Available online at www.sciencedirect.com
ScienceDirect
Mendeleev Commun., 1995, **5(2)**, 49–51

Mendeleev Communications

The Surface Molecular Potential Method for Calculating the Structure–Activity Relationship for Psychotropic Compounds

Igor V. Svitan'ko,*[a] Ivan L. Zyryanov,[b] Mikhail I. Kumskov,[b] Lenor I. Khmel'nitskii,[b] Lyudmila I. Suvorova,[b] Angelina N. Kravchenko,[b] Tatyana B. Markova,[b] Oleg V. Lebedev,[b] Galina A. Orekhova[b] and Svetlana V. Belova[b]

[a] *Higher Chemical College, Russian Academy of Sciences, 125819 Moscow, Russian Federation. Fax: +7 095 135 8860*
[b] *N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 117913 Moscow, Russian Federation. Fax: +7 095 135 5328*

A mathematical model of psychotropic activity has been built based on calculations of the electrostatic potential of molecules of compounds possessing activity of this type; the conclusion has been drawn that this method has general applicability for solving the "structure–property" problem.

In the late seventies, investigators from the N. D. Zelinsky Institute of Organic Chemistry of the RAS and from the Kazan Medical Institute found a new class of compounds possessing psychotropic activity, namely, fused saturated bicyclic bisureas (BBUs). Since then, many BBUs and their heteroanalogues have been synthesized. Their psychotropic activity has been studied, but the dependence of the activity of these compounds on their structure revealed so far is entirely empirical.[1]

In the present work we use the procedure reported previously[2] to try to reveal the "structure–property" relationship with respect to the psychotropic activity of compounds containing urea, guanidine or sulfonylamide moieties. We used this highly complicated method because attempts to reveal a relation between the structure and the psychotropic activity undertaken by the authors using the embedded descriptors of BIBIGON,[6,7] a simple and efficient system, did not afford the expected results.

The study was carried out according to the general method[2] and its particular implementation.[3] This sequence of logical steps is presented below.

1. We used a training set of 47 compounds, nine of which

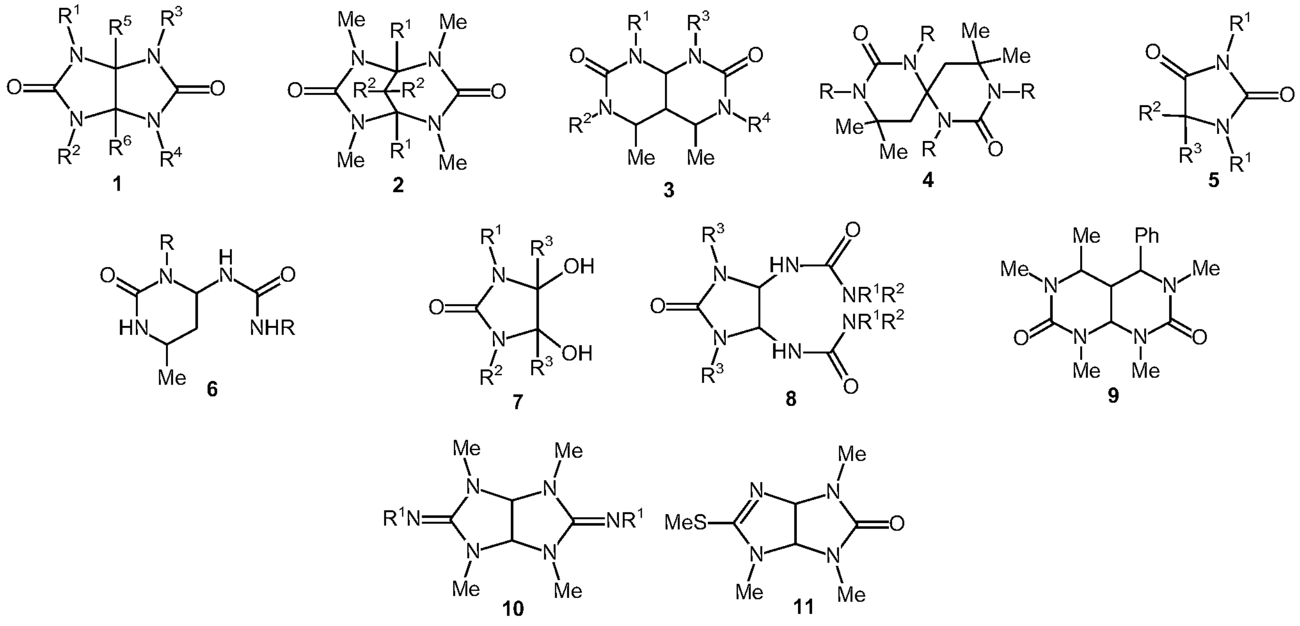

Fig. 1 Basic structures of the training set.

possess pronounced activity, 26 compounds are inactive, and the remaining 12 compounds have intermediate activity. The compounds used were divided into three classes, compounds of class "1" having well-pronounced activity, those of class "0" having low activity or no activity, and class "Z" including compounds with intermediate activity. All compounds included in the set belong to the types shown in Fig. 1. The following substituents were used: H, Me, Et, Pr^n, Pr^i, Bu^n, Ph, CH_2OMe, CH_2OH and CH_2CH_2OH. Furthermore, an additional set was used which contained 21 compounds with unknown activity.

2. Geometry optimization was performed for each molecule of the starting set. The optimization involved two steps: initial optimization by molecular mechanics methods using the MMX force field[4] and final optimization by quantum-chemical methods using the PM3 hamiltonian.[5]

3. Each molecule was surrounded by a surface formed by the centre of a probe sphere (radius 1.6 Å)[3] when it encircles the van der Waals surface. The surface was represented by points with a 0.2 Å mean distance between the nearest neighbours. The molecular electrostatic potentials (MEP) were calculated in each point of the surface. Then, points were found in which the MEP had the maximum or minimum values for the point itself and for its nearest environment. Such a set of points for each molecule was used for subsequent analysis. We shall call these points representation points (RPs), in accordance with ref. 3.

4. The range of the MEP in the RPs was divided into intervals depending on how often the particular values of MEP are met. For this purpose we built a histogram of MEP distribution. Then, necessary boundaries of the intervals were chosen in the points of minima and maxima of the histogram. Furthermore, the high and broad peaks at the histogram were additionally divided in a uniform way.

The range of distances between the RPs (0–18 Å) was divided into intervals (mean step 0.2 Å).

5. For each molecule, all possible RP pairs were tabulated, and a set of descriptors of the type $(E1,E2,R),N$ was constructed, where $E1$ and $E2$ are the numbers of intervals of the MEP range for the first and the second point in a pair; R is the number of intervals in the interval between the representation points; and N is the occurrence of a given descriptor for the molecule.

6. The set of descriptors was determined for the whole collection of molecules. A unique number was assigned to each descriptor belonging to the set. Thus, an $(X_{j1},...,X_{jN})$ vector of attributes was assigned to each molecule, where X_{ij} is the occurrence of the ith descriptor in the jth molecule and N is the number of descriptors in the whole set of descriptors.

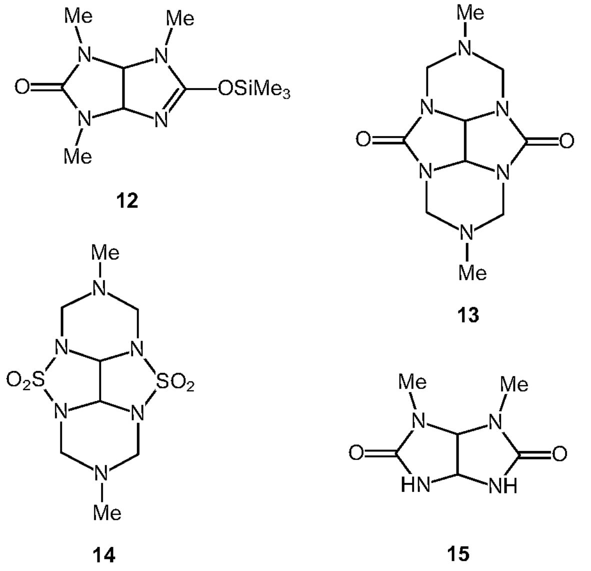

Fig. 2 Structures from the additional set predicted to have high activity.

7. Then, a few significant attributes, *i.e.* those most successfully describing the type of activity studied, were chosen from the total set of descriptors. A particular number was assigned to each significant attribute in the new set, and a vector of significant attributes $(Y_{j1},...,Y_{jM})$, where Y_{ij} is the occurrence of the ith significant attribute in the jth molecule, was assigned to each molecule. Based on these attributes, we constructed linear models describing a property of a molecule:

$$P_j = a_0 + \sum a_i Y_{ij}, \ i = 1 \ldots M$$

where P_j is the magnitude of the given property for the jth molecule in the set; Y_{ij} is the magnitude of the ith significant attribute for the jth molecule; M is the total number of significant attributes; a_0 and a_i are the desired coefficients of the linear model.

The significant attributes were chosen and the linear models were constructed using the BIBIGON system.[6,7] As a

Table 1 Dependence of the number of erroneous attributions on the number of descriptors included in the model.

N_d	$0 \rightarrow Z$	$0 \rightarrow 1$	$Z \rightarrow 0$	$Z \rightarrow 1$	$1 \rightarrow 0$	$1 \rightarrow Z$	Errors
22	0(0)	0(0)	3(3)*	0(0)	0(2)	0(0)	3(5)
20	0(0)	0(0)	3(3)*	0(2)	0(2)	0(0)	3(7)
16	0(0)	0(0)	3(3)*	0(1)	0(2)	0(0)	3(7)
12	1(3)	0(0)	3(4)	0(0)	0(2)	0(1)	4(10)
10	2(3)	0(0)	4(4)	0(0)	0(2)	0(1)	6(10)
8	3(3)	0(0)	5(5)	0(1)	0(2)	1(2)	9(13)

result, we obtained a series of models differing in complexity, which describe the property–structure dependence with different accuracy (see Table 1).

In Table 1, N_d is the number of descriptors in a model; $0 \rightarrow Z$ is the number of wrong attributions of type "0" compounds to type "Z";... ; $1 \rightarrow Z$ is the number of wrong attributions of type "1" compounds to type "Z"; Errors is the overall number of errors. Each cell of Table 1 presents two numbers. They show the number of errors of the type indicated in the column heading at as many descriptors in the model as indicated in the first column of the table. Numbers not in parentheses refer to model testing on the training set. Numbers in parentheses refer to testing on the training set using the cross-validation method. The method consists of the following: first, one compound is excluded from the training set and the other compounds are used to construct a model; the activity of the excluded compound is then predicted based on this model and compared to the actual activity of this compound. Consequently, another compound is excluded from the set, and the same procedure is repeated. This procedure is repeated for each compound of the training set. Asterisks in certain table cells indicate the "non-disappearing" errors that do not disappear as the complexity of the model is increased.

A detailed analysis of the models showed that three of the five erroneously assigned structures (two $Z \rightarrow 0$ errors and one $1 \rightarrow 0$ error) have high conformational mobility (general formula **8**, Fig. 1). Therefore, it is reasonable to assume that the conformation with the optimum energy used in the calculations is inactive, while the compound displays activity in another, less favourable conformation. This allows us to explain why the activity of these compounds is underestimated. A similar explanation is also applicable to the remaining misattributed structure (type $1 \rightarrow 0$) of general formula **1** containing four Et groups at the nitrogen atom. The above assumption is supported by the fact that the number of erroneous classifications decreased sharply when these structures were excluded from the training set.

Evidently, the number of compounds with overestimated activity is very small (see the Errors column in Table 2), and they are absent completely when a sufficient number of descriptors are used. However, structures can exist whose activity will be overestimated, no matter how complex the model is. This should indicate that a molecule, whose active moiety is located on the potential energy surface, cannot be coordinated on the receptor due to steric hindrance. The fact that this phenomenon is not observed in the set studied indicates that the above method for constructing descriptors sufficiently well accounts not only for energetic but also for geometric characteristics of molecules.

Table 2 Number of erroneous attributions on the number of descriptors included in the model. Refined models.

N_d	$0 \rightarrow Z$	$0 \rightarrow 1$	$Z \rightarrow 0$	$Z \rightarrow 1$	$1 \rightarrow 0$	$1 \rightarrow Z$	Errors
14	0(1)	0(0)	0(1)	0(0)	0(0)	0(1)	0(3)
8	0(0)	0(0)	1(2)	0(0)	0(0)	0(2)	1(4)

It is noteworthy that constructing the classification was complicated by the absence of detailed quantitative information on the activity of compounds: only approximate classification with quite diffuse boundaries between the classes was available. For example, the activity factor was less than 2.5 for class "0", more than 3.5 for class "1", and within 2.5–3.5 for class "Z".

The models derived were then used to predict the activity of additional compounds not included in the training set. The analysis of 21 compounds resulted in four structures which are, potentially, highly active (Fig. 2, structures **12**–**15**).

The successful use of the above method to reveal the relation between the structure of BBUs and their psychotropic activity allows us to assume that this method, which is, in essence, a method for determining the complementary surface potentials of molecules, may also be useful for studying other types of biological activity and physicochemical properties.

This study was financially suppported by the Russian Foundation for Fundamental Research (grants no. 94-01-00041 and 94-03-09323).

References

1 O. V. Lebedev, L. I. Khmel'nitskii, L. V. Epishina, L. I. Suvorova, I. V. Zaikonnikova, I. E. Zimakova, S. V. Kirshin, A. M. Karpov, V. S. Chunovskii, M. V. Povstyanoi and V. A. Eres'ko, in *Tselenapravlennyi poisk novykh neirotropnykh preparatov (Directed search for novel neurotropic means)*, Zinatne, Riga, 1983, p. 81 (in Russian).

2 I. V. Svitan'ko, M. I. Kumskov, I. L. Zyryanov and I. A. Suslov, *Mendeleev Commun.*, 1994, 161.

3 M. I. Kumskov, I. L. Zyryanov and I. V. Svitanko, *Pattern Recognition and Image Analysis*, 1995 (in press).

4 U. Burkert and N. L. Ellinger, *Molekulyarnaya Mekhanika (Molecular Mechanics)*, Moscow, Mir, 1986 (in Russian).

5 J. J. P. Stewart, *J. Comput. Chem.*, 1989, 221.

6 M. I. Kumskov, L. A. Ponomareva, E. A. Smolenskii, D. F. Mityushev and N. S. Zefirov, *Izv. Akad. Nauk, Ser. Khim*, 1994, 1391 (*Russ. Chem. Bull.*, 1994, **43**, 1317).

7 M. I. Kumskov, E. A. Smolenskii, L. A. Ponomareva, D. F. Mityushev and N. S. Zefirov, *Dokl. Akad. Nauk*, 1994, **336**, 64 (in Russian).

Received: Moscow, 9th December 1994
Cambridge, 9th January 1995; Com. 4/07741E