STUDIA UNIVERSITATIS BABEŞ-BOLYAI, CHEMIA, LV, 1, 2010

MODELING OF BIOLOGICALLY ACTIVE MOLECULAR STRUCTURES

ERIKA TASNADI[a], GABRIEL KATONA[a], MIRCEA V. DIUDEA[a]

ABSTRACT. In view of finding the best relationship between biological activity and structure using topological indices, a QSAR/QSPR study was performed on a set of 44 benzodiazepin-derivatives. In a similarity test, the highest scored structures were submitted to mono and multi linear regression analysis and the optimal model was drawn out. The activity of several newly designed by us structures, belonging to the same class of molecules, was predicted with the use of the title model.

Keywords: *molecular modeling, biological activity, QSAR-QSPR*

INTRODUCTION

The projection of molecular structures with desired physico-chemical or biological properties is one of the major goals of Molecular Topology. The identification of the crucial factors involved in the structure-property relationship is gained by the comparative analysis of a set of molecules. This analysis is realized by the aid of topological descriptors (indices) [1] and regression analysis, included in various algorithms. The topological characterization of the chemical structures allows their classification based on similarity criteria.

A bioactive molecule interacts with a biological receptor within an *effector-receptor complex*, in which the two partners mutually accommodate their structures and energies such that the resulting complex exists sufficiently long for generating a biological response. The structural features of the bioactive molecules, responsible for a given biological activity (i.e. pharmacological profile), are called *pharmacophore*. It plays an essential role in the recognition process preceding the complex formation. The receptor (i.e. a situs in a biomacromolecule) must contain complementary features to the pharmacophore for the recognition (complexation and ultimately biological response) to occur. In addition to the pharmacophore and its complementary atoms, the effector and biomacromolecule present *areas of nonspecific interactions*, which modulate the biological response.

[a] *Faculty of Chemistry and Chemical Engineering, "Babes-Bolyai" University, Arany Janos Str. 11, 400084, Cluj, Romania, diudea@chem.ubbcluj.ro*

Several 2D- and 3D-QSAR approaches are performed to model the receptor-drug interaction, with ultimate aim the drug design. A quantitative structure-activity relationship (QSAR) correlates the changes in the observed biological activity with the changes in the chemical structure of a series of molecules.

Quantitative Structure–Property Relationship (QSPR) and Quantitative Structure–Activity Relationship (QSAR) studies [2] give a simple and reasonable support for predicting physico-chemical and biological properties. A QSPR/QSAR study looks for connections that link directly the molecular structure to the properties of the chemical compounds [1,2]. By the aid of structural descriptors / topological indices (calculated on molecular graphs), various mathematical models are developed to relate quantitatively the molecular structures and their physico-chemical or biological properties.

This method is inductive and the errors are minimized, verified and validated statistically.

The current methods to quantify the similarity [3] of two or more molecules take into consideration three types of the molecular structure description:

- Topological- by means of molecular graphs
- Configurational- using of atomic coordinates in 3D space
- Quantum mechanic- using data on molecular fields, energy levels of molecular orbitals, etc.

The present report is structured as follows: after an introductive part (see above), the “Method” introduces to the procedure used for optimization and calculation of the indices. The “Results and Discussion” present the results and comments the performance of the used method.

METHOD

The following procedure was used to find the best relationship between structures and biological property:

1. structures are optimized to find a minimum-energy (stable) configuration (HYPER CHEM version 7.52)
2. an index database is generated by using DRAGON 5.0 software [4]
3. an exhaustive search to find the best equation (with the correlation coefficient R> 0.90), by STATISTICA 6.0, software
4. prediction of biological property for newly designed structures

Thus, before computing the indices, it is important to find a minimum-energy (stable) configuration for a molecular system, especially for 3D-type descriptors. We used for geometry optimization the molecular mechanics method MM+, with the Polak-Ribiére conjugate gradient, at RMS lower than 0.009.

The next step was the calculation of indices. The initial database contained 1630 indices which were generated by using Dragon software. From each descriptor block (as provided by DRAGON), the indices showing the best correlation coefficient in monovariate regression against the biological activity, were collected out.

The statistical analysis was performed by using the STATISTICA software package. The procedure to find the best model for the prediction of the biological activity follows the below steps 1-3:

1. search for the best monovariate regression
2. search for the best bi- and tri-variate regression
3. prediction of biological activity based on the best regression equation, of the form: $y_i = a + \sum_j b_j x_{ij}$ where b_j are regression coefficients, a is a constant, y_i is the dependent variable (e.g., the modeled property, log IC_{50}) and x_{ij} are the independent variables (in particular, topological indices), j=1,2,..m for the molecular structures i, i=1,2,…n.

RESULTS AND DISCUSSION

Structure of benzodiazepine derivatives [5]

Benzodiazepines (BZs) play an important role in medicinal chemistry, e.g. as a nucleus for combinatorial synthesis. The benzodiazepines caught the interest of scientists because of their putative role in mammalian central nervous system and for their exceptional tolerance; they form an important class of therapeutic agents with wide-spread applications in the treatment of anxiety and emotional disorders.

In our investigation, we used the set of classical 1,4-BZs derivatives listed in **Table 1**, having the general formula shown in **Figure 1.** The six substituent positions are R^7 and R^8 in A-ring, R^1 and R^3 in B-ring, and $R^{2'}$ and $R^{6'}$ in the C-ring.

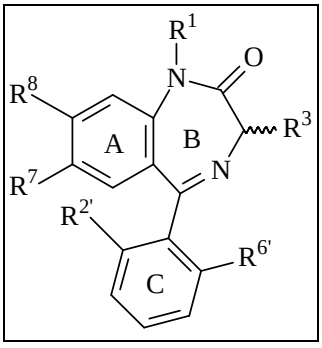


Figure 1. Classical 1,4-BZs

Table 1. Benzodiazepine derivatives and their binding affinities (log IC_{50})

ID	Name	R^7	R^1	$R^{2'}$	$R^{6'}$	R^3	R^8	log IC_{50}
1	Ro 05-3061	F	H	H	H	H	H	1.602
2	Ro 05-4865	F	CH_3	H	H	H	H	1.230
3	Ro 05-6820	F	H	F	H	H	H	0.869
4	Ro 05-6822	F	CH_3	F	H	H	H	0.708
5	Nordazepam	Cl	H	H	H	H	H	0.973
6	Diazepam	Cl	CH_3	H	H	H	H	0.908
7	Ro 05-3367	Cl	H	F	H	H	H	0.301
8	Delorazepam	Cl	H	Cl	H	H	H	0.255
9	Ro 07-9957	I	CH_3	F	H	H	H	0.462
10	Ro 05-2904	CF_3	H	H	H	H	H	1.114
11	Ro 14-3074	N_3	H	F	H	H	H	0.724
12	Nitrazepam	NO_2	H	H	H	H	H	1.000
13	Ro 05-4435	NO_2	H	F	H	H	H	0.176
14	Flunitrazepam	NO_2	CH_3	F	H	H	H	0.580
15	Clonazepam	NO_2	H	Cl	H	H	H	0.255
16	Ro 05-4082	NO_2	CH_3	Cl	H	H	H	0.342
17	Ro 05-3590	NO_2	H	CF_3	H	H	H	0.544
18	Ro 20-7736	NHOH	CH_3	F	H	H	H	1.982
19	Ro 05-3072	NH_2	H	H	H	H	H	2.587
20	Ro 05-4318	NH_2	CH_3	H	H	H	H	2.663
21	Ro 20-1815	NH_2	CH_3	F	H	H	H	1.813
22	Ro 05-4619	NH_2	H	Cl	H	H	H	1.875
23	Ro 05-4528	CN	CH_3	H	H	H	H	2.580
24	Ro 20-2541	CN	CH_3	F	H	H	H	1.477
25	Ro 20-2533	CH_2CH_3	H	H	H	H	H	1.556
26	Ro 20-5747	$CH=CH_2$	H	H	H	H	H	1.380
27	Ro 20-5397	CHO	H	H	H	H	H	1.633
28	Ro 20-3053	$COCH_3$	H	F	H	H	H	1.255
29	Ro 05-2921	H	H	H	H	H	H	2.544
30	Ro 05-4336	H	H	F	H	H	H	1.322
31	Ro 05-4520	H	CH_3	F	H	H	H	1.146
32	Ro 05-4608	H	CH_3	Cl	H	H	H	0.580
33	Ro 07-4419	H	H	F	F	H	H	1.279
34	Ro 07-3953	Cl	H	F	F	H	H	0.204
35	Ro 07-4065	Cl	CH_3	F	F	H	H	0.613
36	Ro 07-5193	Cl	H	Cl	F	H	H	0.477
37	Ro 22-3294	Cl	H	Cl	Cl	H	H	0.845
38	Ro 07-5220	Cl	CH_3	Cl	Cl	H	H	0.740
39	Ro 13-3780	Br	CH_3	F	F	H	H	0.380
40	Ro 07-6198	H	H	F	F	H	Cl	1.447
41	Ro 20-8895	H	H	F	H	H	CH_3	1.279
42	Ro 22-6762	Cl	CH_3	H	H	H	Cl	1.602
43	Ro 20-8065	Cl	H	F	H	H	Cl	0.556
44	Ro 20-8552	CH_3	H	F	H	H	Cl	1.146

Table 2 presents the symbol and description of the indices which show the best correlation with the biological property. In Table 3 the values of these descriptors are presented.

Table 2. Topological indices: the meaning of descriptors (DRAGON 5.0).

Symbol	Descriptor meaning	Descriptor block
EEig04d	**Eig**envalue **04** from **e**dge adj. matrix weighted by **d**ipole moments	Edge adjacency indices
nNH_2	**N**umber of primary amines (aliphatic)	Functional group counts
XMOD	**Mod**ified Randic connectivity index	Connectivity indices
Hy	**Hy**drophilic factor	Molecular properties
IC0	**I**nformation **c**ontent index (neighborhood symmetry of **0**-order)	Information indices
MATS3e	**M**oran autocorrelation – lag **3**/ weighted by **at**omic **S**anderson **e**lectronegativities	2D autocorrelation
BIC0	**B**ond **i**nformation **c**ontent (neighborhood symmetry of **0**-order)	Information indices
S0K	**K**ier **s**ymmetry index	Topological descriptors
RDF045m	**R**adial **d**istribution **f**unction – **4.5**/ weighted by atomic **m**asses	RDF descriptors
G1	**G**ravitational index **G1**	Geometrical descriptors
H1m	Autocorrelation of lag **1**/ weighted by atomic **m**asses	Getaway descriptors

The best obtained monovariate equation was:

$$y=8.72-0.73\times EEig04d \quad (1)$$

R= 0.7310; s= 0.47; F= 48.19; n= 44

where: R - is the correlation coefficient; s - standard error of estimate; F - Fisher's statistics and n- number of molecules and the best bivariate equation:

$$y= 7.95+0.426\times Hy-0.61\times EEig04d \quad (2)$$

R= 0.8378; s= 0.38; F= 48.25; n= 44

Because the statistics of correlation (eqs. 1 and 2) were not satisfactory, we used the program CLUJSIMIL [6,7] to cluster the structures of more than 95 percent similarity with respect to the reference structure (denoted by us RS), by choosing the molecule with the highest log IC_{50}, namely Ro 05-4435 (structure 13).

Table 3. Topological indices: values in the studied BZs

#	EEig 04d	nNH_2	XMOD	Hy	IC0	MATS 3e	BIC0	S0K	RDF 045m	G1	H1m
1	2.465	0	57.729	-0.776	1.618	-0.082	0.304	66.597	3.439	26.97	1.336
10	2.605	0	66.381	-0.717	1.747	-0.016	0.322	79.771	5.59	34.187	1.592
11	2.529	0	67.832	-0.672	1.739	-0.103	0.315	81.164	5.948	34.1	1.25
12	2.515	0	64.226	-0.704	1.669	-0.092	0.306	73.418	4.563	31.23	1.309
13	2.749	0	67.556	-0.672	1.819	-0.133	0.333	82.794	6.404	35.314	1.351
14	2.767	0	70.335	-0.686	1.782	-0.108	0.321	85.44	6.729	37.611	1.285
15	2.756	0	69.865	-0.672	1.819	-0.134	0.333	82.794	11.126	38.472	1.385
16	2.772	0	72.644	-0.686	1.782	-0.104	0.321	85.44	11.452	40.827	1.313
17	2.914	0	76.208	-0.632	1.932	-0.074	0.348	96.605	9.597	50.247	1.419
18	2.547	0	67.465	-0.260	1.724	-0.058	0.312	81.164	5.739	34.856	1.226
19	2.226	1	57.152	0.349	1.517	-0.022	0.281	64.062	3.54	26.703	1.146
2	2.611	0	60.508	-0.787	1.587	-0.051	0.292	69.309	3.765	29.153	1.262
20	2.333	1	59.93	0.319	1.49	0.025	0.271	66.812	3.866	28.878	1.09
21	2.535	1	63.261	0.338	1.642	-0.054	0.299	75.733	5.77	32.699	1.129
22	2.473	1	62.791	0.368	1.676	-0.079	0.311	73.113	10.115	33.208	1.222
23	2.533	1	63.223	0.291	1.466	0.086	0.262	69.595	4.186	30.254	1.065
24	2.589	1	66.553	0.340	1.61	-0.004	0.288	78.402	6.11	34.182	1.097
25	2.392	0	60.091	-0.836	1.406	0.075	0.255	65.621	4.028	27.872	1.083
26	2.237	0	60.091	-0.836	1.417	0.048	0.258	68.042	3.606	27.732	1.14
27	2.571	0	60.798	-0.787	1.531	-0.001	0.282	70.613	3.59	28.099	1.297
28	2.678	0	67.628	-0.762	1.639	-0.048	0.295	79.763	6.379	33.731	1.158
29	2.043	0	54.5	-0.818	1.453	0.041	0.273	60.112	3.324	24.648	1.138
3	2.688	0	61.059	-0.738	1.713	-0.127	0.322	75.992	5.344	30.775	1.379
30	2.4	0	57.83	-0.776	1.618	-0.049	0.304	69.13	5.229	28.217	1.181
31	2.528	0	60.609	-0.787	1.587	-0.019	0.292	71.733	5.554	30.416	1.124
32	2.535	0	62.918	-0.787	1.587	0.013	0.292	71.733	10.224	32.94	1.162
33	2.592	0	61.16	-0.738	1.713	-0.104	0.322	71.992	6.197	31.178	1.266
34	2.714	0	66.699	-0.704	1.869	-0.162	0.351	78.874	10.665	34.867	1.653
35	2.724	0	69.477	-0.717	1.831	-0.133	0.337	81.608	10.991	37.183	1.568
36	2.716	0	69.008	-0.704	1.869	-0.163	0.351	83.074	15.454	37.931	1.687
37	2.717	0	71.317	-0.704	1.777	-0.169	0.334	78.874	20.478	39.523	1.767
38	2.727	0	74.096	-0.717	1.747	-0.128	0.322	81.608	20.804	41.935	1.683
39	2.698	0	74.674	-0.717	1.831	-0.121	0.337	81.608	8.115	41.412	2.061
4	2.699	0	63.838	-0.750	1.68	-0.1	0.31	78.556	5.669	33.017	1.305
40	2.748	0	66.699	-0.704	1.869	-0.162	0.351	78.874	7.78	34.411	1.591
41	2.549	0	60.193	-0.787	1.587	-0.047	0.292	71.733	5.198	29.847	1.124
42	2.635	0	68.457	-0.750	1.68	0.039	0.31	76.011	9.947	34.013	1.741
43	2.785	0	69.008	-0.704	1.869	-0.087	0.351	83.074	11.28	35.793	1.878
44	2.767	0	65.832	-0.750	1.741	-0.09	0.321	78.556	7.363	33.545	1.363
5	2.472	0	60.038	-0.776	1.618	-0.06	0.304	66.597	7.792	27.883	1.528
6	2.619	0	62.817	-0.787	1.587	-0.018	0.292	69.309	8.118	30.095	1.441
7	2.701	0	63.368	-0.738	1.779	-0.115	0.334	75.992	9.697	31.831	1.569
8	2.708	0	65.678	-0.738	1.713	-0.115	0.322	75.992	14.367	34.831	1.603
9	2.632	0	76.54	-0.750	1.741	-0.034	0.321	78.556	10.551	41.728	2.515

Table 4 lists the similarity data of the structures.

The structures bellow 90 percent of similarity are out of our interest and, after rejecting them from the table, 38 structures remained. Next, we performed mono- and multiple- regressions to find the best models.

Table 4. Similarity coefficients with respect to the reference structure (Ro 05-4435)

MolID	Similarity	MolID	Similarity
13	1.00000	25	0.92442
10	0.99968	26	0.92415
38	0.99941	27	0.92093
15	0.99862	22	0.92004
18	0.99842	20	0.91980
39	0.99744	41	0.91945
28	0.99740	8	0.91785
35	0.99663	32	0.91761
24	0.99527	6	0.91760
11	0.99480	7	0.91665
23	0.96508	31	0.91641
14	0.96165	2	0.91640
44	0.96109	33	0.91545
12	0.96103	3	0.91545
16	0.96057	17	0.88737
9	0.96027	19	0.87689
21	0.96027	5	0.87463
37	0.95954	30	0.87340
42	0.95930	1	0.87340
36	0.95838	29	0.82995
43	0.95838		
34	0.95721		
40	0.95721		
4	0.95697		

Monovariate regression:

$$y= 6.14-0.66\times BIC0 \quad (3)$$

R= 0.6626; s= 0.48; F= 28.18; n= 38

Bivariate regression:

$$y= 6.07+0.510\times Hy-0.53\times IC0 \quad (4)$$

R= 0.8248; s= 0.37; F= 37.23; n= 38

To increase even more the correlation coefficient, we have to remove the structures with similarity coefficients bellow 95 percent; thus, only 24 structures remained; on these we performed mono and multi- variate regressions, as above. The best regression equations are:

Monovariate regression:

$$y= 13.61-0.83\times EEig04d \quad (5)$$
$$R= 0.8264;\ s= 0.36;\ F= 47.38;\ n= 24$$

Bivariate regression:

$$y= 11.49-0.67\times EEig04d+0.281\times Hy \quad (6)$$
$$R= 0.8587;\ s= 0.34;\ F= 29.48;\ n= 24$$

Trivariate regression:

$$y= 10.29-1.57\times Hy-0.71\times IC0-1.2\times nNH2 \quad (7)$$
$$R= 0.8933;\ s= 0.33;\ F= 26.33;\ n= 24$$
$$y= 12.32-1.79\times EEig04d+2.19\times Hy-3.03\times IC0-1.18\times nNH2 \quad (8)$$
$$R= 0.9084;\ s= 0.29;\ F= 22.42;\ n= 24$$

The plot for the calculated, by the best regression, [equation (8)] vs. observed biological activity is presented in Figure 2.

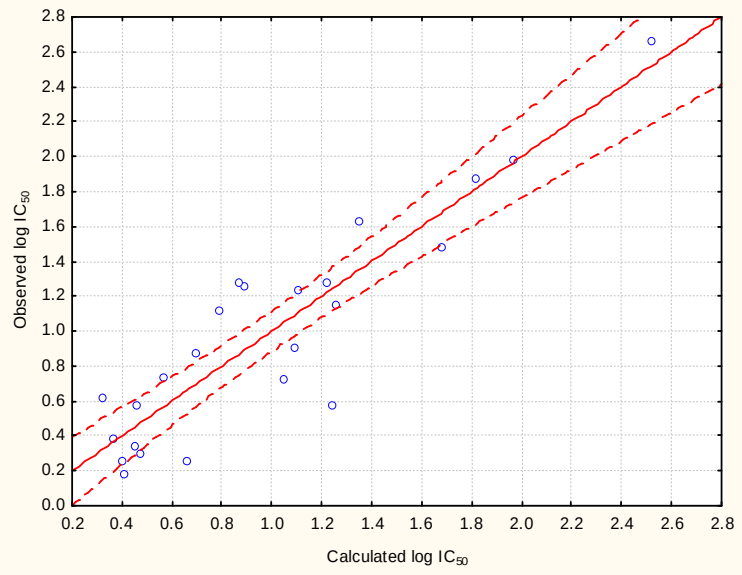
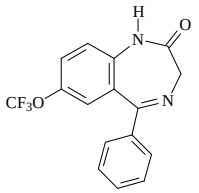
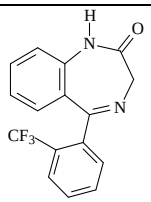
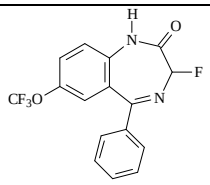
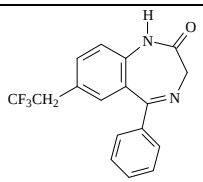
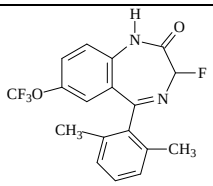
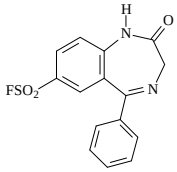
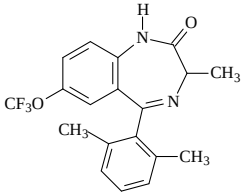
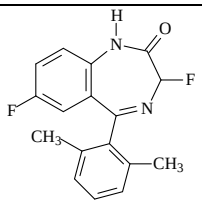


Figure 2. Calculated *vs.* observed log IC_{50} for equation 7

Design of new benzodiazepine derivatives

Information on experimental data shows the structures which have the functional groups NO_2 and Cl in position R_7, as the highest biologically active. For this reason, we tried to build up molecules with the same basic structure of which activities have not been yet measured. These molecules are shown in Table 5.

Table 5. Design of our proposed new structures

No.	Molecule	No.	Molecule
S1.		S5.	
S2.		S6.	
S3.		S7.	
S4.		S8.	

The next step was the calculation of the biological activity for the newly proposed structures by equation (8). The data are presented in Table 7.

Table 7. Calculated indices and evaluated biological activity for the proposed structures (by equation (8)).

Structures	EEig04d	Hy	IC0	nNH_2	Biological activity (BA)
S1	2.73	-0.69	1.83	0	0.38
S2	2.79	-0.66	1.88	0	0.18
S3	2.87	-0.68	1.81	0	0.21
S4	2.86	-0.72	1.73	0	0.38
S5	2.62	-0.72	1.75	0	0.75
S6	2.70	-0.73	1.72	0	0.68
S7	2.58	-0.67	1.89	0	0.50
S8	2.61	-0.76	1.65	0	0.99

From Table 7, one can see the structures S2, S3, S1, S4 show high biological activity, as those in the initial set of molecules; among these, the structure S2 shows the highest value. For this reason, it could be of interest to synthesize these molecules and make biological activity tests.

Notice that the authors [5] reported a biological activity of 0.176 for the most potent structure, which was taken as the reference structure and denoted RS. In comparing the activity of the above structure with our proposed molecules S2 and S3, we calculated using Hyper Chem software package, the value of log P: RS (-3.34); S2 (2.40) and S3 (0.57). It is known that a positive value of log P indicates a hydrophobic character, which is important in the transport of drugs through the biological membranes up to the receptor. Resuming the above data, our structure S2 appears as very promising in the drug design.

CONCLUSIONS

We studied a set of molecular structures belonging to the class of benzodiazepines for which we obtained a good model for the prediction of $\log IC_{50}$ biological activity. The best equation (eq. 8) obtained suggests that the biological activity is highly dependent on structure symmetry and hydrophobic characteristics. Based on the proposed model and the original data, we projected eight new structures for which the activity was calculated.

The procedures for model building proved to be reliable in estimating the biological activity and moreover, to predict the values of activity for so far non-tested structures.

ACKNOWLEDGMENTS

The authors wish to thank for the financial support provided from programs co-financed by The SECTORAL OPERATIONAL PROGRAMME HUMAN RESOURCES DEVELOPMENT, Contract POSDRU 6/1.5/S/3 – „Doctoral studies: through science towards society".

REFERENCES

1. M. V. Diudea, (Ed.), “QSPR/QSAR Studies by Molecular Descriptors”, Nova, New York, **2001**, 438.
2. G. Katona, G. Turcu, A. A. Kiss, O. M. Minailiuc, and M. V. Diudea, “QSAR/QSPR Studies by Cluj and Szeged Descriptors”, *Rev. Roumaine Chim.,* **2001**, *46*, 137.
3. M. Randić, “Design of Molecules with Desired Properties. A Molecular Similarity Approach to Property Optimization. In: Johnson, M. A.; Maggiora, G. M. Eds. “Concepts and Applications of Molecular Similarity”, Wiley, New York, **1990**, Chap. 5, 77-145.
4. R. Todeschini et al., Dragon software, http://www.talete.mi.it.
5. Sung-Sau So and Martin Karplus, “Genetic Neural Networks for Quantitative Structure-Activity Relationships: Improvements and Application of Benzodiazepine Affinity for Benzodiazepine/ GABAA Receptors”, *J. Med. Chem.,* **1996**, *39*, 26.
6. S.H. Bertz and W.C. Herndon, “The Similarity of Graphs and Molecules”. In: Artificial Intelligence Applications in Chemistry, *ACS Symposium Series* 306, Eds.: T. H. Pierce and B. A. Hohne. American Chemical Society, Washington, **1986**, 169–175.
7. O. Ursu, G. Katona, and M. V. Diudea, “Activity prediction by CLUJ-Simil program”, *Rev. Roum. Chim*., **2003**, *48(4*), 321.