Journal of
Medicinal
Chemistry

Volume 43, Number 10 May 18, 2000

Articles

Application of Similarity Matrices and Genetic Neural Networks in Quantitative Structure−Activity Relationships of 2- or 4-(4-Methylpiperazino)pyrimidines: 5-HT_{2A} Receptor Antagonists

Tomasz Borowski,[†] Marcin Król,[‡] Ewa Brocławik,*[,§] Timothy C. Baranowski,[∥] Lucjan Strekowski,[∥] and Maria J. Mokrosz[⊥]

Department of Theoretical Chemistry, Faculty of Chemistry, Jagiellonian University, 30-060 Kraków, Ingardena 3, Poland; Department of Biostatistics and Medicinal Informatics Collegium Medicum, Jagiellonian University, 31-501 Kraków, Kopernika 17, Poland; Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, 30-239 Kraków, Niezapominajek, Poland; Department of Chemistry, Georgia State University, Atlanta, Georgia 30303; and Department of Medicinal Chemistry, Institute of Pharmacology, Polish Academy of Sciences, 12 Smetna Street, 31-343 Kraków, Poland

Received December 1, 1999

Antagonists of the 5-HT_{2A} receptor are being used to treat many psychiatric disorders. The present work focuses on a group of 27 antagonists possessing varying affinities toward the receptor. These are 26 title compounds and clozapine as a reference antagonist. The active conformers of the conformationally flexible ligands were proposed by using the active rigid analogue approach and performing similarity calculations. The calculations involved genetic neural network (GNN) computations deriving QSARs from similarity matrices (SM) with cross-validated correlation coefficients exceeding 0.92. The performance of neural networks with variety of architectures was studied. As the computations were performed for cations and neutral molecules separately, the relevance of the ligand charging is discussed.

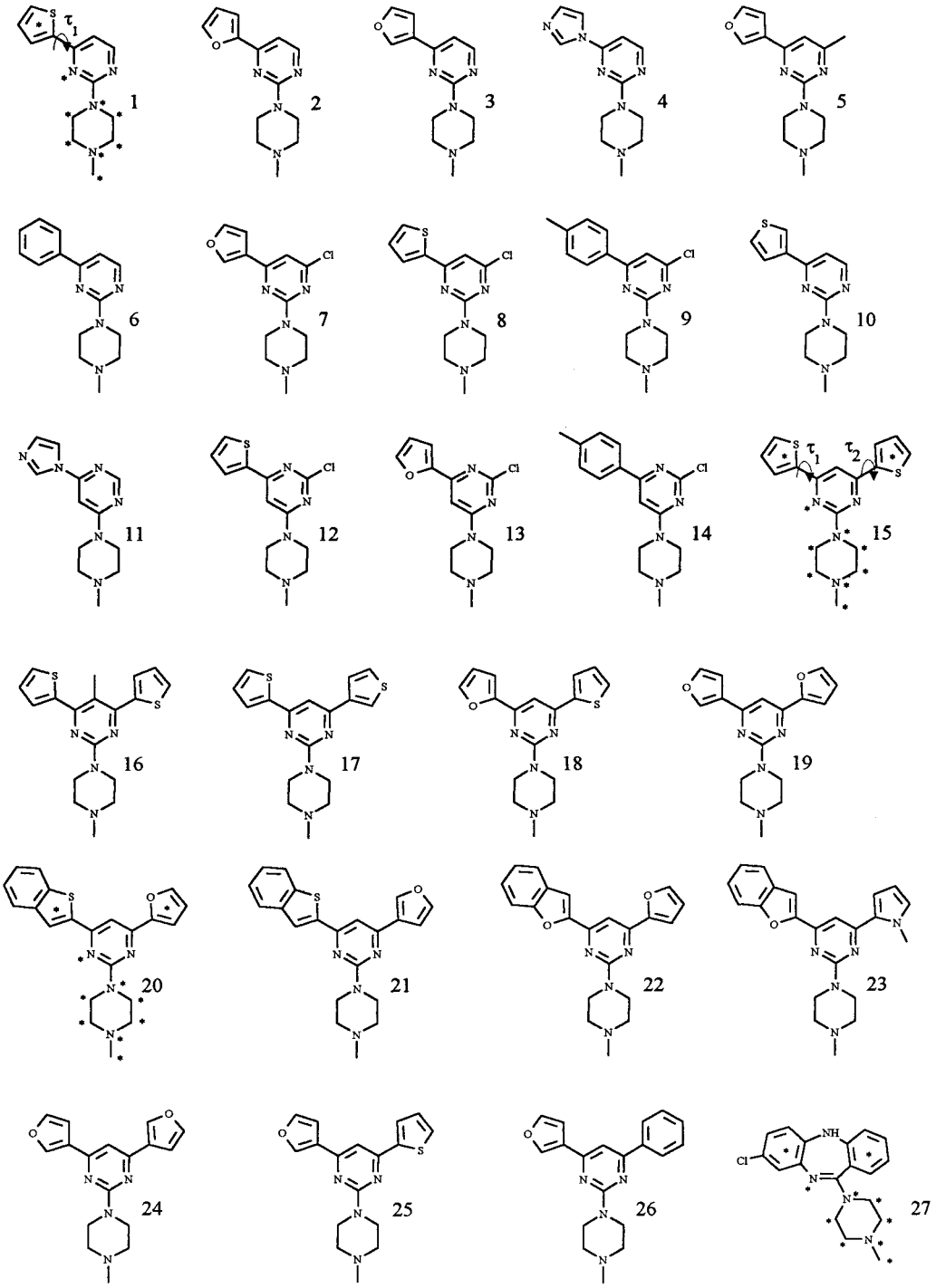
Introduction

The serotonin (5-HT) receptors belong to the superfamily of G-protein coupled transmembrane receptors. Homology studies indicate that the receptors consist of 7 helical segments that span the lipid bilayer of cell membrane.[1] The detailed structures of these proteins remain unknown, which is mostly due to crystallization problems. As these receptors are important from the therapeutic point of view, medicinal chemists try to build pharmacophore or receptor models based on the structure of the 5-HT ligands. For example, Höltje and Jendretzki have constructed a 5-HT_{2A} receptor model on the basis of molecular field computations for receptor ligands and homology studies of G-protein coupled receptors.[2] Their model comprises both agonists and antagonists of the 5-HT_{2A} receptor and defines possible residues participating in the ligand−receptor interaction. A topographic pharmacophore model of 5-HT_{2A} antagonists developed by Mokrosz et al. utilizes three intramolecular distances as crucial features of the structure of 4,6-di(heteroaryl)-2-(*N*-methylpiperazino)-pyrimidines.[3] A similar topographic model of 5-HT_{2A} receptor antagonists has been proposed by Andersen et al.[4] On the basis of a conformational analysis of several indan derivatives as well as cyproheptadine, ritanserin, and danitracen they have proposed a three-point model

* To whom correspondence should be addressed. Tel: (48.12) 425-2814. Fax: (48.12) 425-1923. E-mail: broclawi@turl.ch.uj.edu.pl.
[†] Department of Theoretical Chemistry, Jagiellonian University.
[‡] Department of Biostatistics and Medicinal Informatics Collegium Medicum, Jagiellonian University.
[§] Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences.
[∥] Georgia State University.
[⊥] Institute of Pharmacology, Polish Academy of Sciences.

10.1021/jm9911332 CCC: $19.00
Published on Web 04/29/2000

Chart 1. 5-HT_{2A} Data Set



with two points describing the positions of two benzene rings in relation to a third point that simulates a receptor site interacting with the basic nitrogen atom.

A new approach combining genetic neural networks (GNN) with similarity matrices (SM) has been proposed by So and Karplus for quantitative structure−activity relationship (QSAR) analyses.[5] Along with the formulation the authors have shown the robustness and high efficiency of their method to obtain 3D QSARs for a variety of chemical classes of compounds.[6]

In the present paper we describe an application of the SM/GNN methodology to the series of 27 5-HT_{2A} antagonists. The congeneric set (Chart 1) consists of 18 substituted 2/4-(4-methylpiperazino)pyrimidines reported previously[3,7] and 8 derivatives synthesized as part of this work. Since these compounds possess a considerable conformational freedom,[3,7] an active analogue approach with clozapine (**27**, Chart 1) as a template for molecular superimposition is used. More specifically, this approach is similar to that proposed by Montanari et al.[8] in the sense that the quality of the derived QSAR is used to choose among different possible superimposition modes.

From mutagenesis and comparative studies on the G-protein coupled receptors, the importance of Asp^{155} for high affinity of 5-HT_{2A} antagonists as well as agonists has been established.[9] Thus, it is reasonable to assume that 5-HT_{2A} ligands bind in a cationic form

to the receptor. On the other hand, the interaction of the ligand with an anionic side chain of Asp neutralizes the positive charge on the molecule bound to the receptor. To investigate the influence of the charge on the ligands on the quality of the derived QSARs we performed computations for both cations and neutral molecules.

The results reported in the present paper strongly support the applicability of the SM/GNN methodology as a tool to derive reliable 3D QSARs. In addition, we show how this technique may be used to deal with conformational diversity of flexible ligands. The statistical significance of the models derived here was validated with a standard randomization test. Moreover, these 3D QSARs can be applied to new 5-HT$_{2A}$ antagonists from the same class of ligands studied here, and very high cross-validated correlation coefficients justify their use to predict 5-HT$_{2A}$ affinity of the prospective ligands.

Method

In the SM/GNN methodology the basic concept is that similar compounds exert similar effect on the receptors. By using one of the many available formulas one can quantify the similarity between any two compounds once they are properly superimposed. The similarity index is calculated from three-dimensional fields of the molecules as expressed by their electrostatic potentials (ESPs) and steric factors. Thus, in turn, the similarity between two compounds is a function of their conformations. For a set of N compounds, an $N \times N$ matrix can be constructed, the columns of which consist of similarity indices between the i-th compound and all compounds from the analyzed set. Similarity matrix elements can be interpreted as values of independent variables that can be correlated with biological activities. The role of genetic algorithm (GA) is to select these distinct columns (variables) which generate good quantitative structure−activity relationships (2−7 variables in this work). The choice is made out of many possibilities, and for this task GA serves the best.[10] A neural network is used to establish a relationship between a chosen set of independent variables and biological activity of studied compounds. Neural networks have the advantage in that they can easily incorporate higher order terms in QSAR, the feature that turned up to be important in our study.

As the SM/GNN methodology used here is essentially the adaptation of the So and Karplus proposal; interested readers are referred to the literature for a detailed discussion of the use of genetic neural networks,[11,12] similarity matrices,[13,14] and neural networks[15,16] in the QSAR analysis. In this section only a brief description of the modeling procedure is given.

Generation of Similarity Indices. AM1 Mulliken charges were used to generate electrostatic potential (ESP) around studied molecules using the Coulomb equation and vacuum dielectric constant ($\epsilon = 1.0$). ESP was calculated at equally spaced points forming a rectilinear grid. The spacing of 1 Å and grid with 6 Å extension beyond all atomic coordinates were used. The ESP similarity index between two molecules was computed using the Hodgkin[17] formula (eq 1)

$$H_{AB} = \frac{2\sum P_A P_B}{\sum P_A^2 + \sum P_B^2} \quad (1)$$

where P_A is the ESP produced by molecule A at a given point of the grid. The sums run over all grid points. The calculations were performed either without any cutoff or the potential with an absolute value exceeding 5 or 100 kcal/mol was truncated to the cutoff value (±5 or ±100 kcal/mol). Following the work by So and Karplus,[6] the higher cutoff value was used for cations in order to make similarity indices more discriminating.

The shape similarity was computed using Meyer formula[18] (eq 2),

$$S_{AB} = \frac{U_{AB}}{\sqrt{T_A T_B}} \quad (2)$$

where U is the number of points enclosed by a common volume of the molecules A and B and T_A (T_B) is the number of points inside the van der Waals envelope of the molecule A (B). A regular grid extending 2 Å beyond the molecular boundary and 0.5 Å spacing were used.

Data Sets. Table 1 presents the binding characteristics of the studied ligands toward the 5-HT$_{2A}$ receptor. The binding data of the literature compounds[3,7] **1**−**3**, **5**, **6**, **10**, **15**−**26**, and new ligands **4**, **7**−**9**, **11**−**14** were obtained in the same laboratory and by using the same protocol. Models of the molecules were built from standard fragments within InsightII[19] environment. Both neutral molecules and their cationic forms containing protonated *N*-methyl function of the piperazine were energy minimized using the MOPAC 6.0 semiempirical package.[20] The AM1 Hamiltonian was used to obtain minimum energy structures and Mulliken atomic charges. The centroids were added to the aromatic rings of the models, and a superimposition of the ligands on the clozapine molecule (in minimal energy conformation: piperazine ring in chair conformation, two substituents in equatorial positions) was performed. The QUATFIT[21] program was used to perform the least-squares fitting. The fitted centers are marked for representative compounds **1**, **15**, **20**, **27** with asterisks in Chart 1. All common atoms and aromatic centroids were utilized in the superimposition for similarity indices to be maximal for similar compounds. As the tricyclic moiety of the clozapine molecule is not planar, small adjustments were done to the dihedral angles (τ_1 and τ_2 in Chart 1 for **1** and **15**) of the ligands prior to the superimposition. This ensures that the fitted aromatic rings of the ligand and those of clozapine are coplanar after the superimposition. These small changes in dihedral angles are justified by the results of the previous study,[3,7] which have shown a substantial conformational freedom with energy barriers upon rotation not exceeding a few kcal/mol.

Clozapine was chosen as a template because it is a potent 5-HT$_{2A}$ antagonist. The molecule is considerably rigid and shows high similarity to the studied compounds. Features in common are an *N*-methylpiperazine group, an sp^2 nitrogen atom, and aromatic rings, which (marked with asterisks in Chart 1) were chosen as the points defining the molecular alignment.

Several ways of superimposition of a given molecule onto the clozapine template were realized. For example,

Table 1. 5-HT_{2A} Binding Affinity Data

no.	pK_i	no.	pK_i	no.	pK_i
1[a]	6.68	**10**[a]	6.72	**19**[a]	7.89
2[a]	6.13	**11**	5.00	**20**[a]	7.43
3[a]	8.00	**12**	5.57	**21**[a]	7.26
4	6.42	**13**	5.22	**22**[a]	6.96
5[a]	7.30	**14**	5.37	**23**[a]	6.87
6[a]	5.68	**15**[a]	8.10	**24**[a]	8.10
7	7.39	**16**[a]	7.77	**25**[a]	8.00
8	6.46	**17**[a]	7.68	**26**[a]	8.05
9	6.24	**18**[a]	7.43	**27**[b]	8.80

[a] Taken from refs 3 and 7.[b] Taken from ref 29.

Table 2. Criteria Used To Construct Data Sets: Indices of Similarity to Clozapine

set no.	neutral compounds	cations
1	steric	steric
2	ESP (no cutoff)	ESP (no cutoff)
3	ESP (cutoff = 5 kcal/mol)	ESP (cutoff = 100 kcal/mol)
4[a]	ESP (cutoff = 5 kcal/mol)	ESP (cutoff = 100 kcal/mol)
5[a]	ESP (no cutoff)	ESP (no cutoff)

[a] Sets 4 and 5 differ from 3 and 2 in superimposition mode of compounds **1**−**14**. Only the Cl-substituted ring of clozapine was used as a reference in the superimposition (with sp^2 nitrogen and piperazine ring).

compound **1** has four different possible alignments: two with the 2-thienyl ring superimposed onto the Cl-substituted clozapine ring with the sulfur atom syn and anti relative to the sp^2 nitrogen of **1** and two with the 2-thienyl ring fitted onto the second aromatic ring of clozapine. For each possible alignment the ESP and steric similarity indices between the clozapine molecule and a given ligand molecule were evaluated. The similarity indices fall into the range of 0.27−0.64, 0.12−0.76, and 0.62−0.76 for ESP(no cutoff), ESP(cutoff = 5 kcal/mol), and steric calculations, respectively. These numbers were tabulated and used to construct five distinct sets of conformers most resembling the template clozapine. These five sets proposed for cations and neutral molecules separately are aided by 30 RANDOM sets for a comparative purpose. Each of them consists of randomly chosen conformers − one for each compound. Each of these conformers was aligned onto the template clozapine. The difference between the QSAR performance of a given set and the average QSAR performance of the RANDOM sets gives a clue about the importance of the conformation (resulting from rotations by the dihedrals τ_1 and τ_2) for the 5-HT_{2A} binding affinity. The criteria used to construct the sets are gathered in Table 2. Sets 4 and 5 comprise only these fits of **1**−**14** where the aromatic ring marked with an asterisk is fitted on the Cl-substituted ring of clozapine. This particular choice of the template ring can be rationalized by the fact that such procedure gives good steric similarity to clozapine for these ligands.

The basis for the above fitting and selection procedure is the assumption that clozapine, being a potent 5-HT_{2A} antagonist, produces steric and ESP molecular fields that are complementary to those produced by the receptor. Compounds similar to clozapine when binding to the receptor should adopt a conformation in which they resemble the potent analogue the most. Different criteria of similarity give rise to different proposed sets.

Genetic Neural Network. All computations were performed using an evolutionary programming algorithm[11,12] with 300 individuals and 50 reproductions. The neural network with one hidden layer of the general structure *n*1-*n*2-1 and scaled conjugate gradient algorithm[22] was used. In most cases the 5-2-1 neural network was assumed, as this configuration and number of input units seemed to be appropriate for the set containing 27 compounds. The Pearson correlation coefficient was used as the fitness function during the 50 reproductions. At the end, the cross-validation (leave-one-out) procedure was performed, and all models were sorted with respect to the cross-validated correlation coefficient q^2 (eq 3).

$$q^2 = 1 - \frac{\sum_{i=1}^{N}(y_{i,\text{observed}} - y_{i,\text{predicted}})^2}{\sum_{i=1}^{N}(y_{i,\text{observed}} - \overline{y_{i,\text{predicted}}})^2} \quad (3)$$

The top rank model was taken as a result of a given GNN run. The final selection of the models was based on the value of σ (eq 4), as originally proposed by So and Karplus,[6]

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(y_{i,\text{observed}} - y_{i,\text{predicted}})^2}{N - n - 1}} \quad (4)$$

where N is a number of compounds in the set, n is a number of input units, and $y_{i,\text{predicted}}$ is a cross-validated predicted activity. Using σ as a selection criterion one can discriminate between models that give almost the same correlation coefficients but are different in the number of variables (n). Thus, a compromise between the quality of the model and the risk of overfitting the data can be reached. More specifically, when the neural network has too many adjustable weights compared to the number of training data, the network can "memorize" the training set.

Unless otherwise stated, all computations were performed 30 times with different initial seeds (iseed) because both the starting population for the evolutionary programming algorithm and initial weights in the neural network are obtained using a random number generator. Results reported here are the averages and standard deviations for a single probe.

The validity of the final model was checked with the randomization test. In this procedure the data forming the response vector (binding affinity data) are scrambled by 100 random exchanges in their positions. In this way the data to construct dummy QSARs with the unchanged variance are obtained.

Programs used to construct similarity matrices and to perform GNN calculations were written in FORTRAN77 language in our laboratory in Kraków. They were compiled and run under the IRIX 6.4 operating system on a 250 MHz R10000 Silicon Graphics ORIGIN 2000 supercomputer. The software[23] has been thoroughly tested against the published data.[5]

Results and Discussion

Chemistry. The synthesis of new pyrimidine derivatives **4**, **7**−**9**, and **11**−**14** is given in Scheme 1. In the

Scheme 1. Synthesis of Compounds **4**, **7**−**9**, and **11**−**14**

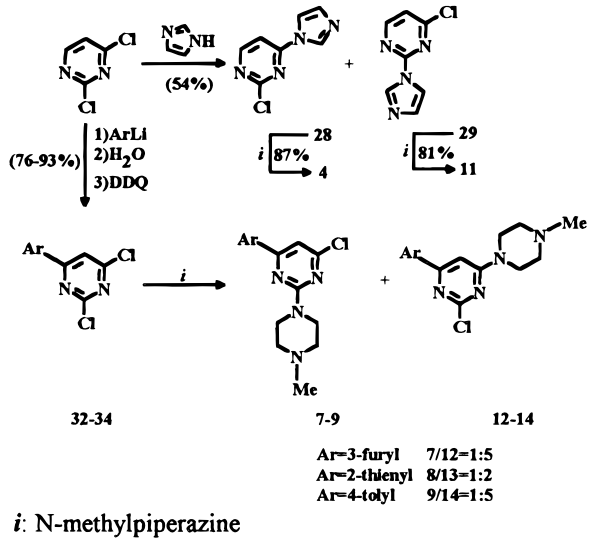


Table 3. 5-HT$_{2A}$ Binding Data for Compounds **4**, **7**−**9**, and **11**−**14**

compd	$K_i \pm$ SEM [nM]	compd	$K_i \pm$ SEM [nM]
4	384 ± 16	**11**	>10000
7	41 ± 1	**12**	2720 ± 500
8	353 ± 63	**13**	5960 ± 140
9	582 ± 3	**14**	4220 ± 60

Table 4. Cross-Validated Correlation Coefficients q^2 for QSARs Derived for Different Sets[a] of Neutral Compounds

set no.	steric	ESP (no cutoff)	ESP (cutoff = 5 kcal/mol)
1	0.87 (0.01)	0.79 (0.03)	0.80 (0.02)
2	0.80 (0.03)	0.85 (0.02)	0.86 (0.02)
3	0.70 (0.04)	0.90 (0.01)	0.87 (0.02)
4	0.93 (0.01)	0.88 (0.02)	0.89 (0.01)
5	0.94 (0.01)	0.90 (0.01)	0.91 (0.01)
RANDOM	0.81 (0.04)	0.77 (0.07)	0.69 (0.09)

[a] The entries are means of 30 results obtained with different iseeds. Standard deviations are given in parentheses.

preparation of **4** and **11**, 2,4-dichloropyrimidine was allowed to react with imidazole followed by separation of the resultant mixture of **28** and **29**, and then treatment of the individual compounds **28** and **29** with *N*-methylpiperazine. In a similar way, compounds **32**−**34**, obtained from 2,4-dichloropyrimidene by adaptation of a published procedure,[24] served as precursors to substituted pyrimidines **7**−**9** and **12**−**14**. The structures of isomeric products **28/29**, **4/11**, **7/12**, **8/13**, and **9/14** were assigned unambiguously by proton NOE experiments.

Pharmacology. The affinity (Table 3) of **4**, **7**−**9**, and **11**−**14** for 5-HT$_{2A}$ receptors of the rat brain cortex was assessed on the basis of their ability to displace [^{3}H]-ketanserin, according to the published procedures.[25]

Comparison of Different Sets. Tables 4 and 5 contain results obtained for neutral molecules and cations, respectively. In all computations the neural network with a 5-2-1 architecture was used. Comparing the results in these tables one can easily notice that the QSARs obtained for neutral ligands are more discriminating than those for cations. This can be rationalized by the fact that most of the sets are based on the ESP similarity index as a selection criterion. The differences

Table 5. Cross-Validated Correlation Coefficients q^2 of QSARs Derived for Different Sets[a] of Cations

set no.	STERIC	ESP (no cutoff)	ESP (cutoff = 100 kcal/mol)
1	0.87 (0.02)	0.70 (0.04)	0.73 (0.03)
2	0.88 (0.01)	0.84 (0.03)	0.84 (0.02)
3	0.88 (0.01)	0.80 (0.03)	0.81 (0.02)
4	0.87 (0.01)	0.80 (0.02)	0.81 (0.02)
5	0.87 (0.01)	0.84 (0.03)	0.85 (0.02)
RANDOM	0.77 (0.07)	0.73 (0.07)	0.72 (0.08)

[a] The entries are means of 30 results obtained with different iseeds. Standard deviations are given in parentheses.

in ESP distribution for different conformers of a given cationic ligand are almost lost due to the large positive electrostatic potential resulting from the protonation. Selections are thus more erratic in comparison to the sets comprising neutral molecules. Moreover, the cation ESP similarity matrices are more flat due to the large contribution of this positive potential to the molecular field. The protonation of the piperazine ring is a prerequisite for binding of a ligand to the 5-HT$_{2A}$ receptor. The ionic interaction between the aspartate residue in the third transmembrane segment (TM3) and the cationic head of the ligand is the driving force for this process. On the other hand, one can expect that the charge distribution around the bound ligand will resemble that for a neutral molecule as the positive charge is neutralized by the negative counterion.[26] Thus the ESP similarity calculations for neutral compounds are more accurate than those for cations and more appropriate for a group of compounds containing the same protonation center.

Another aspect worth noticing is the observation that there is no simple relation between the selection criterion and the quality of the QSAR derived from the SM. For example, one might expect that the set 1 for neutral compounds, which consists of conformers sterically most resembling clozapine, would perform superior in steric SM/GNN QSARs. Actually it follows from the first column in Table 4 that sets 4 and 5 perform better. On the other hand, these two sets 4 and 5 give very good results for both steric and ESP based QSARs. These two sets give almost the same results as they differ only by two conformers. The high value of q^2 for all QSARs derived for set 5 provides a strong support to the conformation selection procedure yielding this set. It is very likely that these conformers are present in the 5-HT$_{2A}$ receptor−ligand complexes. These results also suggest that the long-distance electrostatic interaction plays an important role in the receptor−ligand recognition process. The steric requirements are still important, as can be inferred from the high value of the cross-validated correlation coefficient q^2 for steric SM/GNN QSARs.

A comparison of the results for sets 5 and RANDOM for neutral molecules seems to support the thesis that some conformations of studied ligands are preferred when bound to the 5-HT$_{2A}$ receptor. High correlations obtained for set 5 provide evidence that this particular set may contain conformers present in the drug−receptor interaction.

All further computations were performed for the set 5 of neutral molecules. The conformations of the ligands (proposed for the ligand−receptor systems) in this set are presented in Chart 1.

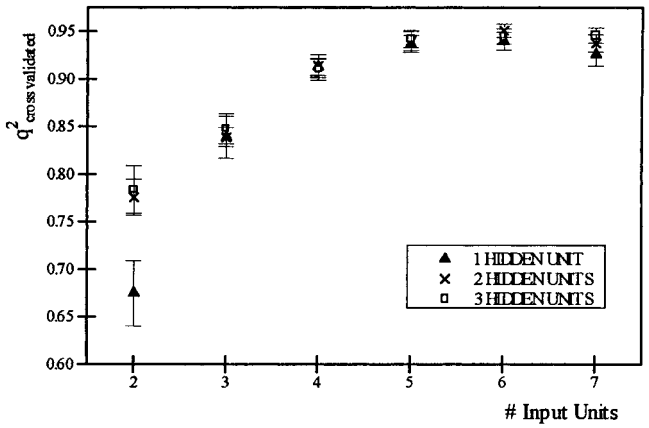


Figure 1. Plot of q^2 as a function of the number of input and hidden nodes for steric SM.

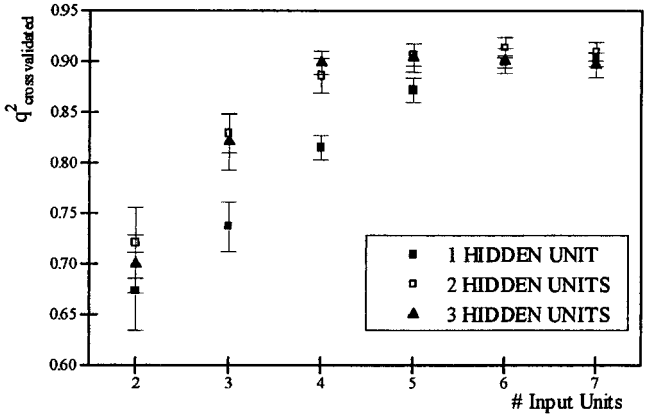


Figure 2. Plot of q^2 as a function of the number of input and hidden nodes for ESP SM with a cutoff of 5 kcal/mol.

Neural Network Architecture. It has been shown previously[5,6] that the number of input units in the neural network is an important parameter in SM/GNN calculations. On the other hand, it is well known that the number of hidden units is closely related to the generalization abilities of the neural network. These units are also responsible for the possible couplings between input variables. An important aspect of neural network computations is that the optimal architecture depends strongly on the characteristic of the problem to be solved. In other words, for some ligand sets two hidden nodes can be the optimal solution; for others, however, different architecture would be necessary to obtain the best predictivity. To study the influence of the neural network architecture on its predictive ability, computations with 1, 2, and 3 hidden units were performed. The number of input neurons varied from 2 to 7. The results of these computations for steric, ESP (cutoff = 5 kcal/mol), and combined SMs are shown in the respective Figures 1−3. Comparing them one can easily notice qualitative differences. In the case of steric SM, the number of hidden units seems to be of a marginal importance for the generalization ability of the neural network. The points for the three sets (1, 2, and 3 hidden units) coincide within an error of the method. A different situation is observed for the ESP SM. Here the change from 1 to 2 hidden units brings a meaningful shift in the q^2 coefficient. This effect is most pronounced for a small number of input units. By contrast, changing the number of hidden units from 2 to 3 does not cause any dramatic change in the fitting performance of the neural network. The importance of hidden units for ESP SM QSARs indirectly implies the presence and importance of higher order terms in the QSAR equation modeled by the neural network. This should be kept in mind when interpreting the functional dependence plots that are discussed in more detail in the next section. For the combined similarity matrix, the conclusions are essentially the same as for the steric one. The reason is that for the combined similarity matrix the GNN procedure gives almost exclusively the steric models. The general conclusion that emerges from these computations is that for the studied group of compounds two hidden units make a good compromise between the generalization abilities of the neural network and the number of adjustable weights. It allows for the higher terms to be included in the QSAR and does not cause the network to "memorize" the training data.

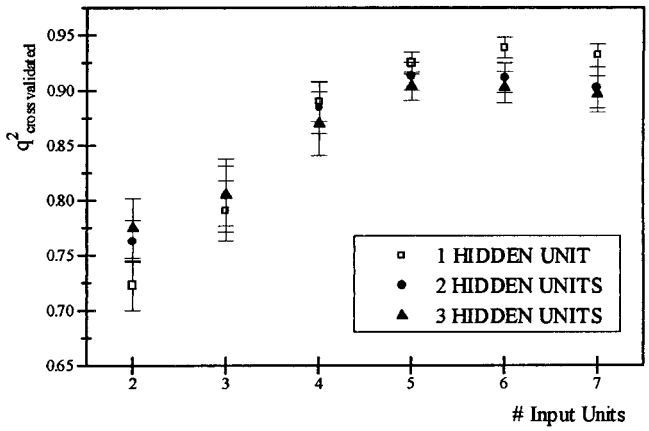


Figure 3. Plot of q^2 as a function of the number of input and hidden nodes by using a combined similarity matrix (steric/ESP).

Models Derived from Set 5. For all QSARs derived from steric, ESP, and combined similarity matrices, the σ parameter was calculated, and the models with the lowest σ values were selected for further analysis. The specification of these models is given in Table 6. For the combined similarity matrix the GNN procedure converges into a steric subspace of solutions. The best model for this case is essentially the same as for steric SM considered alone because the information contained in steric SM seems to screen the ESP derived information. For this reason only the solutions obtained for steric and ESP SMs are discussed. Figures 4 and 5 show the cross-validated versus observed activities for the best models. There are no outliers, and the overall good conformity between the predicted and observed 5-HT_{2A} affinities can be noticed. As stated above, the higher order terms may play an important role in correlating the biological data with similarity indices for the ESP SM. Nevertheless, the functional plots are useful tools when analyzing the results obtained with SM/GNN procedure. When carefully interpreted they can give a valuable insight into QSAR realized by the neural network. Figures 6 and 7 show functional dependences of predicted activity on the similarity indices for our models. The utilized procedure to construct such plots is essentially the same as that proposed by the authors of the method.[5] One similarity index was scanned from 0.0 to 1.0 while all remaining variables were fixed to

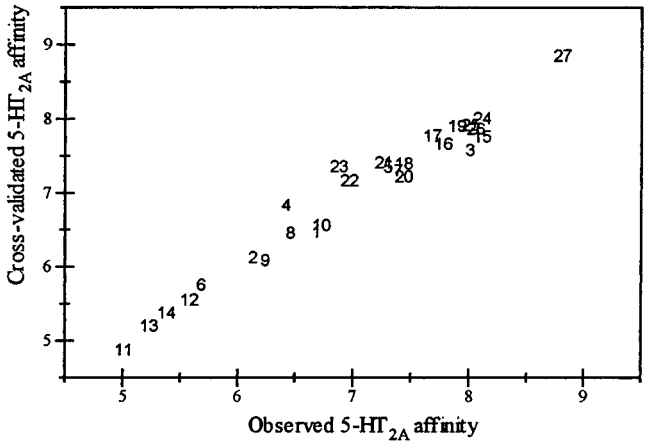


Figure 4. Plot of the cross-validated 5-HT$_{2A}$ affinity against the measured values for the steric model. (For labeling, see Chart 1.)

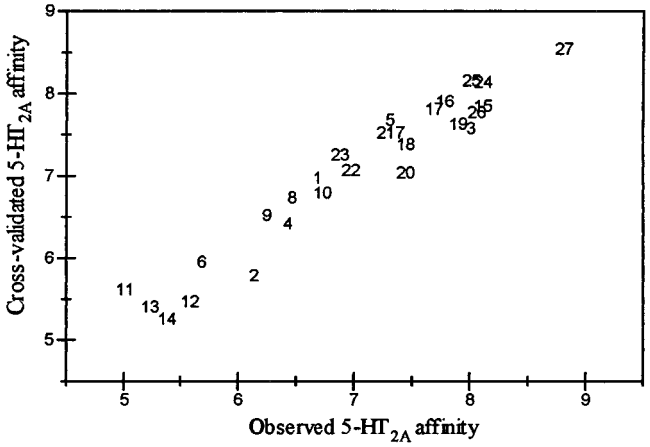


Figure 5. Plot of the cross-validated 5-HT$_{2A}$ affinity against the measured values for ESP model. (For labeling, see Chart 1.)

Table 6. Best Models Obtained for Set 5 of Neutral Molecules

SM	neural network	model	q^2	σ	q^2 for dummy QSARs
STERIC	5-2-1	**2**, **3**, **6**, **11**, **27**	0.963	0.046	0.282 (0.161)
ESP	5-2-1	**4**, **12**, **22**, **24**, **26**	0.928	0.064	0.292 (0.138)
combined	6-1-1	**2**, **3**, **6**, **11**, **16**, **27**[a]	0.955	0.051	0.484 (0.134)

[a] The model comprises solely steric solution.

the value of 0.5. The response of the neural network is the dependent variable on the plot. First, let us consider the plot for steric model (Figure 6). The steric similarities to **3** and **27** are the variables that most positively correlate with the 5-HT$_{2A}$ binding affinity. Indeed, the selected ligands are both very active toward the 5-HT$_{2A}$ receptor. Similarities to **2**, **6**, and **11** show opposite tendencies, and these compounds show low or moderate 5-HT$_{2A}$ affinities. In the ESP case (Figure 7) potent antagonists of 5-HT$_{2A}$ receptor, **24** and **26**, were found among the best reference molecules. They both produce ESP distributions most suitable for effective binding to the 5-HT$_{2A}$ receptor. Ligands **12**, **22**, and, to a lesser extent, **4** are reference molecules the ESP distributions of which are far from being perfect regarding their interaction with the 5-HT$_{2A}$ receptor. Of course, such discussion of one-dimensional sections through many dimensional surfaces is only qualitative, and in order

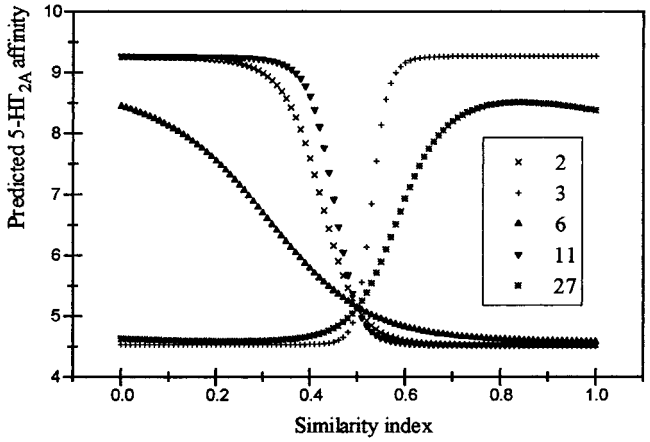


Figure 6. Functional dependence plot for the steric model.

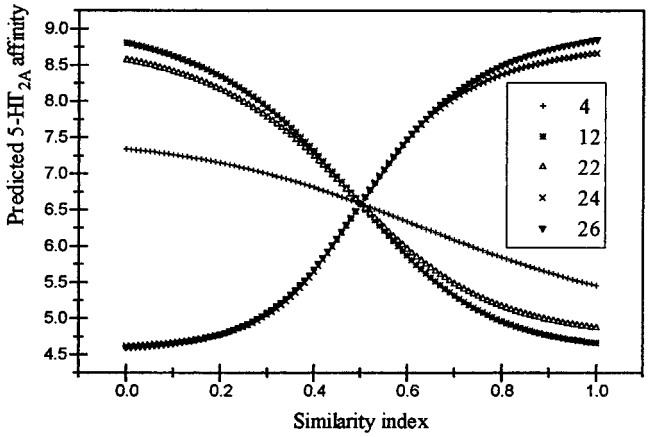


Figure 7. Functional dependence plot for the ESP model.

to predict affinity of a new 5-HT$_{2A}$ ligand one should use a neural network trained with the appropriate data.

Conclusions

This is the first report on 3D QSAR for 5-HT$_{2A}$ antagonists from the group of 2/4-(4-methylpiperazino)-pyrimidines by using SM/GNN calculations. The proposed, straightforward procedure that can be extended on analysis of new ligands includes (i) optimization of geometry and calculation of Mulliken charges for a given compound, (ii) superimposition onto the clozapine template with various modes of alignment, (iii) selection of the conformation giving the highest ESP similarity to the template molecule, followed by (iv) SM/GNN calculations.

Both steric and ESP properties were found to be important for 5-HT$_{2A}$ antagonist activity. Similarity indices calculated for the ligand−template clozapine pair are good criteria for selecting an "active conformation".

The influence of the charge on molecules on quality of derived QSARs was examined. Inclusion of positive charge on ligands in SM/GNN computations causes a high level of noise in the similarity matrices. To avoid this, while constructing SMs for cationic ligands one should consider their deprotonated counterparts whenever possible to improve the SM/GNN QSAR.

As shown by performing computations with diverse architecture of the neural network, the *n*-2-1 GNN variant is a good choice for the case studied. Qualitative differences between steric and ESP SM/GNN QSARs

were found. The latter incorporate higher order terms in QSAR. This finding is especially important when interpreting functional dependence plots, as one should take care drawing conclusions from these one-dimensional scans when higher order terms may play an essential role.

The high quality 3D QSARs may be used in further studies on prospective 5-HT_{2A} antagonists before a subsequent experimental validation. Moreover, set 5, performing superior in our study, along with other potent 5-HT_{2A} ligands, can be used for constructing a pseudo-receptor or other 3D QSAR models. In such a way a deeper insight into specific drug−receptor interactions can be gained, a significant extension of the knowledge derived solely from SM/GNN QSARs.

Experimental Section

General. ^{1}H NMR (300 MHz) and ^{13}C NMR (75 MHz) spectra of free bases were obtained at 28 °C in $CDCl_3$ with Me_4Si as an internal reference. Coupling constants smaller than 0.5 Hz are not reported. The NOE spectra were recorded by using the parameters and conditions reported previously.[27] Melting points (Pyrex capillary) are not corrected.

Chloropyrimidines 28 and 29. A solution of 2,4-dichloropyrimidine (0.75 g, 5 mmol) and imidazole (0.70 g, 10 mmol) in THF (25 mL) under a nitrogen atmosphere was allowed to stand at 23 °C for 12 h. Concentration on a rotary evaporator at 23 °C followed by silica gel chromatography gave **29** (hexanes/AcOEt, 7:3) and then **28** (hexanes/AcOEt, 1:9). Compounds **28** and **29** were crystallized from hexanes/toluene (3:1).

2-Chloro-4-(1-imidazolyl)pyrimidine (28): yield 37%; mp 124−125 °C. ^{1}H NMR δ 8.68 (d, J = 5.3 Hz, 1H), 8.44 (s, 1H), 7.67 (d, J = 1.5 Hz, 1H), 7.25 (d, J = 5.3 Hz, 1H), 7.25 (d, J = 1.5 Hz, 1H). Anal. ($C_7H_5ClN_4$) C, H, N.

4-Chloro-2-(1-imidazolyl)pyrimidine (29): yield 17%; mp 139−140 °C. ^{1}H NMR δ 8.59 (s, 1H), 8.57 (d, J = 5.4 Hz, 1H), 7.86 (s, 1H), 7.24 (d, J = 5.4 Hz, 1H), 7.17 (s, 1H). Anal. ($C_7H_5ClN_4$) C, H, N.

Dichloropyrimidines 32−34. These compounds were prepared from 2,4-dichloropyrimidine by using a general procedure[24] and crystallized from hexanes/toluene (4:1).

2,4-Dichloro-6-(3-furanyl)pyrimidine (32): yield 66%; mp 68−70 °C. ^{1}H NMR δ 8.24 (dd, J = 1.5, 0.9 Hz, 1H), 7.55 (dd, J = 2.1, 1.5 Hz, 1H), 7.34 (s, 1H), 6.87 (dd, J = 2.1, 0.9 Hz, 1H). Anal. ($C_8H_4Cl_2N_2O$) C, H, N.

2,4-Dichloro-6-(2-thienyl)pyrimidine (33): yield 21%; mp 108−109 °C. ^{1}H NMR δ 7.84 (dd, J = 3.8, 1.2 Hz, 1H), 7.63 (dd, J = 5.1, 1.2 Hz, 1H), 7.49 (s, 1H), 7.19 (dd, J = 5.1, 3.8 Hz, 1H). Anal. ($C_8H_4Cl_2N_2S$) C, H, N.

2,4-Dichloro-6-(*p*-tolyl)pyrimidine (34): yield 64%; mp 109−110 °C. ^{1}H NMR δ 7.97 (d, J = 8.4 Hz, 2H), 7.64 (s, 1H), 7.32 (d, J = 8.4 Hz, 2H), 2.44 (s, 3H). Anal. ($C_{11}H_8Cl_2N_2$) C, H, N.

Piperazinopyrimidines 4, 7−9, and 11−14. A solution of a chloropyrimidine **28**, **29** or **32**−**34** (2 mmol) and *N*-methylpiperazine (0.45 mL, 4 mmol) in EtOH (10 mL) under a nitrogen atmosphere was allowed to stand at 23 °C for 6 h and then concentrated at 23 °C on a rotary evaporator. The resultant pair of isomeric products **4/11**, **7/12**, **8/13**, or **9/14** was separated by silica gel chromatography (hexanes/Et_3N/EtOH, 3:2:1), and then the individual products were transformed into hydrobromide salts by using a general procedure.[28] The salts were crystallized from 95% EtOH.

4-(1-Imidazolyl)-2-(4-methypiperazino)pyrimidine (4, from 28): yield 87%; an oil. ^{1}H NMR δ 8.36 (s, 1H), 8.36 (d, J = 5.4 Hz, 1H), 7.60 (s, 1H), 7.17 (s, 1H), 6.50 (d, J = 5.4 Hz, 1H), 3.90 (t, J = 5.1 Hz, 4H), 2.49 (t, J = 5.1 Hz, 4H), 2.35 (s, 3H). ^{13}C NMR δ 161.4, 160.1, 155.3, 135.0, 130.8, 115.6, 96.5, 54.8, 46.1, 43.7. **4·2HBr·H_2O:** yield 86%; mp 229−231 °C. Anal. ($C_{12}H_{16}N_6$·2HBr·H_2O) C, H, N.

2-(1-Imidazolyl)-4-(4-methylpiperazino)pyrimidine (11, from 29): yield 81%; an oil. ^{1}H NMR δ 8.53 (s, 1H), 8.15 (d, J = 6.0 Hz, 1H), 7.82 (s, 1H), 7.12 (s, 1H), 6.38 (d, J = 6.0 Hz, 1H), 3.73 (t, J = 5.2 Hz, 4H), 2.52 (t, J = 5.2 Hz, 4H), 2.37 (s, 3H). ^{13}C NMR δ 162.2, 156.8, 154.3, 136.1, 129.9, 116.5, 100.6, 54.5, 46.0, 43.9. **11·2HBr·H_2O:** yield 77%; mp >300 °C. Anal. ($C_{12}H_{16}N_6$·2HBr·H_2O) C, H, N.

4-Chloro-6-(3-furanyl)-2-(4-methylpiperazino)pyrimidine (7, from 32): yield 16%; an oil. ^{1}H NMR δ 8.06 (dd, J = 1.4, 0.8 Hz, 1H), 7.48 (dd, J = 1.8, 1.4 Hz, 1H), 6.82 (dd, J = 1.8, 0.8 Hz, 1H), 6.64 (s, 1H), 3.89 (t, J = 5.1 Hz, 4H), 2.47 (t, J = 5.1 Hz, 4H), 2.34 (s, 3H). ^{13}C NMR δ 161.4, 161.3, 160.4, 143.9, 143.4. 125.3, 108.4, 104.6, 54.8, 46.1, 43.7. **7·HBr:** yield 76%; mp 290 °C (decomp.). Anal. ($C_{13}H_{15}ClN_4O$·HBr) C, H, N.

4-Chloro-2-(4-methylpiperazino)-6-(2-thienyl)pyrimidine (8, from 33): yield 26%; an oil. ^{1}H NMR δ 7.66 (dd, J = 3.8, 1.2 Hz, 1H), 7.47 (dd, J = 5.1, 1.2 Hz, 1H), 7.12 (dd, J = 5.1, 3.8 Hz, 1H), 6.81 (s, 1H), 3.9 (t, J = 5.2 Hz, 4H), 2.48 (t, J = 5.2 Hz, 4H), 2.35 (s, 3H); ^{13}C NMR δ 161.3, 160.9, 160.4, 142.3, 129.6, 128.0, 127.1, 102.9, 54.7, 46.1, 43.6. **8·HBr:** yield 73%; mp 264−266 °C (decomp.). Anal. ($C_{13}H_{15}ClN_4S$·HBr) C, H, N.

4-Chloro-2-(4-methylpiperazino)-6-(*p*-tolyl)pyrimidine (9, from 34): yield 13%; an oil. ^{1}H NMR δ 7.90 (d, J = 8.4 Hz, 2H), 7.26 (d, J = 8.4 Hz, 2H), 6.91 (s, 1H), 3.94 (t, J = 5.1 Hz, 4H), 2.48 (t, J = 5.1 Hz, 4H), 2.41 (s, 3H), 2.35 (s, 3H). ^{13}C NMR δ 165.8, 161.6, 161.4, 141.2, 133.8, 104.5, 54.8, 46.1, 43.8, 21.4. **9·HBr:** yield 97%; mp 285−286 °C. Anal. ($C_{16}H_{19}ClN_4$·HBr) C, H, N.

2-Chloro-6-(3-furanyl)-4-(4-methylpiperazino)pyrimidine (12, from 32): yield 77%, mp 122−124 °C. ^{1}H NMR δ 8.11 (dd, J = 1.7, 0.9 Hz, 1H), 7.48 (dd, J = 1.8, 1.7 Hz, 1H), 6.78 (dd, J = 1.8, 0.9 Hz, 1H), 6.46 (s, 1H), 3.71 (t, J = 5.0 Hz, 4H), 2.49 (t, J = 5.0 Hz, 4H), 2.35 (s, 3H). ^{13}C NMR δ 163.2, 160.8, 159.6, 143.9, 143.4, 124.9, 108.0, 95.8, 54.4, 46.0, 44.0. **12·HBr:** yield 93%; mp >300 °C. Anal. ($C_{13}H_{15}ClN_4O$·HBr) C, H, N.

2-Chloro-4-(4-methylpiperazino)-6-(2-thienyl)pyrimidine (13, from 33): yield 63%; mp 131.5−133.5 °C. ^{1}H NMR δ 7.70 (dd, J = 3.8, 1.2 Hz, 1H), 7.46 (dd, J = 5.1, 1.2 Hz, 1H), 7.11 (dd, J = 5.1, 3.8 Hz, 1H), 6.64 (s, 1H) 3.72 (t, J = 5.1 Hz, 4H), 2.50 (t, J = 5.1 Hz, 4H), 2.35 (s, 3H). ^{13}C NMR δ 163.1, 160.6, 159.8, 141.6, 129.1, 128.0. 126.8, 94.5, 54.4, 45.9, 44.0. **13·HBr:** yield 91%; mp >300 °C. Anal. ($C_{13}H_{15}ClN_4S$·HBr) C, H, N.

2-Chloro-4-(4-methylpiperazino)-6-(*p*-tolyl)pyrimidine (14, from 34): yield 63%, an oil. ^{1}H NMR δ 7.86 (d, J = 8.1 Hz, 2H), 7.25 (d, = 8.1 Hz, 2H), 6.74 (s, 1H), 3.74 (t, J = 5.1 Hz, 4H), 2.50 (t, J = 5.1 Hz, 4H), 2.40 (s, 3H), 2.35 (s, 3H); ^{13}C NMR δ 165.4, 163.5, 160.8, 140.8, 133.8, 129.3, 126.8, 96.1, 54.4, 45.9, 44.0. **14·HBr:** yield 96%; mp 267−269 °C (decomp.). Anal. ($C_{16}H_{19}ClN_4$·HBr) C, H, N.

Acknowledgment. The computing facilities were supported by a grant from the Polish National Science Fundation (KBN) (KBN/SGI_ORIGIN_2000/UJ/042/1999).

References

(1) Hatrig, P. R. Molecular biology of 5-HT receptors. *Trends Pharmacol. Sci.* **1989**, *10*, 64−69.

(2) Höltje, H.-D.; Briem, H. Theoretical Determination of the Putative Receptor-bound Conformations of 5-HT_2 Receptor Agonists. *Quant. Struct.-Act. Relat.* **1991**, *10*, 193−197. Höltje, H.-D.; Jendretzki, U. Conformational Analysis of 5-HT_2 Receptor Antagonists. *Pharm. Pharmacol. Lett.* **1992**, *1*, 89−92. Höltje, H.-D.; Jendretzki, U. Construction of a Detailed Serotoninergic 5-HT_{2A} Receptor Model. *Arch. Pharm. (Weinheim)* **1995**, *328*, 577−584.

(3) Mokrosz, J. L.; Strekowski, L.; Duszyńska, B.; Harden, D. B.; Mokrosz, M. J.; Bojarski, A. J. Structure−Activity Relationship Studies of CNS Agents. Part 14: Structural Requirements for the 5-HT_{1A} and 5-HT_{2A} Receptor Selectivity of Simple 1-(2-pyrimidinyl)piperazine Derivatives. *Pharmazie* **1994**, *49*, 801−806.

(4) Andersen, K.; Liljefors, T.; Gundertofte, K.; Perregaard, J.; Bogeso, K. P. Development of a Receptor-Interaction Model for Serotonin 5-HT$_2$ Receptor Antagonists. Predicting Selectivity with Respect to Dopamine D$_2$ Receptors. *J. Med. Chem.* **1994**, *37*, 950−962.
(5) So, S.-S.; Karplus, M. Three-dimensional quantitative structure−activity relationships from molecular similarity matrices and genetic neural networks. 1. Method and validations. *J. Med. Chem.* **1997**, *40*, 4347−4359.
(6) So, S.-S.; Karplus, M. Three-dimensional quantitative structure−activity relationships from molecular similarity matrices and genetic neural networks. 2. Applications. *J. Med. Chem.* **1997**, *40*, 4360−4371.
(7) Mokrosz, M. J.; Strekowski, L.; Kozak, W. X.; Duszyńska, B.; Bojarski, A. J.; Kłdzinska, A.; Czarny, A.; Cegła, M. T.; Dereń-Wesołek, A.; Chojnacka-Wójcik, E.; Dove, S.; Mokrosz, J. L. 4,6-Di(heteroaryl)-2-(*N*-methylpiperazino)pyrimidines as New, Potent 5-HT$_{2A}$ Receptor Ligands: A Verification of the Topographic Model. *Arch. Pharm.* **1995**, *328*, 659−666.
(8) Montanari, C. A.; Tute, M. S.; Beezer, A. E.; Mitchell, J. C. Determination of receptor-bound drug conformations by QSAR using flexibile fitting to derive a molecular similarity index. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 67−73.
(9) Edvardsen, O.; Sylte, I.; Dahl, S. Molecular dynamics of Serotonin and Ritanserin Interacting with the 5-HT$_2$ Receptor. *Mol. Brain Res.* **1992**, *14*, 166−178 and references therein.
(10) Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning*; Addison-Wesley Publishing Company, Inc.: Reading, MA, 1989.
(11) So, S.-S.; Karplus, M. Evolutionary Optimization in Quantitative Structure−Activity Relationship: An Application of Genetic Neural Networks. *J. Med. Chem.* **1996**, *39*, 1521−1530.
(12) So, S.-S.; Karplus, M. Genetic Neural Networks for Quantitative Structure−Activity Relationships: Improvement and Application of Benzodiazepine Affinity for Benzodiazepine/GABA$_A$ Receptors. *J. Med. Chem.* **1996**, *39*, 5246−5256.
(13) Good, A. C.; Peterson, S. J.; Richards, W. G. QSAR's from Similarity Matrixes. Technique Validation and Application in the Comparison of Different Similarity Evaluation Methods. *J. Med. Chem.* **1993**, *36*, 2929−2937.
(14) Horwell, D. C.; Howson, W.; Higginbottom, M.; Naylor, D.; Ratcliffe, G. S.; Williams, S. Quantitative Structure−Activity Relationships (QSARs) of N-Terminus Fragments of NK1 Tachykinin Antagonists: A Comparison of Classical QSARs and Three-Dimensional QSARs from Similarity Matrices. *J. Med. Chem.* **1995**, *38*, 4454−4462.
(15) So, S.-S.; Richards, W. G. Application of Neural Networks: Quantitative Structure−Activity Relationships of the Derivatives of 2,4-Diamino-5-(substituted-benzyl)pyrimidines as DHFR Inhibitors. *J. Med. Chem.* **1992**, *35*, 3201−3207.
(16) Andrea, T. A.; Kalayeh, H. Applications of Neural Networks in Quantitative Structure−Activity Relationships of Dihydrofolate Reductase Inhibitors. *J. Med. Chem.* **1991**, *34*, 2824−2836.
(17) Hodgkin, E. E.; Richards, W. G. Molecular similarity based on electrostatic potential and electric field. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1987**, *14*, 105−110.
(18) Meyer, A. M.; Richards, W. G. Similarity of molecular shape. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 421−439.
(19) InsightII release 97.0, Biosym/MSI, San Diego, 1997.
(20) Stewart, J. J. P. MOPAC − A Semiempirical Molecular Orbital Program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1−45.
(21) Heisterberg, D. J., 1990, unpublished results.
(22) Møller, M. F. A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning. *Neural Networks* **1993**, 525−533.
(23) Available from author (T.B.) on request.
(24) Strekowski, L.; Harden, D. B.; Grubb, W. B.; Patterson, S. E.; Czarny, A.; Mokrosz, M. J.; Cegła, M. T.; Wydra, R. L. Synthesis of 2-Chloro-4,6-di(heteroaryl)pyrimidines. *J. Heterocycl. Chem.* **1990**, *27*, 1393−1400.
(25) Bojarski, A. J.; Cegła, M. T.; Charakchieva-Minol, S.; Mokrosz, M. J.; Maækowiak, M.; Misztal, S.; Mokrosz, J. L. Structure−Activity Relationship Studies of CNS Agents Part 9: 5-HT$_{1A}$ and 5-HT$_{2A}$ Receptor Affinity of Some 2- and 3-Substituted 1,2,3,4-Tetrahydro-β-carbolines. *Pharmazie* **1993**, *48*, 289−294.
(26) Brocławik, E.; Borowski, T. Characteristics of the Ligand-Binding Site Interaction for a Series of Arecoline-Derived Muscarinic Agonists: A Quantum Chemical Study. *Comput. Chem.*, in press.
(27) Wydra, R. L.; Harden, D. B.; Strekowski, L.; Battiste, M. A.; Coxon, J. M. *Tetrahedron* **1992**, *48*, 3485−3494.
(28) Strekowski, L.; Mokrosz, J. L.; Honkan, V. A.; Czarny, A.; Cegła, M. T.; Wydra, R. L.; Patterson, S. E.; Schinazi, R. F. Synthesis and Quantitative Structure−Activity Relationship Analysis of 2-(Aryl or Heteroaryl)quinolin-4-amines, a New Class of Anti-HIV-1 Agents. *J. Med. Chem.* **1991**, *34*, 1739−1746.
(29) Leysen, J. E.; Niemegeers, C. J. E.; Van Nueten, J. M.; Laduron, P. M. [^{3}H]Ketanserin (R 41 468), a Selective ^{3}H-Ligand for Serotonin 2 Receptor Binding Sites. Binding Properties, Brain Distribution, and Functional Role. *Mol. Pharmacol.* **1982**, *21*, 301−314.

JM9911332