Virtual screening of organic molecule databases. Design of focused libraries of potential ligands of NMDA and AMPA receptors ✽

I. G. Tikhonova, I. I. Baskin, V. A. Palyulin, and N. S. Zefirov★

Department of Chemistry, M. V. Lomonosov Moscow State University, Leninskie Gory, 119992 Moscow, Russian Federation. Fax: +7 (095) 939 0290. E-mails: tikiri@org.chem.msu.su, zefirov@org.chem.msu.su vap@org.chem.msu.su

A system of virtual screening of organic molecule databases is designed, which permits preprocessing of databases, molecular docking to a three-dimensional model of receptor, and post-processing of the results obtained. Using this screening system, it is possible to reproduce positions of the known ligands in the glutamate sites of the NMDA and AMPA receptors and in the glycine site of the NMDA receptor, to substantially enrich the database with potentially active compounds, and to distinguish between the agonistic and antagonistic character of the action of these compounds in the case of docking to the open and closed forms of the binding sites. Based on the results of screening of a database of low-molecular-weight organic compounds (total of 135,000 structures) using models of the open and closed forms of the glutamate and glycine sites of the NMDA receptor and of the glutamate site of the AMPA receptor, focused libraries of potential agonists and antagonists of these sites were designed.

Key words: virtual screening, *in silico* screening, molecular docking, focused libraries of compounds.

At present, the number of organic compounds is so large that it is highly probable to find potential ligands of particular receptors or their prototypes in the available organic molecule databases. The case in point are the structures that are complementary to binding sites of agonists or antagonists of a biological target under study. Computer-assisted search for such structures can make the design of new active ligands much simpler and cheaper. These efforts can lead to the finding of particular molecular structures or to the design of focused libraries of ligand structures for particular targets.

Virtual screening (or *in silico* screening) of databases of low-molecular-weight compounds is performed by docking molecules to the binding site of a target. The docking procedure is a kind of computational experiment that allows one to determine a possible ligand orientation in the binding pocket of a protein. The results obtained are primarily assessed using the so-called scoring functions (SFs) that allow one to select ligands with best-match structure. The SFs can be calculated using physicochemical characteristics (with inclusion of the energy contributions of hydrophobic, electrostatic, and dispersion interactions, H-bonds, and solvation effects) or based on statistical analysis of the known structures of protein—ligand complexes (the so-called knowledge-based potentials). Then, the best results are selected using various criteria.

One of the first and still most widely used ligand docking programs is the DOCK program.[1] DOCK fills the binding pocket of the target macromolecule with spheres that locally match the macromolecule surface. Then, a database containing three-dimensional structures of organic molecules is screened to find potential ligands (molecules with the interatomic distances equal to the separations between spheres of the negative image of the binding pocket). The original DOCK algorithm included only geometrical matching. Then, additional important features for database processing including molecular force-field scoring, chemical identification of spheres, and the inclusion of hydration were introduced. The latest release of DOCK, version 4,[2] incorporates an algorithm for flexible ligand docking, namely, it is possible to perform a synchronous search with simultaneous variation of all torsion angles or to implement the anchor-first strategy (here, docking of a rigid fragment is followed by incremental construction of flexible fragments).

The DOCK program was successfully used to generate the leader structures of such important targets as HIV-1 protease,[3] dihydropholate reductase,[4] hemagglutinin,[5] malaria plasmodium protease,[6] and thymidilate synthase.[7]

Published in Russian in *Izvestiya Akademii Nauk. Seriya Khimicheskaya,* No. 6, pp. 1282—1291, June, 2004.

1066-5285/04/5306-1335

The known virtual screening systems (VSSs) use different database selection procedures and methods of assessment of the results obtained.[8—12] We designed a VSS, which permits search for new ligands of glutamate receptors based on their structural and other features and excludes potential toxic structures. We also propose to assess the results using QSAR models constructed in the framework of the comparative molecular field analysis (CoMFA) approach.[13]

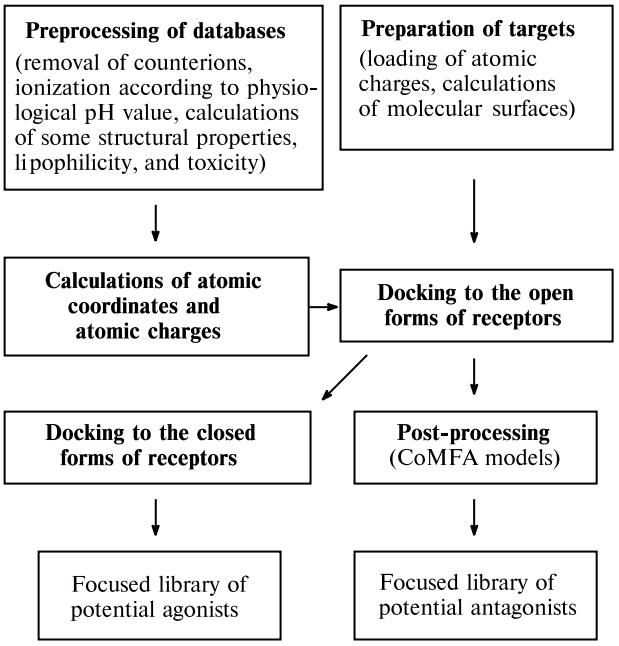
In this work we performed virtual screening of an organic molecule database containing a total of 135,000 structures retrieved from the NCI (http://cactus.cit.nih.gov/ncidb2/download.html), MayBridge (http://www.maybridge.com/html/database_sign1.htm), and ASINEX (http://www.asinex.com/prod/plat.html) databases (all of them were assumed to be public domain) using three-dimensional models of the open and closed forms of the glutamate[14,15] and glycine[16,17] sites of the NMDA receptor (named after the selective ligand, *N*-methyl-D-aspartic acid). In addition, we report on search for potential ligands of the glutamate site of the AMPA receptor (named after the selective ligand, 2-amino-3-(3-hydroxy-5-methylisoxazol-4-yl)propionic acid) using experimental data on the structure[18] of a water soluble AMPA-sensitive protein (AMPASGR).

Screening Strategy and Docking Procedure

Virtual Screening System: General. In the framework of our system, virtual screening of organic molecule databases includes 1) preprocessing of a database (removal of counterions, ionization according to physiological pH value, calculations of structural parameters, lipophilicity, and toxicity and rejection of *a priori* inappropriate structures based on the structural parameters, lipophilicity and toxicity criteria; 2) preparation of targets (loading of atomic charges and performing some procedures necessary for correct operation of particular docking programs); 3) docking of a database; and 4) assessment of the results obtained and post-processing (selection based on the binding energy scoring functions and on the QSAR equations constructed by, *e.g.*, the CoMFA method[13]). The screening block scheme employed in this work (Scheme 1) is shown below.

Scheme 1

Block scheme of screening of organic molecule databases



Preprocessing of databases of low-molecular-weight compounds. Preprocessing of databases included (1) correction of the chemical structures in the databases, namely, removal of counterions, ionization of functional groups (deprotonation of carboxyl groups, protonation of amino groups, *etc.*); 2) filtering of databases based on the values of the structural parameters specific to the ligands of glutamate receptors and based on lipophilicity and toxicity, namely, removal of inorganic compounds; selection based on the molecular weight M (M 50—500); calculations of lipophilicity (log*P*) and selection based on the log*P* value (compounds characterized by the predicted log*P* values lying between 0 and 5 were selected); based on the number of H-bond donors and H-bond acceptors (at most 5 and 10, respectively); based on the number of the base and acid atoms (at most two and three atoms, respectively); based on the number of non-hydrogen atoms in the molecule (at least six atoms); based on the total charge (from 0 to −1); based on toxicity (compounds containing none out of seventy-three "toxophors" were selected); and selection of compounds containing at most eight conformationally flexible bonds, at most ten heteroatoms, and at most five halogen atoms; 3) transformation of two-dimensional coordinates of the structures into three-dimensional coordinates using the CONCORD program,[19] and calculations of charges according to the Gasteiger—Hückel procedure.[19]

The first two tasks were done with a script written in TCL language using the CACTVS command system.[20] Lipophilicity (log*P*) calculations for each structure were carried out using the XLOGP scheme[21] implemented in the script mentioned above.

To assess the toxicity of compounds, we used a database containing a total of seventy-three "toxophors" used by the LigBuilder program.[22] The ranges of parameter values for selection based on the molecular weight, log*P*, and the number of H-bond donors and H-bond acceptors were chosen using the Lipinski rules,[23] according to which compounds of low absorbability are characterized by $M > 500$, $\log P > 5$, by the number of H-bond donors (NH and OH groups) greater than five, and by the number of H-bond acceptors (total number of O and N atoms) greater than ten.

To construct three-dimensional molecular geometries and perform charge calculations in the framework of the Sybyl 6.7.2 molecular graphics package,[19] an SPL[19] script was written.

Docking. Docking of structures from organic molecule databases was carried out using the DOCK4.0 program.

The protein conformations chosen for docking were as follows: 1) closed, agonist-binding forms, namely, a conformation in which the AMPA receptor is bound to glutamate (X-ray dif-

fraction data for AMPASGR—glutamate[18]); a conformation in which the glutamate site of the NMDA receptor is bound to glutamate;[14,15] a conformation in which the glycine site of the NMDA receptor is bound to glycine[16,17] (last two conformations were found earlier in our molecular dynamics calculations); 2) open, antagonist-binding forms, namely, a conformation in which the AMPA receptor is bound to DNQX (X-ray diffraction data for AMPASGR—DNQX);[18] a conformation in which the glutamate site of the NMDA receptor is bound to AP5;[14,15] a conformation in which the glycine site of the NMDA receptor is bound to 7-chloro-6-methyl-5-nitroquinoxaline-2,3-dione[16,17] (the last two conformations were obtained earlier in our molecular dynamics calculations).

The partial atomic charges for protein molecules were taken from the UNITED/KOLLMAN library of the BIOPOLYMER module (Sybyl 6.7.2 molecular graphics package[19]). The sets of spheres describing the binding pockets of the target proteins were calculated using the molecular surfaces pre-calculated with the ms program.[24] The ranges of sphere radii were 1.4—4 Å for the open forms and 1.4—2.5 Å for the closed forms of receptors.

Ligand flexibility was taken into account using the anchor-first search strategy. To optimize the screening parameters, a number of typical agonists and antagonists were docked to the targets under study. Docking was accompanied by variation of the grid spacing, the number of clusters and the number of spheres in clusters, the docking cube size, the number of orientations per conformer, maximum number of orientations (in the case of the anchor-first search strategy), and the energy cut-off distance included in the SF calculations. The SF is an approximation of the binding energy obtained from molecular mechanics calculations and includes the van der Waals and electrostatic components of the energy of the interaction between the ligand (L) and receptor (R) atoms:

$$E = \sum_{i=1}^{L}\sum_{j=1}^{R}\left[A_{ij}/r_{ij}^{a} - B_{ij}/r_{ij}^{b} + 332(q_i q_j)/(Dr_{ij})\right],$$

where E is the intermolecular interaction energy; r_{ij} is the distance from the ith atom of the ligand to the jth atom of the receptor; A_{ij} and B_{ij} are the van der Waals attraction and repulsion parameters; a and b are the empirical exponents of the van der Waals attraction and repulsion functions; q_i and q_j are the point atomic charges of i and j; D is the dielectric constant; and 332 is the coefficient of conversion into kcal mol^{-1}. The coefficients A_{ij} and B_{ij} are calculated as follows

$$A_{ij} = \varepsilon[b/(a-b)] \cdot (2R)^{a}$$

and

$$B_{ij} = \varepsilon[a/(a-b)] \cdot (2R)^{b},$$

where R is the van der Waals atomic radius and ε is the depth of the potential well for the binding energy.

Finally, we obtained a set of parameters, which permits a rather accurate and fast reproduction of both the results of X-ray diffraction studies of AMPA receptor—ligand complexes (RMS ≈ 1 Å) and the structures found from molecular dynamics calculations of complexes of the glycine and glutamate sites of NMDA receptors with agonists and antagonists (RMS ≈ 1.2 Å). These parameters are listed in Table 1 along with the CPU time taken for docking of one structure. Calculations were carried out on an SGI Octane 2 (R12000) workstation.

Results and Discussion

Docking Approbation

The efficiency of this VSS was evaluated taking its ability to retrieve the known ligands from databases as an example. To this end, we used a database containing a total of 1,000 randomly chosen organic compounds including twenty active ligands (ten agonists and ten antagonists) for each target under study (Fig. 1). Compounds from this database were docked to the closed and open forms of the AMPA and NMDA receptors. The corresponding distributions of the binding energy scoring functions are shown in Fig. 2.

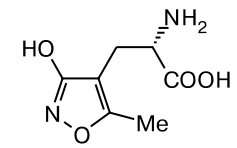
For docking of the randomly chosen compounds to the closed forms of receptors, these values vary over broad ranges. In particular, the range of variation is −38 to 0 kcal mol^{-1} for the AMPA receptor, −24 to 0 kcal mol^{-1} for the glycine site of the NMDA receptor, and −28 to

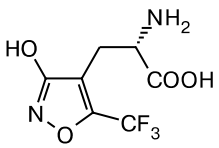
Table 1. Optimum parameters for docking

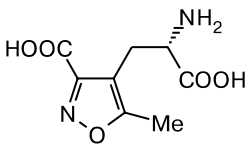
Site	Docking parameters								
	Receptor form	Grid spacing /Å	Number of clusters	Number of spheres per cluster	Docking cube size /Å	a^a	b^b	$d_{max}{}^c$ /Å	τ^d /s
Glutamate site of the AMPA receptor	Open	0.2	32	162	20×20×20	1000	20	20	18
	Closed	0.25	61	26	15×15×15	150	8	10	20
Glycine site of the NMDA receptor	Open	0.35	32	453	20×20×20	2000	25	10	28
	Closed	0.4	30	727	15×15×15	1500	10	10	34
Glutamate site of the NMDA receptor	Open	0.3	40	601	20×20×20	2000	20	20	25
	Closed	0.3	40	601	17×17×17	1000	7	10	26

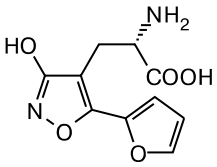
[a] Number of orientations per conformer.
[b] Maximum number of orientations for anchor-first search strategy.
[c] Energy cut-off distance included in the SF calculations.
[d] CPU time taken for docking of one structure.

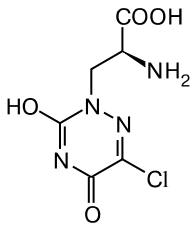
a

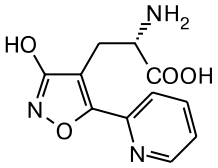

S-AMPA

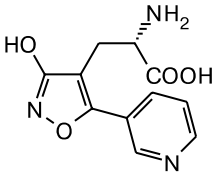

S-F-AMPA

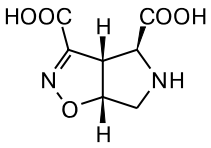

S-ACPA

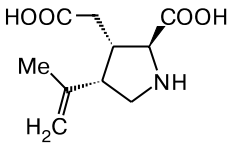

S-2-Fu-AMPA

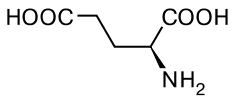

59-A


S-αPyAMPA

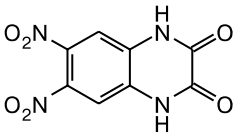

S-βPyAMPA

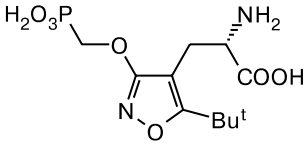

5S-CIP-A

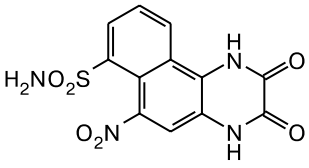

Kainic acid

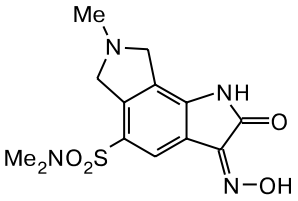

Glutamic acid

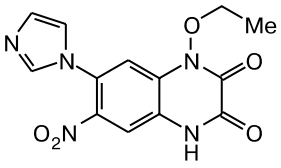
b

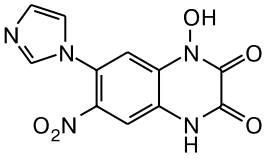

DNQX

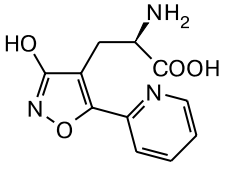

S-ATPO

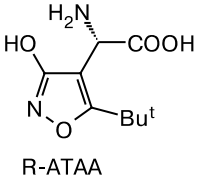

NBQX

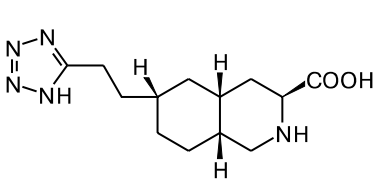

NS-257

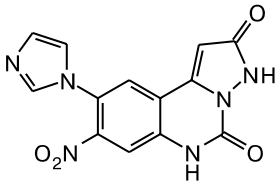

H-1


H-2

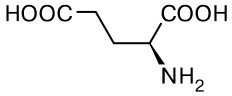

R-αPyAMPA

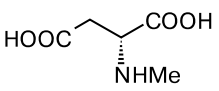

R-ATAA

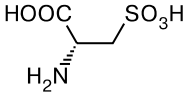

LY-293558

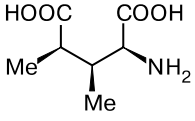

Ro 48-8587

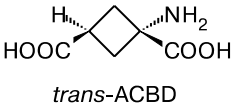
c

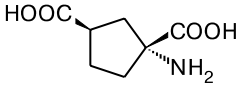

Glutamic acid

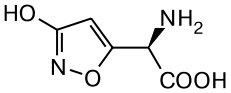

NMDA

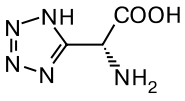

D-HCSA

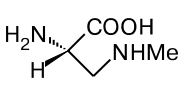

2S,3R,4S-CCG

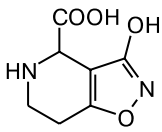

trans-ACBD


1R,1R-ACPD


R-AMAA


D-TZG


B-ALA


4-HPCA

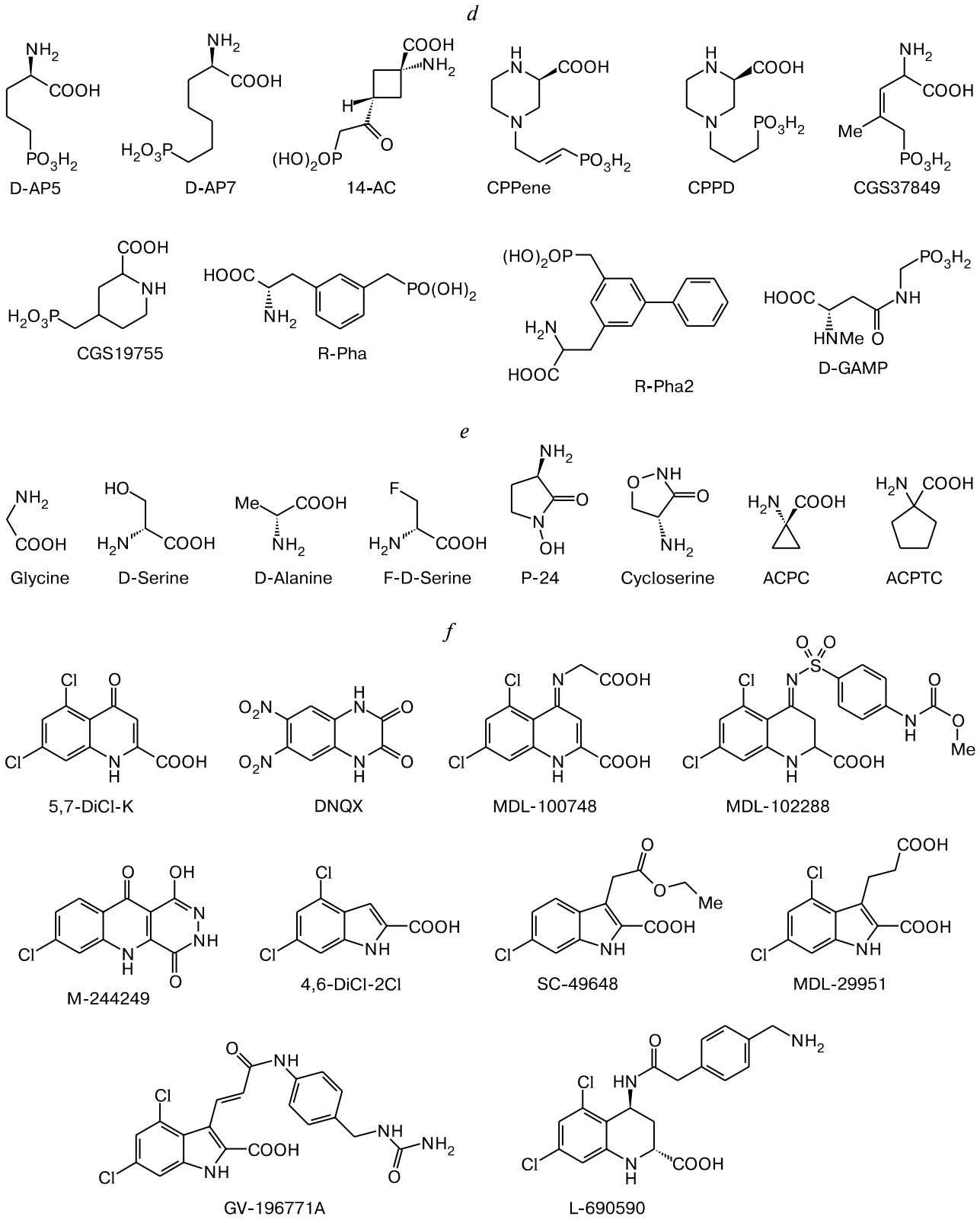


Fig. 1. Agonists (*a*, *c*, *e*) and antagonists (*b*, *d*, *f*) included in screening: agonists and antagonists of the glutamate site of the AMPA receptor (*a*, *b*); agonists and antagonists of the glutamate site of the NMDA receptor (*c*, *d*), and agonists and antagonists of the glycine site of the NMDA receptor (*e*, *f*).

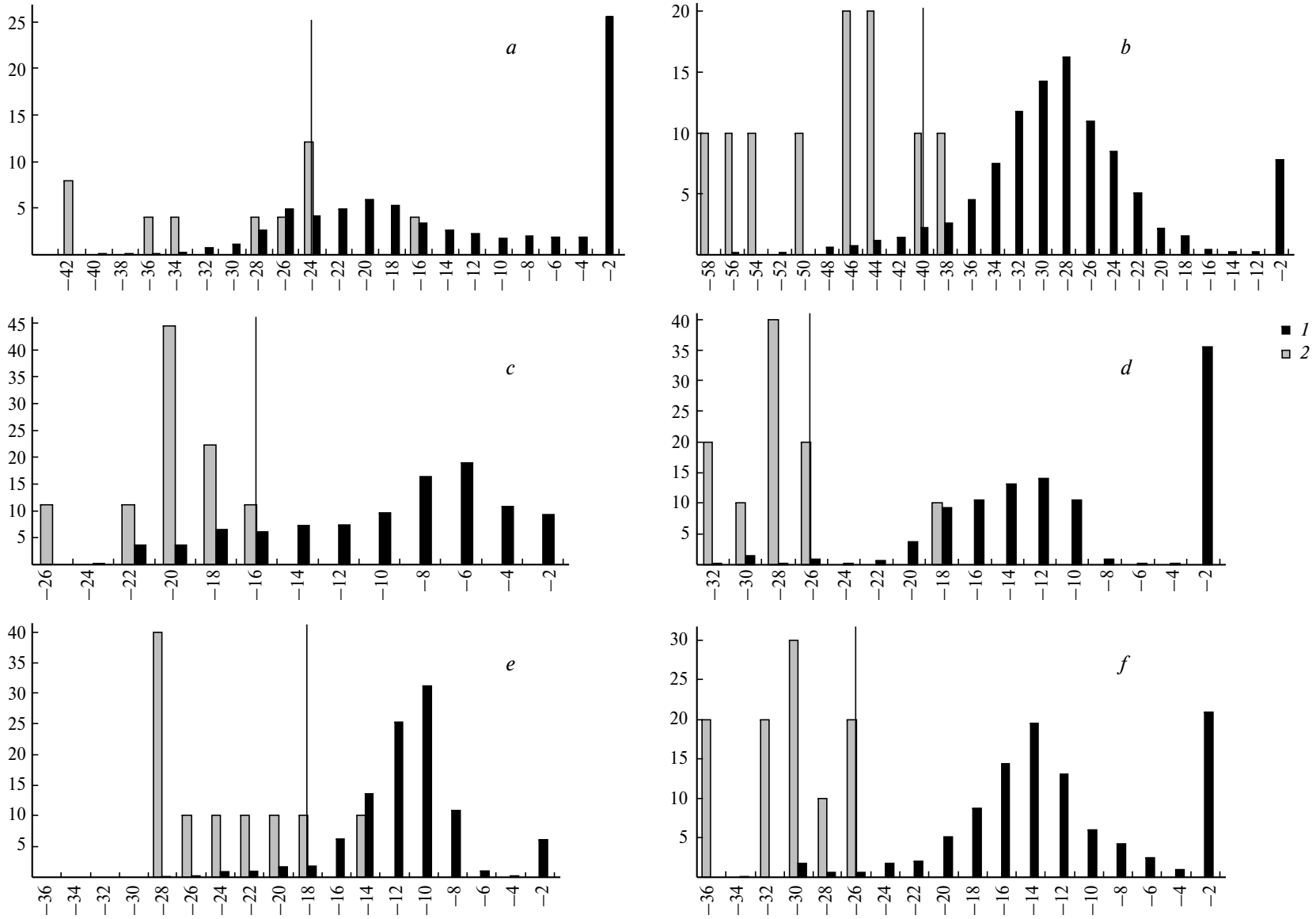


Fig. 2. Distributions of the binding energy scoring functions for docking of a database containing both randomly chosen compounds and active ligands to the closed (*a*, *c*, *e*) and open (*b*, *d*, *f*) forms of the glutamate site of the AMPA receptor (*a*, *b*), of the glycine site of the NMDA receptor (*c*, *d*), and of the glutamate site of the NMDA receptor (*e*, *f*): percentage of randomly chosen compounds in a specified range of SF values (*1*) and percentage of active compounds (*2*).

0 kcal mol^{-1} for the glutamate site of the NMDA receptor. The SF values for nine agonists docked to the AMPA receptor lie between −42 and −22 kcal mol^{-1} (only 14.2% of compounds in the entire database fell into this interval). The SF for the tenth compound (kainic acid, see Fig. 1, *a*) is −15.62 kcal mol^{-1}. Assuming the absence of ligands of the targets studied among all the randomly chosen compounds and setting the SF threshold at −22 kcal mol^{-1} (at greater negative SF values, the compound is treated as a highly affine ligand), we can conclude that the VSS successfully selects 90% of active compounds and 86% of inactive compounds. Note that the percentage of correct selection of inactive compounds seems to be even higher because the database of randomly chosen organic compounds can contain the ligands specific to the target under study. Thus, the VSS allows highly affine ligands of the closed form of the AMPA receptor to be recognized with a 90% probability. Docking of all agonists to the glycine site of the NMDA receptor gave a binding energy range from −21 to −15 kcal mol^{-1}, the fraction of the randomly chosen organic compounds in this interval being 20.15% of the number of compounds in the entire database. Therefore, setting the SF threshold at −15 kcal mol^{-1} leads to correct selection of 100% of active compounds and at least 80% of inactive compounds for the open form of the AMPA receptor.

Docking of agonists to the closed form of the glutamate site of the NMDA receptor led to the SF values in the range from −27.86 to −16.27 kcal mol^{-1} for nine compounds; the energy of the tenth compound, 1R,1R-ACPD (see Fig. 1, *c*) is −12.75 kcal mol^{-1}. The percentage of the randomly chosen organic compounds in this interval is 5.4%. Therefore, at a SF threshold equal to −16 kcal mol^{-1}, the VSS successfully selects 90% of active compounds and at least 94.6% of inactive compounds.

Docking to the open forms of receptors gave the following ranges of the SF values for the randomly chosen compounds: −54 to 0 kcal mol^{-1} for the AMPA receptor, −32 to 0 kcal mol^{-1} for the glycine site of the NMDA receptor, and −32 to 0 kcal mol^{-1} for the glutamate site of the NMDA receptor.

Antagonists docked to the open form of the AMPA receptor fell in the range from −58 to −38 kcal mol^{-1}; 8.9% of randomly chosen compounds were characterized by the SF values less than −38 kcal mol^{-1}. Thus, at the SF threshold of −38 kcal mol^{-1} the VSS correctly recognizes 100% of active compounds and at least 91.1% of inactive compounds.

Docking of the known antagonists to the open form of the glycine site of the NMDA receptor gave a range from −32 to −28 kcal mol^{-1} for the SF values of nine of the ten compounds (the SF value for 5,7-DiCl-2CI (see Fig. 1, *f*) is −17.55 kcal mol^{-1}). Only 1.7% of the total number of compounds in the database fell in this interval. Therefore, setting the SF threshold at −32 kcal mol^{-1} permits a correct selection of 90% of active compounds and 98.3% of inactive compounds for this target. The SFs values obtained for antagonists of the glutamate site of the NMDA receptor lie between −36 and −26 kcal mol^{-1}; only 1.9% of compounds from the entire database fall in this range. Setting the SF threshold at −26 kcal mol^{-1} allowed 100% of active compounds and at least 98.1% of inactive compounds to be correctly selected.

Table 2. Recognizing ability of the virtual screening system

Target	Scoring function (DOCK4) /kcal mol^{-1}	Percentage of successfully recognized active (I) and inactive (II) compounds	
		I	II
Closed form of the glutamate site of the AMPA receptor	−22	90	86
Closed form of the glycine site of the NMDA receptor	−15	100	80
Closed form of the glutamate site of the NMDA receptor	−16	90	94.1
Open form of the glutamate site of the AMPA receptor	−38	100	91.1
Open form of the glycine site of the NMDA receptor	−32	90	98.3
Open form of the glutamate site of the NMDA receptor	−26	100	98.1

Table 2 lists the recognition abilities of the VSS toward ligands of the targets under study.

Analysis of the distribution of the binding energy scoring function values for ligand binding to the open and closed forms shows that the best results (selection of active and inactive ligands) are obtained for the open forms of receptors.

Enrichment rate

One of the main goals of virtual screening is to design focused libraries of compounds, containing a larger proportion of active structures compared to their proportion in the random samples. To characterize the libraries designed quantitatively, the notion "enrichment rate" is used. The enrichment rate is the increase in the proportion of active compounds (hits) found in any given sample of compounds selected from a database of compounds at a

certain SF value, compared with the proportion expected for the entire database

$$W = (n_a/n_n)/(N_a/N_n),$$

where n_a and n_n are respectively the numbers of active and inactive compounds in the sample selected from the database and N_a and N_n are the numbers of active and inactive compounds in the entire database, respectively.

In designing libraries of compounds for high-throughput screening attainment of a higher enrichment rate of active compounds in the sample selected from the whole database seems to be more practically important than sufficiently accurate prediction of the binding energies of particular compounds. To estimate the enrichment rate as function of the chosen range of SF values (this range corresponds to certain proportion of compounds in the entire database), for each target we used the same databases containing a total of a thousand randomly chosen organic compounds and twenty active ligands (ten agonists and ten antagonists to study the binding to the closed and open forms, respectively) as those used in constructing histograms (see Fig. 2).

For each target a change in the enrichment rate as function of the proportion of the sample selected from the entire database (selection was carried out starting from the largest negative SF values) was approximated by a power function (using MS Excel® program) and the enrichment rate was interpolated for three cases, selection of 1, 5, and 10% of the entire database (Table 3). As can be seen in Table 3, the best results were obtained for selection of 1% of compounds contained in the database. The best enrichment rate of the database was found for the closed and open forms of the glutamate site of NMDA receptor (60.4 and 64.8, respectively), somewhat lower is this parameter for the open form of the glycine site of the NMDA receptor (53.6) and for the open form of the glutamate site of the AMPA receptor (47.8), and the lowest values were obtained for the closed forms of the glycine site of the NMDA receptor (22.0) and of the glutamate site of the AMPA receptor (35.9).

Table 3. Enrichment rate as function of the number of compounds in the selected fraction of database

Site	Receptor form	Enrichment rate (SF)*		
		1	5	10
Glutamate site of the AMPA receptor	Closed	35.9 (<−32)	11.2 (<−26)	6.8 (<−24)
Glutamate site of the AMPA receptor	Open	47.8 (<−47)	13.0 (<−41)	7.4 (<−37)
Glycine site of the NMDA receptor	Closed	22.0 (<−21)	8.4 (<−20)	5.5 (<−15)
Glycine site of the NMDA receptor	Open	53.6 (<−28)	13.9 (<−19)	7.8 (<−18)
Glutamate site of the NMDA receptor	Closed	60.4 (<−24)	14.9 (<−17)	8.1 (<−15)
Glutamate site of the NMDA receptor	Open	64.8 (<−29)	15.9 (<−21)	8.7 (<−17)

* Scoring function threshold/kcal mol^{-1} for successfully selected compounds (the percentage of successfully selected compounds in the entire database).

Selectivity

According to the results of trial virtual screening, randomly chosen compounds are characterized by larger negative values of the binding energy scoring functions for ligand binding to the open forms of receptors compared to the binding to the closed forms. These differences are shown in Fig. 3, which demonstrates correlations between the binding energy scoring functions for ligand binding to the closed form and the binding energy scoring functions for ligand binding to the open form. The plots constructed characterize the selectivity of randomly chosen compounds toward the two forms of receptors.

As can be seen in these plots, the points corresponding to agonists and antagonists are grouped into two separate clusters. The cluster of antagonists is characterized by a much lower SF value for the ligand complexes with the open form compared to the SF value for the complex of ligand with the closed form of receptor. The cluster of agonists is characterized by nearly equal SF values for complexes of ligands with the open and closed forms.

Thus, preliminary experiments allowed us to determine the parameters for docking (with the DOCK4 program) that make it possible to reproduce positions of the known ligands in the open and closed forms of the glutamate site of the AMPA receptor as well as in the models of the open and closed forms of the glutamate and glycine sites of the NMDA receptor, to create databases enriched with potentially active compounds, and classify such compounds into potential agonists and antagonists when docking to the open and closed forms of receptors.

Docking of a database of organic compounds

Preprocessing of the database containing a total of 135, 000 compounds allowed nearly 69% of them to be excluded. Then, docking of structures selected from the database thus obtained was carried out using the open and closed forms of receptors in order to design focused libraries of potential agonists and antagonists of the AMPA and NMDA receptors. The results obtained are listed in Table 4.

First, docking to the open forms of receptors was performed. The molecules placed in the binding pockets of the open forms were primarily selected using the known threshold values of the binding energy scoring functions

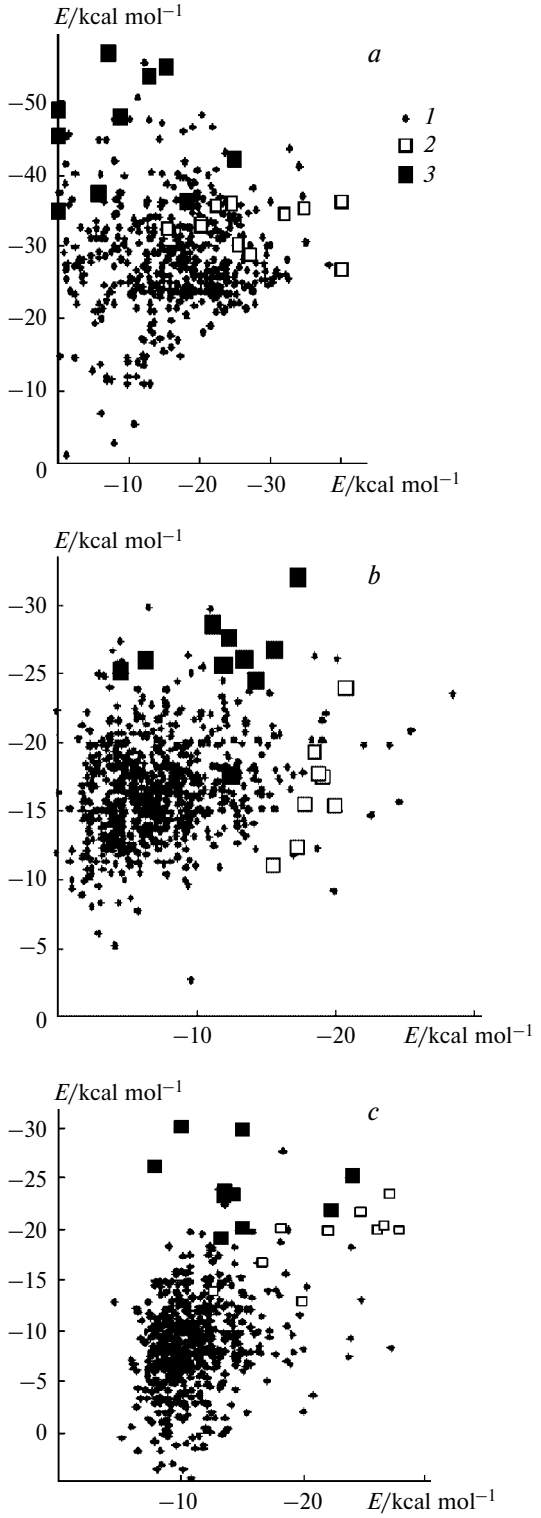


Fig. 3. Correlations between the binding energy scoring functions for ligand binding to the closed form and the binding energy scoring functions for ligand binding to the open form for the glutamate site of the AMPA receptor (*a*), glycine site of the NMDA receptor (*b*), and glutamate site of the NMDA receptor (*c*): randomly chosen compounds (*1*), agonists (*2*), and antagonists (*3*).

Table 4. Results of screening of organic molecule database*

Site	Docking to the open (I) and closed (II) forms of receptors		Post-processing (CoMFA models)
	I	II	
Glutamate site of the AMPA receptor	4167	2154	163
Glycine site of the NMDA receptor	2968	1517	89
Glutamate site of the NMDA receptor	1816	1636	—

* The database contained a total of 135,000 molecules before and 41,338 molecules after preprocessing (selection based on structural parameters, lipophilicity, and toxicity).

(see above). We selected those complexes for which the SF values were less than −38 kcal mol^{-1} for the glutamate site of the AMPA receptor, less than −26 kcal mol^{-1} for the glutamate site of the NMDA receptor, and less than −28 kcal mol^{-1} for the glycine site of the NMDA receptor. The list of compounds thus selected included a total of 4167 compounds for the AMPA receptor, 2968 compounds for the glycine site of the NMDA receptor, and 1816 compounds for the glutamate site of the NMDA receptor.

Further selection of antagonists of the glutamate site of the AMPA receptor (1) and of the glycine site of the NMDA receptor (2) was carried out using an additional filter (post-processing) based on the CoMFA models (see Refs. 17, 25). The statistical characteristics of the models are as follows: (1) $Q^2 = 0.80$ ("leave-one-out" Q^2)), NOC = 5 (optimum number of PLS components), $s = 0.21$ (standard error for cross-validation analysis) and (2) $Q^2 = 0.8$, NOC = 5, $s = 0.38$.

To perform post-processing using the CoMFA models obtained, an SPL script was written, which permits automated prediction of biological activity for the entire database in the CoMFA module of the Sybyl 6.7.2 molecular graphics package.[19] For each structure, the script performed search for a common substructure. Then, all structures were superimposed on the most active structure and the biological activity was predicted. Based on the predicted biological activities, a total of 163 compounds were selected for the glutamate site of the AMPA receptor using the CoMFA model for ligand binding to the AMPA receptor[25] and a total of 89 compounds were selected for the glycine site of the NMDA receptor using the CoMFA model of ligand binding to the NMDA receptor.[17]

Next, the compounds selected after screening using the open forms of receptors were docked to the closed forms of receptors in order to select potential agonists. The selection criteria were the threshold SF values pre-

liminarily determined from docking of a number of the known agonists, namely, less than −22 kcal mol^{-1} for the AMPA receptor, less than −16 kcal mol^{-1} for the glutamate site of the NMDA receptor, and less than −15 kcal mol^{-1} for the glycine site. The number of agonists thus selected was 2,154 for the glutamate site of the AMPA receptor, 1,517 for the glycine site of the NMDA receptor, and 1,636 for the glutamate site of the NMDA receptor.

Thus, docking of a database of randomly chosen organic compounds to the open and closed forms of the glutamate and glycine sites of the NMDA receptor and of the glutamate site of the AMPA receptor permitted nearly 98—99% of compounds whose structures are certainly inappropriate for binding to the targets under study to be excluded.

Eventually, the system of virtual screening of organic molecule databases described in this work allowed us to design focused libraries of potential agonists and antagonists of the AMPA and NMDA receptors. The libraries designed can find application in revealing and optimizing particular leader structures using methods of combinatorial chemistry and high-throughput screening.

An additional selection stage using the CoMFA models for the glutamate site of the AMPA receptor and for the glycine site of the NMDA receptor made it possible to substantially reduce the number of potential ligands and to find particular compounds as the best candidates for further studies. From the results obtained it also follows that the models of domains of the NMDA receptors constructed based on homology are at least competitive in screening as compared to the domains of the AMPA receptor with the experimentally determined geometry. Since the three-dimensional structure of the vast majority of receptors is unknown but can be theoretically designed based on homology, this opens wide prospects for application of the system designed to search for novel pharmaceuticals.

References

1. M. L. Connolly, *J. Appl. Crystallogr.*, 1983, **16**, 548.
2. T. J. A. Ewing and I. D. Kuntz, *J. Comput. Chem.*, 1997, **18**, 1175.
3. R. L. DesJarlais and J. S. Dixon, *J. Comput.-Aided Molec. Design*, 1994, **8**, 231.
4. D. A. Gschwend, W. Sirawaraporn, D. V. Santi, and I. D. Kuntz, *Proteins*, 1997, **29**, 59.
5. L. R. Hoffman, I. D. Kuntz, and J. M. White, *J. Virol.*, 1997, **71**, 8808.
6. R. Li, X. Chen, B. Gong, P. M. Selzer, Z. Li, E. Davidson, G. Kurzban, R. E. Miller, E. O. Nuzum, J. H. McKerrow, R. J. Fletterick, S. A. Gillmor, C. S. Craik, I. D. Kuntz, F. E. Cohen, and G. L. Kenyon, *Bioorg. Med. Chem.*, 1996, **4**, 1421.
7. B. K. Shoichet, R. M. Stroud, D. V. Santi, I. D. Kuntz, and K. M. Perry, *Science*, 1993, **259**, 1445.
8. C. Bissantz, G. Folkers, and D. Rognan, *J. Med. Chem.*, 2000, **43**, 4759.
9. S. Gruneberg, M. T. Stubbs, and G. Klebe, *J. Med. Chem.*, 2002, **45**, 3588.
10. B. B. Waszkowycz, T. D. J. Perkins, R. A. Sykes, and J. Li, *IBM Systems J.*, 2001, **40**, 360.
11. T. Langer and E. M. Krovat, *Curr. Opin. Drug Discov. Devel.*, 2003, **6**, 370.
12. O. Schwardt, H. Kolb, and B. Ernst, *Curr. Top. Med. Chem.*, 2003, **3**, 1.
13. R. D. Cramer III, D. E. Patterson, and J. D. Bunce, *J. Am. Chem. Soc.*, 1988, **110**, 5959.
14. I. G. Tikhonova, I. I. Baskin, V. A. Palyulin, and N. S. Zefirov, *Dokl. Akad. Nauk*, 2002, **382**, 550 [*Dokl. Chem.*, 2002 (Engl. Transl.)].
15. I. G. Tikhonova, I. I. Baskin, V. A. Palyulin, N. S. Zefirov, and S. O. Bachurin, *J. Med. Chem.*, 2002, **45**, 3836; 5608.
16. I. G. Tikhonova, I. I. Baskin, V. A. Palyulin, and N. S. Zefirov, *Dokl. Akad. Nauk*, 2002, **382**, 840 [*Dokl. Chem.*, 2002 (Engl. Transl.)].
17. I. G. Tikhonova, I. I. Baskin, V. A. Palyulin, and N. S. Zefirov, *J. Med. Chem.*, 2003, **46**, 1609.
18. N. Armstrong and E. Gouaux, *Neuron*, 2000, **28**. 165.
19. *Sybyl 6.7.2*, Tripos Assoc., St. Louis (MO).
20. W. D. Ihlenfeldt, Y. Takahashi, H. Abe, and S. Sasaki, *J. Chem. Inf. Comp. Sci.*, 1994, **34**, 109.
21. R. Wang, Y. Gao, and L. Lai, *Pers. Dr. Discov. Des.*, 2000, **19**, 47.
22. R. Wang, Y. Gao, and L. Lai, *J. Mol. Model.*, 2000, **6**, 498.
23. C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, *Adv. Drug Delivery Rev.*, 1997, **23**, 23.
24. *Quantum Chemistry Program Exchange (QCPE), Creative Arts Bldg.* 181 Indiana University, Bloomington (IN 47405).
25. I. G. Tikhonova, I. I. Baskin, V. A. Palyulin, and N. S. Zefirov, *Dokl. Akad. Nauk*, 2003, **389**, 123 [*Dokl. Chem.*, 2003 (Engl. Transl.)].

Received March 25, 2004