, the active centers of the less active molecules are coordinated with water molecules, which makes the interaction with the enzyme hardly possible.

The antiviral activity of the compounds can also be estimated using CoCon eq 11. The cross-validation quality is equal to a value of  $Q = 0.89$  ( $R = 0.91$ ,  $S = 2.6$ ).

$$BA = 25.4 + 9.33 \sum_{i=1}^{N_{(R)}} \sum_{j=1}^{N_{(\text{ray})}} E_{ij(R)} - 7.55 \sum_{k=1}^{N_{(W)}} \sum_{l=1}^{N_{(\text{ray})}} E_{kl(W)}$$

The relationship shows that the stabilization of the "neuraminidase-ligand" complexes by water increases the antiviral activity of the compounds, and this relationship is demonstrated in Figure 19. The amino acids whose parameters of interaction with the antiviral compounds are presented in the equation are Arg118, Glu119, Ile149, Asp151, Arg152, Glu196, Asn221, Ile222, Arg224, Glu227, Ala246,



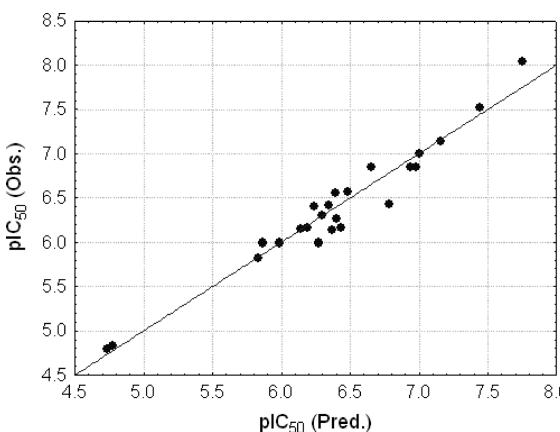
**Figure 19.** Computed with CoCon and experimental activities for the neuraminidase inhibitors.

Glu276, Glu277, Lys292, Arg371, Tyr406, and Pro431. It can be hypothesized that these amino acids will be the most important in the interaction with the antiviral compounds.

**5.3. The Study of p38 MAP Kinase Inhibitors.** Twenty-four inhibitors of the p38 MAP kinase (Table 4) with known inhibitory concentrations were studied.<sup>78</sup> The "inhibitor-p38 MAP kinase" complexes were modeled using the complex of **1bl7**<sup>14</sup> from the Protein Data Bank.<sup>73</sup> The quality of the docking method was estimated by the comparison of the modeled complexes of the compounds presented in Table 8 to existing X-ray crystallographic data. The rms does not exceed 1.00 Å. The modeled complexes showed that the receptor site consists of the amino acids Val30, Gly31, Tyr35, Val38, Ala51, Val52, Lys53, Arg67, Glu71, Leu75, Ile84, Thr106, Leu108, Met109, Ala111, Asp112, Asn115, Asp150,

**Table 4.** p38 MAP Kinase Inhibitors and Their Activities<sup>78</sup>

No	Structure	pIC <sub>50%</sub>	No	Structure	pIC <sub>50%</sub>
1		6.85	13		6.85
2		7.00	14		6.30
3		6.85	15		7.16
4		6.59	16		7.52
5		6.43	17		6.00
6		6.41	18		8.05
7		6.57	19		6.17
8		6.27	20		6.14
9		6.42	21		4.84
10		6.17	22		6.16
11		4.80	23		6.00
12		5.82	24		6.00

**Figure 20.** Computed with CoCon and experimental activities for the p38 MAP kinase inhibitors.

Lys152, Ser154, Asn155, Ala157, Leu167, Asp168, Leu171, Arg173, Thr175, Glu178, and Thr185.

For many compounds, the phenyl of the molecules is placed in parallel with the phenyl of the Tyr35 of the enzyme so that the centers of the aromatic rings are displaced with respect to each other. In addition, the hydrogen atom of the hydroxyl of Tyr35 forms a hydrogen bond with an oxygen of the compounds.

The phenyl and methylene groups of the most active molecules ( $pIC_{50\%} > 8$ ) very often form hydrophobic interactions with the lipophilic part of the enzyme (methyl and methylene groups of Val38, Ala51, Lys53, Leu75, Ile84, Leu108, Met109, Leu167, and Leu171). The distance between the contacting atoms is equal to or greater than 2 Å. The most active molecules form hydrogen bonds, and their most typical feature is a hydrogen bond between the carbonyl oxygen of the dihydrobenzimidazol-2-one system with a hydrogen atom of Met109. Furthermore, the aromatic ring of the most active compounds penetrates deeper into the receptor cavity than the less active compounds.

We determined that the best relationship obtained with CoCon is eq 8, which reproduces the  $pIC_{50\%}$  value of the compounds with a cross-validation quality value equal to  $Q = 0.86$  ( $R = 0.90$ ,  $S = 0.4$ ) (Figure 20)

$$pIC_{50\%} = 13.41 + 1.78 \sum_i E_i$$

The experimental activity of the compounds was determined *in vitro*. At the same time we must keep in mind the ADMET problem. For example, the compounds can be metabolized in a living organism. Therefore, we predicted the possibility of the metabolism of the p38 MAP kinase inhibitors at different isoforms (such as 1A1, 1A2, 2A6, 2B6, 2C18, 2C19, 2D6, 2E1, 2C9, 3A3, 3A4, 3A5, and 3A7) of cytochrome P450. For this aim we used the rules determined previously. One of the rules has been described in item 5.1 with the example of CYP3A4 substrates. Our results suggest that the compounds of the set are metabolized with a high probability in the 2D6 and 3A7 isoforms. In addition, some are metabolized in different isoforms of 3A subfamilies.

**5.4. The Study of Antitumor Dihydrofolatereductase Inhibitors.** Antitumor dihydrofolatereductase (DHFR) inhibitors shown in Table 5 were also considered. Their activities were found in the National Cancer Institute database.<sup>79,80</sup> The molecules were docked to the receptor

**Table 5.** Set of DHFR<sup>a</sup> Inhibitors and Their Biological Activities<sup>79,80b</sup>

Inhibitor	pGI <sub>50%</sub>	Structure	Inhibitor	pGI <sub>50%</sub>	Structure
NSC 623017	6.92		NSC 352122	6.72	
NSC 102816	5.99		NSC 132483	5.94	
NSC 148958	3.17		NSC 264880	4.62	
NSC 143095	5.26		NSC 139105	5.89	
NSC 126771	5.08		NSC 184692	6.82	
NSC 19893	4.59		NSC 174121	8.20	
1rg7	6.60		NSC 163501	5.48	
NSC 368390	5.82		NSC 153353	5.18	
			NSC 134033	6.45	

<sup>a</sup> Number of molecules corresponds to their number in the National Cancer Institute [http://dtp.nci.nih.gov/docs/3d\\_database/](http://dtp.nci.nih.gov/docs/3d_database/) Structural\_information/structural\_data.html. <sup>b</sup> The name of the molecule shown with a ‡ is taken from the Protein Data Bank.<sup>73</sup>

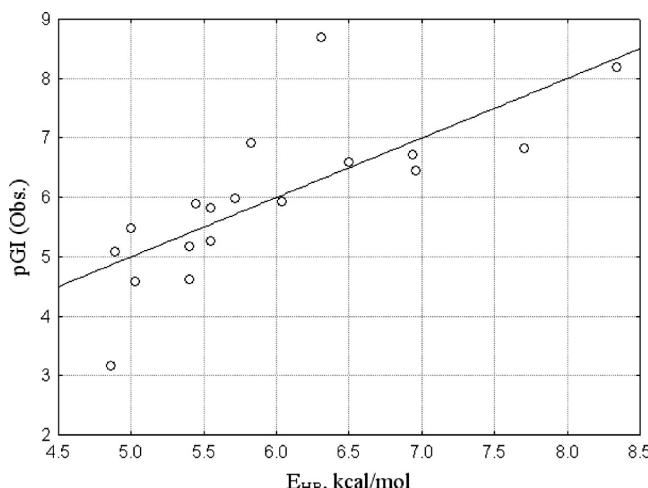
cavity with the BiS/MC algorithm. The DHFR in complex with methotrexate (**1rg7**<sup>81</sup>) was used as the template for the docking. The predicted complex was compared to the experimental one (**1hfp**) found in the Protein Data Bank,<sup>73</sup> and the rms deviations of the computed and experimental coordinates are displayed in Table 2.

The active molecules are larger in size than the less active ones. The modeled complexes revealed that the active molecules fill the receptor cavity more effectively when able to deeply penetrate into the receptor pocket with their large sizes. The less active compounds remain at the inlet of the cavity. It can prevent their effective interaction with the site.

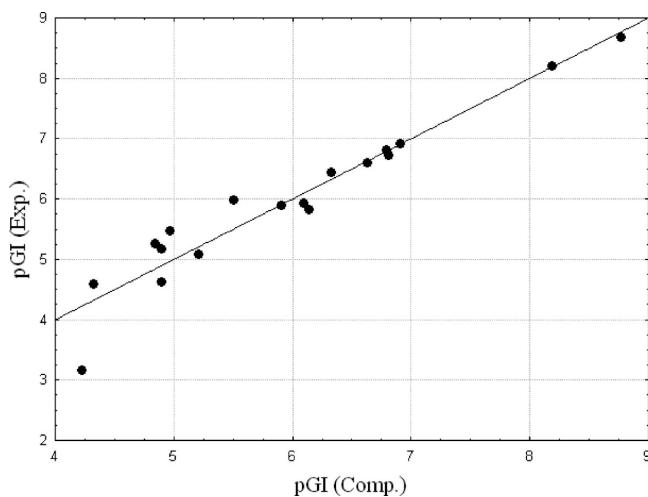
Amino acid residues Ala19, Asn23, Leu28, Asp27, Phe31, Thr46, Ile50, Arg52, and Ile94 form the greatest amount of

short contacts and hydrogen bonds with the inhibitors. A few water molecules included in the receptor site also play important roles in forming short contacts with the ligands. The aromatic system of the molecules is placed in parallel with the phenyl group of Phe31 of the enzyme. Many lipophilic H...H short contacts were found in the modeled "receptor-ligand" complexes. In most cases, the hydrogens of the methyl and/or methylene groups of Ala19, Leu28, Ile50, and Leu54 of the DHFR form short contacts with the lipophilic groups of the molecules -methylene, naphthyl, and phenyl.

The contacts between the electronegative atom (in most cases oxygen) of an inhibitor and the hydrogen of the amino group of Asn23 of the enzyme or/and between the hydrogen



**Figure 21.** The relationship between hydrogen bond energy and antitumor activity for the DHFR inhibitors.



**Figure 22.** Computed with CoCon and experimental activities for the DHFR inhibitors.

of the carboxylic group of an inhibitor with the oxygen atoms of the carboxylic group of the Asp27 of the enzyme are typical of the most active compounds ( $pGI_{50} > 6.9$ ). For the less active compounds, the hydrophilic parts of the molecules form ineffective interactions with the lipophilic parts of the enzyme, such as the methyl and methylene hydrogens of Thr46, Ile94, etc. This may be a cause for the activity decreasing.

The relationships between biological activity and the parameters of the interaction in the “receptor-ligand” complexes were studied with CoCon. The  $pGI_{50}$  values of the compounds depend on the energies of the hydrogen bonds calculated with CoCon eqs 13 and 14. Equation 14 defines the  $pGI_{50\%}$  with a correlation coefficient of  $R = 0.756$  ( $F=21$ ,  $S = 0.88$ ) (Figure 21)

$$pGI_{50\%} = 4.57 + 0.0094 \cdot E_{HB}$$

The best relationship showed that the biological activity of the compounds can be determined with a cross-validation quality equal to  $Q = 0.91$  ( $R = 0.96$ ,  $S = 0.10$ ) using the CoCon eq 11 (Figure 22)

$$pGI_{50\%} = 2.94 - 0.152 \sum_{i=1}^{N(R)} \sum_{j=1}^{N(ray)} E_{ij(R)} + \\ 0.0836 \sum_{k=1}^{N(W)} \sum_{l=1}^{N(ray)} E_{kl(W)}$$

This method allowed for the selection of active centers of the enzyme, which are mostly hydrogens. The amino acids of the active centers are Ile5, Asn23, Pro25, Leu28, Trp30, Phe31, Lys32, Gly43, Arg44, Ser49, Ile50, Gly51, Leu 54, Ile94, and Gly96. This proves the necessity of short contacts between the ligand atoms and the amino acids of the receptor for the biological action of the compounds. The compounds have no pronounced pharmacophoric fragments. Almost the whole structure of the compounds determines their biological activity and interaction with the receptor.

## 6. CONCLUSIONS

A new methodology for docking and study of “receptor-ligand” complexes was suggested. The methodology is based on a combination of the 3D/4D QSAR BiS/MC and CoCon algorithms. The BiS/MC algorithm performs the restricted docking of compounds to receptor pockets taking into account weaknesses of experimental data on 3D structures of biological receptors. The CoCon algorithm is based on the usage of new nonparametrical formalism for description of intermolecular interactions and for the selection of active centers of the ligands and the receptor. This algorithm determines the relationships between the bioactivity and the parameters of interactions in the “receptor-ligand” complexes, including a new formalism for estimating hydrogen bond energies. The methodology has been tested on different sets of compounds. The restricted docking of approximately 70 compounds (rhinovirus HRV14, thermolysin, CDK2, elastase, DHFR, p38 MAP-kinase inhibitors, DNA-antimethabolites) showed that modeled complexes are in good agreement with experimental ones determined by X-ray crystallography. The relationships between the biological activity and interaction energies/forces were carried out for DHFR, neuraminidase, p38 MAP kinase inhibitors, and substrates of cytochrome P450 (3A4 isoform). The relationships determined the activities of the compounds with cross-validation qualities not less than 0.86. Moreover, the suggested methodology allows for the selection of the protein areas that are responsible for the interaction with molecules. The rest of the protein site plays a less important role in the interaction with the ligands, and it can be in any conformational or tautomeric state. Easily interpretable results are obtained and provide information for fundamental research on biological mechanisms. Moreover, this methodology can provide quantitative models that can be used for predicting the bioactivity of new compounds.

## ACKNOWLEDGMENT

The work is supported by RFBR (grants 07-03-96041, 07-04-96053), the Human Capital Foundation, and the supercomputer SKIF-GRID initiative. In addition, we are thankful to Prof. Vladimir L. Rusinov for the data on the bioactivities of antiviral compounds and to Nadezhda Vishnyakova for thoughtful discussions.

**Supporting Information Available:** Modeled complexes considered in item 5 and marked active centers (\*.hin files) and table with the probability of the metabolism of the considered p38 MAP kinase inhibitors in the metabolism\_p38map.doc. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## REFERENCES AND NOTES

- (1) Evrard, G. X.; Langer, G. G.; Perrakis, A.; Lamzin, V. S. Assessment of automatic ligand building in ARP/wARP. *Acta Crystallogr.* **2007**, *D63*, 108–117.
- (2) Rawal, R. K.; Kumar, A.; Siddiqi, M. I.; Katti, S. B. Molecular docking studies on 4-thiazolidinones as HIV-1RT inhibitors. *J. Mol. Model.* **2007**, *13*, 155–161.
- (3) Cai, W.; Xia, B.; Shao, X.; Guo, Q.; Maigret, B. Pan, Zh. Effect of  $\alpha$ -cyclodextrin coating on electronic properties of molecular wires. *Chem. Phys. Lett.* **2001**, *342*, 387–396.
- (4) Dadras, O. G.; Sobhani, A. M.; Shafee, A.; Mahmoudian, M. Flexible ligand docking studies of matrix metalloproteinase inhibitors using Lamarckian genetic algorithm. *Daru* **2004**, *12*, 1–10.
- (5) Gorse, A.-D.; Gready, J. E. Molecular dynamics simulations of the docking of substituted N5-deazapterins to dihydrofolate reductase. *Protein Eng.* **1997**, *10*, 23–30.
- (6) Laskowski, R. A.; Rullmann, J. A.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **1996**, *8*, 477–486.
- (7) Westbrook, J.; Feng, Z.; Jain, Sh.; Bhat, N.; Thanki, N.; Ravichandran, V.; Gilliland, G. L.; Bluhm, W.; Weissig, H.; Greer, D. S.; Bourne, Ph. E.; Berman, H. M. The Protein Data Bank: unifying the archive. *Nucleic Acids Res.* **2002**, *30*, 245–248.
- (8) Venclovas, C.; Ginalski, K.; Kang, Ch. Sequence-structure mapping errors in the PDB: OB-fold domains. *Protein Sci.* **2004**, *13*, 1594–1602.
- (9) Sippl, M. J. Recognition of errors in three-dimensional structures of proteins. *Proteins* **1993**, *17*, 355–362.
- (10) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
- (11) Hooft, R. W.; Sander, C.; Vriend, G. Verification of Protein Structures: Side-Chain Planarity. *J. Appl. Crystallogr.* **1996**, *29*, 714–716.
- (12) Hooft, R. W.; Vriend, G.; Sander, C.; Abola, E. E. Errors in protein structures. *Nature* **1996**, *381*, 272–272.
- (13) Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Biol.* **2002**, *9*, 268–272.
- (14) Wang, Z.; Canagarajah, B. J.; Boehm, J. C.; Kassisa, S.; Cobb, M. H.; Young, P. R.; Abdel-Meguid, S.; Adams, J. L.; Goldsmith, E. J. Structural basis of inhibitor selectivity in MAP kinases. *Structure* **1998**, *6*, 1117–1128.
- (15) Shewchuk, L.; Hassell, A.; Wisely, B.; Rocque, W.; Holmes, W.; Veal, J.; Kuyper, L. F. Binding mode of the 4-anilinoquinazoline class of protein kinase inhibitor: X-ray crystallographic studies of 4-anilino-quinazolines bound to cyclin-dependent kinase 2 and p38 kinase. *J. Med. Chem.* **1999**, *43*, 133–138.
- (16) Iwase, K.; Hirono, Sh. Estimation of active conformations of drugs by a new molecular superposing procedure. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 499–512.
- (17) Lee, M. C.; Yang, R.; Daun, Y. Comparison between Generalized-Born and Poisson-Boltzmann methods in Physics Based Scoring Functions for Protein Structure Prediction. *J. Mol. Model.* **2005**, *12*, 101–110.
- (18) Koehler, R. T.; Villar, H. O. Statistical relationships among docking scores for different protein binding sites. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 23–37.
- (19) Olsen, L.; Petterson, I.; Hemmingsen, L. Docking and scoring of metallo-beta-lactamase inhibitors. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 287–302.
- (20) Ha, S.; Andreani, R.; Robbins, A.; Muegge, I. Evaluation of docking/scoring approaches: a comparative study based on MMP3 inhibitors. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 435–448.
- (21) Kubinyi, H. Combinatorial and computational approaches in structure-based drug design. *Curr. Opin. Drug Discovery Dev.* **1998**, *1*, 16–27.
- (22) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
- (23) Kubinyi, H.; Hamprecht, F. A.; Mietzner, T. Three-dimensional quantitative similarity-activity relationships (3D QSAR) from SEAL similarity matrices. *J. Med. Chem.* **1998**, *41*, 2553–2564.
- (24) Golbraikh, A.; Tropsha, A. Beware of q2! *J. Mol. Graph. Modell.* **2002**, *20*, 269–276.
- (25) Novellino, E.; Fattorusso, C.; Greco, G. Use of comparative molecular field analysis and cluster analysis in series design. *Pharm. Acta Helv.* **1995**, *70*, 149–154.
- (26) Norinder, U. Single Domain Mode Variable Selection in 3D QSAR Applications. *J. Chemom.* **1996**, *10*, 95–105.
- (27) Kubinyi, H. A general view on similarity and QSAR studies. In *Computer-Assisted Lead Finding and Optimization*, 1st ed.; van de Waterbeemd, H., Testa, B., Folkers, G., Eds.; Wiley-VCH: Zurich, Switzerland, 1997; pp 9–28.
- (28) Cosgrove, D. A.; Bayada, D. M.; Johnson, A. P. A novel method of aligning molecules by local surface shape similarity. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 573–591.
- (29) Potemkin, V. A.; Bartashevich, E. V.; Grishina, M. A.; Guccione, S. In *Rational Approaches to Drug Design*; Proceedings of the 13th European Symp. on Quantitative Structure-Activity Relationships, Dusseldorf, Germany, 2000; Holtje, H.-D.; Sippl, W., Eds.; Prous Science Publishers: Barcelona, Spain, 2001; pp 349–353.
- (30) Potemkin, V. A.; Grishina, M. A.; Bartashevich, E. V. Modeling of drug molecule orientation within a receptor cavity in the BiS algorithm framework. *J. Struct. Chem.* **2007**, *48*, 155–160.
- (31) Potemkin, V. A.; Arslambekov, R. M.; Bartashevich, E. V.; Grishina, M. A.; Belik, A. V.; Perspicace, S.; Guccione, S. Multiconformational Method for Analyzing the Biological Activity of Molecular Structures. *J. Struct. Chem.* **2002**, *43*, 1045–1049.
- (32) Potemkin, V. A.; Grishina, M. A. A new paradigm for pattern recognition of drugs. *J. Comput.-Aided. Mol. Des.* **2008**, *22*, 489–505.
- (33) Potemkin, V. A.; Bartashevich, E. V.; Belik, A. V. New approaches to prediction of thermodynamic parameters of substances using molecular data. *Russ. J. Phys. Chem.* **1996**, *70*, 411–416.
- (34) Mikuchina, K.; Potemkin, V.; Grishina, M.; Laufer, S. 3D QSAR analysis and pharmacophore modelling of p38 MAP kinase inhibitors using BiS algorithm. *Arch. Pharm. Pharm. Med. Chem.* **2002**, *335*, C74.
- (35) Grishina, M. A.; Mikushina, K. M.; Potemkin, V. A. In *Proceedings of the 16th European Symposium on Quantitative Structure-Activity Relationships and Molecular Modelling*; Proceedings of the 16th European Symposium on Quantitative Structure-Activity Relationships and Molecular Modelling, Mediterranean Sea, Italy, 2006; Cruciani G., Ed.; ITALGRAF: Rome, 2006; p 241.
- (36) Tsirelson, V.; Stash, A.; Zhurova, E.; Zhurov, V.; Pinkerton, A. A.; Zavodnik, V.; Shutalev, A.; Gurskaya, G.; Rykounov, A.; Potemkin, V. Electron-density Properties of the Functionally-substituted Hydro-pyrimidines. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **2005**, *A61*, C427.
- (37) Grant, J. A.; Pickup, B. T. A Gaussian description of molecular shape. *J. Phys. Chem.* **1995**, *99*, 3503–3510.
- (38) Grant, J. A.; Gallardo, M. A.; Pickup, B. T. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* **1996**, *17*, 1653–1666.
- (39) Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548–558.
- (40) Connolly, M. L. Computation of molecular volume. *J. Am. Chem. Soc.* **1985**, *107*, 1118–1124.
- (41) Mortier, W. J.; Van Geenchten, K.; Gasteiger, J. Electronegativity equalization-application and parametrization. *J. Am. Chem. Soc.* **1985**, *107*, 829–835.
- (42) Grishina, M. A.; Bartashevich, E. V.; Potemkin, V. A.; Belik, A. V. Genetic Algorithm for Predicting Structures and Properties of Molecular Aggregates in Organic Substances. *J. Struct. Chem.* **2002**, *43*, 1040–1043.
- (43) Bartashevich, E. V.; Potemkin, V. A.; Grishina, M. A.; Belik, A. V. A Method for Multiconformational Modeling of the Three-Dimensional Shape of a Molecule. *J. Struct. Chem.* **2002**, *43*, 1033–1039.
- (44) Kirkwood, J. G. Quantum Statistics of Almost Classical Assemblies. *Phys. Rev.* **1933**, *44*, 31–37.
- (45) Voorman, R. L.; Maio, S. M.; Hauer, M. J.; Sanders, P. E.; Payne, N. A.; Ackland, M. J. Metabolism of Delavirdine, a Human Immunodeficiency Virus Type-1 Reverse Transcriptase Inhibitor, by Microsomal Cytochrome P450 in Humans, Rats, and Other Species: Probable Involvement of CYP2D6 and CYP3A. *Drug Metab. Dispos.* **1998**, *26*, 631–639.
- (46) Kudo, S.; Okumura, H.; Miyamoto, G.; Ishizaki, T. Cytochrome P-450 Isoforms Involved in Carboxylic Acid Ester Cleavage of Hantzsch Pyridine Ester of Pranidipine. *Drug Metab. Dispos.* **1999**, *27*, 303–308.
- (47) Patten, C. J.; Thomas, P. E.; Guy, R. L.; Lee, M.; Gonzalez, F. J.; Guengerich, F. P.; Yang, C. S. Cytochrome P450 enzymes involved in acetaminophen activation by rat and human liver microsomes and their kinetics. *Chem. Res. Toxicol.* **1993**, *6*, 511–518.

- (48) Wienkers, L. C.; Steenwyk, R. C.; Sanders, P. E.; Pearson, P. G. Biotransformation of tirilazad in human: I. Cytochrome P450 3A-mediated hydroxylation of tirilazad mesylate in human liver microsomes. *J. Pharmacol. Exp. Ther.* **1996**, *277*, 982–990.
- (49) Ghosal, A.; Satoh, H.; Thomas, P. E.; Bush, E.; Moore, D. Inhibition and kinetics of cytochrome P4503A activity in microsomes from rat, human, and cdna-expressed human cytochrome P450. *Drug Metab. Dispos.* **1996**, *24*, 940–947.
- (50) Nielsen, T. L.; Rasmussen, B. B.; Flinois, J. P.; Beaune, P.; Brosen, K. In Vitro Metabolism of Quinidine: The (3*S*)-3-Hydroxylation of Quinidine Is a Specific Marker Reaction for Cytochrome P-4503A4 Activity in Human Liver Microsomes. *J. Pharmacol. Exp. Ther.* **1999**, *289*, 31–37.
- (51) Rochat, B.; Amey, M.; Gillet, M.; Meyer, U. A.; Baumann, P. Identification of three cytochrome P450 isozymes involved in N-demethylation of citalopram enantiomers in human liver microsomes. *Pharmacogenetics* **1997**, *7*, 1–10.
- (52) Ishiguro, N.; Senda, C.; Kishimoto, W.; Sakai, K.; Funae, Y.; Igarashi, T. Identification of CYP3A4 as the predominant isoform responsible for the metabolism of ambroxol in human liver microsomes. *Xenobiotica* **2000**, *30*, 71–80.
- (53) Olesen, O. V.; Linnet, K. Metabolism of the tricyclic antidepressant amitriptyline by cDNA-expressed human cytochrome P450 enzymes. *Pharmacology* **1997**, *55*, 235–243.
- (54) Haaz, M. C.; Rivory, L.; Richâa, C.; Vernillet, L.; Robert, J. Metabolism of CPT-11 by Human Hepatic Microsomes: Participation of Cytochrome P-450 3A and Drug Interactions. *Cancer. Res.* **1998**, *58*, 468–472.
- (55) Koudriakova, T.; Iatsimirskaia, E.; Utkin, I.; Gangl, E.; Vouros, P.; Storozhuk, E.; Orza, D.; Marinina, J.; Gerber, N. Metabolism of the Human Immunodeficiency Virus Protease Inhibitors Indinavir and Ritonavir by Human Intestinal Microsomes and Expressed Cytochrome P4503A4/3A5: Mechanism-Based Inactivation of Cytochrome P4503A by Ritonavir. *Drug Metab. Dispos.* **1998**, *26*, 552–561.
- (56) Shou, M.; Martinet, M.; Korzekwa, K. R.; Krausz, K. W.; Gonzalez, F. J.; Gelboin, H. V. Role of human cytochrome P450 3A4 and 3A5 in the metabolism of taxotere and its derivatives: enzyme specificity, interindividual distribution and metabolic contribution in human liver. *Pharmacogenetics* **1998**, *8*, 391–401.
- (57) Yaïch, M.; Popon, M.; Mädard, Y.; Aigrain, E. J. In-vitro cytochrome P450 dependent metabolism of oxybutynin to N-deethoxybutynin in humans. *Pharmacogenetics* **1998**, *8*, 449–451.
- (58) Dupont, I.; Berthou, F.; Bodenez, P.; Bardou, L.; Guiriec, C.; Stephan, N.; Dreano, Y.; Lucas, D. Involvement of cytochromes P450 2E1 and 3A4 in the 5-hydroxylation of salicylate in humans. *Drug Metab. Dispos.* **1999**, *27*, 322–326.
- (59) Tracy, T. S.; Korzekwa, K. R.; Gonzalez, F. J.; Wainer, I. W. Cytochrome P450 isoforms involved in metabolism of the enantiomers of verapamil and norverapamil. *Br. J. Clin. Pharmacol.* **1999**, *47*, 545–552.
- (60) Zhang, Q. Y.; Dunbar, D.; Kaminsky, L. Human cytochrome P-450 metabolism of retinals to retinoic acids. *Drug Metab. Dispos.* **2000**, *28*, 292–297.
- (61) Ganterbein, M.; Attolini, L.; Bruguerolle, B.; Villard, P. H.; Puyouou, F.; Durand, A.; Lacarelle, B.; Hardwigsen, J.; Le-Treut, Y. P. Oxidative metabolism of bupivacaine into pipecolylxylidine in humans is mainly catalyzed by CYP3A. *Drug Metab. Dispos.* **2000**, *28*, 383–385.
- (62) Prakash, C.; Kamel, A.; Cui, D.; Whalen, R. D.; Miceli, J. J.; Tweedie, D. Identification of the major human liver cytochrome P450 isoform(s) responsible for the formation of the primary metabolites of ziprasidone and prediction of possible drug interactions. *Br. J. Clin. Pharmacol.* **2000**, *49*, 35S–42S.
- (63) Tang, C.; Shou, M.; Mei, Q.; Rushmore, T. H.; Rodrigues, A. D. Major role of human liver microsomal cytochrome P450 2C9 (CYP2C9) in the oxidative metabolism of celecoxib, a novel cyclooxygenase-II inhibitor. *J. Pharmacol. Exp. Ther.* **2000**, *293*, 453–459.
- (64) Bourque, B.; Lambert, N.; Boukaiba, R.; Martinet, M. In vitro metabolism and drug interaction potential of a new highly potent anti-cytomegalovirus molecule, CMV423 (2-chloro 3-pyridine 3-yl 5,6,7,8-tetrahydroindolizine 1-carboxamide). *Br. J. Clin. Pharmacol.* **2001**, *52*, 53–63.
- (65) Molden, E.; Asberg, A.; Christensen, H. Desacetyl-Diltiazem Displays Severalfold Higher Affinity to CYP2D6 Compared with CYP3A4. *Drug Metab. Dispos.* **2002**, *30*, 1–3.
- (66) Zhou, Q.; Yao, T. W.; Yu, Y. N.; Zeng, S. Stereoselective metabolism of propafenone by human liver CYP3A4 expressed in transgenic Chinese hamster CHL cells lines. *Acta Pharmacol. Sin.* **2001**, *22*, 944–948.
- (67) Zhao, G.; Allis, J. W. Kinetics of bromodichloromethane metabolism by cytochrome P450 isoenzymes in human liver microsomes. *Chem. Biol. Interact.* **2002**, *140*, 155–168.
- (68) Kalgutkar, A. S.; Taylor, T. J.; Venkatakrishnan, K.; Isin, E. M. Assessment of the contributions of CYP3A4 and CYP3A5 in the metabolism of the antipsychotic agent haloperidol to its potentially neurotoxic pyridinium metabolite and effect of antidepressants on the bioactivation pathway. *Drug Metab. Dispos.* **2003**, *31*, 243–249.
- (69) Zhang, W.; Ramamoorthy, Y.; Tyndale, R. F.; Sellers, E. M. Interaction of buprenorphine and its metabolite norbuprenorphine with cytochromes p450 in vitro. *Drug Metab. Dispos.* **2003**, *31*, 768–772.
- (70) Rodrigues, A. D.; Mulford, D. J.; Lee, R. D.; Surber, B. W.; Kukulka, M. J.; Ferrero, J. L.; Thomas, S. B.; Shet, M. S.; Estabrook, R. W. In vitro metabolism of terfenadine by a purified recombinant fusion protein containing cytochrome P4503A4 and NADPH-P450 reductase. Comparison to human liver microsomes and precision-cut liver tissue slices. *Drug Metab. Dispos.* **1995**, *23*, 765–775.
- (71) Wang, R. W.; Newton, D. J.; Scheri, T. D.; Lu, A. Y. H. Human cytochrome P450 3A4-catalyzed testosterone 6 $\beta$ -hydroxylation and erythromycin N-demethylation: competition during catalysis. *Drug Metab. Dispos.* **1997**, *25*, 502–507.
- (72) Ekroos, M.; Sjogren, T. Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 13682–13687.
- (73) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissenig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (74) Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. Crystal structures of human cytochrome P450 3A4 bound to metyrapone and progesterone. *Science* **2004**, *305*, 683–686.
- (75) Rozhansky, V. A.; Tsendin, L. D. *Transport Phenomena in Partially Ionized Plasma*; Taylor & Francis: London and New York, 2001; p 469.
- (76) Sinyayev, A. A.; Grishina, M. A.; Potemkin, V. A. Theoretical study of solvent influence on the regiospecificity of the reaction of 3-phenyl-s-tetrazine with ketene-N,N-aminal. *ARKIVOC* **2004**, *XI*, 43–52.
- (77) Varghese, J. N.; Smith, P. W.; Sollis, S. L.; Blick, T. J.; Sahasrabudhe, A.; McKimm-Breschkin, J. L.; Colman, P. M. Drug design against a shifting target: A structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. *Structure (London)* **1998**, *6*, 735–746.
- (78) Dombroski, M. A.; Letavic, M. A.; McClure, K. F.; Barberia, J. T.; Cartt, T. J.; Cortina, S. R.; Csiki, C.; Dipesa, A. J.; Elliott, N. C.; Gabel, C. A.; Jordan, C. K.; Labasi, J. M.; Martin, W. H.; Peese, K. M.; Stock, I. A.; Svensson, L.; Sweeney, F. J.; Yu, C. H. Benzimidazolone p38 inhibitors. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 919–923.
- (79) National Cancer Institute database. DTP/2D and 3D Structural Information. [http://dtp.nci.nih.gov/docs/3d\\_database/Structural\\_information/structural\\_data.html](http://dtp.nci.nih.gov/docs/3d_database/Structural_information/structural_data.html) (accessed December 18, 2008).
- (80) van Osdol, W. W.; Myers, T. G.; Paull, K. D.; Kohn, K. W.; Weinstein, J. N. Use of the Kohonen self-organizing map to study the mechanisms of action of chemotherapeutic agents. *J. Natl. Cancer Inst.* **1994**, *86*, 1853–1859.
- (81) Sawaya, M. R.; Kraut, J. Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: crystallographic evidence. *Biochemistry* **1997**, *36*, 586–603.

CI800405N