9th International Electronic Conference on Synthetic Organic Chemistry. ECSOC-9. 1-30 November 2005.

http://www.usc.es/congresos/ecsoc/9/ECSOC9.HTM & http://www.mdpi.net/ecsoc/

[G010]

QSTR study of small organic molecules against Tetrahymena pyriformis

F. A. Pasha[a], H. K. Srivastava[b], A. Srivastava[c] P. P. Singh[b*]

[a]Department of Chemistry, Bareilly College Bareilly, (U.P.), INDIA
Email: fpasha@rediffmail.com

[b]Department of Chemistry M. L. K. (P. G.) College Balrampur, (U.P.), INDIA
Email: hemantkrsri@rediffmail.com

[c] Department of Botany M. L. K. (P. G.) College Balrampur, (U.P.), INDIA
Email: anupma.srivastava@rediffmail.com

Keywords- QSAR, QSTR, Phenol, Tetrahymena pyriformis, Toxicity

ABSTRACT- The quantitative structure Toxicity relationships (QSTR) models of various organic compound like Phenols, alcohols, esters, aldehydes and ketenes have been made with the help of constitutional descriptors, geometrical descriptors, BCUT descriptors, empirical descriptors, topological descriptors, galvez topol. charge indices. In first step all the descriptors have been applied on 46 phenol derivatives. The Molecular modeling is carried out with CAChe pro software. The geometry optimization to correspond a minimum energy conformation. The Calculation of descriptors done by Dragon software. The topological descriptors provide better results than other and have much high reliability as clear indicated by its correlation coefficient and cross validation coefficient values. The final QSTR models for test set of compounds are made with the help of topological descriptors. The values obtained by these QSTR models are in close range with observed toxicity.

*Author for correspondence

Introduction: Computer simulation techniques have gained significance in bridging the gap between the experimental and theoretical evidences. Modeling macroscopic processes in the realistic environment is one of the most challenging problems in theoretical and computational chemistry. The synthesis of novel pharmacologically active molecules with reduced toxicity is of prime interest. Recently, QSAR has gained importance in the field of pharmacological sciences [1]. The QSAR methodologies save resources and expedite the process of development of new molecules and drugs. Quantitative structure–activity relationship (QSAR) techniques increase the probability of success and reduce time and cost involvement in drug discovery process [2-3]. Success of QSAR not only rests on the development of new drug molecules but also in exploring the toxicological and ecotoxicological characteristics of molecules.[4] Quantitative structure Toxicity relationships (QSTR) are predictive tools for a preliminary evaluation of the hazards of chemical compounds by using computer aided models

In present works we have taken various small organic compounds with their observed toxicity and comparative QSTR models have been made. Previously a number of chemical descriptors [5-10, 52] have been identified with in the framework of QSAR [11] techniques. Here few important groups of descriptors like constitutional descriptors, geometrical descriptors, BCUT descriptors, empirical descriptors, topological descriptors, galvez topol. charge indices have been tested and final QSAR model has been made with the help of most significant descriptors.

Theory: The following group of descriptors has been tested to describe the toxicity of the compound of training set and the values of these descriptors have been calculated on Dragon software.

1- Constitutional descriptors[12]

The following descriptors of this class have been taken in study MW (molecular weight), AMW (average molecular weight), Sv (sum of atomic vander Waals volumes, scaled on Carbon atom), Se (sum of atomic Sanderson electronegativities, scaled on carbon atom), Sp (sum of atomic polarizabilities, scaled on Carbon atom), Ss (sum of Kier-Hall electrotopological states), Mv (mean atomic van der Waals volume, scaled on Carbon atom), Me (mean atomic Sanderson electronegativity, scaled on Carbon atom), Mp (mean atomic polarizability, scaled on Carbon atom), Ms (mean electrotopological state), RBF (rotatable bond fraction).

2- Geometrical descriptors[12-27]

The following descriptors of this class have been taken in study W3D (3D-Wiener index), J3D (3D-Balaban index), H3D (3D-Harary index), AGDD (average geometric distance degree), DDI (D/D index), ADDD (average, distance/distance degree), G1 (gravitational index G1), G2 (gravitational index G2), Rgyr (radius of gyration, mass weighted), SPAN (span R), SPAM (average span R), PJI3 (3D Petijean shape index), L/Bw (length-to-breadth ratio by WHIM), SEig (absolute eigen value sum on geometry matrix) and DISPm (d COMMA2 value / weighted by atomic masses).

3- BCUT descriptors[28-32]

The following descriptors of this class have been taken in study BEHm1, BEHm2, BEHm3, BEHm4 is the highest eigenvalue n. 1 to 4 of Burden matrix / weighted by atomic masses, BELm1, BELm2, BELm3, BELm4 is the lowest eigenvalue n. 1 to 4 of Burden matrix / weighted by atomic masses. The BEHv1, BEHv2, BEHv3, BEHv4is the highest eigenvalue n. 1 to 4 of Burden matrix / weighted by atomic van der Waals volumes and the BELv1, BELv2, BELv3, BELv4, is the lowest eigenvalue n. 1 to 4 of Burden matrix / weighted by atomic van der Waals volumes.

4- Empirical descriptors [12,33]

The following descriptors of this class have been taken in study Hy (hydrophilic factor), ARR (aromatic ratio), MR (Ghose-Crippen molar refractivity), and MLOGP (Moriguchi octanol-water partition coeff, (logP), MW (molecular weight)

5- Topological descriptors[34-46]

The following descriptors of this class have been taken in study ISIZ (information index on molecular size), IAC (total information index of atomic composition), ZM1V (first Zagreb index by valence vertex degrees), Snar (Narumi simple topological index, log), Hnar (Narumi harmonic topological index), Xt (otal structure connectivity index), Dz (Pogliani index), Pol (olarity number), GMTI (Gutman Molecular Topological Index), Xu (Xu index), PI (uperpendentic index), W (Wiener W index), Har (Harary H index), w (detour index), J (Balaban J index), DELS (molecular electrotopological variation), S0K (Kier symmetry index), PHI (Kier flexibility index), BLI (Kier benzene-likeliness index), Lop (Lopping centric index).

6- Galvez topol. charge indices[47-49]

The following descriptors of this class have been taken in study GGI1, GGI2, GGI3, GGI4 is the topological charge index of order 1 to 4, and the JGI1, JGI2, JGI3, JGI4, is the mean topological charge index of order1 to 4.The JGT is the global topological charge index.

Materials and Method:

The different group of compounds such as phenol, alcohol, acids, ester, aldehydes, ketone and amines have been chosen with their toxicity values [50-51] in terms of 50% inhibitory growth concentration (IGC501)) against the ciliate T. pyriformis. The T. pyriformis is one of the generally used ciliated protozoa for laboratory research. In this ciliate species, diverse endpoints can be used to originate the cytotoxic effects and xenobiotics. Experimental determination of toxicological and biochemical endpoints as well as the human health endpoints is a difficult task .Hence QSAR modeling of the toxicity of compounds on the T. pyriformis is of vital importance in investigating its toxicity in terms of its inhibitory growth concentration.

For the purpose all the molecules have been drawn and designed with the help of CAChe pro. Software in SYBYL Mol2 (.Mol2) format. For the calculation of all the descriptors values the Dragon software is used

Results: To develop QSTR model the molecules of training set i.e. 46 Phenol derivatives have been drawn on CAChe and their. MOL2 format structure has been used for descriptors calculation. The six QSAR model has been made with six different group of descriptors

First model:

The first model has been made with 11 constitutional descriptors as described in theory . The values of all the descriptors have been calculated with the help of Dragon software and the MLR analysis have been made by using all descriptors jointly against observed toxicity. The total 2048 regression models have been made and the regression summery of best fitted 10 model for forward step regression analysis is presented here

All Subsets Results

Index	VC	R2	RSS	Std. Error	Cp	Variables
1	1	0.6870076344	5.490593426	0.3659465983	104.8924072	MW
2	2	0.8276327919	3.023710364	0.2749413739	42.2426117	Se,Mv
3	3	0.8782622565	2.135555137	0.2340038974	20.96665886	Se,Me,Mp
4	4	0.9037659973	1.688161888	0.2107731312	11.24178934	Ss,Mp,Ms,RBF
5	5	0.9213881853	1.379028885	0.1930571184	5.140316796	AMW,Ss,Mp,Ms,RBF
6	6	0.9249464288	1.316609253	0.1912393245	5.504480654	AMW,Sv,Sp,Mp,Ms,RBF
7	7	0.9291527896	1.242820179	0.1884387115	5.570684696	AMW,Sv,Sp,Ss,Me,Mp,RBF
8	8	0.9301997718	1.224453745	0.1897718247	7.089354138	MW,AMW,Sv,Sp,Me,Mp,Ms,RBF
9	9	0.9322881925	1.187818126	0.1897221354	8.129241623	MW,AMW,Sv,Se,Sp,Ss,Me,Mp,RBF
10	10	0.9325104396	1.183919409	0.1923472941	10.02706765	MW,AMW,Sv,Se,Sp,Ss,Mv,Me, Mp,RBF

The model number 5 of this class is significant and noteworthy because of high correlation coefficient and low standard error as well as only 5 descriptors is used. The following regression equation –1 has been developed for model-5

$PT_{Const.}$ =-0.0907756*AMW+0.118817*Ss+10.6822*Mp-0.469315*Ms+6.64428*RBF-8.44418 -1

r^2_{CV}=0.900721 r^2=0.921388

The Predicted toxicity (PT) alongwith the values of descriptors used in study is presented in table-2. The reliability of this model is very high as clear from the value of cross validation coefficient rcv2 and correlation coefficient r2.

Second model:

The Second QSTR model has been developed with the help of 15 geometrical descriptors as described in theory . The calculation of all the descriptors is carried out with the help of Dragon software and the regression analysis has been made by using descriptors values jointly against observed toxicity. The total 65536 regression models have been made. The regression summary of best fitted 10 models for forward step regression analysis is presented here

All Subsets Results with geometrical descriptors

No.	VC	R2	RSS	Std. Error	Cp	Variables
1	1	0.6992155331	5.076187573	0.3518656747	46.09099893	G1
2	2	0.7317081825	4.527825535	0.336445595	38.89892852	G2,SPAM
3	3	0.7626563042	4.005529712	0.3204775968	32.14380025	ADDD,SPAN,SPAM
4	4	0.8422979003	2.661458709	0.2646476659	11.61343811	W3D,AGDD,DDI,ADDD
5	5	0.8813315388	2.0027077	0.2326526061	2.570942025	J3D,H3D,SPAN,SPAM,SEIG
6	6	0.8835383097	1.965465142	0.2336584129	3.946653345	W3D,J3D,AGDD,DDI,ADDD,SEIG
7	7	0.8979427383	1.722368873	0.2218344861	1.871684873	W3D,J3D,AGDD,DDI,ADDD,RGYR,SEIG
8	8	0.8995974642	1.6944429	0.2232410264	3.403568011	W3D,J3D,H3D,AGDD,DDI,ADDD,SPAM,SEIG
9	9	0.9040411636	1.619448829	0.2215269892	4.146459144	W3D,J3D,H3D,AGDD,DDI,ADDD,SPAM,PJI3,SEIG
10	10	0.905604683	1.593062101	0.2231214706	5.70414432	W3D,J3D,H3D,AGDD,DDI,ADDD,RGYR, SPAM,PJI3,SEIG

The model number 5 of this class is significant and noteworthy because of high correlation coefficient and low standard error as well as only 5 descriptors is used. The following regression equation –2 has been developed for model-5

S=1.38819*J3D-0.222105*H3D-2.06457*SPAN+42.106*SPAM+0.460147*SEig-26.7576 -2

r^2_{CV}=0.780879 r^2=0.881332

The Predicted toxicity (PT) alongwith the values of descriptors used in study is presented in table-3. The reliability of this model is very high as clear from the value of cross validation coefficient r^2_{cv} and correlation coefficient r^2.

Third model:

The third model for QSTR study has been developed with the help of 16 BCUT descriptors as described in theory . After the

calculation of all the descriptors on Dragon software the regression analysis have been made by using descriptor values jointly against observed toxicity. The total 2048 regression models have been made. The regression summery of best fitted 10 models for forward step regression analysis is presented here

All Subsets Results

No.	VC	R2	RSS	Std. Error	Cp	Variables
1	1	0.6985383702	5.260513516	0.3581971632	70.60048517	BEHM4
2	2	0.8047344233	3.407389545	0.2918642469	33.99146236	BELM1, BEHV1
3	3	0.8436329912	2.728608494	0.2645076943	21.84935773	BELM1, BEHV1, BEHV2
4	4	0.8689026636	2.287652033	0.2453596733	14.66222384	BEHM3, BELM1, BEHV1,BELV2
5	5	0.8838352613	2.027077803	0.2340638515	11.23327437	BEHM3,BEHM4, BELM1, BEHV2, BEHV3
6	6	0.8980650158	1.778768204	0.2222841153	8.059853109	BEHM3,BEHM4, BELM1, BELM4, BEHV2, BEHV3
7	7	0.918189821	1.427589815	0.2019610864	2.743205834	BEHM3,BEHM4, BELM1, BELM4, BEHV2, BEHV3, BEHV4
8	8	0.9229160976	1.345116163	0.1989025894	3.024903569	BEHM3,BEHM4, BELM1, BELM4, BEHV2, BEHV3, BEHV4,BELV2
9	9	0.9244737003	1.317935953	0.1998435716	4.458615857	BEHM3,BEHM4, BELM1, BELM4, BEHV2, BEHV3, BEHV4,BELV2,BELV4
10	10	0.9252657254	1.304115094	0.2018752008	6.170664351	BEHM2,BEHM3,BEHM4, BELM1, BELM4, BEHV2, BEHV3, BEHV4,BELV2,BELV4

The model number 5 of this class is of prime interest because of high correlation coefficient and low standard error as well as only 5 descriptors is used. The following regression equation –3 has been developed for model-5

S=-0.293959*BEHm3+0.626733*BEHm4-12.8675*BELm1+2.39878*BEHv2+2.71087*BEHv3+10.3177 rCV^2=0.806447 r^2=0.883835 -3

The Predicted toxicity (PT) alongwith the values of descriptors used in study is presented in table-4. The reliability of this model is very high as clear from the value of cross validation coefficient r^2_{cv} and correlation coefficient r^2.

Fourth model:

The Fourth model for QSTR study has been constructed by using 5 most common empirical descriptors as described in theory. After the calculation of all the descriptors on Dragon software the regression analysis have been made by using descriptors values jointly against observed toxicity. The regression summary of best fitted 5 models for forward step regression analysis is presented here

All Subsets Results

Index	VC	R2	RSS	Std. Error	Cp	Variables
1	1	0.8486315187	2.644332592	0.2539604632	47.13272155	MLOGP
2	2	0.8988069185	1.767793144	0.2102256612	20.58157469	Hy,MLOGP
3	3	0.913825115	1.50543267	0.1964709589	14.03581848	Hy,MR,MW
4	4	0.9357836127	1.121828564	0.1718191035	3.540844986	Hy,ARR,MR,MW
5	5	0.9373467493	1.094521339	0.1719931912	4.65137835	Hy,ARR,MR,MLOGP,MW

The model number 4 of this class is of prime interest. This model has high validation because it involves only four descriptors and provides much better result. The following regression equation –4 has been developed for model-4 of empirical series

J=6.48893*Hy-1.3904*ARR+0.139817*MR-0.0076914*MW-1.87399 -4
rCV^2=0.915446 r^2=0.935784

The Predicted toxicity (PT) alongwith the values of descriptors used in study is presented in table-5. The reliability of this model is very high as clear from the value of cross validation coefficient r^2_{cv} and correlation coefficient r^2.

Fifth model:

The Fifth model for QSTR study involves 20 topological descriptors as described in theory. The calculation of all the descriptors is carried out with the help of Dragon software and the regression analysis has been made by using descriptors values jointly against observed toxicity. The regression summary of best fitted 10 models for forward step regression analysis is presented here

All Subsets Results

Index	V C	R2	RSS	Std. Error	Cp	Variables
1	1	0.7967603419	3.495949844	0.2956327893	86.87988764	PHI
2	2	0.8620032441	2.373698824	0.2467064499	48.79161762	ISIZ,PHI
3	3	0.8867970639	1.947217341	0.2263681992	35.55714288	ISIZ,IAC,PHI
4	4	0.9217768945	1.345525062	0.190697515	16.06390998	ZM1V,Har,DELS,BLI
5	5	0.9405200806	1.023121259	0.1685824278	6.547225667	ZM1V,Xu,DELS,BLI,Lop
6	6	0.9488037775	0.8806323915	0.1586219577	3.45733749	ZM1V,Dz,Pol,Xu,BLI,Lop
7	7	0.954097696	0.7895710605	0.1523898889	2.204507993	ZM1V,Dz,Xu,J,S0K,PHI,BLI
8	8	0.9590640125	0.704144852	0.1460743742	1.152972417	ISIZ,ZM1V,Dz,Xu,W,J,BLI,Lop
9	9	0.9623746394	0.6471983598	0.1422144463	1.11876945	ISIZ,ZM1V,Dz,Xu,W,J,S0K,BLI,Lop
10	10	0.9640928954	0.617642433	0.1411522209	2.062993199	ISIZ,ZM1V,Xt,Dz,Pol,Xu,J,S0K,BLI,Lop

The model number 5 of this class is of prime interest. The following regression equation –5 has been developed for model-5 of topological series.

W=-0.0166322*ZMIV+0.316043*Xu+0.152982*DELS+2.78691*BLI-0.872823*Lop-3.69202 -5
rCV^2=0.914528 r^2=0.94052

The Predicted toxicity (PT) alongwith the values of descriptors used in study is presented in table-6. The reliability of this model is very high as clear from the value of cross validation coefficient r^2_{cv} and correlation coefficient r^2.

Sixth model:

The sixth QSTR has been developed by using 9 Galvez topol. charge indices as descriptors. Calculations of all the descriptors have been made on Dragon software and the regression analysis have been made by using descriptors values jointly against observed toxicity. The regression summary of best fitted 9 models for forward step regression analysis is presented here

All Subsets Results (512)

No.	VC	R2	RSS	Std. Error	Cp	Variables
1	1	0.5557556596	7.077213992	0.420630895	15.14823373	GGI3
2	2	0.6046588382	6.298142143	0.4018592094	11.29758505	GGI3,GGI4
3	3	0.6161957192	6.114349198	0.4011283165	11.91734271	GGI3,JGI3,JGI4
4	4	0.6606385781	5.406334274	0.3822527208	8.600315035	GGI1,GGI3,JGI1,JGI3
5	5	0.7018801825	4.749318228	0.3632155095	5.666277089	GGI2,GGI3,JGI2,JGI3,JGT
6	6	0.7065131079	4.675511538	0.3654942461	7.112005982	GGI2,GGI3,GGI4,JGI2, JGI3,JGT
7	7	0.724741049	4.385123956	0.3591298289	6.931262672	GGI1,GGI2,GGI3,GGI4,JGI1, JGI3,JGT
8	8	0.7297030508	4.306074781	0.3612300023	8.337621725	GGI1,GGI2,GGI3,GGI4,JGI1, JGI2,JGI3,JGT
9	9	0.7325250927	4.261117103	0.364910824	10	GGI1,GGI2,GGI3,GGI4, JGI1, JGI2,JGI3,JGI4,JGT

The model number 5 of this class is significant. The following regression equation –6 has been developed for model-5 of Galvez topol. charge indices series.

L=-5.6804*GGI2+12.0491*GGI3+43.3685*JGI2-89.0108*JGI3+2.46196*JGT-1.07331 –6
rCV^2=0.494713 r^2=0.70188

The Predicted toxicity (PT) alongwith the values of descriptors used in study is presented in table-7. The reliability of this model is clear from the value of cross validation coefficient r^2_{cv} and correlation coefficient r^2.

After such an study the topological descriptor has been identified as a major tool to describe the toxicity of phenol. The methodology may not be generalized only by the study of phenols, to make the proposed methodology more general and to certify its reliability we have tested the most significant model, which is based on 20 topological descriptor with other series of compounds whose activities are previously known. The further study involves alcohols, aldehydes, ketone, ester, acids and amine. The QSTR models of these series of organic molecules have been made with the help of 20 Topological descriptors and the total 1048576 regression models have been develop for each set. The predicted toxicity of each set with the help of most significant models is reported in various tables.

Amino alcohols: The 18 Amino alcohols have been taken for study alongwith their same type of toxicity. The summary of top 10 significant models are presented here

Top-10 models

No.	VC	R2	RSS	Std. Error	Cp	Variables
1	1	0.3460284831	2.687134147	0.4232520248	0	BLI
2	2	0.7257509716	1.126874657	0.2837094411	0	ZM1V,POL
3	3	0.8822195824	0.4839534654	0.1929434882	0	ISIZ,DZ,DELS
4	4	0.9224569831	0.3186201282	0.1629468544	0	SNAR,Xu,DELS,BLI
5	5	0.9645054872	0.1458450633	0.1151461772	0	XT, Xu, W, DELS, BLI
6	6	0.9796092169	0.08378464233	0.09153395126	0	SNAR,XT,Xu,W,DELS,BLI
7	7	0.9840308277	0.0656164788	0.08538571231	0	XT,POL,GMTI,Xu,Har,DELS,BLI
8	8	0.9916597108	0.03426980413	0.06545017583	0	ZM1V,HNAR,XT,DZ,POL,XuDELS,BLI
9	9	0.9947736807	0.02147466794	0.05538781189	0	ZM1V,SNAR,HNAR,DZ,POL,Xu,Har,DELS,PHI
10	10	0.9975073332	0.0102422352	0.04131633091	0	ZM1V,SNAR,HNAR,DZ,POL,Xu, Har,DELS,S0K,PHI

The model No.5 involves 5 variables namely XT, Xu, W, DELS, BLI. For the predicted toxicity of this set by model 5 the regression equation 7 has been developed and predicted toxicity is presented in table-8. The reliability of this model is high as clear from its value of squire of correlation coefficient (r^2 = 0.964505) and squire of cross validation coefficient (rCV^2 = 0.530897).

W=27.7764*Xt+2.64136*Xu-0.088688*W+0.413688*DELS+9.47142*BLI-35.8612 -7
rCV^2=0.530897 r^2=0.964505

Acetylenic alcohols The 10 Acetylenic alcohols are used as study material alongwith their observed toxicity values. The summary of top 6 significant models are presented here

Top-6 subsets

No.	VC	R2	RSS	Std. Error	Cp	Variables
1	1	0.8403370638	0.2160964397	0.1897790117	0	SPI
2	2	0.9436461627	0.07627232653	0.1235089685	0	SNAR,LOP
3	3	0.9797628222	0.02739008924	0.08274975716	0	DZ,POL,W
4	4	0.9983347099	0.00225389356	0.02740981308	0	J,DELS,S0K,LOP
5	5	0.9999766405	3.161599143E-005	0.003975927026	0	ZM1V,SPI,W,DELS,LOP
6	6	0.9999999996	4.760778787E-010	2.181920894E-005	0	ISIZ,XT,W,Har,BLI,LOP

The model No.4 involves 4 variables namely J,DELS,S0K,LOP. For the predicted toxicity of this set by model 4 the regression equation 8 has been developed and predicted toxicity is presented in table-9. The compounds 2 and 5 are outliers. This model has high reliability and its value of squire of correlation coefficient (r^2 = 0.998335) and squire of cross validation coefficient (rCV^2 = 0.96536).

W=0.639793*Q-0.323816*R+0.328531*S-1.94098*V-1.69144 -8
rCV^2=0.96536 r^2=0.998335

Diols The next set contains 8 Diols with their observed toxicity values. The summary of top 6 significant models are presented below

Top-6 subsets

No.	VC	R2	RSS	Std. Error	Cp	Variables
1	1	0.859145244	0.1303891883	0.1474161842	6	POL
2	2	0.9889827667	0.01019864826	0.04516336626	6	POL,J
3	3	0.9962641796	0.003458247345	0.02940343239	6	ZM1V,Xu,Har
4	4	0.9989463305	0.00097538142	0.01803128226	6	ZM1V,DZ,SPI,Har
5	5	0.9999953081	4.343282175E-006	0.001473648902	6	ISIZ,IAC,XT,SPI,Har
6	6	1	8.830425215E-012	2.971603139E-006	6	HNAR,XT,POL,Xu,S0K,LOP

The model No.3 involves 3 variables namely ZM1V, Xu, Har. For the predicted toxicity of this set by model 3 the regression equation 9 has been developed and predicted toxicity is presented in table-10. The reliability of this model is high as is clear from its value of squire of correlation coefficient (r^2=0.996264) and squire of cross validation coefficient (rCV^2=0.675103).

W=-0.525404*ZM1V-1.3731*Xu+2.82797*Har+22.2782 -9
rCV^2=0.675103 r^2=0.996264

Halogenated alcohols This set involves halogenated alcohols with their observed toxicity values. The regression summary of 8 significant models are presented below

Top-8 subsets

No.	VC	R2	RSS	Std. Error	Cp	Variables
1	1	0.5880041292	1.406196107	0.4192547119	0	PHI
2	2	0.8849250259	0.3927660255	0.2368742962	0	IAC,DZ
3	3	0.9677738617	0.1099920495	0.1353957468	0	IAC,DZ,J
4	4	0.9765252861	0.08012228692	0.1265877458	0	IAC,DZ,SPI,J
5	5	0.9937802923	0.02122868051	0.07285032689	0	ISIZ,XT,DZ,BLI,LOP
6	6	0.9954206883	0.01562979336	0.07217985259	0	IAC,ZM1V,POL,W,PHI,LOP
7	7	0.999887562	0.0003837655787	0.01385217634	0	SNAR,XT,DZ,SPI,J,S0K,PHI
8	8	1	5.669772087E-012	2.381128322E-006	0	ISIZ,IAC,HNAR,XT,SPI,Har,DELS,PHI

The model No.4 involves 4 variables namely ISIZ, XT, DZ, BLI, LOP. For the predicted toxicity of this set by model 3 the regression equation 10 has been developed and predicted toxicity is presented in table-11. The model 4 is reliable as is clear from the value of squire of correlation coefficient (r^2 = 0.976525) and squire of cross validation coefficient (rCV^2 = 0.775319).

W=1.09771*IAC-1.74597*Dz+0.119486*SPI-1.02252*J+0.0852569 -10
rCV^2=0.775319 r^2=0.976525

Saturated alcohols 20 saturated alcohols are used as study material alongwith their observed toxicity values. The summary of top 10 significant models are presented here

Top-10 subsets

No.	VC	R2	RSS	Std. Error	Cp	Variables
1	1	0.9527635226	0.3533651468	0.1401120874	0	S0K
2	2	0.9701324212	0.2234324394	0.1146432932	0	IAC,DELS
3	3	0.9791167115	0.1562230441	0.09881265231	0	SNAR,J,DELS
4	4	0.986641257	0.09993366213	0.08162257128	0	DZ,J,DELS,BLI
5	5	0.9937547849	0.04671900742	0.05776739529	0	SNAR,DZ,SPI,DELS,BLI
6	6	0.9948435346	0.03857432322	0.05447252181	0	HNAR,Xu,J,DELS,S0K,BLI
7	7	0.9968629828	0.02346729884	0.04422225951	0	ISIZ,SNAR,XT,DZ,DELS,S0K,BLI
8	8	0.9973262555	0.02000166357	0.04264191661	0	IAC,XT,DZ,GMTI,DELS,S0K,BLI,LOP
9	9	0.9982226046	0.01329628336	0.03646406911	0	ISIZ,IAC,ZM1V,DZ,POL,GMTI,W,DELS,S0K
10	10	0.9989913957	0.007545134859	0.02895424524	0	ISIZ,IAC,ZM1V,DZ,POL,GMTI,SPI,W,DELS,S0K

The model No.5 involves 5 variables namely SNAR, DZ, SPI, DELS, BLI. For the predicted toxicity of this set by model 5 the

regression equation 10 has been developed and predicted toxicity is presented in table-11. The reliability of this model is high as is clear from its value of squire of correlation coefficient (r^2 = 0.993755) and squire of cross validation coefficient (rCV^2 = 0.962475).

W=-1.76146*SNar+1.00977*Dz-0.0477655*SPI-0.757903*DELS+4.00497*BLI-11.4942 -11
rCV^2=0.962475 r^2=0.993755

Unsaturated alcohols 24 unsaturated alcohols are used as study material alongwith their observed toxicity values. The summary of top 10 significant models are presented here.

Top-10 Subsets

No.	VC	R2	RSS	Std. Error	Cp	Variables
1	1	0.8190142646	1.404845292	0.252698643	-19.9958062	ISIZ
2	2	0.8898133628	0.8552893862	0.2018119571	-17.99744676	ISIZ,IAC
3	3	0.9278552014	0.5600014857	0.1673322273	-15.99832826	Xu,DELS,LOP
4	4	0.9375717708	0.484579648	0.1597003193	-13.99855341	SNAR,HNAR,DELS,LOP
5	5	0.9463863553	0.4161591859	0.1520524737	-11.99875766	IAC,Har,DELS,S0K,LOP
6	6	0.9556285398	0.3444196133	0.1423375469	-9.998971825	IAC,GMTI,W,DELS,S0K,LOP
7	7	0.9621056121	0.2941433607	0.1355874627	-7.999121911	SNAR,XT,Xu,SPI,Har,J,DELS
8	8	0.9676944935	0.2507614132	0.1292958915	-5.999251417	ISIZ,IAC,ZM1V,Xu,Har,J,S0K,LOP
9	9	0.9771092217	0.1776825236	0.1126570407	-3.999469575	ISIZ,IAC,ZM1V,XT,POL,Xu,SPI,Har,J
10	10	0.9795176741	0.1589876634	0.1105885178	-1.999525384	ISIZ,IAC,ZM1V,XT,POL,Xu,SPI,Har,J,DELS

The model No.5 involves 5 variables namely IAC, Har, DELS, S0K, LOP For the predicted toxicity of this set by model 5 the regression equation 10 has been developed and predicted toxicity is presented in table-12. The reliability of this model is high as is clear from its value of squire of correlation coefficient (r^2 = 0.946386) and squire of cross validation coefficient (rCV^2 = 0.869735).

W=0.368063*D-0.76613*O-0.205815*R+0.352465*S-1.27437*V-4.19772 -12
rCV^2=0.869735 r^2=0.946386

Monoesters 10 monoesters are used as study material alongwith their observed toxicity values. The summary of top 8 significant models are presented here

Top-8 subsets

No.	VC	R2	RSS	Std. Error		Cp	Variables
1	1	0.8186318369	0.5696224131	0.266838531	6		GMTI
2	2	0.8555441702	0.4536919655	0.2545842452	6		HNAR,SPI
3	3	0.9551864382	0.1407458111	0.1531588561	6		IAC,ZM1V,PHI
4	4	0.9676661485	0.1015508245	0.1425137359	6		ISIZ,ZM1V,DZ,PHI
5	5	0.9779643652	0.06920724813	0.1315363525	6		ISIZ,SNAR,J,PHI,BLI
6	6	0.9961058253	0.01223042214	0.0638498816	6		ISIZ,SNAR,DZ,POL,Xu,PHI
7	7	0.9999043324	0.0003004630144	0.01225689631	6		IAC,ZM1V,DZ,POL,SPI,Har,J
8	8	0.9999999917	2.591921358E-008	0.000160994452	6		IAC,SNAR,XT,Har,J,DELS,S0K,BLI

The model No.4 involves 4 variables namely ISIZ, ZM1V, DZ, PHI For the predicted toxicity of this set by model 4 the regression equation 10 has been developed and predicted toxicity is presented in table-13. The reliability of this model is high as is clear from its value of squire of correlation coefficient (r^2 = 0.967666) and squire of cross validation coefficient (rCV^2 = 0.795532).

W=0.405415*ISIZ+0.864327*ZM1V-5.93915*Dz+2.15043*PHI-26.7516 -13
rCV^2=0.795532 r^2=0.967666

Aldehydes 11 aldehydes are used as study material alongwith their observed toxicity values. The summary of top 8 significant models are presented here

Top-8 subsets

No.	VC	R2	RSS	Std. Error		Cp	Variables
1	1	0.7509266084	0.1120523304	0.1183492345	0		GMTI
2	2	0.7943303675	0.09252598789	0.1149695537	0		POL,PHI
3	3	0.9059896957	0.04229305111	0.08395738117	0		HNAR,XT,J
4	4	0.9882829807	0.005271214679	0.03246910741	0		SNAR,GMTI,J,PHI
5	5	0.9934307291	0.002955362319	0.02718162209	0		ISIZ,POL,PHI,BLI,LOP
6	6	0.9996392159	0.0001623083758	0.007355460008	0		IAC,XT,J,PHI,BLI,LOP
7	7	0.999993111	3.099195929E-006	0.001244828488	0		IAC,XT,Har,J,S0K,BLI,LOP
8	8	1	4.509451777E-014	2.12354698E-007	0		IAC,HNAR,DZ,SPI,Har,S0K,BLI,LOP

The model No.4 involves 4 variables namely SNAR, GMTI, J, PHI For the predicted toxicity of this set by model 4 the regression equation 10 has been developed and predicted toxicity is presented in table-14. The compound 4 is outlier. The reliability of this model is high as is clear from its value of squire of correlation coefficient (r^2 = 0.988283) and squire of cross validation coefficient (rCV^2 =

0.886292).

W=-2.61671*SNar+0.00849455*GMTI+2.94892*J+1.09521*PHI-5.53238 -14
rCV^2=0.886292 r^2=0.988283

Ketone 9 ketones are used as study material alongwith their observed toxicity values. The summary of top 7 significant models are presented here

Top-7 subsets

No.	VC	R2	RSS	Std. Error	Cp	Variables
1	1	0.9827005802	0.05995091248	0.09254413038	6	SNAR
2	2	0.9936665593	0.02194845566	0.06048202992	6	IAC,Xu
3	3	0.9980002709	0.006930035035	0.03722911504	6	HNAR,POL,SPI
4	4	0.999468154	0.001843105477	0.02146570216	6	HNAR,POL,SPI,DELS
5	5	0.9999234736	0.0002652011558	0.009402147906	6	ISIZ,IAC,HNAR,XT,BLI
6	6	0.9999998577	4.931424044E-007	0.0004965593642	6	ISIZ,HNAR,POL,Xu,J,PHI
7	7	1	1.450763606E-011	3.808889084E-006	6	ZM1V,HNAR,DZ,GMTI,Xu,W,DELS

The model No.3 involves 3 variables namely HNAR, POL, SPI For the predicted toxicity of this set by model 3 the regression equation 10 has been developed and predicted toxicity is presented in table-15. The reliability of this model is high as is clear from its value of squire of correlation coefficient (r^2 = 0.998) and squire of cross validation coefficient (rCV^2 = 0.996362).

W=5.78794*HNar-0.132708*Pol+0.0692124*SPI-9.38798 -15
rCV^2=0.996362 r^2=0.998

Amines The 18 amines are used as study material alongwith their observed toxicity values. The summary of top 10 significant models are presented here.

Top-10 subsets

No.	VC	R2	RSS	Std. Error	Cp	Variables
1	1	0.7124914337	0.1347219225	0.09809686266	0	S0K
2	2	0.866945803	0.06234707176	0.06925264325	0	ISIZ,ZM1V
3	3	0.917706956	0.03856120613	0.05668715766	0	GMTI,DELS,BLI
4	4	0.9578630027	0.01974472397	0.04236714418	0	POL,GMTI,SPI,DELS
5	5	0.9664697359	0.01571174625	0.03963804517	0	POL,GMTI,SPI,J,DELS
6	6	0.9760679443	0.0112141791	0.03529900707	0	ISIZ,XT,GMTI,Xu,SPI,DELS
7	7	0.9803067882	0.00922792453	0.03396307651	0	ZM1V,SNAR,XT,POL,SPI,W,DELS
8	8	0.9896015837	0.004872531826	0.02638325179	0	ISIZ,ZM1V,DZ,POL,Xu,J,S0K,PHI
9	9	0.9962353779	0.001764041775	0.01714663123	0	ISIZ,ZM1V,SNAR,DZ,POL,Xu,J,S0K,PHI
10	10	0.9983570438	0.0007698630524	0.01240857004	0	ISIZ,ZM1V,SNAR,DZ,POL,Xu,J,DELS,S0K,PHI

The model No.5 involves 5 variables namely POL,GMTI,SPI,J,DELS For the predicted toxicity of this set by model 3 the regression equation 16 has been develop and predicted toxicity is presented in table-12. The compounds 8 and 13 are outliers. The reliability of this model is high as clear from its value of squire of correlation coefficient (r^2=0.900639) and squire of cross validation coefficient (rCV^2=0.96647).

W=0.0981751*Pol+0.0034793*GMTI-0.126441*SPI+0.170896*J+0.139643*DELS-1.14907 -16
rCV^2=0.900639 r^2=0.96647

Conclusion: On the basis of present study it has been concluded that the topological descriptors have sufficient reliability to describe the toxicity of any small organic molecule against Tetrahymena pyriformis. On the basis of known toxicity of library of any series of compounds one can predict the toxicity of several hypothetical molecules prior than their synthesis, which reduces the coast and time efforts of discovery of molecule of interest.

Acknowledgment: We are thankful to Prof. P. K. Chattaraj (Indian Institute of Technology Kharagpur, W. B., INDIA) for valuable suggestions.

The Manuscript Includes following 17 Tables, click on the corresponding link to see the table:

Table:8. The values of topological descriptors of Amino alcohols and their observed and predicted toxicity against *Tetrahymena pyriformis*[51]

Table:9. The values of topological descriptors of a Acetylenic alcohols and their observed and predicted toxicity against *Tetrahymena pyriformis* [51]

Table:10. The values of topological descriptors of Diols and their observed and predicted toxicity against *Tetrahymena pyriformis* [51]

Table:11. The values of topological descriptors of Halogenated alcohols and their observed and predicted toxicity against *Tetrahymena pyriformis* [51]

Table:12. The values of topological descriptors of Saturated alcohols and their observed and predicted toxicity against *Tetrahymena pyriformis* [51]

Table:13. The values of topological descriptors of unsaturated alcohols and their observed and predicted toxicity against *Tetrahymena pyriformis* [51]

Table:14. The values of topological descriptors of monoesters and their observed and predicted toxicity against *Tetrahymena pyriformis* [51]

Table:15. The values of topological descriptors of Aldehydes and their observed and predicted toxicity against *Tetrahymena pyriformis* [51]

Table:16. The values of topological descriptors of Ketones and their observed and predicted toxicity against *Tetrahymena pyriformis* [51]

Table:17. The values of topological descriptors of Amines and their observed and predicted toxicity against *Tetrahymena pyriformis* [51]

References:

1- Smeyers, Y.G., Bouniam, L., Smeyers, N.J., Ezzamarty, A., Hernandez-Laguna,A., Sainz-Diaz,C.I., *Eur.J. Med. Chem*.,33 ,103, 1998.

2- Choplin, Computers and the medicinal chemist; in: Comprehensive Medicinal Chemistry, vol. 4, Eds. C.Hansch, P. G. Sammes, and J. B. Taylor, Pergamon Press, Oxford, 1990, pp. 33–58.

3- R. Franke, Theoretical Drug Design Methods, Elsevier, Amsterdam, 1984.

4- Roy DR, Parthasarathi R, Maiti B, Subramanian V, Chattaraj PK. Bioorg Med Chem. 16;13(10):3405-12, 2005.

5- P. P. Singh, F. A. Pasha, H. K. Srivastava, *QSAR & Combi Sci.*, 22(8), 843-851, 2003

6- P. P. Singh, F. A. Pasha, H. K. Srivastava, *Indian Journal of Chem.-B*, 43B, 983-991, 2004

7- P. P. Singh, H. K. Srivastava, F. A. Pasha, *Bioorg. Med. Chem.*, 12(1), 171-177, 2004

8- F. A. Pasha, H. K. Srivastava, P. P. Singh, *Molecular Diversity*, 9 (1-3), 215-220, 2005

9- F. A. Pasha, H. K. Srivastava, P. P. Singh, *Intr. J. Quant. Chem.*, 104 (1), 87-100, 2005.

10- H. K. Srivastava, F. A. Pasha, P. P. Singh, *Intr. J. Quant. Chem.*, 103 (3), 237-245, 2005.

11- Hansch, C.; Fujita, T. *J Am Chem Soc*, 86, 1616-1626, 1964.

12- Todeschini, R. &-. Consonni, V., Handbook of Molecular Descriptors. Wiley-VCH, Weinheim (Germany), (2000).

13- Randic,M. Int .J. Quantum Chem. Quant. Biol. Symp., (1966), 15, 201-208.

14- Nikolic, S., Trinajstic, N., Mihalic, Z. & Carter, S. Chem. Phys. Lett., (1991), 179,21-23.

15- Diudea, M.V, Horvath, D. & Graovac, A. J. Chem. inf. Comput. Sci., (1995), 35,129-135.

16- Balaban, A.T. J. Chem. inf. Comput. Sci, (1997), 37, 645-650.

12- Mekenyan, O., Peitchev, D., Bonchev, D., Trinajstic, N. & Bangov, I.P. Arzneim. Forsch., (1986), 36,176-183.

13- Mihalic, Z., Nikolic, S. & Trinajstic, N. J. Chem. inf. Comput. Sci, (1992), 32, 26-37.

14- Randic, M, Kleiner, A.F. & DeAlba, L.M. J. Chem. inf. Comput. Sci., (1994), 34, 277-286.

15- Katritzky, A.R, Mu, L., Lobanov, V.S. & Karelson, M. J. Phys. Chem, (1996), 100,10400-10407.

16- Tanford, C. Physical Chemistry of Macromolecules. Wlley, New York (NY), (1961).

17- Volkenstein, MV. Configurational Statistics of Polymeric Chains. Wiley-lnterscience, New York (NY), (1963).

18- Arteca, G..A. Reviews in Computational Chemistry - Vol. 9 (Lipkowitz, KB. and Boyd, D., eds.), VCH Publishers, New York (NY), (1991), pp. 191-253.

19- Robinson, D.D., Barlow, T.W. & Richards, W.G. J. Chem.i nf. Comput. Sci., (1997), 37, 939-942.

20- Bath, P.A, Poirrette, A.R., Willett, P. & Alien, F.H. J. Chem. inf. Comput. Sci., (1995), 35, 714-716.

21- Kruszewski, J. & Krygowski, T.M. Tetrahedron Lett., (1972), 36, 3839-3842.

22- Jug, K. J.Org.Chem., (1983), 48,1344.

23- Burden. F.R., J.Chem. inf. Comput. Sci., (1989), 29, 225-227.

24- Burden. F.R., Quant. Struct .-Act. Relat., (1997), 16, 309-314.

25- Pearlman, R.S. & Smith, KM, 3D-QSAR in Drug Design - V.2 (H.Kubinyi et al.), Kluwer/Escom, Dordrecht (The Netherlands), (1998), pp. 339-353.

26- Pearlman, R.S. & Smith, KM, J.Chem. inf. Comput. Sci., (1999), 39, 28-35.

27- Benigni, R., Passerini, L., Pino, A. & Giuliani, A., Quant. Struct. -Act. Relat., (1999), 15, 449-455.
28- Moriguchi, I., Hirono, S., Liu, Q., Nakagome, I. & Matsushita,Y. Chem. Pharm. Bull., (1992), 40, 127-130
29- Narumi, H. MATCH (Comm. Math. Comp. Chem.), (1987), 22,195-207.
30- Needham, D.E.Wei, I.C. & Seybold, P.O. J. Am. Chem. Soc., (1988), 110, 4186-4194.Pogliani, L. J. Phys. Chem., (1996), 100, 18065-18077.
31- Platt.J.R. J. Chem. Phys., (1947), 15, 419-420.
32- Gutman,I. J. Chem. Inf. Comput. Sci., (1994), 34, 1037-1039.
33- Gupta, S., Singh. M. & Madan, A.K. J. Chem. Inf. Comput. Sci., (1999), 39, 272-277.
34- Wiener, H. J. Am. Chem. Soc., (1947), 69, 17-20.
35- vanciuc, O., Balaban, T.-S. & Balaban, A.T. J. Math. Chem, (1993), 12, 309-318.
36- Amic, D. & Trinajstic, N. Croat. Chem. Acta, (1995), 68, 53-62.
37- Balaban A.T. Chem. Phys. Lett., (1982), 89, 399-404.
38- Gramatica, P., Corradi, M., Consonni, V. Chemosphere, (2000), 41,783-777.
39- Kier, L.B. Quant. Struct. -Act. Relat., (1987), 6, 8-12.
40- Kier, L.B. Quant. Struct. -Act. Relat., (1989), 8, 221-224.
41- Kier, L.B . & Hall, L.H. Molecular Connectivity in Structure-Activity Analysis. RSP-Wlley, Chichester (UK), (1986).
42- Galvez, J., Garcia, R., Salabert, M.T. & Soler, R. J. Chem. Inf. Comp. Sci., (1994), 34, 520-525.
43- Galvez, J., Garcia-Domenech, R., De Julian-Ortiz, V. & Soler, R. J. Chem. Inf. Comput. Sci., (1995), 35, 272-284.
44- Galvez, J., Garcia-Domenech, R., de Gregorio Alapont, C., De Julian-Ortiz, V. & Popa, L. J. Mol. Graphics, (1996), 14, 272-276.
45- Rios-Santamarina, I., Garcia-Domenech, R., Galvez, J., Cortijo, J., Santamaria, P. & Marcillo, E. Bioorg. Med. Chem. Lett., (1998), 8, 477-482.
46- L. H. Hall and T. A. Vaughn, *Me*d. *Che*m. *Re*s., 7, 407–416, 1997.
47- Schultz, T. W. *Toxicol. Methods*,7 ,289, 1997.
48- Akers,K.S.;Sinks,G.D.;Schultz,T.W.Environ.Toxicol.Pharmacol.,7 ,33, 2003.
49- Dimitrov,S.D.;Mekenyan,O.G.;Sinks,G.D.;Schultz,T.W.J.Mol.Struct.(Theochem) ,622 ,63, 2003.
50- Hall,L.H.;Vaughn,T.A. Med.Chem.Res.,7 ,407, 1997.
51- Schultz,T.W.Toxicol.Methods ,7 ,289, 1997.
52- F. A. Pasha, H. K. Srivastava, P. P. Singh, *Bioorg. & Med. Chem*, (inpress 2005)