International Journal of ChemTech Research
CODEN(USA): IJCRGG ISSN : 0974-4290
Vol.4, No.2, pp 662-668, April-June 2012

Topological Investingation and Modeling of Antimalarial Activity of Chalcone

Sudhanshu Dhar Dwivedi[1], Arpan Bharadwaj[2], Anis Ghuraiya[2], Amit Shrivastava[2]*

[1]Govt. Sc. and Com. College Benazeer, Bhopal (M.P.),India.

[2]Govt. Madhav Science P.G. College, Ujjain (M.P.),India.

*Corres.author: amit_chem@yahoo.com
Ph. Number: 09691431437

Abstract: A set of chalcone derivatives were tested for their antimalarial activities[1]. Quantitative structure activity relationship (QSAR) analysis was applied to forty-two of the abovementioned derivatives using a combination of various topological descriptors. A multiple linear regression (MLR) procedure was used to model the relationships between molecular descriptors and the antimalarial activity of the chalcone derivatives. The stepwise regression method was used to derive the most significant models as a calibration model for predicting the antimalarial activity of this class of molecules. The best QSAR models were further validated by the calculation of statistical parameters for the established theoretical models. High agreement between experimental and predicted activity values, obtained in the validation procedure, indicated the good quality of the derived QSAR models.
Keywords: QSAR; chalcone derivatives; multiple linear regressions; statistical parameters.

INTRODUCTION

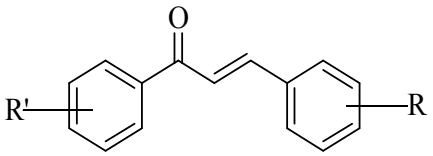
Quantitative structure–activity relationship and Quantitative structure–property relationship (QSAR/QSPR) studies are indubitably of great importance in modern chemistry and biochemistry. To obtain a significant correlation, it is essential that appropriate descriptors are employed, for such considerations the molecular structure is often represented as a simple mathematical object, such as a number, sequence, or a set of selected invariants of matrices, generally referred to as molecular descriptors[2-4]. Multiple regression analysis is usually used in such studies in the hope that it might point to structural factors that influence a particular property. It may help one in model building and assist in the design of molecules with prescribed desirable properties, which is an important goal in drug research. In chemistry, anything that can be said about the magnitude of the property and its dependence upon changes in the molecular structure[5] depends on the chemist's capability to establish valid relationships between structure and property. In many physical-chemistry, organic, biochemical and biological areas, it is increasingly necessary to translate those general relations into quantitative associations expressed in useful algebraic equations known as Quantitative Structure-Activity (-Property) Relationships (QSAR/QSPR)[6]. To get an insight into the structure-activity relationship we need molecular descriptors that can effectively characterize molecular size, molecular branching or

the variations in molecular shapes, and can influence the structure and its activities. Many descriptors reflect simple molecular properties and thus they can provide some meaningful insights into the physical chemistry nature of the activity/property under consideration. Chemical graph theory[7] advocates an alternative approach to QSAR/QSPR studies based on mathematically derived molecular descriptors. Such descriptors often referred to as topological indices[8]. Many descriptors reflect simple molecular properties and can thus provide insight into the physicochemical nature of the activity/ property under consideration. If molecular structure is critical for understanding of a particular structure-activity and property-activity relationship, then one should consider structural invariants derived from molecular structure[9]. Several graph theoretical invariants have been generalized so that they produce structure-dependent descriptors[10-13]. Ideally, the activities and properties are connected by some known mathematical function, F: Biological activity = F [structure (in present study topological & physicochemical descriptors are used as the structural parameters)] Biological activity can be any measure such as log1/C, Ki, IC50, ED50, EC50, log K and Km.

The relationship or function is more often than not a mathematical expression derived by statistical or related techniques. In present study the multiple linear regression (MLR) technique is used. The parameters describing structural properties are used as independent variables and the biological activities are dependent variables.

In the present investigation a QSAR study is performed over a set of 42 chalcone derivatives. Their Biological activity is measure as IC_{50} [a](μM). For simplification of mathematics calculation we take Log IC_{50} [a](μM). This study based on the application of topological parameters in QSAR.

Data set of chalcone derivative (table 1)



Ring B Ring A

Compound	R'	R	IC_{50} [a](μM)	Log IC_{50} [a](μM)
1	2',3',4'-trimethoxy	2,4- dichloro	5.4	0.73239
2	2',3',4'-trimethoxy	4-dimethylamino	18	1.25527
3	2',3',4'-trimethoxy	4-trifluoromethyl	3	0.47712
4	2',3',4'-trimethoxy	2,4-dimethoxy	16.5	1.21748
5	2',3',4'-trimethoxy	4-methyl	25.6	1.40823
6	2',3',4'-trimethoxy	4-ethyl	16.5	1.21748
7	2',3',4'-trimethoxy	3-quinolinyl	2	0.30103
8	2',3',4'-trimethoxy	4-methoxy	25	1.39794
9	2',3',4'-trimethoxy	4-fluoro	9.5	0.97772
10	2',3',4'-trimethoxy	4-phenyl	26.2	1.41830
11	2',3',4'-trimethoxy	4-nitro	22.5	1.35218
12	2',3',4'-trimethoxy	3,4-dichloro	14.5	1.16136
13	2',3',4'-trimethoxy	4-chloro	14.5	1.16136
14	2',3',4'-trimethoxy	2-chloro	41.5	1.61805
15	2',3',4'-trimethoxy	3-chloro	24.4	1.38738
16	2',3',4'-trimethoxy	H	15.8	1.19865

17	4'-butoxy	2,4-dimethoxy	108	2.03342
18	2',4'-dimethoxy	2,4-dichloro	18.8	1.27415
19	2',4'-dimethoxy	4-trifluromethyl	5.9	0.77085
20	2',4'-dimethoxy	2,4-difluoro	6.2	0.79239
21	2',4'-dimethoxy	2,4-dimethoxy	2.1	0.32222
22	2',4'-dimethoxy	4-dimethylamino	70	1.84509
23	2',4'-dimethoxy	4-cyano	94.5	1.97543
24	2',4'-dimethoxy	H	55.5	1.74429
25	4'-ethoxy	2'4-difluoro	28.1	1.44871
26	4'-ethoxy	4-methoxy	33	1.51851
27	4'-ethoxy	3-quinolinyl	24.9	1.39619
28	4'-ethoxy	4-fluoro	24.1	1.38202
29	4'-ethoxy	2,4-dichloro	96	1.98227
30	4'-ethoxy	4-trifluromethyl	24	1.38021
31	4'-ethoxy	2,4-dimethoxy	30	1.47712
32	4'-ethoxy	4-methyl	38	1.57978
33	4'-ethoxy	4-nitro	39	1.59106
34	4'-ethoxy	4-dimethylamino	30	1.47712
35	4'-ethoxy	H	43	1.63347
36	2',4'-dihydroxy	2,4-difluoro	16	1.20412
37	2',4'-dimethoxy	3-quinolinyl	2.2	0.34242
38	2',4'-dimethoxy	4-quinolinyl	27	1.43136
39	2',4'-dimethoxy	4-methoxy	128	2.10890
40	2',4'-dimethoxy	4-dimethylamino	55.3	1.74273
41	4'-methoxy	4-methoxy	21.7	1.33645
42	4'-methoxy	4-methyl	70	1.84509

MATERIAL AND METHOD

We studied a series of chalcones with the activity express as IC_{50} [a] (μM) was taken from the literature. These chacone derivatives with their activity are presented in table 1. Topology Indices: All the topological indices used are calculated from the hydrogen suppressed molecular graph though their calculations are exclusively discussed in the literature. Topology indices are used for convert structure property into numerical form. Calculated topological descriptors included wiener index[14-15] (W), mean distance degree deviation (MDDD), schultz molecular topological index[16] (SMTI), (Xu), kier flexibility index (PHI), eccentric connectivity index[17] (CSI), mean information contain on the distance degree magnitude (IDDM), unipolarity (UNIP), high per detour index (WW), log of product of row sum (LPRS), gutman molecular topological index (GMTI), solvation connectivity index chi-0 (χ^0Sol), polarity no. [P_3], (DECC), total information content on distant magnitude[18] (IDMT), mean information content on distance equality[18] (IDE), solvation connectivity index chi-2 (χ^2Sol), randic connectivity indices[19] fifth order (χ^5A), average randic connectivity indices[19] first order (χ^1A), mean information contain on the distance degree equality (IDDE), second mohar index (TI2), mean wiener index (WA), harary index (HAR1), first zegreb index (Zm1), randic connectivity indices chi- 5 (χ^5), randic connectivity indices[19] chi-3 (χ^3), total walk count (TWC), average randic connectivity indices[19] zero order (χ^0A),

Topological molecular descriptors are used in QSAR studies because of their accessibility, being easily computed by available software programs. The set of molecular descriptor which are used in the study are calculated by DRAGON software[20]. Stepwise multi regression analysis method was used to perform QSAR analysis. The stepwise multiple linear regressions (MLR) are a commonly

used variant of MLR. Each variable is added to the equation at a time and a new regression is performed. The new term is retained only if the equation passes a test for significance. This regression method is especially useful when the number of variables is large and when the key descriptors are not known. This is the basis of maximum-R^2 method for deriving most appropriate QSAR model. When the number of independent variables is greater than the number of molecules, multiple linear regressions cannot be applied then we applied stepwise multiple regression.

The best model was selected on the basis of various statistic parameter such as PSE (Prediction square error), K (Degree of lack of relationship), E (Index of forecasting efficiency), PE (Probable error of estimation), T test, Adjusted R^2, Q (Quality of proposed model), Spress (Uncertainty of prediction), PRESS (The Expression of PRESS), F test, SEE (Standard error of estimation) these are statistic parameter show the predictivity and significance of the model.

RESULT AND DISCUSSION

When the data was subjected to stepwise multiple linear regression analysis, in order to develop QSAR between antimalarial activity of various compound as dependents variables and topology indices as independent variables, several equation is obtained.

Various widely used topological indices tested in the present study. In the proposing QSAR model for the modeling the antimalarial activity of compound we used the maximum R^2 method. We used the cross validation parameter for investigating predictive power of various parameters and prove our finding. For the QSAR study of the same series we tested the multivariate combination of the parameter. The result obtain from the multivariate combination are encouraging and better model are show below with their statistics.

Model -1

Log $IC_{50}{}^{a}(\mu M)$= -14.764(±8.719)+0.186(±0.129)*Ramification+0.282 (±0.029)* χ^0Sol-0.216(±0.047)*Polarity+0.347 (±0.057)*MDDD-1.485(±0.270)*Xu-0.711(± 0.139) *PHI+0.259(±0.129)*S2K-1.719(±0.583)*IDDM-0.00006(0.00002)*IDMT +0.466(±0.129)*TWC -0.604(±0.260)*χ^5 -0.009(±0.003)*UNIP +0.013(±0.007)* VDA

…… (1)

R^2= 0.912, E=0.70.41, PE= 0.0089, T_{test}=20.414, Adjusted R^2=0.872, Q=5.947, S_{PRESS}=0.1335, PSE=0.1314, PRESS=0.726, F_{test}=1433.92, SEE=0.164, K=0.295, R=0.955

Model -2

Log $IC_{50}{}^{a}(\mu M)$= 10.555(±2.179) - 0.549(±0.103)*Ramification -0.288(±0.035)* χ^0Sol -0.116(±0.037)*XMOD +0.309(±0.060)*MDDD -0.103(±0.029)*LPRS-0.00006 (±0.00002)*W +0.0003(±0.00009)*GMTI -2.019(±0.604)*IDDM-0.008 (±0.003)* UNIP-0.661(± 0.152)*PHI +0.649(±0.189)*S2K +0.031(±0.014)*MR

…… (2)

R^2= 0.900, E=0.68.38, PE= 0.0101, T_{test}=19.209, Adjusted R^2=0.859, Q=5.627, S_{PRESS}=0.1401, PSE=0.138, PRESS=0.799, F_{test}=1157.98, SEE=0.1168, K=0.316, R=0.949

Model -3

Log $IC_{50}{}^{a}(\mu M)$= 70.1841(±9.684) +0.485(±0.074)*χ^3Sol -0.0004(±0.0002)*BP-0.384 (±0.066)*DECC+3.557(±0.425)*χ^1A-0.637 (±0.093)* PHI-13.079(±2.645)* AECC+ 1.219 (±0.329)*Qindex-0.00005 (±0.00002)*HVcpx-90.859(±14.243)*WW +0.255 (±0.192)*TPC +0.019(±0.014)*IDDE

…… (3)

R^2= 0.887, E=0.66.43, PE= 0.0114, T_{test}=17.744, Adjusted R^2=0.846, Q=5.352, S_{PRESS}=0.1452, PSE=0.143, PRESS=0.858, F_{test}=954.53, SEE=0.176, K=0.335, R=0.942

Model - 4

Log $IC_{50}{}^{a}(\mu M)$= 6.487(±14.847) - 0.577(±0.169)*Qindex +3.591(±0.405)*χ^2Sol +0.369(±0.068)*DECC-0.599(±0.094)* PHI+0.750 (±0.328)*HVcpx-13.213 (±2.733)*AECC +1.186 (±0.353)*χ^3-0.00003(±0.00001)*WW -104.906 (±30.623) *χ^1A -0.011(0.008)*BP +0.008(±0.006)*Tc

…… (4)

R^2= 0.884, E=0.66.02, PE= 0.0117, T_{test}=17.502, Adjusted R^2=0.842, Q=5.278, S_{PRESS}=0.1478, PSE=0.1454, PRESS=0.889, F_{test}=916.277, SEE=0.178, K=0.339, R=0.941

Model - 5

Log IC_{50}^{a}(μM)= 74.979(±15.002) -0.550(±0.173)* Qindex +3.554(±0.409)*χ^2Sol +0.375(±0.069)* DECC -0.588(±0.092)*PHI +0.583(±0.316)* HVcpx-11.821(±2.7 07)*AECC+1.014(±0.351)*χ^3-0.0003(±0.0001)*WW-106.101(±31.293)*χ^1A

……..…(5)

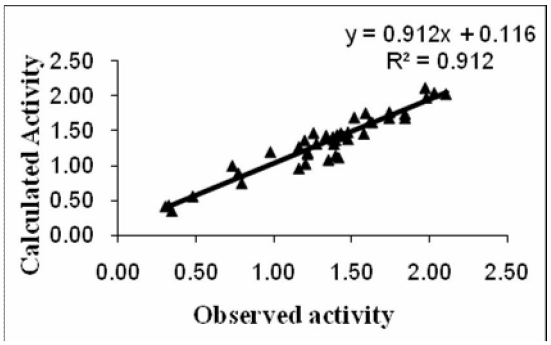
R^2= 0.869, E=0.63.93, PE= 0.0132, T_{test}=16.354, Adjusted R^2=0.833, Q=5.094, S_{PRESS}=0.1486, PSE=0.146, PRESS=0.898, F_{test}=753.453, SEE=0.183, K=0.361,R=0.932

In above equation, + sign indicate that activity is proportional in successive regression analysis. We have carried out several multi parametric regression analysis. In all such multi parametric regression analysis better result are obtained than the mono parametric model. The observed and calculated activities of these models are given in table 2 and their graph between observed and calculated activity value are recorded in Figure 1, 2 3, 4 and 5.

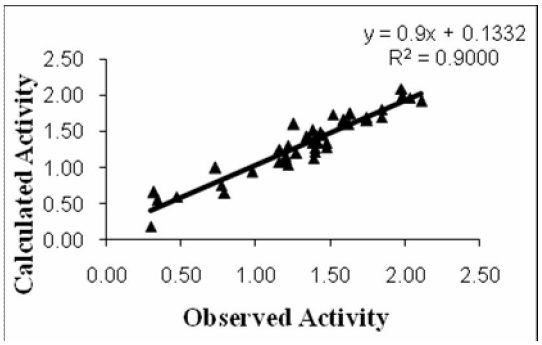
Table (2) for observed and calculated activity

Comp. No.	Observed Log IC_{50} ^{a}(μM)	Predicted Activity in Log IC_{50} a(μM) for Model 1	Predicted Activity in Log IC_{50} a(μM) for Model 2	Predicted Activity in Log IC_{50} a(μM) for Model 3	Predicted Activity in Log IC_{50} a(μM) for Model 4	Predicted Activity in Log IC_{50} a(μM) for Model 5
1	0.73239	0.9978	0.99666	0.99759	1.02913	1.06522
2	1.25527	1.46697	1.59981	1.46179	1.53754	1.57907
3	0.47712	0.55979	0.59466	0.49739	0.52806	0.55622
4	1.21748	1.16828	1.0356	1.09872	1.10637	1.13582
5	1.40824	1.44031	1.44451	1.33157	1.34074	1.34961
6	1.21748	1.18812	1.30485	1.30559	1.24011	1.23926
7	0.30103	0.41465	0.18564	0.15981	0.27513	0.28676
8	1.39794	1.14473	1.22248	1.23973	1.18454	1.17873
9	0.97772	1.19427	0.94349	1.13909	1.12504	1.11659
10	1.4183	1.11419	1.34749	1.40843	1.53304	1.49221
11	1.35218	1.0805	1.40638	1.22205	1.30795	1.04601
12	1.16137	0.9582	1.07924	1.13515	1.02692	1.06287
13	1.16137	1.26684	1.25873	0.99924	1.06121	1.07028
14	1.61805	1.62399	1.60081	1.47603	1.44598	1.4352
15	1.38739	1.31197	1.13015	1.1835	1.22453	1.27428
16	1.19866	1.35988	1.15905	1.42323	1.4415	1.41619
17	2.03342	2.04203	1.96679	2.01631	2.01501	2.05903
18	1.27416	1.3129	1.20152	1.39696	1.43922	1.4189
19	0.77085	0.88946	0.75411	0.72936	0.68469	0.70603
20	0.79239	0.74558	0.64936	0.60302	0.60259	0.6059
21	0.32222	0.42927	0.65848	0.76119	0.68906	0.73208
22	1.8451	1.68306	1.702	1.64118	1.68817	1.72887
23	1.97543	2.11367	2.0936	2.08807	1.99615	2.05936
24	1.74429	1.76422	1.65434	1.72445	1.6075	1.54734
25	1.44871	1.42459	1.45309	1.46474	1.42312	1.42134
26	1.51851	1.69044	1.73178	1.45479	1.41019	1.41464
27	1.3962	1.36605	1.26874	1.22644	1.13352	1.13009
28	1.38202	1.37433	1.51481	1.5011	1.54349	1.53733
29	1.98227	1.97304	1.99304	1.84272	1.84616	1.85796
30	1.38021	1.40929	1.33879	1.44284	1.49592	1.50748
31	1.47712	1.37919	1.28034	1.23243	1.2145	1.25365
32	1.57978	1.45985	1.61402	1.69358	1.7606	1.77036

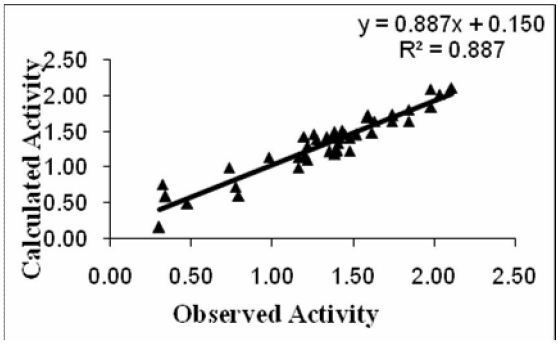
33	1.59106	1.75271	1.65423	1.72903	1.63292	1.57997
34	1.47712	1.47942	1.34417	1.40772	1.51066	1.57997
35	1.63347	1.61914	1.74522	1.63921	1.68309	1.63676
36	1.20412	1.0251	1.11016	1.15642	1.14387	1.19816
37	0.34242	0.3483	0.5522	0.59237	0.48192	0.50091
38	1.43136	1.4652	1.47564	1.51797	1.48964	1.43618
39	2.1089	2.02525	1.92239	2.10451	2.07182	2.05936
40	1.74273	1.68306	1.702	1.64118	1.68817	1.72887
41	1.33646	1.43191	1.42628	1.42908	1.44056	1.3879
42	1.8451	1.73988	1.80079	1.80186	1.81714	1.75466



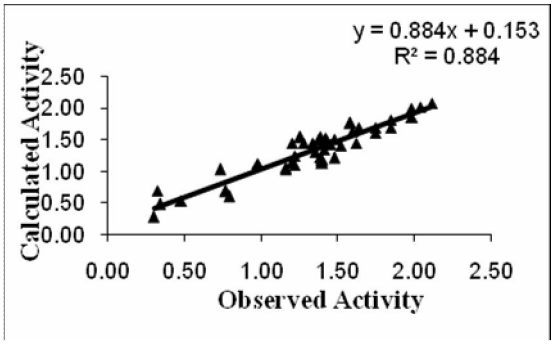
Graph between observed and calculated value of $LogIC_{50}$ [a](µM) Model-1 (Figure-1)



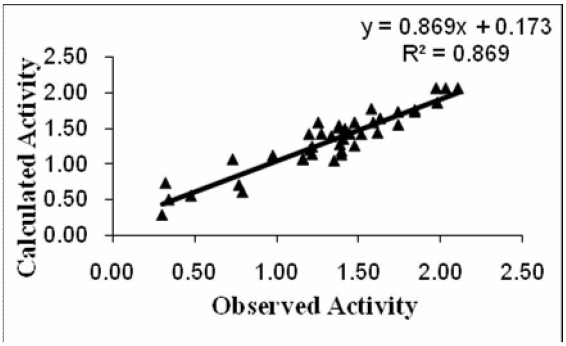
Graph between observed and calculated value of $LogIC_{50}$ [a](µM) Model-2 (Figure-2)



Graph between observed and calculated value of $LogIC_{50}$ [a](µM) Model-3 (Figure-3)



Graph between observed and calculated value of $LogIC_{50}$ [a](µM) Model-4 (Figure-4)



Graph between observed and calculated value of $LogIC_{50}$ [a](µM) Model-5 (Figure-5)

The model expressed by equation 1, this model has the highest R^2 value with good statistics. The best model is one which has the best statistics as well as best predicting power. Thus we have obtained predictive correlation coefficient (R^2) for the model express Equation (1-5) by correlating observed activity with calculated one. The R^2 obtained are presented in figure (1-5). This show the model expressed by Equation 1 is most appropriate model for modeling activity. The predictive power of the model can also be justify by calculating PSE, K, E, PE, T_{test}, F_{test}, PRESS, S_{PRESS}, SEE, Adjusted R^2.The above all statics for the model mentioned with the 1-5 models. The comparative analysis of statistic associated with model show that the model based on Equation 1 is most suitable for modeling of activity. These models are also help to design of new molecule of desired activity. These topological parameters are related to steric, and electronic attitude of molecule. We predict value of activity of unknown compound by the use of value of topological indices. The above discussion supported the utility of topological parameters which are used in research work.

ACKNOWLEDGEMENT

We are grateful to Department of Chemistry, Govt. Madhav Science P.G. College, Ujjain (M.P.) for the use of the research facilities.

REFFERENCES

1. *J. Med. Chem. 2001,44, 4443-4452.*
2. Randic´, M. *J. Chem. Inf. Comput. Sci.* 1997, 37, 672.
3. Balaban, A. T. et al; *Topics Curr. Chem.* 1983, 114, 21.
4. Balaban, A. T. Historical developments of topological indices. *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds; Gordon and Breach: Amsterdam, the Netherlands, 1999; p 403.
5. Balaban, A. T. *J. Mol. Struct. (THEOCHEM)* 1988, 165, 243.
6. Basak, S. C. Information theoretic indices of neighborhood complexity and their applications. In Topological Indices and Related Descriptors in QSAR and QSPR;Devillers, J., Balaban, A. T., Eds; Gordon and Breach: Amsterdam, the Netherlands,1999; p 563.
7. Trinajstic ´, N. In *Chemical Graph Theory*; CRC Press: Boca Raton, FL, 1992; p225.
8. Basak. S. C. et al. Use of graph-theoretic geometrical molecular descriptors in structure-activity relationships. In *From Chemical Topology to Threedimensional Geometry*; Plenum Press: New York, 1977; p 73.
9. Randic´, M.; Razinger, M. On characterization of 3D molecular structure. In *From Chemical Topology to Three-dimensional Geometry*; Plenum Press: New York, 1977;p 159.
10. Randic´, M. *Int. J. Quantum Chem:* Quantum Biol. Symp. 1988, *15*, 201.
11. Randic´, M. et al. *Computer Chem*. 1990, *14*, 237.
12. Randic´, M. *J. Chem. Inf. Comput. Sci.* 1995, *35*, 373.
13. Randic´, M. *New J. Chem*. 1995, *196*, 781.parameters in QSAR.
14. Wieney, H.J. Am. Chem. Soc., 69. (1947), 2636-2638.
15. Wieney, H.J. Phys. Chem., 15, (1947), 766.
16. Schultz, H.P., J. Chem. Inf. Compute. Sci., 29, (1989) 227-228.
17. Sharma, V. et al. Chem. Inf. Comput. Sci., 37 (1997), 273-282.
18. Banchev, D. Information Theoretic Indices fro characterization of chemical structureRSP-Wiley, Chichetsey (U.K.) (1983).
19. Randic´, M. *Stud. Phys. Theor. Chem*. 1988, *54*, 101.
20. *Dragon 5* software, http://www.disat.unimib.it /chm/Dragon.htm

*****