Multi linear regression analysis of carcinogencity of aromatic amines with the help of quantum chemical and energy descriptors

P.P.Singh*, S.A.Khan, Prabhat Kumar Verma
M.L.K. (P.G.) College, Balrampur - 271 201, U.P., (INDIA)
E-mail : dr_ppsingh@sify.com
Received: 1st April, 2010 ; Accepted: 11th April, 2010

ABSTRACT

QSAR models of 73 derivatives of aniline which have carcinogenic property have been developed with the help of quantum chemical and energy descriptors such as heat of formation, global hardness, total energy, HOMO energy, LUMO energy, absolute hardness and chemical potential. The qualities of the models have been adjudged by the value of cross-validation and correlation coefficients evaluated by multi linear regression analysis. The models have been arranged in a table in descending order of predictive power. The best model has correlation coefficient about 0.90 and has been developed with the help of heat of formation, total energy, HOMO energy and chemical potential. The most widely used descriptor is total energy.
© 2010 Trade Science Inc. - INDIA

INTRODUCTION

The aromatic amines are one of the chemical classes in which the structural and molecular basis of carcinogenicity is most clearly under stood[1]. This class of molecules offers the unique possibility of covering all the investigation levels, ranging from physicochemical properties to epidemiological findings in human populations, with rational explanations.

They 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[2]. 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.

Most of the above studies refer to exposures to mixtures of aromatic amines. For 2-naphthylamine, o-toluidine, benzidine and 4-aminobiphenyl, it has been possible to select cohorts of individuals experiencing exposure that can be reasonably considered as single agent exposure[2], thus providing formal demonstration of the carcinogenic potential of these agents for humans. In the case of 4-aminobiphenyl, there are molecular epidemiology studies[2,3] that were able to identify a specific DNA adduct identified as a derivative of 4-aminobiphenyl.

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[4,6]. The evidence in experimental animals has been crucial in the classification of some aromatic amines for their carcinogenicity to humans. Benzidine-based dyes and MOCA (4,4′-methylene bis-2chloroaniline) were classified by the international agency for research on cancer (IARC) as probable carcinogens based on the strong evidence in animals before epidemiological evi-
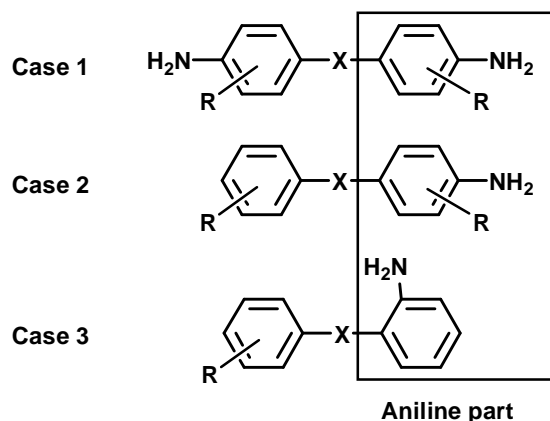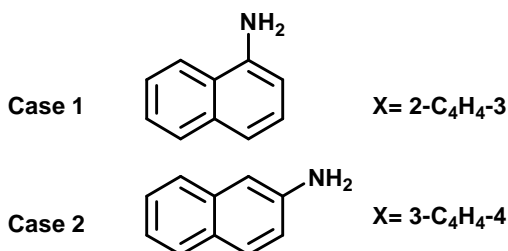
Figure 1 : Treatment of biphenylamine

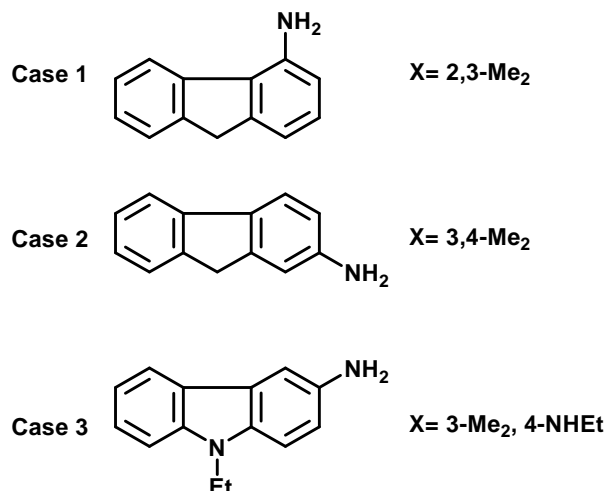**Figure 2 : Treatment of naphthylamines**

Figure 3 : Treatment of aminofluorenes

dence was available[2].

Although the major concern posed by the aromatic amines drives from their carcinogenic potential, the number of QSAR studies is quite limited[5], hence needs a comprehensive study on QSAR of aromatic amines, whose biological activity is reported. In this paper we propose to make QSAR studies on carcinogenicity of aromatic amines with the help of quantum chemical and energy 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 pre-

dicting the activity of any new compound.

Theory

A series of quantities, which are readily used while considering chemical reactivity, appear in a most natural way within the framework of quantum chemical techniques[12-14]. For example, a new theoretical basis is found for the use of the frontier molecular orbital (FMO), being the highest occupied molecular orbital (HOMO), and lowest unoccupied molecular orbital (LUMO), as reactivity indices. This concept was introduced by Fukui in the FMO Theory[15].

Parr et al.[16] define electronegativity as the negative of chemical potential.

$$\chi = -\mu = -(\partial E / \delta N)v(\Upsilon) \tag{1}$$

The absolute hardness, η, is defined as[20]:

$$\eta = 1/2 \ (\delta\mu / \delta N)v(\Upsilon) = 1/2 \ (\delta 2E / \delta N2)v(\Upsilon) \tag{2}$$

where E is the total energy, the number of electrons of the chemical species, and $v(\Upsilon)$ the external potential.

The corresponding global softness, S, which bears an inverse relationship with the hardness, is defined as:

$$S = 1/2\eta = (\partial N / \partial\delta)v(\Upsilon) \tag{3}$$

The operational definition of absolute hardness, global softness, and electronegativity is defined as:

$$\eta = (IP - EA)/2 \tag{4}$$

$$S = IP - EA \tag{5}$$

$$\chi = -\mu = (IP + EA)/2 \tag{6}$$

where IP and EA are the ionization potential and electron affinity, respectively, of the chemical species. According to Koopman's theorem, the IP is simply the eigenvalue of the HOMO with change of sign and the EA is the eigenvalue of the LUMO with change of sign[16]; hence, Eq. (4-6) can be written as:

$$\eta = (\varepsilon LUMO - \varepsilon HOMO)/2 \tag{7}$$

$$S = 1/(\varepsilon LUMO - \varepsilon HOMO) \tag{8}$$

$$\chi = -(\varepsilon LUMO + \varepsilon HOMO)/2 \tag{9}$$

In the matter of describing QSAR of a chemical system a more useful quantity is the heat of formation of the compound from its elements in their standard state. This is obtained when the energy required to ionize the valence electrons of the atoms involved. The heat of formation is defined as:

$$\Delta Hf = E_{elect} + E_{nuc} - E_{isol} + E_{atom} \tag{10}$$

where E_{elect} is the electronic energy, E_{nuc} is the nuclear-nuclear repulsion energy, E_{isol} is the energy required to

Full Paper

TABLE 1 : Structures of carcinogenic compounds

Comp	Ring	AnX	Bridge X	R	log P	Comp	Ring	AnX	Bridge X	R	log P
1	N	3-C$_4$H$_4$-4		H	2.27	38	B	4-Ph-4-F		H	3.09
2	B	4-Ph-4-NH$_2$		H	2.16	39	B	4-Ph-4-F		COMe	2.72
3	F	3,4-Me$_2$		COMe	2.61	40	F	3,4-Me$_2$		COCF$_3$	3.73
4	B	2-Cl,4-Ph-3-Cl,4-NH$_2$	CH$_2$	H	3.60	41	B	2-Cl,4-Ph-3-Cl,4-NH$_2$		H	3.20
5	A	2-Me		H	1.73	42	B	4-SO$_2$-Ph-4-NHCOMe	NH$_2$	COMe	0.57
6	B	4-C(? NH)-Ph-4N(Me)$_2$	C ? NH$_2$	Me$_2$	3.02	43	A	4-OEt		COMe	0.99
7	B	2-Ph		H	2.95	44	A	4-F		Me,NO	1.83
8	A	2,6-Cl$_2$,4-NH$_2$		H	1.52	45	A	H		Me,NO	1.69
9	A	2-NO$_2$,4-N(C$_2$H$_4$OH)$_2$		Me	0.34	46	A	2-NH$_2$		H	0.48
10	B	4-CH$_2$-Ph-4-NH$_2$	CH$_2$	H	2.56	47	B	2-NH$_2$,4-Ph-3,4-(NH$_2$)$_2$		H	0.60
11	A	4-Cl		CONMe$_2$	1.64	48	A	2,4,5,6-F$_4$,3-NH$_2$		H	1.04
12	B	4-O-Ph-4-NH$_2$	O	H	1.91	49	A	2,4,6-Me$_3$		H	2.67
13	A	2-OEt,5-NHCOMe	H		0.20	50	A	H		Me	1.84
14	F	3-Me,4-NEt		H	2.39	51	A	4-Me		H	1.73
15	A	3-NO$_2$,4-OH		H	0.93	52	A	2-OH,5-NO$_2$		H	0.93
16	A	H		H	1.26	53	A	2,4,6-Cl$_3$		H	2.82
17	A	2-OMe		H	1.01	54	A	3-Me		H	1.73
18	A	4-Cl		H	1.78	55	B	2-OMe,4-Ph-3-OMe,4-NH$_2$			1.66
19	A	2Cl,5-NH$_2$		H	1.00	56	B	2-Me,4-Ph-3M$_3$,4-NH$_2$		H	2.53
20	A	2NH$_2$,4Cl		H	1.00	57	A	2,5-Cl$_2$,3-COOH		H	2.00
21	A	2Me,4-OMe		H	1.48	58	B	2-Me,4-CH$_2$-Ph-3-Me,4-NH$_2$	CH$_2$	H	3.50
22	A	2-OMe,5-Me		H	1.48	59	A	3-Cl		COOiPr	2.79
23	B	4-SO$_2$-Ph-4-NH$_2$	SO$_2$	H	1.31	60	A	2-M2, 3-NH$_2$		H	0.95
24	A	2-OMe,5-NH$_2$		H	0.23	61	A	2-COOH		H	0.96
25	B	4-CH$_2$-Ph-4-N(Me)$_2$	CH$_2$	H	3.71	62	A	4-COCH$_2$Cl		COMe	0.80
26	B	4-CO-Ph-4-N(Me)$_2$	CO	H	2.85	63	A	2-Cl, 4-NH$_2$		H	1.00
27	N	2-C$_3$H$_3$C(NH$_2$)-3		H	1.48	64	A	2,4-OMe$_2$		H	0.76
28	A	3-NO2,4-OEt		COMe	0.94	65	A	2,6-OMe$_2$, 4-OCONMe		Me$_2$	2.25
29	A	2-OMe,5-NO$_2$		H	0.96	66	N	2-C$_4$H$_4$-3		C$_2$H$_4$NH$_2$	1.69
30	A	2-NO$_2$,4-NH$_2$		H	0.43	67	A	2-COOH, 5-NO$_2$		H	0.92
31	B	4-S-Ph-4-NH$_2$	S	H	2.25	68	A	2-NH$_2$, 4-NO$_2$		H	0.43
32	A	2,6-(NO$_2$)2,4-CF$_3$		(nPr)$_2$	4.25	69	A	4-NH$_2$		H	0.48
33	A	2,4,5-Me$_3$		H	2.67	70	A	4-NH-Ph-4-NH$_2$	NH	H	2.88
34	B	4-Ph		H	2.95	71	A	H		CSNH$_2$	1.86
35	A	2-OH,4-NO$_2$		H	0.93	72	A	2-Me, 4-NH$_2$		H	0.95
36	A	2-OH,5-NH$_2$		H	0.20	73	A	2-Cl, 4-Me		H	2.25
37	B	4-Ph		COMe	2.58						

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

strip all the valence electrons of all the atoms in the system, and E$_{atom}$ is the total heat of atomization of all the atoms in the system.

Total energy of a molecular system is the sum of the total electronic energy of internuclear repulsion, E$_{nr}$.

The total electronic energy of the system is given by:

$$E = P(H+F)/2 \tag{11}$$

where P is the density matrix and H is the one-electron matrix. These parameters and the charges on atoms

Full Paper

TABLE 2 : Value of the quantum chemical descriptors of the carcinogenic compounds under study

Comp.	Heat of formation (kcal/mole)	Global hardness	Total energy (Hartree)	HOMO energy (eV)	LUMO energy (eV)	Absolute hardness	Chemical potential	log P Obsd.
2	79.453	-8.319	-130.557	-8.430	-0.111	4.160	4.271	2.160
5	13.068	-8.951	-54.221	-8.539	0.412	4.476	4.063	1.730
6	127.116	-8.564	-188.424	-8.725	-0.161	4.282	4.443	3.020
10	66.662	-8.456	-144.871	-8.426	0.031	4.228	4.197	2.560
12	18.946	-8.380	-154.876	-8.662	-0.282	4.190	4.472	1.910
15	58.491	-7.110	-90.936	-8.839	-1.729	3.555	5.284	0.930
16	25.683	-8.683	-47.065	-8.067	0.615	4.341	3.726	1.260
17	-14.521	-8.755	-66.403	-8.396	0.359	4.378	4.019	1.010
19	20.308	-8.404	-68.304	-7.784	0.620	4.202	3.582	1.000
20	13.908	-8.492	-68.236	-8.383	0.109	4.246	4.137	1.000
21	-52.445	-8.529	-85.767	-8.308	0.221	4.265	4.044	1.000
22	-24.241	-8.786	-73.584	-8.524	0.262	4.393	4.131	1.480
25	163.865	-8.465	-173.666	-8.373	0.092	4.232	4.140	3.710
26	14.558	-8.859	-193.983	-9.276	-0.416	4.430	4.846	2.850
28	19.885	-7.118	-129.870	-8.804	-1.686	3.559	5.245	0.940
29	64.445	-7.112	-98.082	-8.873	-1.761	3.556	5.317	0.960
30	105.508	-6.707	-88.163	-8.442	-1.736	3.353	5.089	0.430
31	99.087	-7.902	-148.951	-8.579	-0.677	3.951	4.628	2.250
35	60.886	-6.787	-90.961	-8.453	-1.666	3.393	5.060	0.930
39	-7.917	-8.664	-161.662	-8.944	-0.280	4.332	4.612	2.720
44	6.467	-8.784	-89.675	-9.272	-0.488	4.392	4.880	1.830
45	49.650	-8.965	-73.765	-9.171	-0.205	4.483	4.688	1.690
48	-150.364	-8.321	-120.144	-9.061	-0.740	4.161	4.900	1.040
50	19.924	-8.897	-61.286	-8.443	0.455	4.449	3.994	1.840
51	16.498	-8.589	-54.252	-7.945	0.644	4.295	3.650	1.730
52	56.742	-7.139	-90.937	-8.914	-1.776	3.569	5.345	0.930
54	16.406	-8.632	-54.253	-8.029	0.603	4.316	3.713	1.730
56	66.180	-8.252	-144.918	-8.280	-0.028	4.126	4.154	2.530
61	-65.431	-8.267	-76.778	-8.729	-0.462	4.134	4.595	0.960
63	22.584	-7.792	-68.293	-7.361	0.431	3.896	3.465	1.000
64	-52.487	-8.515	-85.774	-8.177	0.338	4.258	3.919	0.760
65	-75.109	-9.093	-133.909	-9.105	-0.012	4.547	4.558	2.250
66	40.165	-7.854	-94.378	-8.241	-0.387	3.927	4.314	1.690
67	14.699	-7.080	-108.460	-9.222	-2.142	3.540	5.682	0.920
69	29.798	-7.943	-56.525	-7.214	0.730	3.972	3.242	0.480
70	86.068	-8.215	-149.302	-8.451	-0.236	4.107	4.344	2.380
72	21.822	-7.888	-63.707	-7.181	0.707	3.944	3.237	0.950

where obtained from PM3[14] calculations.

MATERIALS AND METHODS

Seventy three derivatives of aniline, biphenylamine, naphthylamine and aminofluorenes, as listed in TABLE 1 are the study material of this paper. All the above amines have been treated as substituted anilines. For the biphenylamines (Figure 1) substituted in the aniline part are characterized as in substituted anilines. In cases

Full Paper

TABLE 3 : All the MLR equations alongwith the values of cross-validation and regression coefficients

MLR Equations	rCV^2	r^2
APA1 = 0.00818513*Hf-0.819146*GH-5.34296	0.529418	0.586678
APA2 = 0.00243696*Hf-0.0126867*TE+0.221916	0.498608	0.558285
APA3 = 0.00569993*Hf-0.478271*HOMO-2.61858	0.185509	0.253033
APA4 = 0.00594995*Hf+0.290903*LUMO+1.51031	0.089713	0.244916
APA5 = 0.00818513*Hf+1.63829*AH-5.34296	0.529418	0.586678
APA6 = 0.00526362*Hf-0.0889628*CP+1.83613	-0.033016	0.156812
APA7 = -0.565073*GH-0.0135256*TE-4.4325	0.702436	0.758690
APA8 = -0.577188*GH-0.380143*HOMO-6.37039	0.278374	0.313206
APA9 = -0.957331*GH-0.380143*LUMO-6.37039	0.278374	0.313206
APA10 = 1*GH+2*AH+1.5827	Failed	Failed
APA11 = -0.767259*GH+0.380143*CP-6.37039	0.278374	0.313206
APA12 = -0.014547*TE+0.133643*HOMO+1.22741	0.478459	0.534285
APA13 = -0.0161412*TE+0.432857*LUMO+0.0526467	0.669929	0.724980
APA14 = -0.0135256*TE+1.13015*AH-4.4325	0.702436	0.758690
APA15 = -0.0165127*TE-0.482416*CP+2.00556	0.576361	0.649069
APA16 = -0.957331*HOMO+0.577188*LUMO-6.37039	0.278374	0.313206
APA17 = -0.380143*HOMO+1.15438*AH-6.37039	0.278374	0.313206
APA18 = -1.53452*HOMO-1.15438*CP-6.37039	0.278374	0.313206
APA19 = -0.380143*LUMO+1.91466*AH-6.37039	0.278374	0.313206
APA20 = 1.53452*LUMO+1.91466*CP-6.37039	0.278374	0.313206
APA21 = 1.53452*AH+0.380143*CP-6.37039	0.278374	0.313206
APA22 = 0.00542614*Hf-0.723238*GH-0.0110801*TE-5.61899	0.867392	0.890619
APA23 = 0.00870984*Hf-0.822327*GH-0.486161*HOMO-9.50343	0.605358	0.691065
APA24 = 0.00870984*Hf-1.30849*GH-0.486161*LUMO-9.50343	0.605358	0.691065
APA25 = 0.00818513*Hf-0.5*GH+0*AH-2.72709	0.512642	0.513054
APA26 = 0.00870984*Hf-1.06541*GH+0.486161*CP-9.50343	0.605358	0.691065
APA27 = -0.567315*GH-0.0144184*TE+0.151012*HOMO-3.26219	0.716441	0.766673
APA28 = -0.416303*GH-0.0144184*TE+0.151012*LUMO-3.26219	0.716441	0.766673
APA29 = 0*GH-0.0135256*TE+0*AH+0.199095	Failed	Failed
APA30 = -0.491809*GH-0.0144184*TE-0.151012*CP-3.26219	0.716441	0.766673
APA31 = -0.0144184*TE-0.416303*HOMO+0.567315*LUMO-3.26219	0.716441	0.766673
APA32 = -0.0144184*TE+0.151012*HOMO+1.13463*AH-3.26219	0.716441	0.766673
APA33 = -0.0144184*TE-0.983618*HOMO-1.13463*CP-3.26219	0.716441	0.766673
APA34 = -0.75*HOMO+0.5*LUMO+0.5*AH-6.68371	Failed	Failed
APA35 = -0.890625*HOMO+0.648438*LUMO+0.148438*CP-6.4349	0.244554	0.313151
APA36 = -1.1875*LUMO+2.84375*AH-1.0625*CP-5.75224	-0.282463	0.260056
APA37 = -0.375*GH-0.5*HOMO+0.125*LUMO-5.69413	0.257913	0.308873
APA38 = 1*GH-0.380143*HOMO+2*AH-1.63949	Failed	Failed
APA39 = -1.34375*GH+0.875*HOMO+1.375*CP-8.03436	0.183853	0.301921
APA40 = 0*GH-0.380143*LUMO+4*AH-14.9166	Failed	Failed
APA41 = -0.792969*GH-0.210938*LUMO+0.320313*CP-6.37819	-0.211306	0.286360
APA42 = -0.754883*GH+0.42312*AH+0.380143*CP-8.00299	Failed	Failed
APA43 = -0.0144184*TE+0.151012*LUMO+0.832606*AH-3.26219	0.716441	0.766673
APA44 = -0.0144184*TE+0.983618*LUMO+0.832606*CP-3.26219	0.716441	0.766673

MLR Equations	rCV^2	r^2
APA45 = -0.0144184*TE+0.983618*AH-0.151012*CP-3.26219	0.716441	0.766673
APA46 = 0.00231686*Hf-0.0130139*TE+0.0463849*HOMO+0.584719	0.451580	0.558965
APA47 = 0.00314158*Hf-0.0148923*TE+0.456366*LUMO+0.105784	0.731350	0.774680
APA48 = 0.00542614*Hf-0.0110801*TE+1.44648*AH-5.61899	0.867392	0.890619
APA49 = 0.00219073*Hf-0.0154873*TE-0.471202*CP+2.00474	0.588733	0.673454
APA50 = 0.00870984*Hf-1.30849*HOMO+0.822327*LUMO-9.50343	0.605358	0.691065
APA51 = 0.00870984*Hf-0.486161*HOMO+1.64465*AH-9.50343	0.605358	0.691065
APA52 = 0.00558232*Hf-0.0106885*TE-1.50831*HOMO-1.45397*CP-6.07427	0.859928	0.891543
APA53 = 0.00870984*Hf-0.486161*LUMO+2.61698*AH-9.50343	0.605358	0.691065
APA54 = 0.00870984*Hf+2.13082*LUMO+2.61698*CP-9.50343	0.605358	0.691065
APA55 = 0.00870984*Hf+2.13082*AH+0.486161*CP-9.50343	0.605358	0.691065
APA56 = -0.21875*HOMO+1.23438*AH+0.15625*CP-6.0143	0.280482	0.312136
APA57 = 0.00558232*Hf-0.726984*GH-0.0106885*TE-0.0543415*HOMO-6.07427	0.859928	0.891543
APA58 = 0.00558232*Hf-0.781325*GH-0.0106885*TE-0.0543415*LUMO-6.07427	0.859928	0.891543
APA59 = 0.00542614*Hf-0.5*GH-0.0110801*TE+0*AH-3.78923	0.863376	0.854596
APA60 = 0.00558232*Hf-0.754154*GH-0.0106885*TE+0.0543415*CP-6.07427	0.859928	0.891543
APA61 = 0.00870984*Hf-0.25*GH-1.125*HOMO+0.75*LUMO-10.0175	0.258272	0.663401
APA62 = 0.00870984*Hf-0.5*GH-0.486161*HOMO+0*AH-6.86149	0.622593	0.615966
APA63 = 0.00870984*Hf-1.0625*GH+0.078125*HOMO+0.4375*CP-8.60434	0.589492	0.683627
APA64 = 0.00870984*Hf-1.5*GH-0.486161*LUMO+0*AH-11.0731	0.533968	0.664554
APA65 = 0.00870984*Hf-1.23438*GH-0.34375*LUMO+0.0546875*CP-9.09553	-7.723340	0.686531
APA66 = 0.00870984*Hf-1.35662*GH-0.785624*AH+0.486161*CP-8.67063	0.624812	0.683603
APA67 = 0.5*GH-0.0144184*TE-1*HOMO+1*LUMO-3.99046	0.276086	0.753537
APA68 = 0*GH-0.0144184*TE+0.151012*HOMO+0*AH+1.38779	-1.277000	0.534030
APA69 = -0.703125*GH-0.0144184*TE+0.375*HOMO+0.328125*CP-3.91331	0.699326	0.761341
APA70 = 0*GH-0.0144184*TE+0.151012*LUMO+2*AH-8.04643	Failed	Failed
APA71 = -0.59375*GH-0.0144184*TE-0.195313*LUMO-0.273438*CP-3.61641	-1.759190	0.763627
APA72 = 0.269127*GH-0.0144184*TE+2.45078*AH-0.151012*CP-7.06904	-716.4830	0.610745
APA73 = -0.0144184*TE-0.5*HOMO+0.75*LUMO+0*AH-3.92051	0.547533	0.738791
APA74 = -0.0144184*TE-0.679688*HOMO+0.335938*LUMO-0.46875*CP-3.50724	0.672054	0.766360
APA75 = -0.0144184*TE-0.0234375*HOMO+0.929688*AH-0.148438*CP-3.2511	0.717045	0.765818
APA76 = -0.554688*HOMO+0.140625*LUMO+0.679688*AH+0.0234375*CP-5.96775	-0.548980	0.303449
APA77 = 0.00558232*Hf-0.0106885*TE-0.781325*HOMO+0.726984*LUMO-6.07427	0.859928	0.891543
APA78 = 0.00558232*Hf-0.0106885*TE-0.0543415*HOMO+1.45397*AH-6.07427	0.859928	0.891543
APA79 = -0.00201587*Hf-0.00350203*TE+1.64093*HOMO-1.6642*CP-2.1452	-0.233293	0.242577
APA80 = 0.00558232*Hf-0.0106885*TE-0.0543415*LUMO+1.56265*AH-6.07427	0.859928	0.891543
APA81 = 0.00558232*Hf-0.0106885*TE+1.50831*LUMO+1.56265*CP-6.07427	0.859928	0.891543
APA82 = 0.00870984*Hf-0.5*HOMO+0*LUMO+2*AH-11.077	-0.408135	0.668075
APA83 = 0.00870984*Hf- 1.23438*HOMO +0.898438*LUMO +0.140625 *CP-9.46959	0.525351	0.691001
APA84 = 0.00519515*Hf+1.44842	0.147212	0.152001
APA85 = -0.586369*GH-3.22345	0.234944	0.248533
APA86 = -0.013756*TE+0.17552	0.489551	0.528030
APA87 = -0.402638*HOMO-1.83016	0.015339	0.072622
APA88 = 0.214358*LUMO+1.64268	-0.052310	0.052193
APA89 = 1.17274*AH-3.22345	0.234944	0.248533
APA90 = -0.0516357*CP+1.80877	-0.148755	0.001630

Full Paper ⬤━━━━━━━

TABLE 4 : Predicted activities in the decreasing order of regression coefficient

P. A.	rCV^2	r^2	Descriptors used in MLR analysis
APA52	0.859928	0.891543	Heat of Formation, Total Energy, HOMO Energy, Chemical Potential
APA57	0.859928	0.891543	Heat of Formation, Global Hardness, Total Energy, HOMO Energy
APA58	0.859928	0.891543	Heat of Formation, Global Hardness, Total Energy, LUMO Energy
APA60	0.859928	0.891543	Heat of Formation, Global Hardness, Total Energy, Chemical Potential
APA77	0.859928	0.891543	Heat of Formation, Total Energy, HOMO Energy, LUMO Energy
APA78	0.859928	0.891543	Heat of Formation, Total Energy, HOMO Energy, Absolute Hardness
APA80	0.859928	0.891543	Heat of Formation, Total Energy, LUMO Energy, Absolute Hardness
APA81	0.859928	0.891543	Heat of Formation, Total Energy, LUMO Energy, Chemical Potential
APA22	0.867392	0.890619	Heat of Formation, Global Hardness, Total Energy
APA48	0.867392	0.890619	Heat of Formation, Total Energy, Absolute Hardness
APA59	0.863376	0.854596	Heat of Formation, Global Hardness, Total Energy, Absolute Hardness
APA47	0.73135	0.77468	Heat of Formation, Total Energy, LUMO Energy
APA27	0.716441	0.766673	Global Hardness, Total Energy, HOMO Energy
APA28	0.716441	0.766673	Global Hardness, Total Energy, LUMO Energy
APA30	0.716441	0.766673	Global Hardness, Total Energy, Chemical Potential
APA31	0.716441	0.766673	Total Energy, HOMO Energy, LUMO Energy
APA32	0.716441	0.766673	Total Energy, HOMO Energy, Absolute Hardness
APA33	0.716441	0.766673	Total Energy, HOMO Energy, Chemical Potential
APA43	0.716441	0.766673	Total Energy, LUMO Energy, Absolute Hardness
APA44	0.716441	0.766673	Total Energy, LUMO Energy, Chemical Potential
APA45	0.716441	0.766673	Total Energy, Absolute Hardness, Chemical Potential
APA74	0.672054	0.76636	Global Hardness, Total Energy, LUMO Energy, Chemical Potential
APA75	0.717045	0.765818	Total Energy, HOMO Energy, Absolute Hardness, Chemical Potential
APA69	0.699326	0.761341	Global Hardness, Total Energy, HOMO Energy, Chemical Potential
APA7	0.702436	0.75869	Heat of Formation, Total Energy, LUMO Energy
APA14	0.702436	0.75869	Total Energy, Absolute Hardness
APA67	0.276086	0.753537	Global Hardness, Total Energy, HOMO Energy, LUMO Energy
APA73	0.547533	0.738791	Global Hardness, Total Energy, LUMO Energy, Absolute Hardness
APA13	0.669929	0.72498	Total Energy, LUMO Energy
APA23	0.605358	0.691065	Heat of Formation, Global Hardness, HOMO Energy
APA24	0.605358	0.691065	Heat of Formation, Global Hardness, LUMO Energy
APA26	0.605358	0.691065	Heat of Formation, Global Hardness, Chemical Potential
APA50	0.605358	0.691065	Heat of Formation, HOMO Energy, LUMO Energy
APA51	0.605358	0.691065	Heat of Formation, HOMO Energy, Absolute Hardness
APA53	0.605358	0.691065	Heat of Formation, LUMO Energy, Absolute Hardness
APA54	0.605358	0.691065	Heat of Formation, LUMO Energy, Chemical Potential
APA55	0.605358	0.691065	Heat of Formation, Absolute Hardness, Chemical Potential
APA83	0.525351	0.691001	Heat of Formation, HOMO Energy, LUMO Energy, Chemical Potential
APA63	0.589492	0.683627	Heat of Formation, Global Hardness, HOMO Energy, Chemical Potential
APA66	0.624812	0.683603	Heat of Formation, Global Hardness, Absolute Hardness, Chemical Potential
APA49	0.588733	0.673454	Heat of Formation, Total Energy, Chemical Potential
APA82	-0.40814	0.668075	Heat of Formation, HOMO Energy, LUMO Energy, Absolute Hardness

P. A.	rCV^2	r^2	Descriptors used in MLR analysis
APA64	0.533968	0.664554	Heat of Formation, Global Hardness, LUMO Energy, Absolute Hardness
APA61	0.258272	0.663401	Heat of Formation, Global Hardness, HOMO Energy, LUMO Energy
APA15	0.576361	0.649069	Total Energy, Chemical Potential
APA62	0.622593	0.615966	Heat of Formation, Global Hardness, HOMO Energy, Absolute Hardness
APA1	0.529418	0.586678	Heat of Formation, Global Hardness
APA5	0.529418	0.586678	Heat of Formation, Absolute Hardness
APA46	0.45158	0.558965	Heat of Formation, Total Energy, HOMO Energy
APA2	0.498608	0.558285	Heat of Formation, Total Energy
APA12	0.478459	0.534285	Total Energy, HOMO Energy
APA86	0.489551	0.528030	Total Energy
APA25	0.512642	0.513054	Heat of Formation, Global Hardness, Absolute Hardness

P.A. = Predicted Activity

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 (Figure 2) two situations are possible. They are treated as anilines substituted by -C₄H₄-. Amino fluorenes are only three in the list of 73 carcinogenic compounds at serial **(3)**, **(14)** and **(40)**. Their structural formula is shown in figure 3.

Overall hydrophobicity is expressed in term of Log P computed from the program Tsar (Oxford Molecular). For QSAR prediction, the 3D modeling and geometry optimization of all the carcinogenic compounds of TABLE 1 has been done with the help of Cache Pro Software of Fujitsu using the semiempirical PM3 Hamiltonian. The MOPAC calculations have been performed using same software. The values of quantum chemical descriptors like absolute hardness, global softness, electronegativity, and chemical potential have been calculated by solving the equations given in the theory. The value of energy descriptors have been evaluated from the PM3 method. 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 in terms of log P.

RESULT AND DISCUSSION

Values of quantum chemical descriptors, heat of formation, global hardness, total energy, HOMO energy, LUMO energy, absolute hardness and chemical potential for the carcinogenic compounds listed in TABLE 1 are given in TABLE 2. Outlier carcinogenic compounds are **(1)**, **(3)**, **(4)**, **(7)**, **(8)**, **(9)**, **(11)**, **(13)**, **(14)**, **(18)**, **(23)**, **(24)**, **(27)**, **(32)**, **(33)**, **(34)**, **(36-38)**, **(40-43)**, **(46)**, **(47)**, **(49)**, **(53)**, **(55)**, **(57-60)**, **(62)**, **(68)**, **(71)** and **(73)**. We have done 90 MLR analyses in which predicted activities form APA1 to APA90 are calculated using the different combinations of quantum chemical and energy descriptors. All the MLR equations are included in TABLE 3. Predicted activities, cross-validation coefficients, regression coefficients and descriptors used in MLR analysis are given in the TABLE 4 in decreasing order of regression coefficient. It means the predicted activity listed first has the highest predictive power and the predicted activity listed at the last has the lowest predictive power.

MLR analysis using the combination of the following quantum chemical descriptors gives the best prediction of the activity of the carcinogenic compounds in terms of log P. Regression coefficients of all these MLR equations are 0.891543 and the cross-validation coefficients are 0.859928.

1   Heat of Formation, Total Energy, HOMO Energy and Chemical Potential.
2   Heat of Formation, Global Hardness, Total Energy and HOMO Energy.
3   Heat of Formation, Global Hardness, Total Energy and LUMO Energy.
4   Heat of Formation, Global Hardness, Total Energy and Chemical Potential.
5   Heat of Formation, Total Energy, HOMO Energy and LUMO Energy.
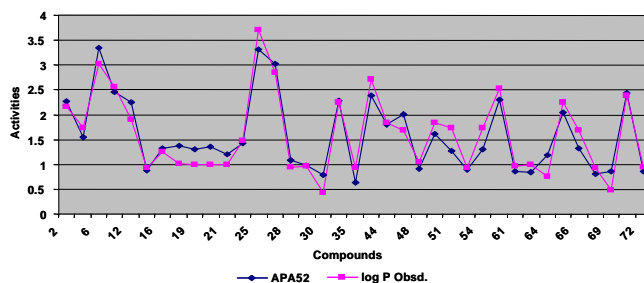6   Heat of Formation, Total Energy, HOMO Energy and Absolute Hardness.

Full Paper

TABLE 5 : Values of the predicted activities in terms of log P for the carcinogenic compounds

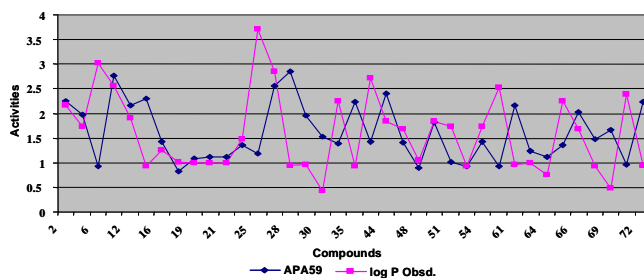| Comp.	APA22	APA27	APA47	APA52	APA59	APA86
2	2.276	2.067	2.249	2.271	2.248	1.971
5	1.527	1.308	1.142	1.55	1.358	0.921
6	3.353	2.996	3.238	3.35	3.27	2.767
10	2.464	2.352	2.487	2.452	2.406	2.168
12	2.26	2.417	2.343	2.249	2.219	2.306
15	0.848	0.748	0.855	0.874	1.091	1.426
16	1.321	1.124	1.168	1.323	1.213	0.823
17	1.37	1.394	1.213	1.376	1.245	1.089
19	1.326	1.315	1.47	1.302	1.28	1.115
20	1.354	1.273	1.215	1.362	1.288	1.114
21	1.215	1.558	1.319	1.202	1.141	1.355
22	1.419	1.496	1.245	1.427	1.288	1.188
25	3.317	2.78	3.249	3.305	3.257	2.564
26	3.017	3.16	2.85	3.025	2.869	2.844
28	1.076	1.319	1.333	1.078	1.316	1.962
29	0.961	0.847	0.965	0.986	1.203	1.525
30	0.781	0.539	0.958	0.791	1.113	1.388
31	2.284	2.073	2.326	2.281	2.35	2.224
35	0.628	0.623	0.891	0.631	0.942	1.427
39	2.395	2.633	2.36	2.394	2.291	2.399
44	1.762	1.614	1.239	1.81	1.631	1.409
45	1.952	1.503	1.267	2.007	1.78	1.19
48	0.914	1.822	1.085	0.912	0.887	1.828
50	1.603	1.394	1.289	1.619	1.447	1.019
51	1.284	1.193	1.26	1.274	1.196	0.922
52	0.859	0.753	0.828	0.888	1.096	1.426
54	1.314	1.205	1.24	1.309	1.217	0.922
56	2.314	2.259	2.459	2.293	2.302	2.169
61	0.856	1.217	0.833	0.866	0.84	1.232
63	0.896	1.031	1.391	0.846	0.986	1.115
64	1.205	1.571	1.373	1.184	1.134	1.355
65	2.034	2.452	1.859	2.043	1.834	2.018
66	1.325	1.31	1.461	1.316	1.401	1.474
67	0.783	0.926	0.79	0.815	1.032	1.667
69	0.914	0.97	1.374	0.863	0.97	0.953
70	2.444	2.275	2.492	2.433	2.44	2.229
72	0.91	1.047	1.446	0.853	0.979	1.052

APA22 and APA48 give same value of predicted activities. APA52, APA57, APA58, APA60, APA77, APA78, APA80 and APA81 give same value of predicted activities. APA27, APA28, APA30, APA32, APA33, APA43, APA44 and APA45 give same value of predicted activities

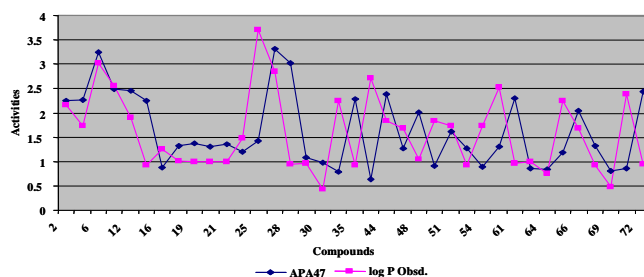7   Heat of Formation, Total Energy, LUMO Energy and Absolute Hardness.

APA52, APA57, APA58, APA60, APA77, APA78, APA80 and APA81 give same value of predicted activities

Graph 1 : Line graph between predicted activity APA52 and observed activity

Graph 2 : Line graph between predicted activity APA59 and observed activity

Graph 3 : Line graph between predicted activity APA47 and observed activity

8   Heat of Formation, Total Energy, LUMO Energy and Chemical Potential.

9   Heat of Formation, Global Hardness and Total Energy.

10   Heat of Formation, Total Energy and Absolute Hardness.

    MLR equations obtained using the above descriptors are given below:

1   **APA52 = 0.00558232*Hf-0.0106885*TE-1.50831* HOMO-1.45397*CP-6.07427**

2   **APA57 = 0.00558232*Hf-0.726984*GH-0.0106885* TE-0.0543415* HOMO-6.07427**

3   **APA58 = 0.00558232*Hf-0.781325*GH-0.0106885* TE-0.0543415* LUMO-6.07427**

4   **APA60 = 0.00558232*Hf-0.754154*GH-0.0106885* TE+0.0543415* CP - 6.07427**

5   **APA77 = 0.00558232*Hf-0.0106885*TE-0.781325***

Full Paper

HOMO+0.726984* LUMO-6.07427

6   **APA78 = 0.00558232*Hf-0.0106885*TE-0.0543415* HOMO+1.45397* AH- 6.07427**

7   **APA80 = 0.00558232*Hf-0.0106885*TE-0.0543415* LUMO+1.56265*AH-6.07427**

8   **APA81 = 0.00558232*Hf-0.0106885* TE+1.50831* LUMO+1.56265*CP-6.07427**

9   **APA22 = 0.00542614*Hf-0.723238*GH-0.0110801* TE-5.61899**

10 **APA48 = 0.00542614*Hf-0.0110801* TE+1.44648* AH-5.61899**

Values of predicted activities of the carcinogenic compounds from above MLR equations are same. Graph between observed activities and predicted activities are given in the graph 1.

MLR analysis using the combination of the quantum chemical descriptors viz. heat of formation, global hardness, total energy and absolute hardness gives the second best prediction of the activity of the carcinogenic compounds in terms of log P. Regression coefficients of the MLR equation is 0.854596 and the cross-validation coefficient is 0.863376. MLR equation obtained using the above descriptors are given below:

APA59 = 0.00542614*Hf-0.5*GH-0.0110801*TE+0*AH-3.78923

Graph between observed activities and predicted activities are given in the graph 2.

MLR analysis using Heat of Formation, Total Energy and LUMO Energy as the combination of the quantum chemical descriptors gives the third best prediction of the activity of the carcinogenic compounds in terms of log P. Regression coefficients of the MLR equation is 0.77468 and the cross-validation coefficient is 0.73135. MLR equation obtained using the above descriptors are given below:

APA47 = 0.00314158*Hf-0.0148923*TE +0.456366* LUMO+0.105784

Graph between observed activities and predicted activities are given in the graph 3.

CONCLUSION

Single best descriptor is total energy because all the combinations of descriptors having total energy as one of the descriptor give the value of regression coefficient greater than 0.5 and even the single descriptor total energy has good predictive power (regression coefficient = 0.528030 and cross-validation coefficient = 0.489551) as shown by MLR equation APA86.

REFERENCES

[1] D.Y.Lai, Y.T.Woo, M.F.Argus, J.C.Arcos; In: 'Designing Safar Chemicals', Green Chemistry for Pollution Prevention; S.C.DeVito, R.L.Garrett, Eds.; American Chemical Society: Washington, D.C., 1-13 **(1996)**.

[2] P.Vineis, Pirastu, R.Carner; Causes Control, **8**, 346-355 **(1997)**.

[3] H.Bartsch, C.Malaveille, M.Friesen, F.F.Kadlubar, P.Vineis; Eur.J.Cancer, A, **29**, 1199-1207 **(1993)**.

[4] J.Huff; Environ.Health Perspect., **100**, 201-210 **(1993)**.

[5] R.Benigni, A.Giutiani, R.Franke, A.Gruska; Chem.Rev., **100**, 3697 **(2000)**.

[6] L.S.Gold, T.H.Slone, N.B.Manley, G.B.Garfinkel, E.S.Hudes, L.Rohrbach, B.N.Ames; Environ. Health Perspect., **96**, 11-15 **(1991)**.

[7] P.P.Singh, F.A.Pasha, H.K.Srivastava; QSAR & Comb.Sci., **22**, 843 **(2003)**.

[8] P.P.Singh, H.K.Srivastava, F.A.Pasha; Bioorg.Med. Chem., **12**, 171 **(2004)**.

[9] P.P.Singh, F.A.Pasha, H.K.Srivastava; Indian J.Chem.B, **43**, 983-991 **(2004)**.

[10] F.A.Pasha, H.K.Srivastava, P.P.Singh; Mol.Diver., **9(1-3)**, 215 **(2005)**.

[11] G.Klopman; J.Am.Chem.Soc., **90**, 223 **(1964)**.

[12] M.J.S.Dewar, W.J.Thiel; J.Am.Chem.Soc., **99**, 4899 **(1977)**.

[13] M.J.S.Dewar, E.G.Zoebisch, E.F.Healy, J.J.P.Stewart; J.Am.Chem.Soc., **107**, 3902-3909 **(1985)**.

[14] J.J.P.Stewart; J.Comp.Chem., **10**, 209-220 **(1989)**.

[15] K.Fukui; Science (Washington DC), **218**, 747 **(1982)**.

[16] R.G.Parr, R.A.Donnelly, M.Levy, W.E Palke; J.Chem.Phys., **68**, 3801 **(1978)**.