International Journal of Life science & Pharma Research

TOPOLOGICAL DESCRIPTORS BASED MULTILINEAR REGRESSION ANALYSIS OF CARCINOGENIC PROPERTIES OF ANILINE DERIVATIVES

S. A. KHAN*, PRABHAT KUMAR VERMA, ROLI TEWARI

M. L. K. (P. G.) College, Balrampur-271201, U.P., India

ABSTRACT

QSAR models of 73 derivatives of aniline, biphenylamine, naphthylamine and aminofluorenes which have carcinogenic property have been developed with the help of topological and energy descriptors such as log P calculated, connectivity index (order 1), valence connectivity index (order 0), shape index (order 1), dipole moment, solvent accessibility surface area and molar refractivity. The qualities of the models have been adjudged by the value of cross-validation and correlation coefficients evaluated by multi linear regression analysis. The best model has correlation coefficient 0.854963 and has been developed with the help of descriptors Log P calculated, connectivity index (order 1), valence connectivity index (order 0) and shape index (order 1).

Key words: Carcinogenicity, connectivity index, log P, dipole moment, solvent accessibility surface, molar refractivity.

INTRODUCTION

Aromatic amines have been reported to be powerful carcinogens and mutagens, and/or hemotoxicants. Updated review on the toxicology of aromatic amines and their mechanisms of action has been illustrated by Woo [1]. Given their importance and the large amount of data available, the toxicity of the aromatic amines has been studied also with methods based on structure–activity relationship (SAR) and quantitative SAR (QSAR) concepts. [2, 3] Several QSAR studies on the aromatic amines have been reported, mainly regarding their mutagenic properties. [4-8] surprisingly, very sporadic and limited QSAR studies of their carcinogenic properties existed until recently, in spite of the fact that several of them had been bioassayed thus providing the necessary database [9-10].

Aromatic amines [11] are a common contaminant in several working environments, including the chemical and mechanical industries, and arylamines based dyes are widely used in textile industries, and cosmetics [12-13]. The wide use of aromatic amines together with the presence of relatively, very high exposure permitted the development of epidemiological knowledge unparalleled for other chemical classes. The evidence regarding the carcinogenic potential of aromatic amines in animals was available before formal epidemiologic studies were conducted: in this sense, arylamines are one of the best examples of the predictivity of animal experiments for human risk [14].

Although the major concern posed by the aromatic amines drives from their carcinogenic potential, the number of QSAR studies is quite limited [15], hence needs a comprehensive study on QSAR of aromatic amines, whose biological activity is reported. In this paper we have made

QSAR studies on carcinogenicity of aromatic amines with the help of topological descriptors, and to evaluate the quality of QSAR by multi linear regression analysis. Once the quality is established, the best descriptors can be chosen for predicting the activity of any new compound.

MATERIAL AND METHOD

Our main objective is to make QSAR/MLR analysis of the compound listed in Table-1, with the help of following topological descriptors. [17-22]

1. Log P Calculated By PM3 Method LPC
2. Connectivity Index (order 1, standard) CI
3. Valence Connectivity Index (order 0, standard) VCI
4. Shape Index (kappa alpha, order 1) SI
5. Dipole Moment DM
6. Solvent Accessibility Surface Area SASA
7. Molar Refractivity MR

73 derivatives of aniline, biphenylamine, naphthylamine and aminofluorenes, as listed in Table-1 are the study material of the paper. For the biphenylamines (Fig-1) substituted in the aniline part are characterized as in substituted anilines. In cases 1 and 2 the second part of the molecule (second phenyl ring plus substituents at this ring) is then treated as a para substituents, where the bridge X may be present or absent. In case 3, the non-aniline part appears as the ortho substituent. In the case of naphthylamine (Fig-2) two situations are possible. They are treated as anilines substituted by -C_4H_4-. Amino fluorenes are only three in the list of 58 compounds at serial 3, 14 and 40. Their structural formula is shown in fig-3. The carcinogenic potency data that are shown in the Table-1 are the TD_{50} (mg/kg/day) values calculated by Gold [16]. The TD_{50} is the daily dose rate required to halve the probability of an experimental animal of remaining tumorless to the end of its standard life span.

For QSAR prediction, the 3D modeling and geometry optimization of all the compounds of Table-1, have been done with the help of Cache software using the semiempirical PM3 Hamiltonian. For regression analysis, we have used the Project leader program associated with Cache Pro Software of Fujitsu. Various regression equations have been developed for the prediction of activity of carcinogenic compounds.

Table 1. ***Structures of carcinogenic compounds***

Comp	Ring	AnX	Bridge X	R	log P
1	N	3-C_4H_4-4		H	2.27
2	B	4-Ph-4-NH_2		H	2.16
3	F	3,4-Me_2		COMe	2.61
4	B	2-Cl,4-Ph-3-Cl,4-NH_2	CH_2	H	3.60
5	A	2-Me		H	1.73
6	B	4-C(=NH)-Ph-4N$(Me)_2$	C=NH_2	Me_2	3.02
7	B	2-Ph		H	2.95
8	A	2,6-Cl_2,4-NH_2		H	1.52
9	A	2-NO_2,4-N$(C_2H_4OH)_2$		Me	0.34
10	B	4-CH_2-Ph-4-NH_2	CH_2	H	2.56
11	A	4-Cl		$CONMe_2$	1.64

Comp	Ring	AnX	Bridge X	R	log P
12	B	4-O-Ph-4-NH_2	O	H	1.91
13	A	2-OEt,5-NHCOMe	H		0.20
14	F	3-Me,4-NEt		H	2.39
15	A	3-NO_2,4-OH		H	0.93
16	A	H		H	1.26
17	A	2-OMe		H	1.01
18	A	4-Cl		H	1.78
19	A	2Cl,5-NH_2		H	1.00
20	A	2NH_2,4Cl		H	1.00
21	A	2Me,4-OMe		H	1.48
22	A	2-OMe,5-Me		H	1.48
23	B	4-SO_2-Ph-4-NH_2	SO_2	H	1.31
24	A	2-OMe,5-NH_2		H	0.23
25	B	4-CH_2-Ph-4-N$(Me)_2$	CH_2	H	3.71
26	B	4-CO-Ph-4-N$(Me)_2$	CO	H	2.85
27	N	2-$C_3H_3C(NH_2)$-3		H	1.48
28	A	3-NO2,4-OEt		COMe	0.94
29	A	2-OMe,5-NO_2		H	0.96
30	A	2-NO_2,4-NH_2		H	0.43
31	B	4-S-Ph-4-NH_2	S	H	2.25
32	A	2,6-(NO_2)2,4-CF_3		$(nPr)_2$	4.25
33	A	2,4,5-Me_3		H	2.67
34	B	4-Ph		H	2.95
35	A	2-OH,4-NO_2		H	0.93
36	A	2-OH,5-NH_2		H	0.20
37	B	4-Ph		COMe	2.58
38	B	4-Ph-4-F		H	3.09
39	B	4-Ph-4-F		COMe	2.72
40	F	3,4-Me_2		$COCF_3$	3.73
41	B	2-Cl,4-Ph-3-Cl,4-NH_2		H	3.20
42	B	4-SO_2-Ph-4-NHCOMe	NH_2	COMe	0.57
43	A	4-OEt		COMe	0.99
44	A	4-F		Me,NO	1.83
45	A	H		Me,NO	1.69
46	A	2-NH_2		H	0.48
47	B	2-NH_2,4-Ph-3,4-$(NH_2)_2$		H	0.60
48	A	2,4,5,6-F_4,3-NH_2		H	1.04
49	A	2,4,6-Me_3		H	2.67
50	A	H		Me	1.84
51	A	4-Me		H	1.73

Comp	Ring	AnX	Bridge X	R	log P
52	A	2-OH,5-NO_2		H	0.93
53	A	2,4,6-Cl_3		H	2.82
54	A	3-Me		H	1.73
55	B	2-OMe,4-Ph-3-OMe,4-NH_2			1.66
56	B	2-Me,4-Ph-3M_3,4-NH_2		H	2.53
57	A	2,5-Cl_2,3-COOH		H	2.00
58	B	2-Me,4-CH_2-Ph-3-Me,4-NH_2	CH_2	H	3.50
59	A	3-Cl		COOiPr	2.79
60	A	2-M2, 3-NH_2		H	0.95
61	A	2-COOH		H	0.96
62	A	4-$COCH_2Cl$		COMe	0.80
63	A	2-Cl, 4-NH_2		H	1.00
64	A	2,4-OMe_2		H	0.76
65	A	2,6-OMe_2, 4-OCONMe		Me_2	2.25
66	N	2-C_4H_4-3		$C_2H_4NH_2$	1.69
67	A	2-COOH, 5-NO_2		H	0.92
68	A	2-NH_2, 4-NO_2		H	0.43
69	A	4-NH_2		H	0.48
70	A	4-NH-Ph-4-NH_2	NH	H	2.88
71	A	H		$CSNH_2$	1.86
72	A	2-Me, 4-NH_2		H	0.95
73	A	2-Cl, 4-Me		H	2.25

[α]A = anilines; B = biphenylamines; N = naphthylamines; F = aminofluorenes. Bridge: bridge between the phenyl rings in biphenylamines if present. AnX: ring substituent (all compounds described as substituted anilines. R = substituent at the functional amino group.

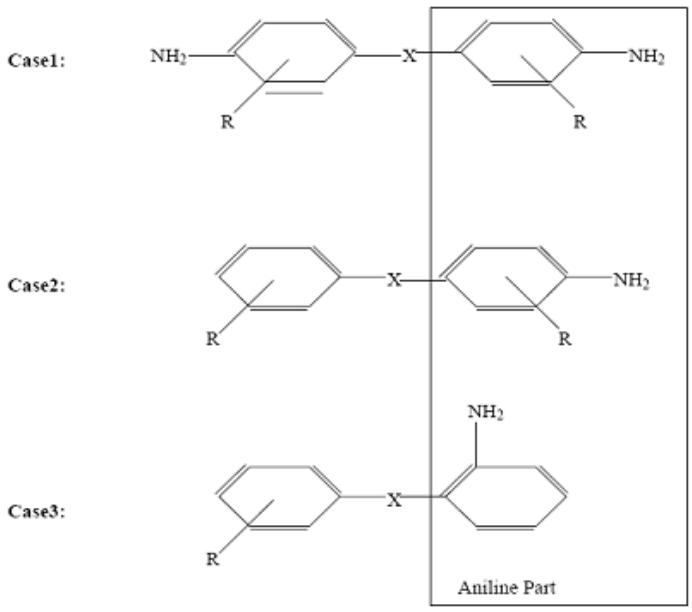


Figure 1.*Treatment of biphenylamine.*

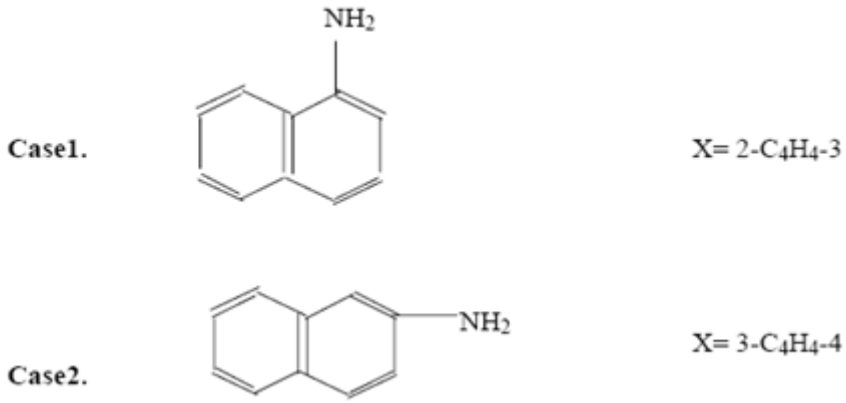


Figure 2. ***Treatment of naphthylamines***

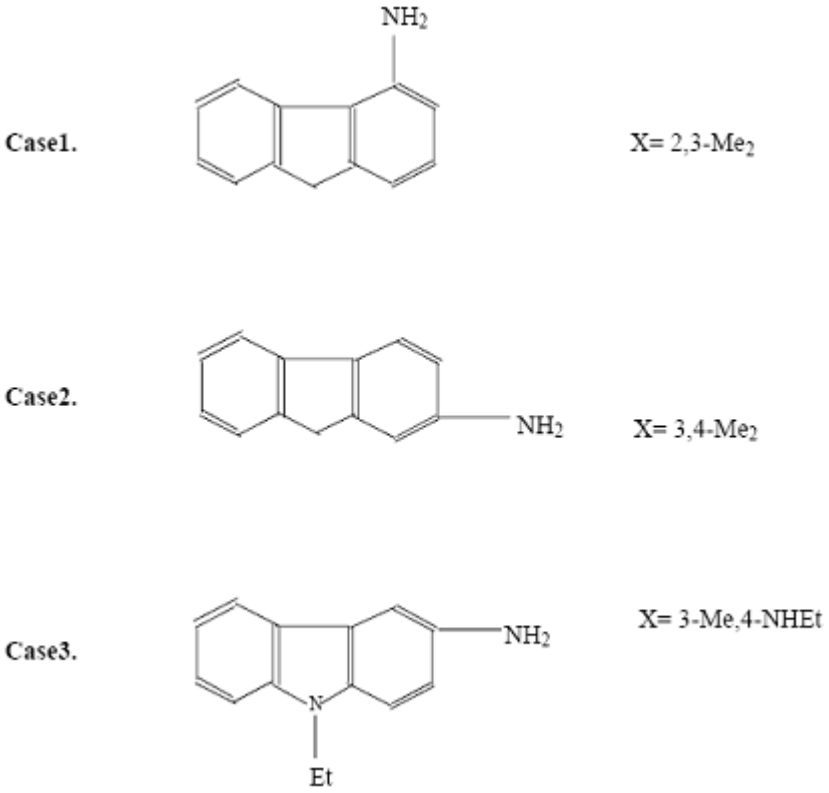


Figure 3. ***Treatment of aminofluorenes.***

RESULT AND DISCUSSION

Values of topological descriptors of carcinogenic compounds have been evaluated with the help of Cache Software using PM3 Hamiltonian and are included in Table-2 alongwith reported activity in terms of Log P. The values of descriptors in different combinations have been used for development of QSAR models. Nine models providing correlation coefficient above 0.80 have been chosen, which are presented below. The outliers are the compounds 4, 8, 9, 11, 27, 32, 42, 58.

Table 2. ***Values of topological descriptors of carcinogenic compounds alongwith observed activities in terms of log P***

Compound	Log P Calculated By PM3 Method	Connectivity Index (order 1, standard)	Valence Connectivity Index (order 0, standard)	Shape Index (kappa alpha, order 1)	Dipole Moment (debye)	Solvent Accessibility Surface Area (angstromsquare)	Molar Refractivity	Observed Activity In terms of log P
1	2.266	5.360	6.119	6.375	1.000	82.000	47.209	2.270
2	1.956	9.689	11.290	12.853	2.000	127.000	86.503	2.160
3	3.548	9.075	11.502	12.122	3.000	122.000	78.038	2.260
5	1.731	3.805	4.887	5.319	1.000	69.000	35.800	1.730
6	3.190	13.365	17.030	19.928	3.000	162.000	121.293	3.020
7	4.632	9.360	10.583	11.696	1.000	124.000	81.031	2.950
10	2.604	10.646	12.704	14.764	1.000	137.000	95.858	2.560
12	2.213	10.646	12.107	14.687	1.000	136.000	89.200	1.910
13	0.202	6.630	8.333	10.848	3.000	105.000	55.052	0.200
14	2.859	8.220	10.196	10.510	2.000	110.000	72.859	2.390
15	1.027	5.236	5.520	7.803	7.000	82.000	39.376	0.930
16	1.263	3.394	3.964	4.342	2.000	62.000	30.758	1.260
17	1.011	4.343	5.295	6.263	1.000	74.000	37.222	1.010
18	1.781	3.788	5.021	5.604	3.000	74.000	35.563	1.780
19	0.998	4.198	5.521	6.549	1.000	78.000	40.264	1.000
20	0.998	4.198	5.521	6.549	1.000	79.000	40.264	1.000
21	0.758	5.274	6.626	8.199	1.000	85.000	43.685	1.000
22	1.478	4.736	6.218	7.249	0.000	81.000	42.263	1.480
23	0.978	12.587	15.372	18.839	3.000	154.000	107.944	1.310
24	0.227	4.736	5.795	7.210	1.394	80.000	41.922	0.230
25	2.689	12.491	16.651	16.407	1.549	152.000	101.934	3.710
26	2.847	13.365	16.846	19.928	5.646	159.000	117.179	2.850
28	1.033	7.668	9.019	12.440	5.304	114.000	57.275	0.940
29	1.058	5.774	6.481	8.793	6.160	89.000	44.145	0.960
30	0.528	5.236	5.650	7.803	5.498	83.000	42.382	0.430
31	2.899	10.646	13.740	15.436	2.821	150.000	102.087	2.250
33	2.665	4.609	6.732	7.289	1.186	81.000	45.882	2.670
34	4.632	9.343	10.583	11.696	1.255	124.000	81.031	2.950
35	1.027	5.236	5.520	7.803	5.927	81.000	39.376	0.930
36	0.196	4.198	4.834	6.223	2.453	72.000	37.153	0.200

Compound	Log P Calculated By PM3 Method	Connectivity Index (order 1, standard)	Valence Connectivity Index (order 0, standard)	Shape Index (kappa alpha, order 1)	Dipole Moment (debye)	Solvent Accessibility Surface Area (angstromsquare)	Molar Refractivity	Observed Activity In terms of log P
37	4.264	10.737	12.414	14.237	3.289	141.000	89.413	2.580
38	2.906	9.689	11.091	12.824	1.875	124.000	83.153	3.090
39	2.538	11.083	12.922	15.379	2.823	145.000	91.535	2.720
40	4.660	10.270	12.136	14.783	3.725	131.000	79.048	3.730
41	1.510	10.538	13.403	15.321	2.346	141.000	95.935	3.200
43	0.985	6.220	7.833	9.895	1.876	101.000	50.351	0.990
44	1.830	5.236	5.990	7.813	2.247	82.000	40.045	1.830
45	1.691	4.843	5.689	6.894	3.114	78.000	39.828	1.690
46	0.480	3.805	4.464	5.280	0.219	68.000	35.459	0.480
47	-1.092	10.538	12.290	14.687	1.532	133.000	95.726	0.600
48	1.038	5.464	5.667	8.952	2.550	79.000	36.324	1.040
49	2.665	4.609	6.732	7.289	1.197	82.000	45.882	2.670
50	2.311	4.305	5.834	6.302	1.183	75.000	40.487	1.840
51	1.731	3.788	4.887	5.319	1.383	70.000	35.800	1.730
52	1.027	5.236	5.520	7.803	5.928	82.000	39.376	0.930
53	2.817	4.609	7.134	8.150	1.459	94.000	45.173	2.820
54	1.731	3.788	4.887	5.319	1.410	70.000	35.800	1.730
55	-0.031	11.614	13.952	16.612	1.543	140.000	99.252	1.660
56	2.262	10.538	13.135	14.764	0.602	133.000	95.067	2.530
57	1.998	5.520	7.355	9.478	3.208	96.000	47.126	2.000
59	2.794	6.575	8.837	11.176	1.607	111.000	54.735	2.790
60	0.947	4.215	5.387	6.263	2.506	73.000	40.500	0.950
61	0.962	4.715	5.242	6.924	1.615	75.000	37.517	0.960
62	0.795	6.630	8.467	10.888	4.516	109.000	54.296	0.800
63	0.998	4.198	5.521	6.549	0.934	79.000	40.264	1.000
64	0.758	5.274	6.626	8.199	1.486	85.000	43.685	0.760
65	3.044	7.952	10.827	13.794	2.064	120.000	64.125	2.250
66	1.832	6.915	8.033	9.186	0.699	102.000	60.652	1.690
67	1.009	6.147	6.429	9.458	3.913	90.000	44.440	0.920
68	0.528	5.236	5.650	7.803	6.155	83.000	42.382	0.430
69	0.480	3.788	4.464	5.280	2.435	69.000	35.459	0.480
70	1.862	10.646	12.290	14.687	3.147	140.000	95.176	2.380
71	2.197	4.788	6.189	7.338	5.560	87.000	47.288	1.860
72	0.947	4.198	5.387	6.263	2.424	75.000	40.500	0.950

Compound	Log P Calculated By PM3 Method	Connectivity Index (order 1, standard)	Valence Connectivity Index (order 0, standard)	Shape Index (kappa alpha, order 1)	Dipole Moment (debye)	Solvent Accessibility Surface Area (angstromsquare)	Molar Refractivity	Observed Activity In terms of log P
73	2.249	4.198	5.943	6.588	1.443	80.000	40.604	2.250

Table 3. ***Values of predicted activities PA1 to PA9 of carcinogenic compounds***

Comp	PA1	PA2	PA3	PA4	PA5	PA6	PA7	PA8	PA9
1	1.864	1.788	1.951	1.900	1.934	1.776	1.982	1.868	1.983
2	2.013	1.894	2.027	1.987	2.024	1.967	2.128	1.944	2.124
3	3.131	3.096	3.159	3.063	3.196	3.001	3.101	3.099	3.141
5	1.548	1.565	1.599	1.594	1.539	1.513	1.553	1.576	1.545
6	3.220	3.429	3.289	3.378	3.079	3.297	3.123	3.416	3.110
7	3.339	3.243	3.400	3.412	3.390	3.348	3.532	3.350	3.505
10	2.456	2.420	2.470	2.511	2.410	2.503	2.545	2.454	2.511
12	2.034	1.994	2.071	2.153	2.081	2.133	2.202	2.074	2.149
13	0.813	0.758	0.662	0.678	0.796	0.877	0.742	0.660	0.732
14	2.649	2.605	2.730	2.621	2.709	2.486	2.667	2.648	2.708
15	0.825	0.884	0.901	0.800	0.888	0.716	0.802	0.968	0.900
16	1.160	1.153	1.249	1.224	1.193	1.096	1.219	1.209	1.219
17	1.106	1.114	1.148	1.190	1.135	1.131	1.154	1.132	1.118
18	1.590	1.533	1.535	1.463	1.566	1.516	1.528	1.484	1.557
19	1.175	1.121	1.100	1.105	1.118	1.192	1.145	1.057	1.124
20	1.175	1.102	1.081	1.084	1.118	1.194	1.147	1.029	1.124

Comp	PA1	PA2	PA3	PA4	PA5	PA6	PA7	PA8	PA9
21	1.053	1.086	1.053	1.130	1.073	1.131	1.063	1.067	1.015
22	1.496	1.523	1.492	1.557	1.478	1.556	1.489	1.491	1.443
23	1.733	1.853	1.776	1.839	1.700	1.797	1.703	1.864	1.694
24	0.693	0.673	0.682	0.710	0.686	0.720	0.714	0.659	0.685
25	3.455	3.386	3.505	3.310	3.694	3.168	3.346	3.341	3.411
26	2.973	3.241	3.107	3.093	2.893	2.920	2.809	3.268	2.880
28	1.096	1.119	0.998	1.012	1.142	1.168	1.032	1.081	1.058
29	0.978	1.025	1.011	0.933	1.033	0.901	0.935	1.069	1.013
30	0.618	0.623	0.651	0.577	0.646	0.544	0.620	0.678	0.683
31	2.917	2.712	2.674	2.578	2.761	2.872	2.818	2.560	2.841
33	2.325	2.403	2.308	2.332	2.227	2.335	2.208	2.326	2.196
34	3.342	3.246	3.400	3.405	3.390	3.344	3.527	3.349	3.505
35	0.825	0.886	0.903	0.845	0.888	0.762	0.832	0.971	0.900
36	0.540	0.530	0.570	0.563	0.540	0.522	0.573	0.556	0.572
37	3.254	3.138	3.237	3.184	3.329	3.220	3.354	3.193	3.376
38	2.430	2.426	2.539	2.560	2.475	2.432	2.573	2.535	2.557
39	2.341	2.216	2.278	2.263	2.411	2.357	2.429	2.234	2.425
40	3.365	3.559	3.496	3.582	3.460	3.448	3.388	3.655	3.392
41	2.107	2.026	2.009	1.964	2.042	2.073	2.045	1.948	2.052
43	1.230	1.174	1.102	1.144	1.246	1.314	1.197	1.092	1.166
44	1.419	1.523	1.511	1.590	1.478	1.491	1.472	1.589	1.449
45	1.429	1.470	1.500	1.478	1.472	1.386	1.442	1.517	1.463
46	0.745	0.718	0.793	0.837	0.746	0.770	0.836	0.749	0.787
47	0.421	0.305	0.431	0.418	0.399	0.397	0.537	0.340	0.510
48	0.690	0.974	0.891	1.110	0.768	0.908	0.793	1.116	0.731
49	2.325	2.390	2.296	2.316	2.227	2.334	2.207	2.307	2.196
50	1.986	2.045	2.028	2.042	1.956	1.973	1.947	2.030	1.936
51	1.551	1.549	1.582	1.569	1.539	1.514	1.551	1.549	1.545
52	0.825	0.881	0.899	0.839	0.888	0.762	0.832	0.964	0.900

Comp	PA1	PA2	PA3	PA4	PA5	PA6	PA7	PA8	PA9
53	2.443	2.359	2.172	2.179	2.308	2.512	2.248	2.135	2.229
54	1.551	1.551	1.583	1.570	1.539	1.513	1.550	1.552	1.545
55	1.148	1.200	1.244	1.295	1.185	1.170	1.212	1.254	1.180
56	2.467	2.511	2.535	2.585	2.434	2.490	2.502	2.532	2.464
57	1.755	1.774	1.609	1.641	1.688	1.837	1.624	1.658	1.622
59	2.310	2.293	2.088	2.192	2.247	2.496	2.198	2.132	2.146
60	1.132	1.153	1.164	1.128	1.090	1.072	1.096	1.149	1.113
61	0.901	0.986	1.012	1.094	0.947	0.967	0.988	1.065	0.952
62	1.165	1.068	0.956	0.897	1.154	1.165	1.032	0.937	1.067
63	1.175	1.113	1.092	1.113	1.118	1.210	1.157	1.046	1.124
64	1.053	1.084	1.051	1.101	1.073	1.103	1.045	1.064	1.015
65	2.595	2.772	2.471	2.641	2.531	2.826	2.418	2.632	2.365
66	1.708	1.577	1.686	1.693	1.740	1.722	1.831	1.600	1.794
67	0.772	0.887	0.865	0.954	0.845	0.877	0.857	0.988	0.846
68	0.618	0.617	0.646	0.543	0.646	0.514	0.601	0.670	0.683
69	0.748	0.694	0.768	0.712	0.746	0.673	0.772	0.711	0.787
70	1.924	1.783	1.870	1.828	1.889	1.907	2.015	1.809	2.023
71	1.834	1.766	1.727	1.574	1.730	1.696	1.692	1.695	1.786
72	1.136	1.117	1.128	1.086	1.090	1.079	1.098	1.094	1.113
73	1.978	1.964	1.902	1.900	1.911	1.980	1.891	1.878	1.882

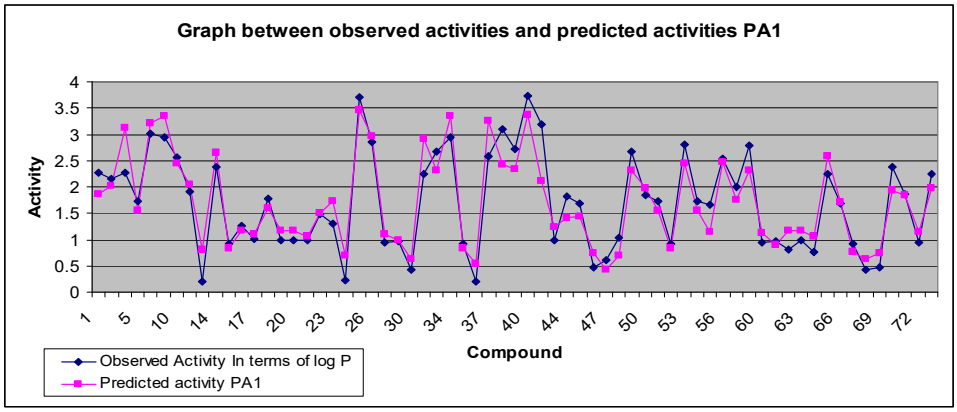
QSAR model PA1

This is best QSAR model and has been developed using the descriptors Log P calculated, connectivity index (order 1, standard), valence connectivity index (order 0, standard) and shape index (kappa alpha, order 1). Value of correlation coefficient is 0.854963 and cross-validation coefficient is 0.764958. These values of correlation and cross-validation coefficients indicate the best predictive power of this QSAR model. With the help of this QSAR model one can efficiently predict the activity of any carcinogenic compound of this group by substituting the values of descriptors in the following MLR equation.

PA1=0.5157*LPC-0.187965*CI+0.386048*VCI-0.134906*SI+0.201453 rCV^2=0.764958 r^2=0.854963

Graph between observed activities and predicted activities by QSAR model PA1 shown in Graph-1 which indicates that the observed and predicted values of activities are very close. Predicted activities PA1 of carcinogenic compounds are listed in Table-3.

Graph 1. ***Graph between predicted activities PA1 and observed activities of carcinogenic compounds in terms of Log p***



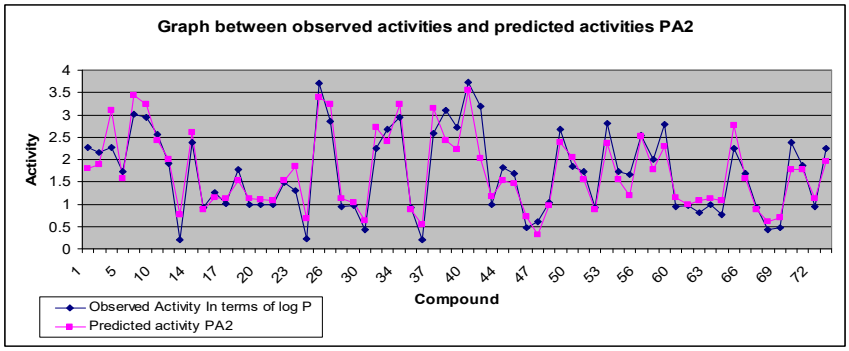
QSAR model PA2

This QSAR model has been developed using the descriptors Log P calculated, connectivity index (order 1, standard), valence connectivity index (order 0, standard) and solvent accessibility surface area. Value of correlation coefficient is 0.854566 and cross-validation coefficient is 0.755597. These values of correlation and cross-validation coefficients indicate the best predictive power of this QSAR model. With the help of this QSAR model one can efficiently predict the activity of any carcinogenic compound of this group by substituting the values of descriptors in the following MLR equation.

PA2=0.562587*LPC-0.172762*CI+0.397175*VCI-0.0227345*SASA+0.872491 rCV^2=0.755597 r^2=0.854566

Graph between observed activities and predicted activities by QSAR model PA2 shown in Graph-2 which indicates that the observed and predicted values of activities are very close. Predicted activities PA2 of carcinogenic compounds are listed in Table-3.

Graph 2. ***Graph between predicted activities PA2 and observed activities of carcinogenic compounds in terms of Log p***



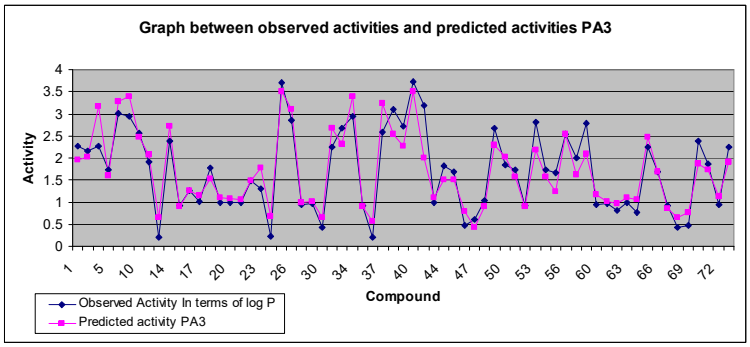
QSAR model PA3

This QSAR model has been developed using the descriptors Log P calculated, valence connectivity index (order 0, standard), shape index (kappa alpha, order 1) and solvent accessibility surface area. Value of correlation coefficient is 0.853765 and cross-validation coefficient is 0.764894. These values of correlation and cross-validation coefficients indicate the best predictive power of this QSAR model. With the help of this QSAR model one can efficiently predict the activity of any carcinogenic compound of this group by substituting the values of descriptors in the following MLR equation.

PA3=0.539375*LPC+0.377096*VCI-0.115161*SI-0.0213018*SASA+0.901655
rCV^2=0.764894
r^2=0.853765

Graph between observed activities and predicted activities by QSAR model PA3 shown in Graph-3 which indicates that the observed and predicted values of activities are very close. Predicted activities PA3 of carcinogenic compounds are listed in Table-3.

Graph 3. ***Graph between predicted activities PA3 and observed activities of carcinogenic compounds in terms of Log p***



QSAR model PA4

This QSAR model has been developed using the descriptors Log P calculated, valence connectivity index (order 0, standard), dipole moment and solvent accessibility surface area. Value of correlation coefficient is 0.852475 and cross-validation coefficient is 0.765162. These values of correlation and cross-validation coefficients indicate the best predictive power of this QSAR model. With the help of this QSAR model one can efficiently predict the activity of any carcinogenic compound of this group by substituting the values of descriptors in the following MLR equation.

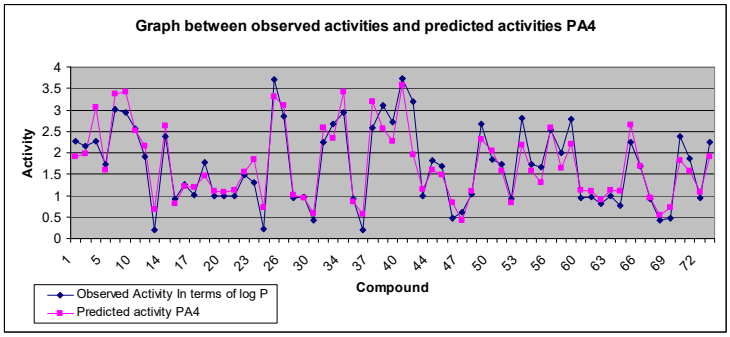
PA4=0.566893*LPC+0.292298*VCI-0.0417343*DM-0.0271546*SASA+1.10852

rCV^2=0.765162

r^2=0.852475

Graph between observed activities and predicted activities by QSAR model PA4 shown in Graph-4 which indicates that the observed and predicted values of activities are very close. Predicted activities PA4 of carcinogenic compounds are listed in Table-3.

Graph 4. ***Graph between predicted activities PA4 and observed activities of carcinogenic compounds in terms of Log p***



QSAR model PA5

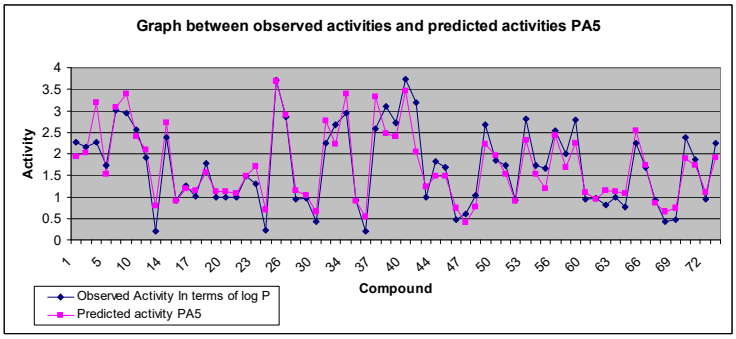
This QSAR model has been developed using the descriptors Log P calculated, valence connectivity index (order 0, standard), molar refractivity and shape index (kappa alpha, order 1). Value of correlation coefficient is 0.851616 and cross-validation coefficient is 0.737379. These values of correlation and cross-validation coefficients indicate the best predictive power of this QSAR model. With the help of this QSAR model one can efficiently predict the activity of any carcinogenic compound of this group by substituting the values of descriptors in the following MLR equation.

PA5=0.501563*LPC+0.423969*VCI-0.205358*SI-0.015675*MR+0.252305

rCV^2=0.737379
r^2=0.851616

Graph between observed activities and predicted activities by QSAR model PA5 shown in Graph-5 which indicates that the observed and predicted values of activities are very close. Predicted activities PA5 of carcinogenic compounds are listed in Table-3.

Graph 5. ***Graph between predicted activities PA5 and observed activities of carcinogenic compounds in terms of Log p***



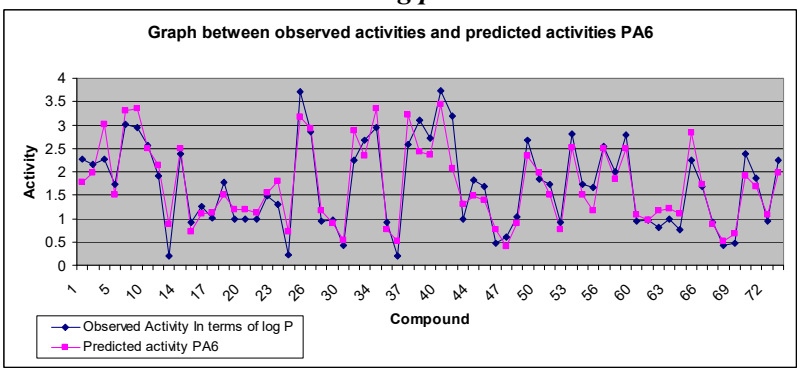
QSAR model PA6

This QSAR model has been developed using the descriptors Log P calculated, connectivity index (order 1, standard), valence connectivity index (order 0, standard) and dipole moment. Value of correlation coefficient is 0.850572 and cross-validation coefficient is 0.76901. These values of correlation and cross-validation coefficients indicate the best predictive power of this QSAR model. With the help of this QSAR model one can efficiently predict the activity of any carcinogenic compound of this group by substituting the values of descriptors in the following MLR equation.

PA6=0.544382*LPC-0.22888*CI+0.266206*VCI-0.0453737*DM+0.200643
rCV^2=0.76901
r^2=0.850572

Graph between observed activities and predicted activities by QSAR model PA6 shown in Graph-6 which indicates that the observed and predicted values of activities are very close. Predicted activities PA6 of carcinogenic compounds are listed in Table-3.

Graph 6. ***Graph between predicted activities PA6 and observed activities of carcinogenic compounds in terms of Log p***



QSAR model PA7

This QSAR model has been developed using the descriptors Log P calculated, valence connectivity index (order 0, standard), dipole moment and shape index (kappa alpha, order 1). Value of correlation coefficient is 0.847989 and cross-validation coefficient is 0.76274. These values of correlation and cross-validation coefficients indicate the best predictive power of this QSAR model. With the help of this QSAR

model one can efficiently predict the activity of any carcinogenic compound of this group by substituting the values of descriptors in the following MLR equation.

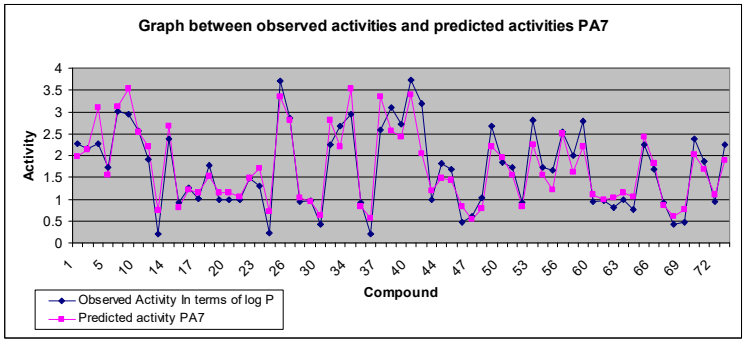
PA7=0.516342*LPC+0.258321*VCI-0.156395*SI-0.0288465*DM+0.26689

rCV^2=0.76274

r^2=0.847989

Graph between observed activities and predicted activities by QSAR model PA7 shown in Graph-7 which indicates that the observed and predicted values of activities are very close. Predicted activities PA7 of carcinogenic compounds are listed in Table-3.

Graph 7. ***Graph between predicted activities PA7 and observed activities of carcinogenic compounds in terms of Log p***



QSAR model PA8

This QSAR model has been developed using the descriptors Log P calculated, valence connectivity index (order 0, standard) and solvent accessibility surface area. Value of correlation coefficient is 0.8477 and cross-validation coefficient is 0.767428. These values of correlation and cross-validation coefficients indicate the best predictive power of this QSAR model. With the help of this QSAR model one can efficiently predict the activity of any carcinogenic compound of this group by substituting the values of descriptors in the following MLR equation.

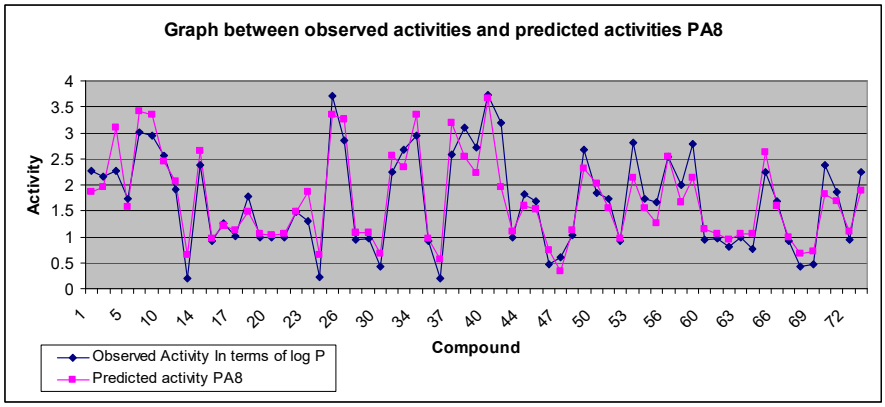
PA8=0.576771*LPC+0.331662*VCI-0.0323539*SASA+1.18364

rCV^2=0.767428

r^2=0.8477

Graph between observed activities and predicted activities by QSAR model PA8 shown in Graph-8 which indicates that the observed and predicted values of activities are very close. Predicted activities PA8 of carcinogenic compounds are listed in Table-3.

Graph 8. ***Graph between predicted activities PA8 and observed activities of carcinogenic compounds in terms of Log p***



QSAR model PA9

This QSAR model has been developed using the descriptors Log P calculated, valence connectivity index (order 0, standard) and shape index (kappa alpha, order 1). Value of correlation coefficient is 0.846163 and cross-validation

coefficient is 0.773126. These values of correlation and cross-validation coefficients indicate the best predictive power of this QSAR model. With the help of this QSAR model one can efficiently predict the activity of any carcinogenic compound of this group by substituting the values of descriptors in the following MLR equation.

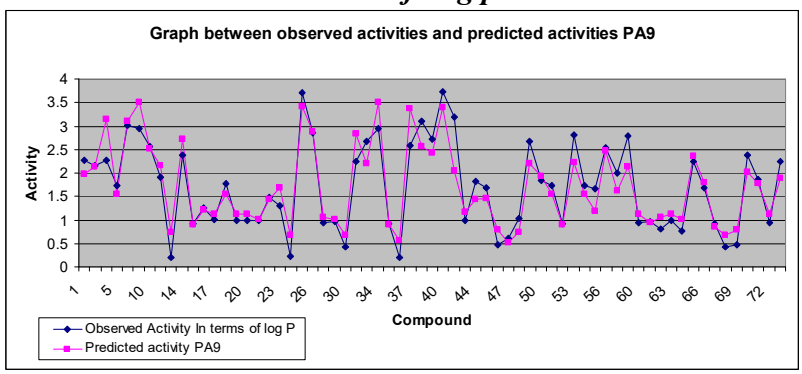
PA9=0.512168*LPC+0.295946*VCI-0.190043*SI+0.22365

rCV^2=0.773126

r^2=0.846163

Graph between observed activities and predicted activities by QSAR model PA9 shown in Graph-9 which indicates that the observed and predicted values of activities are very close. Predicted activities PA9 of carcinogenic compounds are listed in Table-3.

Graph 9. ***Graph between predicted activities PA9 and observed activities of carcinogenic compounds in terms of Log p***



Predicted activities in decreasing order of correlation coefficient i. e. predictive power are given in Table-4 which contains cross-validation coefficient, correlation coefficient and descriptors used in the QSAR model.

Table 4. ***Predicted Activities in decreasing order of regression coefficient***

S. No.	Predicted Activity	rCV^2	r^2	Descriptors used in MLR analysis
1	PA1	0.764958	0.854963	Log P Calcd., Connectivity Index (order 1, standard), Valence Connectivity Index (order 0, standard), Shape Index (Kappa alpha, order 1)
2	PA2	0.755597	0.854566	Log P Calcd., Connectivity Index (order 1, standard), Valence Connectivity Index (order 0, standard), Solvent Accessibility Surface Area
3	PA3	0.764894	0.853765	Log P Calcd., Valence Connectivity Index (order 0, standard), Shape Index (Kappa alpha, order 1), Solvent Accessibility Surface Area
4	PA4	0.765162	0.852475	Log P Calcd., Valence Connectivity Index (order 0, standard), Dipole Moment, Solvent Accessibility Surface Area
5	PA5	0.737379	0.851616	Log P Calcd., Valence Connectivity Index (order 0, standard), Shape Index (Kappa alpha, order 1), Molar Refractivity

S. No.	Predicted Activity	rCV^2	r^2	Descriptors used in MLR analysis
6	PA6	0.769010	0.850572	Log P Calcd., Connectivity Index (order 1, standard), Valence Connectivity Index (order 0, standard), Dipole Moment
7	PA7	0.762740	0.847989	Log P Calcd., Valence Connectivity Index (order 0, standard), Shape Index (Kappa alpha, order 1), Dipole Moment
8	PA8	0.767428	0.847700	Log P Calcd., Valence Connectivity Index (order 0, standard), Solvent Accessibility Surface Area
9	PA9	0.773126	0.846163	Log P Calcd., Valence Connectivity Index (order 0, standard), Shape Index (Kappa alpha, order 1)

CONCLUSION

QSAR model PA1 developed using the descriptors Log P calculated, connectivity index (order 1, standard), valence connectivity index (order 0, standard) and shape index (kappa alpha, order 1) is the best QSAR model. Value of correlation coefficient is 0.854963 and cross-validation coefficient is 0.764958.

REFERENCES

1. Y.T. Woo, D.Y. Lai, Aromatic amino and nitro-amino compounds and their halogenated derivatives, in: E. Bingham, B. Cohrssen, C.H. Powell (Eds.), Patty's Toxicology, Wiley, New York, **2001**, pp. 969–1105.
2. C. Hansch, A. Leo, Exploring QSAR. Part 1. Fundamentals and Applications in Chemistry and Biology, American Chemical Society, Washington, DC, **1995**.
3. T. Fujita, *Q. Struct. Act. Relat*., **1997**, 16, 107–112.
4. C. Hansch, D. Kim, A.J. Leo, E. Novellino, C. Silipo, A. Vittoria, *Rev. Toxicol.,* **1989**, 19. 185–226.
5. H. J. M. Verhaar, J. Solbe, J. Speksnijder, C. J. van Leeuwen, J. L. M. Hermens, *Chemosphere,* **2000**, 40, 875–883.
6. H. Hong, W. Tong, H. Fang, L.M. Shi, Q. Xie, J. Wu, R. Perkins, J.D. Walker, W. Branham, D.M. Sheehan, *Environ. Health Perspect*., **2002**, 110, 29–36.
7. R. Benigni, A. Giuliani, *Med. Res. Rev*., **1996**, 16, 267–284.
8. A.M. Richard, R. Benigni, *SAR QSAR Environ. Res*., **2002**, 13 (1), 1–19.
9. R. Benigni, A. Giuliani, R. Franke, A. Gruska, *Chem. Rev*., **2000**, 100, 3697–3714.
10. R. Franke, A. Gruska, A. Giuliani, R. Benigni, *Carcinogenesis*, **2001**, 22, 1561–1571.
11. Lai, D. Y.; Woo, Y. T.; Argus, M. F.; Arcos, J. C. *In Designing safar chemicals. Green chemistry for pollution prevention;* DeVito, S. C., Garrett, R. L., Eds.; *American Chemical Society: Washington, D. C.,* **1996**; pp 1-13.
12. Vineis, P.; Pirastu, R. Carner; *Causes Control,* **1997**, 8, 346-355.
13. Bartsch, H.; Malaveille, C.; Friesen, M.; Kadlubar, F. F.; *Vineis, P. Eur. J. Cancer,* **1993**, 29A, 1199-1207.
14. Huff, J.; *Environ. Health Perspect***., 1993**, 100, 201-210.

15. Benigni, R,; Giutiani, A.; Franke, R.; Gruska, A.; *Chem. Rev.,* **2000**, 100, 3697.
16. Gold, L. S.; Slone, T. H.; Manley, N. B.; Garfinkel, G. B.; Hudes, E. S.; Rohrbach, L.; Ames, B.N. *Environ. Health Perspect.,* **1991**, 96, 11-15.
17. Singh, P. P.; Pasha, F. A.; Srivastava, H. K. *QSAR & Comb. Sci.,* **2003**, *22*, 843.
18. Singh, P. P.; Srivastava, H. K.; Pasha, F. A.; *Bioorg Med Chem,* **2004**, 12, 171.
19. Singh, P. P.; Pasha, F. A.; Srivastava, H. K. *Indian J. Chem-B,* **2004**, *43B*, 983-991.
20. Pasha, F. A.; Srivastava, H. K.; Singh, P. P.; *Mol Diver,* **2005**, 9(1-3), 215.
21. Srivastava, H.K., Pasha, F.A., Singh, P. P., *Int. J. Quan. Chem.*, **2005**, 103, 237
22. Singh, P. P.; Khan, S. A.; Verma, P. K. Organic Chemistry: An Indian Journal (OCAIJ), **2010**, 6(3), 230-239.