Acta Universitatis Cibiniensis
Seria F Chemia 8(2005-2):33-45

Modeling of Physico-chemical Properties by Topological Indices

G. Katona[1*] and Teodora Panea[2]

Abstract

A study of the quantitative structure-property and structure-activity relationship QSPR/QSAR is presented for a set of aryloxiacetic acid derivatives. Some QSAR properties were also calculated and correlated with topological descriptors. As molecular descriptors the Cluj and Cluj type, Szeged and some classical indices were used. Cluj and Szeged properties indices considering the electronegativity, mass and 3D-geometry were also used. Some models for predicting LogP, hydration energy and Ka of the above set of acids are proposed.

Keywords: modelling, quantitative structure-property and structure-activity.

I. Introduction

A bioactive compound, when entered into an organism, generates a response of that organism that is a function of the structure and chemical identity of that compound. The interaction between the bioactive and the organism occurs, at the molecular level, in the so-called biological receptor (*i.e.*, the active site of a protein).[1-4] The early "key in lock" interaction of rather rigid complementary partners shifted to a more flexible mutual accommodation, within the effector-receptor complex, [*E...R*]. This complex will generate the biological response.

---

[1]*Faculty of Chemistry an Chemical Engineering, Babeş-Bolyai University, Cluj-Napoca, România*
[2]*Research and Production Centre "BIOS" Cluj-Napoca, România*
*To whom the correspondence should be addressed.

In a quantitative model, HANSCH[5-8] looks at the biological response as a function of pasive transport, *T*, of the bioactive compound to the receptor, and of electronic, *El*, steric, *S* and hydrophobic, *H*, interactions of the effector with the receptor

$$-\log([E]_0) = T + f(El, S, H) \quad (1)$$

The passive transport of the effector (through the lipidic membranes of cells till the receptor site) can be described by a parabolic function of log *P*

$$T = a(\log P) + b(\log P)^2 + c \quad (2)$$

where *P* is the *partition coefficient* of the bioactive in the system *octanol/water*. This parameter also accounts for the hydrophobic interactions between effector and receptor. Values of log *P* may be measured experimentally or calculated by additive models.[1, 5-11]

The affinity of the receptor vs. the effector (see the term *f*(*El*, *S*, *H*) in eq 1) can be described e.g., by a HAMMETT formalism.[1]

Cluj and Szeged type Indices

The graph-theoretical descriptors[12-17] *CJ*, *CF* and *SZ* represent the theoretical ground for counting the fragmental property indices. They are derived from the cardinality of the vertex sets defined by:

$$\boldsymbol{CJ_{i,j,p} = \{v \mid v \in V(G);\ di(G)_{v,i} < di(G)_{v,j};\ \text{and}\ \exists\, w \in W_{v,i},\ V(w) \cap V(p) = \{i\}\}} \quad (1)$$

$$\boldsymbol{CF_{i,j,p} = \{v \mid v \in V(G);\ di(G_p)_{v,i} < di(G_p)_{v,j};\ G_p = G - p} \quad (3)$$

$$\boldsymbol{SZ_{i,j} = \{v \mid v \in V(G);\ di(G)_{v,i} < di(G)_{v,j}} \quad (4)$$

In the above relations, $\boldsymbol{G_p = G - p}$ is the spanning subgraph, resulted by deleting the path $\boldsymbol{p}$ joining the vertices $\boldsymbol{i}$ and $\boldsymbol{j}$ (except its endpoints). $\boldsymbol{di(G)}$ and $\boldsymbol{di(G_p)}$ denote the topological distances measured in $\boldsymbol{G}$ and $\boldsymbol{G_p}$, respectively.

The sets $CJ_{i,j,p}$ and $CF_{i,j,p}$ represent subgraphs (connected or not) in *G,* referred to the endpoint *i* and related to *j* and path *p*.

In defining *Cluj indices*, the *path p* plays the central role in selecting the subgraphs (eqs 2 and 3), particularly in cycle-containing graphs, where more than one path could join the pair (*i,j*). In such graphs, more than one subgraph (i.e. fragment), referred to *i*, can be counted. For this reason, the nondiagonal entries $[\mathbf{UM}]_{ij}$ in Cluj matrices are defined as the *maximum cardinality* of the sets supplied by eq 2 or 3

$$[\mathbf{UM}]_{ij} = \max_{p} \left| V_{i,j,p} \right| \quad (5)$$

where $\boldsymbol{V_{i,j,p}}$ is either $\boldsymbol{CJ_{i,j,p}}$ or $\boldsymbol{CF_{i,j,p}}$ and consists of vertices, *v*, lying *closer* to the vertex *i* than to the vertex *j*. When $\boldsymbol{p \in Di(G)}$, (i.e. the set of all topological distances, or geodesics in $\boldsymbol{G}$) **M = CJDi** (Cluj-Distance) or **CFDi** (Cluj-Fragmental-Distance)**.** **W**hen $\boldsymbol{p \in De(G)}$, (i.e. the set of all topological detours, or the longest distances in $\boldsymbol{G}$) **M = CJDe** (Cluj-Detour) or **CFDe** (Cluj-Fragmental-Detour). The diagonal entries are zero. The Cluj matrices are square arrays, of dimension NxN**,** usually *unsymmetric* (excepting some symmetric regular graphs).

Szeged property matrices are defined[18], by analogy to $\mathbf{SZ}_u$,

$$[\mathbf{SZ_uP}]_{ij} = P_{i,(i,j)} \quad (6)$$

$$P_{i,(i,j)} = f(P_v) \mid v \in V(\Gamma);\ D_{iv} < D_{jv} \mid \quad (7)$$

$$f(P_v) = m \sum_v P_v \quad (8)$$

$$f(P_v) = (\Pi_v P_v)^{1/N} \quad (9)$$

The summation and product in eqs 8 and 9 run over all vertices in the graph.

The entries in a Szeged property matrix (see eq 6), are defined by a function $f(P_v)$, evaluated on vertices v, which obey the Szeged index condition. Consequently, $P_{i,(i,j)}$ can be viewed as a fragmental property. $P_{i,(i,j)}$ is mainly a topological (local) property (e.g., a topological index) but other physico-chemical properties are also considered (e.g., atomic mass or group electronegativities – see below).

The multiplicative function (eq 9) was used for group electronegativities: $P_v = X_v$; ($\mathbf{SZ_uX}$ matrix). X_v local electronegativity considered is the EC group electronegativity[13] (a Sanderson-type group electronegativity) for heteroatoms and fragments. It is defined by

$$\mathbf{EC_v = ES_v / (mlc_v EC_C)} \quad (10)$$

where ES_v is the Sanderson group electronegativity of vertex v, mlc_v is the mean length of covalent bond (relative to the tetraconnected carbon atom) around the vertex v and EC_C = 2.746/1.4996. For the tetraconnected carbon atom, $EC_v = 1$. Appendix contains EC_v values for some frequently used groups.

Topological indices on these matrices are calculated by the general relation

$$TI_{e/p} = \sum_{e/p} [\mathbf{SZ_uP}]_{ij} [\mathbf{SZ_uP}]_{ji} \ ;\ TI = SZ;\ SZP;\ SZNP;\ SZA;\ SZX \quad (11)$$

When summation goes over all edges in Γ, e = all $(i,j) \in E(\Gamma)$, the TI_e is an edge-defined index (e.g., SZ_e index). When it goes over all paths in Γ, p = all $(i,j) \in P(\Gamma)$, the TI_p is a path-defined index (or a hyper-index, e.g., SZ_pA index).

Wiener Indices

Wiener[19] was the first who used a topological descriptor in structure-properties relationship (see the recent reviews[20-21]). The index was called Wiener, i.e. W and, is defined, by

$$\mathbf{W = W(\Gamma) = \sum_{all\,(i,j) \in E(\Gamma)} N_{i,(i,j)} N_{j,(i,j)}} \quad (12)$$

where $N_{i,(i,j)}$ and $N_{j,(i,j)}$ denote the number of vertices lying on the two sides of the edge $(i,j) \in E(\Gamma)$, $E(\Gamma)$ being the set of edges in a connected graph, Γ. The summation runs over all edges in Γ. The product $N_{i,(i,j)} N_{j,(i,j)}$ represents the contribution of edge (i,j) to the global index, W. It is just the (i,j)-entry in the edge-defined Wiener matrix

$$[\mathbf{W}_e]_{ij} = N_{i,(i,j)} N_{j,(i,j)} \ ;\ (i,j) \in E(\Gamma) \quad (13)$$

from which W can be calculated as half sum of its entries

$$W = (1/2) \sum_i \sum_j [\mathbf{W}_e]_{ij} \quad (14)$$

Note that for non-adjacent vertices i and j, $(i,j) \notin E(\Gamma)$, the non-diagonal entries in $\mathbf{W}_e$ are zero. When (i,j) represents a path, $(i,j) \in P(\Gamma)$, with $P(\Gamma)$ being the set of paths in graph, then a relation similar to (1) will define the *hyper-Wiener* index, WW

$$\mathbf{WW = WW(\Gamma) = \sum_{all\ (i,j)\in P(\Gamma)} N_{i,(i,j)} N_{j,(i,j)}} \qquad (15)$$

The summation runs over all paths in Γ. $N_{i,(i,j)}$ and $N_{j,(i,j)}$ mean now the number of vertices lying on the two sides of the path $(i,j)\in P(\Gamma)$ and the product $N_{i,(i,j)}N_{j,(i,j)}$ is the contribution of the path (i,j) to the global index, WW.

In cycle-containing graphs, Wiener matrices are not defined. Wiener indices are here calculated by means of the distance-type matrices.

III. Results and discussions

Aryloxyacetic acids have a large applicability in agriculture, such as selective systemic herbicide and plant growth regulators[22]. This is the reason which determines obtaining through syntheses a large number of different aryloxyacetic acids, some of which have interesting biological activity[23]. Also, these compounds have been characterised[24-25]. From Table 1 we can see the Ka values increasing from phenoxyacetic acids unsubstituted to substituted acids with halogens, respectively with nitro groups. When another halogen group is introduced, the Ka increases more . It is interesting to see that the Ka of metilphenoxyacetic acids is more or less that of phenoxyacetic acids. The ortho and meta isomers have the less acid character. Also, we can see that the flor and chlor have bigger influences than iod, when the phenil group is substituted.

We consider here a set of 24 aryloxiacetic acids derivatives. The studied properties are LogP, hydratation energy (E_H) and the constant of acidity, K_a (see Table 1).

The molecular descriptors used are: Cluj type, Wiener type and Szeged type indices, extended by electronegativity, mass, and 3D-distance. The 3D-structure were performed by the HyperChem package program, release 6, with the Polak-Ribiere gradient. A version of the SZEGED program was used for extracting the 3d-distance matrix (actually the **G** matrix) from the HyperChem output file. Table 2 and 3 show the Cluj and Szeged extended indices, used in regression.

Table 1. Aryloxy acids derivatives and K_a, $R—O—CH_2—COOH$

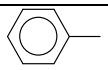
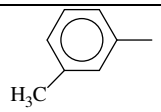
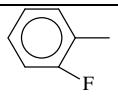
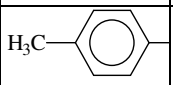
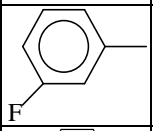
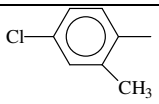
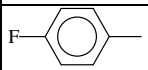
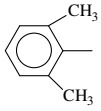
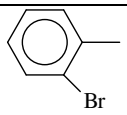
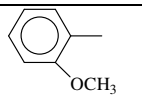
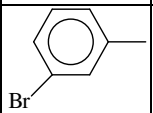
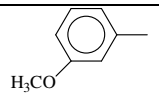
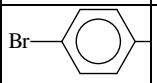
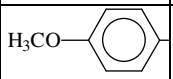
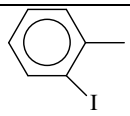
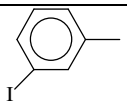
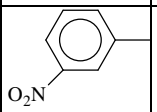
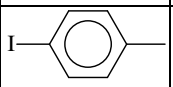
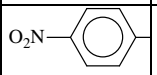
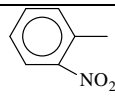
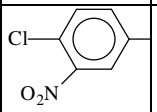
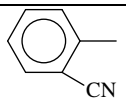
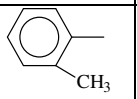
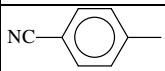
Nr.	R	LogP	E_H (kcal/mol)	$Kx10^4$	Nr.	R	LogP	E_H (kcal/mol)	$Kx10^4$
1		1.34	-9.12	6.75	13		1.80	-7.92	6.27
2		1.48	-8.85	8.72	14		1.80	-7.88	6.09
3		1.48	-8.81	8.28	15		2.32	-7.83	5.20
4		1.48	-8.81	7.42	16		2.27	-6.42	4.41
5		2.13	-8.79	7.53	17		1.08	-9.43	5.88
6		2.13	-8.74	8.03	18		1.08	-10.68	7.22
7		2.13	-8.74	7.37	19		1.08	-10.80	6.13
8		2.59	-8.79	6.72	20		2.40	-8.37	7.44
9		1.29	-13.93	11.2	21		2.59	-8.74	6.94
10		1.29	-14.19	12.8	22		1.29	-11.85	12.7
11		1.81	-13.07	11.0	23		1.37	-11.89	10.6
12		1.80	-8.15	5.93	24		1.37	-13.28	11.7

Table 2. Cluj and Szeged type indices.

No	D_3D	CJ_De	CJ_Di	SZp*3D	SZe*3D	SZp*1/3D	SZe*1/3D
1	191.3735	218.5	159.5	20756.4915	459.1730	207.8223	132.8464
2	231.7914	267.5	204	27857.3948	556.4819	265.3179	161.8660
3	242.1715	278.5	201.5	32337.8646	574.4213	268.4377	167.4286
4	244.6661	282	197	33843.1165	592.5268	277.6339	172.9501
5	236.2340	267.5	204	28476.3137	577.1529	258.5203	158.5387
6	248.0176	278.5	201.5	33668.2692	594.6555	262.7071	164.1989
7	250.8122	282	197	35482.5227	612.9049	271.5949	169.6582
8	237.7613	267.5	204	28663.1995	584.5660	257.2429	158.0663
9	375.8533	434.5	312.5	69736.5847	858.5915	425.6816	250.1257
10	375.8533	440	299	79829.9565	912.6308	452.1726	266.7514
11	435.1483	510	364	95495.3668	1037.875	513.9983	293.9400
12	233.2039	267.5	204	28089.0059	563.0495	262.1048	160.2053
13	244.0741	278.5	201.5	32811.9738	580.6466	265.8159	165.8460
14	246.6693	282	197	34444.7867	598.7101	274.8729	171.3935
15	275.8585	320	250	45079.7769	722.4784	335.7959	200.6805
16	273.8051	320	250	36507.2199	679.0968	325.0228	190.5377
17	285.3854	336	263.5	36998.6106	681.5055	333.6449	195.7435
18	304.8546	355.5	256	47699.7066	716.9080	340.4105	207.0194
19	306.3330	360	247	54725.6726	752.8924	357.8085	218.1314
20	252.8395	282	197	34115.8233	602.3194	261.5628	163.6621
21	252.8395	282	197	36034.302	620.2881	270.5417	169.2371
22	332.5447	406.5	325	51036.0847	805.6341	415.2400	233.2042
23	308.3649	355.5	256	38013.9764	669.8614	335.4146	199.8348
24	314.0283	360	247	54949.3601	741.9720	361.6754	222.0293

Table 3. Szeged index with electronegativity.

No	W	WW	SZpE	SZeE	SZpE*D	SZpE*3D	SZeE*3D	SZpE*1/3D	SZeE*1/3D
1	174	446	71.157	14.291	919.667	1073.340	26.132	13.422	7.914
2	213	548	112.302	21.729	1407.673	1586.026	39.417	21.285	12.094
3	218	577	116.516	22.378	1535.458	1828.231	40.572	21.668	12.458
4	223	611	118.713	23.027	1648.220	1919.231	41.744	22.247	12.817
5	213	548	106.517	20.337	1355.442	1589.711	43.686	18.256	10.24
6	218	577	109.997	20.872	1473.220	1853.369	44.560	18.584	10.548
7	223	611	111.790	21.407	1579.823	1955.468	45.514	19.030	10.842
8	213	548	102.967	19.503	1321.909	1565.427	43.397	17.406	9.781
9	332	954	162.504	24.706	2747.278	3199.513	44.636	25.601	13.924
10	347	1066	125.877	19.579	2435.614	2742.334	35.357	19.909	11.029
11	392	1145	149.163	21.598	2675.168	3152.547	41.005	22.137	11.776
12	213	548	83.506	15.243	1114.869	1264.983	28.406	14.729	8.306
13	218	577	83.786	15.275	1186.799	1430.039	28.439	14.540	8.328
14	223	611	83.928	15.307	1264.617	1495.190	28.498	14.545	8.345
15	266	723	104.590	17.620	1566.627	1849.799	34.531	16.922	9.281
16	256	660	96.799	16.193	1326.002	1466.167	30.684	16.105	8.693
17	264	701	102.995	17.127	1503.976	1557.579	31.604	17.295	9.385
18	274	764	104.017	17.281	1669.519	1926.923	31.858	16.917	9.477
19	284	837	104.471	17.434	1846.118	2135.831	32.126	16.973	9.565
20	218	577	105.979	19.966	1433.242	1831.194	44.197	17.656	10.039
21	223	611	107.522	20.428	1535.864	1934.460	45.024	18.042	10.298
22	317	857	123.911	19.157	1898.051	1934.043	34.645	20.350	10.784
23	264	701	103.453	17.431	1502.211	1596.594	31.068	17.778	9.951
24	284	837	104.892	17.730	1843.173	2143.864	31.603	17.552	10.120

First, we looked for the best correlation in mono variable regression. The best set of descriptors was tested for prediction by the cross validation procedure (leave one out method).

Table 4. Monovariable regression.

Variable	R	s	F
LogP			
CJ_Di	0.36321	0.466	27.040
E_H			
CJ_De	0.81813	1.260	44.533
D_3D	0.82478	1.239	46.805
Ka			
W	0.650173	1.852	16.110
WW	0.662997	1.825	17.255
SZpE_1/D	0.706543	1.725	21.929
SZp_D	0.673430	1.802	18.256

From Table 4 we can see a good correlation only in the case of hydration energy. This means that the properties logP and Ka are described more by interaction. Figure 1 shows the graphical representation for hydration energy.

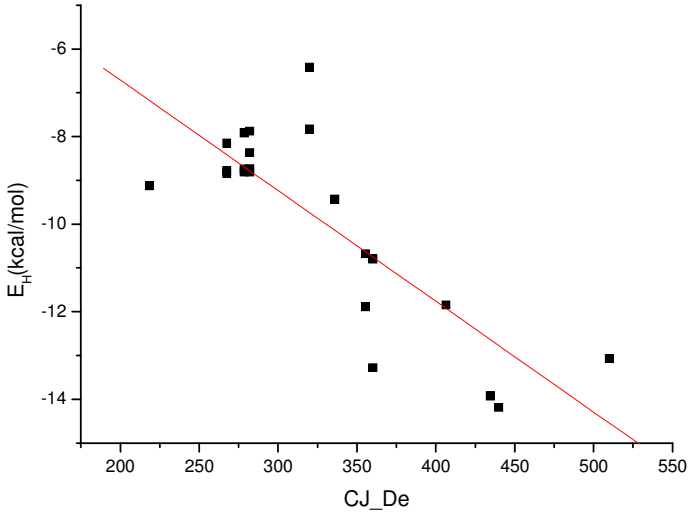


Fig. 1. Plot E_H vs CJ_Di.

Table 5 presents bivariable regression for the considered properties.

Table 5. Bivariable regression.

Variable	R	s	F
LogP			
SZpE_3D SZpE_D	0.65600	0.38680	7.93
SZe_1/3D SZe_3D	0.87411	0.24892	34.01
Ka			
SZe SZe_1/3D	0.781023	1.558532	16.42287
SZpE_1/D SZpE_3D	0.742342	1.672131	12.88902
WW SZeE_1/3D	0.741077	1.675323	12.79165
E_H			
SZe_1/3D SZe_3D	0.87212	1.09750	33.36
SZpE SZpED	0.86015	1.14400	29.86

In bivariable regression the best correlation is produced using the Szeged indices extended with 3D and electronegativity. Graphical representation of calculated vs measured hydration energy is presented in figure 2.

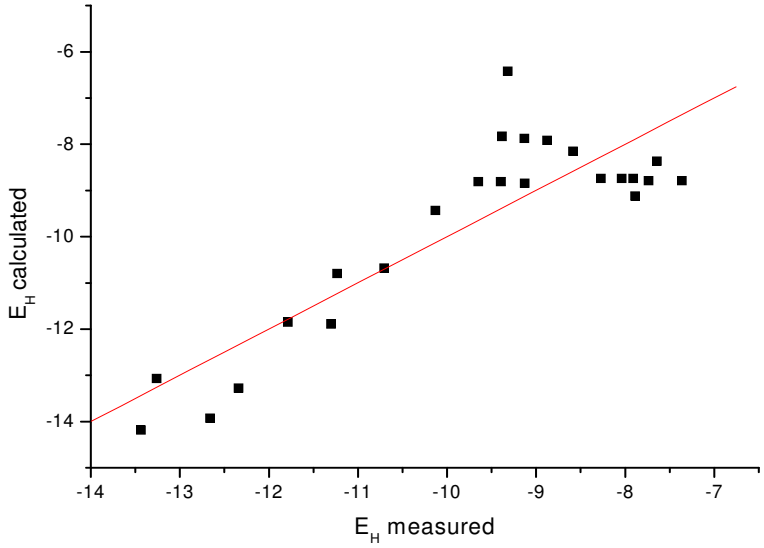


Fig. 2. Calculated vs. measured hydration energy.

When more than two descriptors are used, the correlation is improved. The best results for tri, four and five variable are represented in table 6.

Table 6. Multivariable regression.

Variable	**R**	**s**	**F**
(1)	**(2)**	**(3)**	**(4)**
LogP			
CJ_Di SZp_E SZpE_1/3d SZpE_D W	0.90265	0.23823	15.84
CJ_De SZe_3D SZp_1/3D W WW	0.94229	0.18533	28.52
CJ_De SZe_3D	0.95147	0.17528	27.08

SZeE SZeE_1/3D SZeE_3D WW			
(1)	**(2)**	**(3)**	**(4)**
E_H			
CJ_De SZe_3D SZp_3D	0.90236	0.99055	29.22
Cj_Di SZe_3D SZeE SZeE_3D W	0.96216	0.66017	44.88
SZe_1/3D SZe_3D SZeE SZeE_3D SZpE_D W	0.97386	0.56626	52.08
Ka			
SZe_3D SZe SZe_1/3D	0.905760	1.083750	30.45297
SZe_3D SZe SZe_1/3D SZp	0.935745	0.925323	33.43878
SZe_1/3D SZe_3D SZeE SZeE_3D SZp_1/3D SZp_3D	0.95580	0.81701	29.94

When the property is logP, the best correlation (0.95) is used for evaluation (based on regression equation)(see figure 3). If the property is Ka, the graphical plot is represented in figure 4.

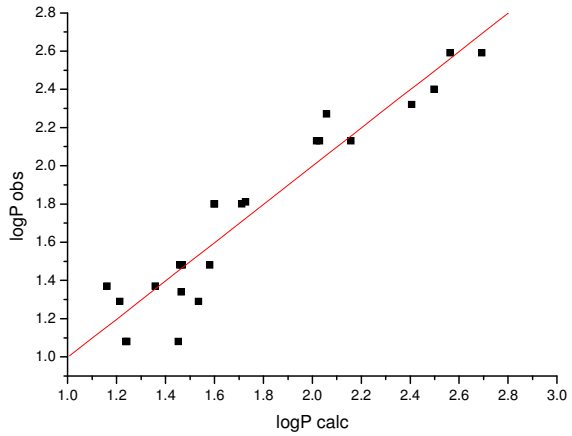


Fig. 3. Plot LogP obs. vs. LogP calc.

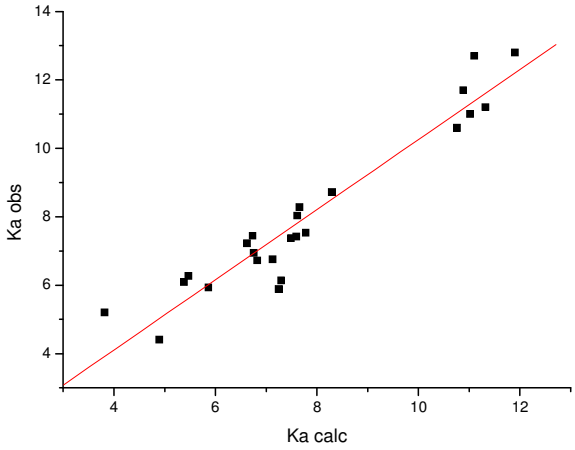


Fig. 4. Plot Ka obs. vs. LogP Ka calc.

IV. Conclusions

The biological response induced by chemical compounds can be properly modeled by the aid of topological indices. Some biological and physico-chemical properties of a set of aryloxiacetic acids derivatives were modeled by the theoretical and extended indices. Cluj and Szeged extended indices are very useful tools for modeling the logP, hydration energy and K_a for arylcarboxylic acid derivatives. Note that the correlation of Szeged index extended by mass was very low. The using of the 3D matrix in construction of the Szeged extended indices, results in an increased correlation for all properties considered.

This suggests that the spatial arrangement of the structures has some influence on properties. They demonstrate the usefulness of our descriptors in modeling biological and physical properties of organic compounds. Among these descriptors, the Szeged fragmental indices appear to be adequate for structures containing multiple bonds and hetero-atoms.

Acknowledgement

We thanks to Prof.Dr. M.V. Diudea for helpful discussion.

V. References

1. Moţoc, I., Structura moleculelor şi activitatea biologică, Ed. Facla, Timişoara, **1980**.
2. Ariëns, E.J., *Med. Res. Rev*., **1986**, *6*, 451.
3.. Ariëns, E.J, *Med. Res. Rev*., **1987**, *7*, 367.
4. Ariëns, E.J., *Med. Res. Rev*., **1988**, *8*, 309.
5. Hansch C., and Fujita, T., *J. Amer. Chem. Soc.,* **1964**, *86*, 1616.
6. Hansch, C., *Acc. Chem. Res.,* **1968**, *8*, 232.
7. Hansch, C., Drug design, Ed. E. J. Ariens, Academic Press, New York, **1971**, *16*, 271.
8. Kubinyi, H., *Struct.-Act. Relat.,* **1988**, ***7***, 121-133.
9. Rekker, R.F., The Hydrophobic Fragmental Constant. Its derivation and Applications. A Means of Characterizing Membrane System, Elsevier, Amsterdam, **1977**.
10. Rekker, R.F., Mannhold, R., Calculation of Drug Lipophilicity. The Hydrophobic Fragmental Constant Approach, VCH Weinheim, **1992**.
11. Kubinyi, H., *J. Med. Chem.,* **1977**, *20*, 1991.
12. Diudea, M. V., *J.Chem.Inf.Comput.Sci.* **1997**, *37*, 300-305.
13. Diudea, M. V., *Commun. Math. Comput. Chem. (MATCH),* **1997**, *35*, 169-183.
14. Diudea, M., Minailiuc V. O., Katona G. and Gutman I. *Commun. Math. Comput. Chem. (MATCH)* **1997**, *35*, 129-143.
15. Diudea, M. V., Pârv, B. and Topan, M. I. *J. Serb.Chem.Soc*., **1997**, *62*, 267-276.
16. Minailiuc O., Katona G., Diudea, M. V., Strunje, M., Graovac, A. and Gutman, I. *Croat. Chem. Acta.,* **1998**, *71*, 473-488.
17. Jäntschi, L., Katona, G. and Diudea, M. V. *Commun. Math. Comput. Chem. (MATCH),* **2000**, *41*, 151-188.
18. Szeged fragmental indices were first put forward on a lecture held by M.V.D. at the Rudger Boškoviæ Institute, Zagreb, Croatia, on July 7, 1996.
19. H. Wiener, *J.Am.Chem.Soc.* **1947**, *69*, 17-20.
20. Gutman, I., Yeh, Y.N., Lee, S.L., Luo, Y.L., *Indian J.Chem*. **1993**, *32A* 651-661.
21. Nikolić, S., Trinajstić, N., Mihalić, Z., *Croat.Chem.Acta.* **1995**, *68* 105-129.
22. Bovey, R.W., The Science of 2, 4, 5-T, and Associated Phenoxy Herbicides, Ed. John Wiley & Sons; **1980**.
23. Krause, H.P., Kocur, J., de Una, J.M., Bickers, U., Hacker, E., Schnabel, G., *US 20020137634*.

24. Lynch, D. E., Cooper, C. J., Chauhan, V., Smith, G., Healy, P., Parsons, S., *Aust. J. Chem.*, **1999**, *52*, 695-703.
25. Cserhati, T., Forgacs, E., Deyl, Z., Miksik, I., *Journal of Chromatography B*, **2001**, *753(1)*, 87-92.